



OPEN

## Identification and validation of a prognostic risk-scoring model based on the level of *TIM-3* expression in acute myeloid leukemia

Wanxue Huang<sup>1</sup>, Shasha Zheng<sup>2</sup>, Qi Wang<sup>1</sup>, Na Zhao<sup>1✉</sup> & Zhiguo Long<sup>1✉</sup>

Acute myeloid leukemia (AML) is characterized by an unfavorable prognosis due to the presence of self-renewing leukemic stem cells (LSCs). The presence of T-cell immunoglobulin mucin-3 (*TIM-3*) on the surface of LSCs has been observed in various types of human AML, exerting an impact on the prognostic outcome. Exploring the hub genes associated with varying levels of *TIM-3* expression offers a valuable approach to enhance our understanding of the underlying mechanisms involving *TIM-3* and to identify potential prognostic indicators in AML. Nevertheless, to date, no research studies have reported a prognostic model that relies on the level of *TIM-3* expression. In our study, we screen the hub-genes based on different expression level of *TIM-3* through WGCNA. The prognostic risk-scoring model was constructed based on hub-genes. The results show the risk prognostic model has extraordinary ability to predict prognosis in both the training and validation sets. The high-risk group present poor prognosis with mutation of *NPM1*, *TP53* (Multiple Hit) and *FLT3*(multiple hit), while *IDH2* (Missense Mutation), *MUC16* (Multiple Hit/Missense Mutation) occur mutation in low-risk group presenting favorite prognosis than high-risk group. Leukocyte cell–cell adhesion, regulation of T cell activation and I- $\kappa$ B kinase/NF- $\kappa$ B signaling enriched in high-risk group, involving in HSCs or LSCs anchoring to BM, which implicated in LSCs survival and chemotherapy resistance. *B7-H3* (*CD276*) and *CD276* would be the potential immune targets in high-risk group. The risk score model may help in distinguishing immune and molecular characteristics, predicting patient outcomes.

Acute myeloid leukemia (AML) is a disease characterized by an unfavorable prognosis, originating from self-renewing leukemic stem cells (LSCs)<sup>1</sup>. A major contributing element to AML's incurable nature is resistance to treatment<sup>2</sup>. LSCs endure in individuals suffering from AML, exhibiting resilience against traditional therapeutic methods. These LSCs display varied clones and function as reservoirs, thereby contributing to the recurrence, relapse, or intensification of the ailment into more aggressive phenotypes<sup>3</sup>. Throughout a relapse, there exists a proliferation of the LSC populace<sup>4</sup>. Leukemia development is facilitated by reciprocal interactions between leukemic cells and the surrounding microenvironment<sup>5</sup>. Enacting a treatment approach centered on aiming at LSCs harbors potential for diminishing the likelihood of AML reappearance and potentially enhancing the overall recovery rate.

T-cell immunoglobulin mucin domain-containing protein 3 (*Tim-3*) operates as an inhibitor of inherent and adaptable immune reactions<sup>6–8</sup>. *TIM-3* expression has been observed on the surface of LSCs in various types of human AML, whereas hematopoietic stem cells (HSCs) lack expression<sup>9,10</sup>. A recent study revealed that *Tim-3* protein is functionally expressed by CD34 + CD38—LSCs in AML, while it remains absent in healthy HSCs, as well as in myeloerythroid and lymphoid progenitor populations<sup>10</sup>. Furthermore, by analyzing somatic mutations, it has been determined that multistep leukemogenesis arises from self-renewing HSCs, highlighting *Tim-3* as a potential target for selectively eliminating LSCs while preserving residual HSCs<sup>11</sup>. In a recent investigation involving 302 AML patients, *Tim-3* was detected in LSCs at the time of initial diagnosis in 78.5% of cases<sup>12,13</sup>.

<sup>1</sup>Department of Hematology, Fudan University Affiliated Pudong Medical Center, Shanghai, China. <sup>2</sup>Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, China. ✉email: yiyi\_8147@163.com; ZgLong1976@126.com

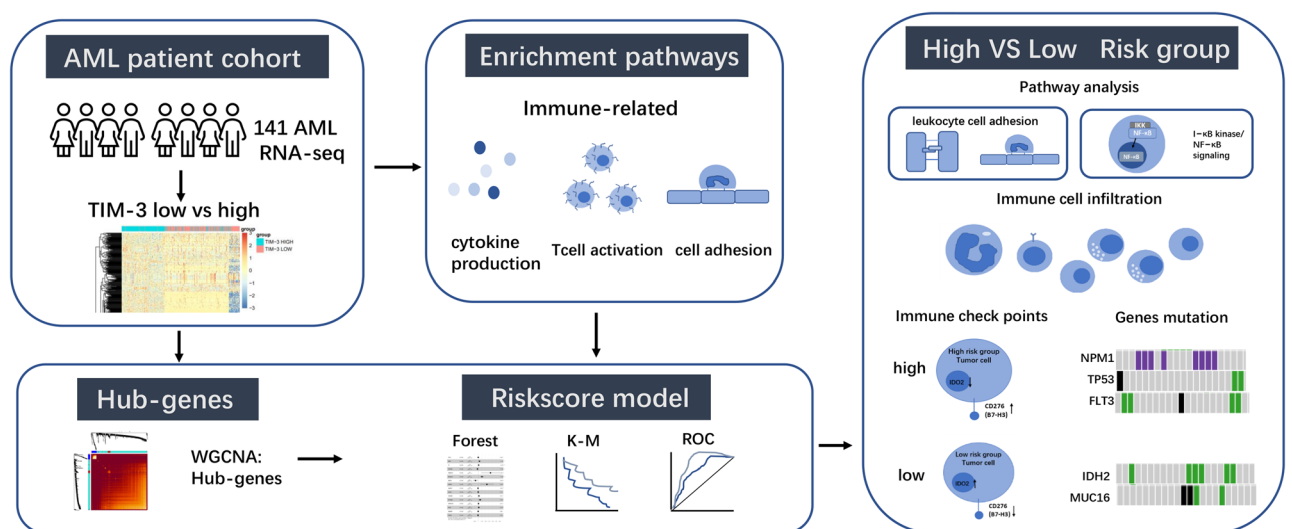
Moreover, *Tim-3* serves as an immune checkpoint and plays a critical role in immune responses during AML<sup>14</sup>. Therefore, *Tim-3* holds promise as a potential therapeutic marker in AML.

Though a study reveals the expression of *Tim-3* may affect prognosis in AML patients<sup>13</sup>. The precise mechanisms through which *TIM-3* impacts the prognosis of individuals with AML are currently lacking in comprehensive understanding. The therapeutic potential and underlying mechanisms of *Tim-3* in AML necessitate further investigation. Despite extensive research aimed at identifying prognostic markers in AML, a validated biomarker capable of accurately predicting response to immunotherapy and overall survival (OS) has yet to be identified. This emphasizes the imperative of discovering a prognostic biomarker specifically for immunotherapy in AML<sup>15</sup>. Investigating the hub genes associated with varying levels of *TIM-3* expression presents an opportunity to gain insights into the underlying mechanisms of *TIM-3* and identify prognostic markers in AML. However, there is no prognostic model based on level of *TIM-3* expression. In our study, we employed weighted gene co-expression network analysis (WGCNA) to identify hub genes in the high and low *TIM-3* expression groups. Subsequently, we developed a risk prognostic model based on the identified hub genes. The risk prognostic model has extraordinary ability to predict prognosis in both the training and validation sets, closing to European Leukemia Net (ELN) recommendations in decision curve analysis (DCA). The high-risk group exhibits a poor prognosis associated with mutations in *NPM1*, *TP53* (Multiple Hit), and *FLT3* (Multiple Hit), whereas the low-risk group is characterized by mutations in *IDH2* (Missense Mutation) and *MUC16* (Multiple Hit / Missense Mutation). Leukocyte cell–cell adhesion, regulation of T cell activation, and I- $\kappa$ B kinase/NF- $\kappa$ B signaling are enriched in the high-risk group, and these processes are involved in the anchoring of HSCs or LSCs to the bone marrow (BM), implicating their role in LSC survival and chemotherapy resistance. Additionally, *B7-H3* (*CD276*) and *IDO2* may serve as potential immune targets in the high-risk group. The risk score model demonstrated superior predictive power in AML with high *TIM-3* expression. The prognostic models built upon these hub genes have the potential to offer enhanced predictive insights (Fig. 1).

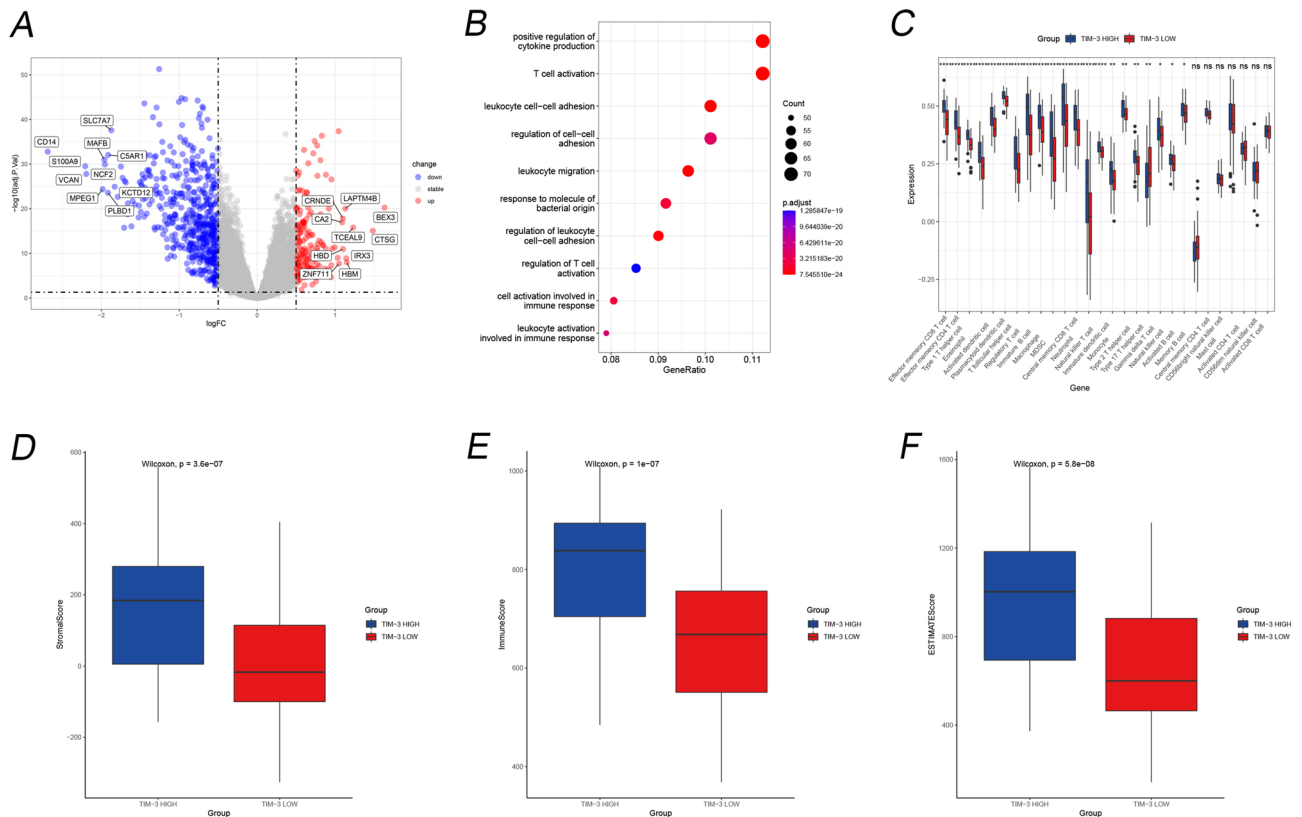
## Results

**DEGs of *TIM-3* expression.** In the analysis of differential gene expression between high and low *TIM-3* expression groups, a total of 7652 genes showed significant differential expression (Fig. 2A). Supplementary Fig. S1A provides the corrected data for the top 20 DEGs. Following the analysis of enrichment pathways for the DEGs (Fig. 2B), we identified the top three pathways: positive regulation of cytokine production, T cell activation, and leukocyte cell–cell adhesion. Building upon these discoveries, we subsequently examined the immune cells and scores of the immune microenvironment in the two groups exhibiting high and low *TIM-3* expression. We detected disparities in the manifestation of diverse T cells between the groups displaying high and low *TIM-3* expression (Fig. 2C). Additionally, the immune scores varied among the distinct groups, with increased estimateScores, StromalScore, and immuneScores discerned in the group exhibiting high *TIM-3* expression. Remarkably, immune-associated genes held a crucial position among the DEGs linked to *TIM-3* manifestation (Fig. 2D–F).

**Immune-related hub genes.** To gain comprehensive insights into immune-related hub genes, we performed WGCNA on the set of candidate genes. By employing WGCNA, we aimed to elucidate the intricate relationships and co-expression patterns among these genes, thereby facilitating a deeper understanding of their involvement in immune processes. The ideal soft-thresholding exponent was established as 5 (Supplementary Fig. S1 B–C). Subsequently, we detected a cumulative of 11 modules (Fig. 3A). By assessing the Pearson correlation coefficient between each module and the sample feature, we noticed that the blue and turquoise



**Figure 1.** Flowchart of this article.



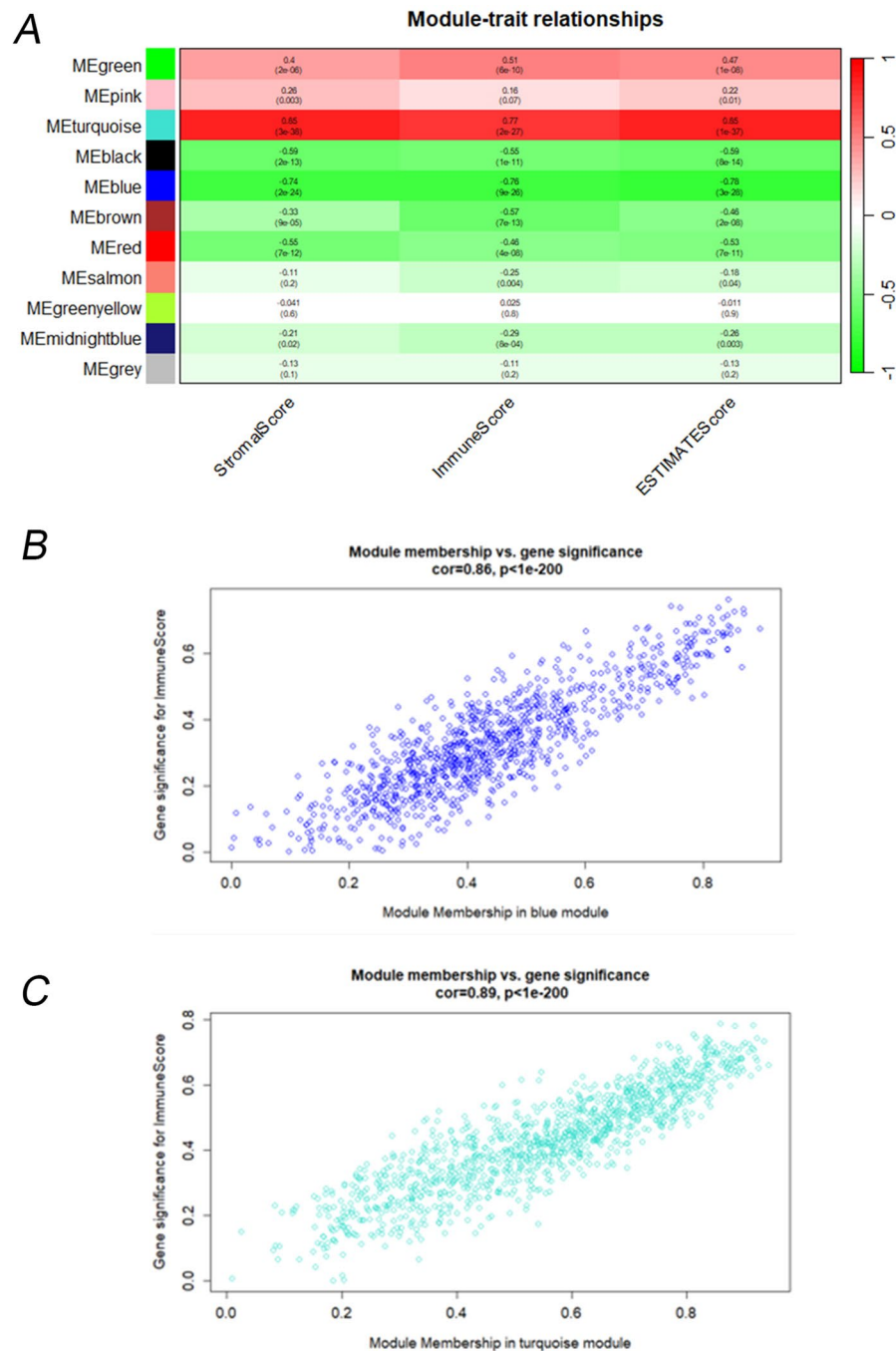
**Figure 2.** Exploration based on the high and low expression of *TIM-3*. (A) The Volcano plot of DEGs based on expression of *TIM-3*. The top 10 genes for up- and down-regulation are labeled separately. (B) The pathway of DEGs. (C) Immune cells infiltration between high vs low group in expression of *TIM-3*. (D) StromalScore in high versus low group based on *TIM-3* expression. (E) ImmuneScore in high versus low group based on *TIM-3* expression. (F) ESTIMATEScore in high versus low group based on *TIM-3* expression.

module demonstrated a robust correlation with immune scores. Consequently, we chose the genes within two modules as hub genes for subsequent analysis (Fig. 3B,C).

**Construction of risk score model.** To identify independent prognostic genes, we conducted multivariate survival analysis (Kaplan–Meier,  $p < 0.05$ ) on immune-related hub genes, specifically examining overall survival (OS). Among the hub genes, we identified significant OS influencers, which were utilized to develop a model through multivariate Cox regression analysis. Subsequently, we selected 22 genes from the immune-related hub genes to create the risk score model (Fig. 4A). This model served as a prognostic index for all cancer samples, computed using a formula (Supplementary Table 1). By employing the median risk score as the cut-off, patients with a low-risk score displayed superior overall survival compared to those with a high-risk score ( $P = 0.00031$ , Fig. 4B). Additionally, to fully validate the reliability of the model, we chose different databases: GSE71014 (Fig. 4D,E), GSE12417 (Fig. 4E,G), GSE146173 (Fig. 4H,I) as external validation. This included samples of cytogenetically normal AML (CN-AML) and secondary AML, treatment-resistant AML, with a view to approximating real-world. Summary of patients' clinical characteristics from 4 datasets was shown in Table 1. The results exhibited consistent results with the TCGA dataset, confirming that patients with a low-risk score experienced significantly better prognosis than those with a high-risk score. Moreover, the receiver operating characteristic (ROC) curves in both the training and validation sets demonstrated the robust predictive prognostic capability of the risk score model at 1, 3, and 5-year survival durations, particularly at the 5-year mark (Fig. 4C,E).

Correlation analysis of clinical parameters results showed that the risk score was correlated with patients' Gene mutation (*FLT3*, *IDH1*, *NPM1c*), no significant correlation was found between risk score and other clinical parameters (Table 2). Multivariate survival analysis found that risk prognostic model was the independent prognostic factor for AML patients (Table 3).

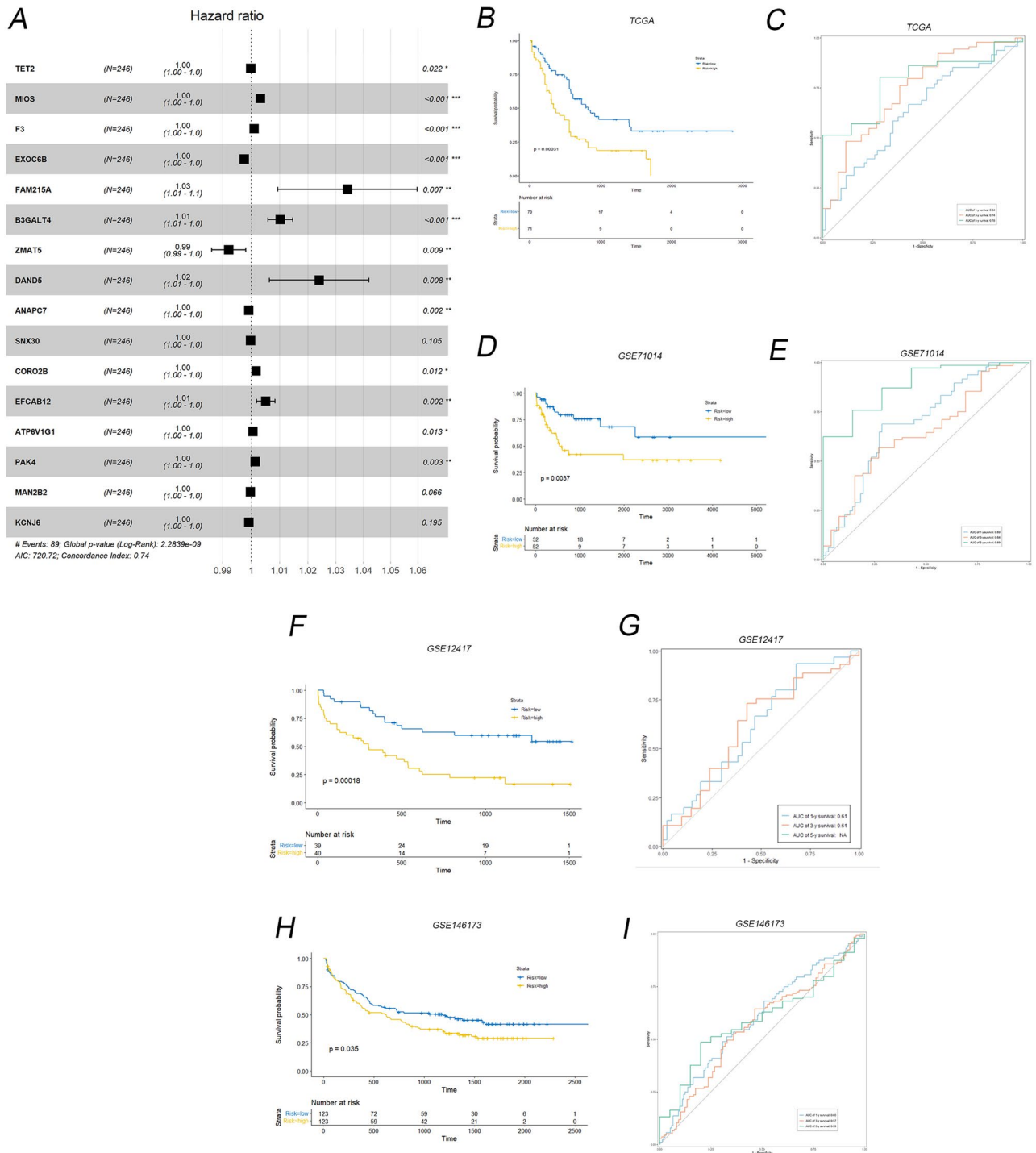
**Molecular characteristics of different risk score model.** GO, GSEA was performed to determine the gene sets enriched in different risk groups. The DEGs of different risk groups enrich in leukocyte cell–cell adhesion, regulation of T cell activation and I- $\kappa$ B kinase/NF- $\kappa$ B signaling (Fig. 5A). The results of GSEA are consistent with the result of functional enrichment analysis (Fig. 5B–D). Next, we analyzed gene mutations to gain further biological insight into the risk score groups. We found *NPM1* (Frameshift insertions), *TP53* (Multiple



**Figure 3.** Immune-related hub genes based on DEGs of expression of *TIM-3* through WGCNA. (A) Module trait relationship. (B) The correlation between module membership in blue module and ImmuneScore. (C) The correlation between module membership in turquoise module and ImmuneScore.

Hit) present with high-risk group, while *IDH2* (Missense Mutation), *MUC16* (Multiple Hit/Missense Mutation) in low-risk group (Fig. 5E). We also found that *DNMT3A* occurs frameshift insertions and nonsense mutation in the high-risk group, while multiple hit and splice site in the low-risk group; *RUNX1* occurs multiple hit in the high-risk group, while nonsense mutation in the low-risk group; *FLT3* occurs multiple hit in the high-risk group and *DNAH11* occurs (Frameshift Deletion) in the low-risk group. The Protein-Protein Interaction (PPI) Networks analysis for DEGs (high vs low risk groups) were shown in Fig. 5F,G.

**Immune characteristics of different risk score model.** After analyzing the composition of immune cells in different risk groups, we found that Monocyte, CD56dim natural killer cell, Type1 T helper cell, central memory CD8, central memory CD4 T cells and natural killer (NK) cells were more abundant in the high-risk subgroup (Fig. 6A).



**Figure 4.** Construction and validation of risk score model. (A) 22 genes in immune-related hub genes to form the risk score model. (B) Kaplan Meier (K–M) plot of TCGA based on high versus low risk group. (C) Receiver operating characteristic (ROC) curves of TCGA. (D) K–M plot of GSE71014 based on high versus low risk group. (E) ROC curves of GSE71014. (F) K–M plot of GSE12417 based on high versus low risk group. (G) ROC curves of GSE12417. (H) K–M plot of GSE146173 based on high versus low risk group. (I) ROC curves of GSE146173.

Next, we explore the relationship between risk group and immune checkpoint (IC) expression. As a result, the following seven ICs differentially expressed in different risk groups *LAIR1*, *CD276*, *CD70*, *CD160*, *ADORA2AM*, *IDO2*, *FGL1* (Fig. 6B). Among them, *CD276* and *IDO2* were the most statistically significant, with *CD276* highly expressed in the high-risk group and *IDO2* highly expressed in the low-risk group ( $P < 0.01$ , Fig. 6C).

Patient's parameters	TCGA	GSE71014	GSE146173	GSE12417
Age, years/n(%)				
< 65	98(69.5)	–	172(69.9)	47(59.5)
≥ 65	43(30.5)	–	74(30.1)	32(40.5)
WBC, × 10 <sup>9</sup> /L/n(%)				
< 10	53(37.6)	–	103(41.9)	–
≥ 10	87(61.7)	–	143(58.1)	–
NA	1(0.7)	–	–	–
PB blasts, %/n(%)				
< 40	16(11.3)	–	–	–
≥ 40	125(88.7)	–	–	–
BM blasts, %/n(%)				
< 50	88(62.4)	–	–	79(100)
≥ 50	53(37.6)	–	–	–
Risk/n(%)				
Favorable	30(21.3)	–	65(26.4)	–
Intermediate/normal	77(54.6)	–	127(51.6)	–
Poor	32(22.7)	–	46(18.7)	–
NA	2(1.4)	–	8(3.2)	–
OS/n(%)				
Alive	52(36.9)	68(65.4)	97(39.4)	33(41.8)
Dead	89(63.1)	36(34.6)	149(60.6)	46(58.2)
Survival rates, year/(%)				
1	56	52.9	62.2	59.5
3	18.4	24	41.1	26.6
5	5	14.4	8.1	–
AUC, year/(%)				
1	0.64	0.69	0.60	0.61
3	0.74	0.64	0.57	0.61
5	0.78	0.89	0.59	–
Platform	–	GPL10558	IlluminaHiSeq 1500	GPL570
Total	141	104	246	79

**Table 1.** Summary of patients' clinical characteristics from 4b datasets.

**Nomogram and DCA of risk score model.** The DCAs demonstrated that the performance of the nomogram closely resembled that of the French American British (FAB) classification or European Leukemia Net (ELN) recommendations in AML patients (Fig. 6D). Alongside the risk score, there exist multiple well-known prognostic factors for genes, including *FLT3* mut, *NPM1*, and *IDH2*. To predict 1-, 3-, and 5-year overall survival (OS), we developed a nomogram by incorporating the risk signature and the aforementioned genes. Within the nomogram, each signature was assigned points based on its risk contribution to OS (Fig. 6E).

## Discussion

AML is a hematological malignancy originating from hematopoietic stem/progenitor cells. The heterogeneous nature of AML is characterized by a wide range of clinical manifestations and prognoses. Notably, advancements in AML chemotherapy response have been made by targeting specific markers, particularly those expressed on LSCs or associated signaling pathways. Consequently, investigating leukemia markers, specifically those expressed on LSCs, is a crucial area of ongoing research. Among the recognized surface markers on LSCs, *Tim-3* has emerged as a significant player in AML progression<sup>9,10</sup>. Meta-analyses exploring *Tim-3* have revealed a correlation between *Tim-3* overexpression and an unfavorable prognosis across various cancer types<sup>16</sup>. However, further investigation is required to understand the clinical and biological characteristics of *Tim-3* specifically in the context of AML. A recent study highlighted the potential significance of upregulated *Tim-3* as a prognostic marker, indicating an unfavorable prognosis for individuals diagnosed with AML. Moreover, the biological properties of *Tim-3* were found to be associated with immune responses and signaling pathways involved in the regulation of LSCs<sup>13</sup>. The identification of hub genes displaying differential expression levels in relation to *Tim-3* may offer valuable opportunities to deepen our understanding of prognosis and the role of LSCs within the Tumor Immune Microenvironment (TME) of AML.

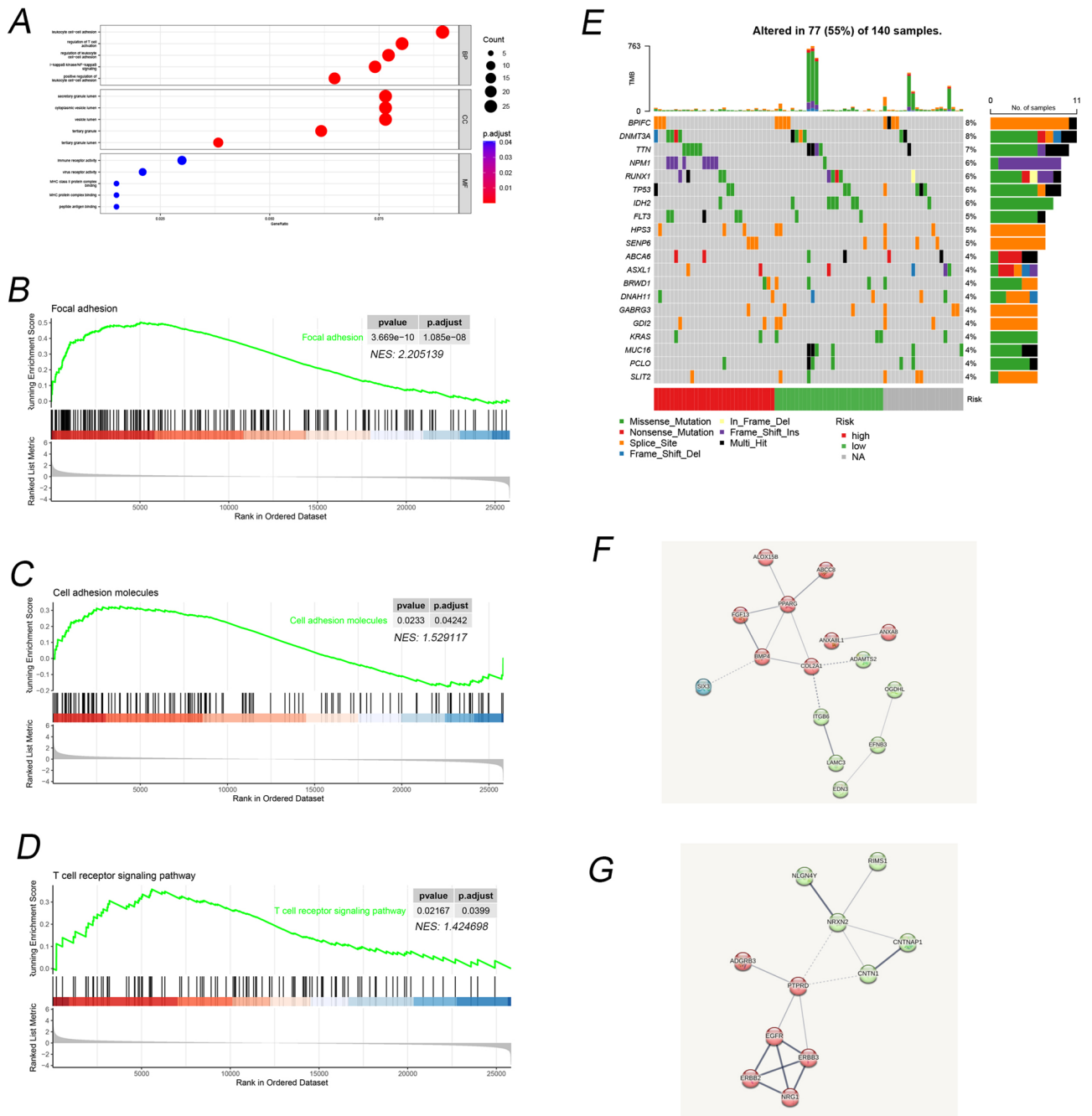
In our study, DEGs grouped by high and low *TIM-3* expression were enriched in the following 3 pathways: positive regulation of cytokine production, T cell activation, leukocyte cell–cell adhesion. We then explored the immune cells and immune microenvironment scores in the two groups with high and low *TIM-3* expression and found that they differed in both groups. Based on our conjecture that variations in the TME play a dominant

Patient's parameters	Total	Low (n = 70)	High (n = 71)	P*
Age, years/n(%)				
< 65	98(69.5)	50(61.4)	48(46.5)	0.7565
≥ 65	43(30.5)	20(38.6)	23(53.5)	
WBC, × 10 <sup>9</sup> /L/n(%)				
< 10	53(37.6)	28(40)	25(35.2)	0.7275
≥ 10	87(61.7)	42(60)	45(63.4)	
NA	1(0.7)	–	1(1.4)	
PB blasts, %/n(%)				
< 40	16(11.3)	10(14.3)	6(8.5)	0.4084
≥ 40	125(88.7)	60(85.7)	65(91.5)	
BM blasts, %/n(%)				
< 50	88(62.4)	47(67.1)	41(57.7)	0.3281
≥ 50	53(37.6)	23(32.9)	30(42.3)	
Risk/n(%)				
Favorable	30(21.3)	19(26.8)	11(15.5)	0.1917
Intermediate/normal	77(54.6)	34(48.6)	43(60.6)	
Poor	32(22.7)	17(24.3)	15(21.1)	
NA	2(1.4)	–	2(2.8)	
FAB classifications/n(%)				
M0	14(9.9)	8(11.4)	6(8.5)	0.7437
M1	32(22.7)	17(24.3)	15(21.1)	
M2	34(24.1)	14(20)	20(28.2)	
M3	14(9.9)	8(11.4)	6(8.5)	
M4	28(19.9)	15(21.4)	13(18.3)	
M5	15(10.7)	5(7.1)	10(14)	
M6	2(1.4)	1(1.4)	1(1.4)	
M7	1(0.7)	1(1.4)	–	
NA	1(0.7)	1(1.4)	–	
Gene mutation				
FLT3	41(29)	14(20)	27(38)	0.01851
IDH1	26(18.4)	18(25.7)	8(11.3)	0.0461
NPM1c	33(23.4)	10(14.3)	23(32.4)	0.01927
OS/n(%)				
Alive	52(36.9)	36(51.4)	16(22.5)	< 0.001
Dead	89(63.1)	34(48.6)	55(77.5)	

**Table 2.** Comparison of TCGA patients' clinical characteristics in two groups. WBC, white blood cell; BM, bone marrow; PB, peripheral blood; NA, not applicable. \*Chi-square test.

Variable	Univariate analysis			Multivariate analysis		
	Coef	HR(95%CI)	P	Coef	HR(95%CI)	P
Risk model	2.7	1.9–3.9	< 0.001	3.2	1.8–5.5	< 0.001
Age(≥ 65 vs. < 65)	2.9	1.9–4.5	< 0.001	2.4	1.5–3.7	< 0.001
WBC(≥ 10 × 10 <sup>9</sup> /L vs < 10)	4.6	1.1–19	0.033	2.7	0.7–11.2	0.16
PB blasts(≥ 40 vs. < 40)	1.3	0.68–2.4	0.44	–	–	–
BM blasts(≥ 50 vs. < 50)	1.3	0.86–2	0.21	–	–	–
ELN risk stratification	1.8	1.3–2.5	< 0.001	1.696	1.2–2.4	0.004
Gene mutation						
FLT3 (mutated vs. wild)	1.3	0.86–2.1	0.19	–	–	–
IDH1 (mutated vs. wild)	0.72	0.31–1.7	0.44	–	–	–
NPM1c (mutated vs. wild)	1.1	0.22	0.64	–	–	–

**Table 3.** Univariate and multivariate overall survival analysis in the TCGA. WBC, white blood cell; BM, bone marrow; PB, peripheral blood.

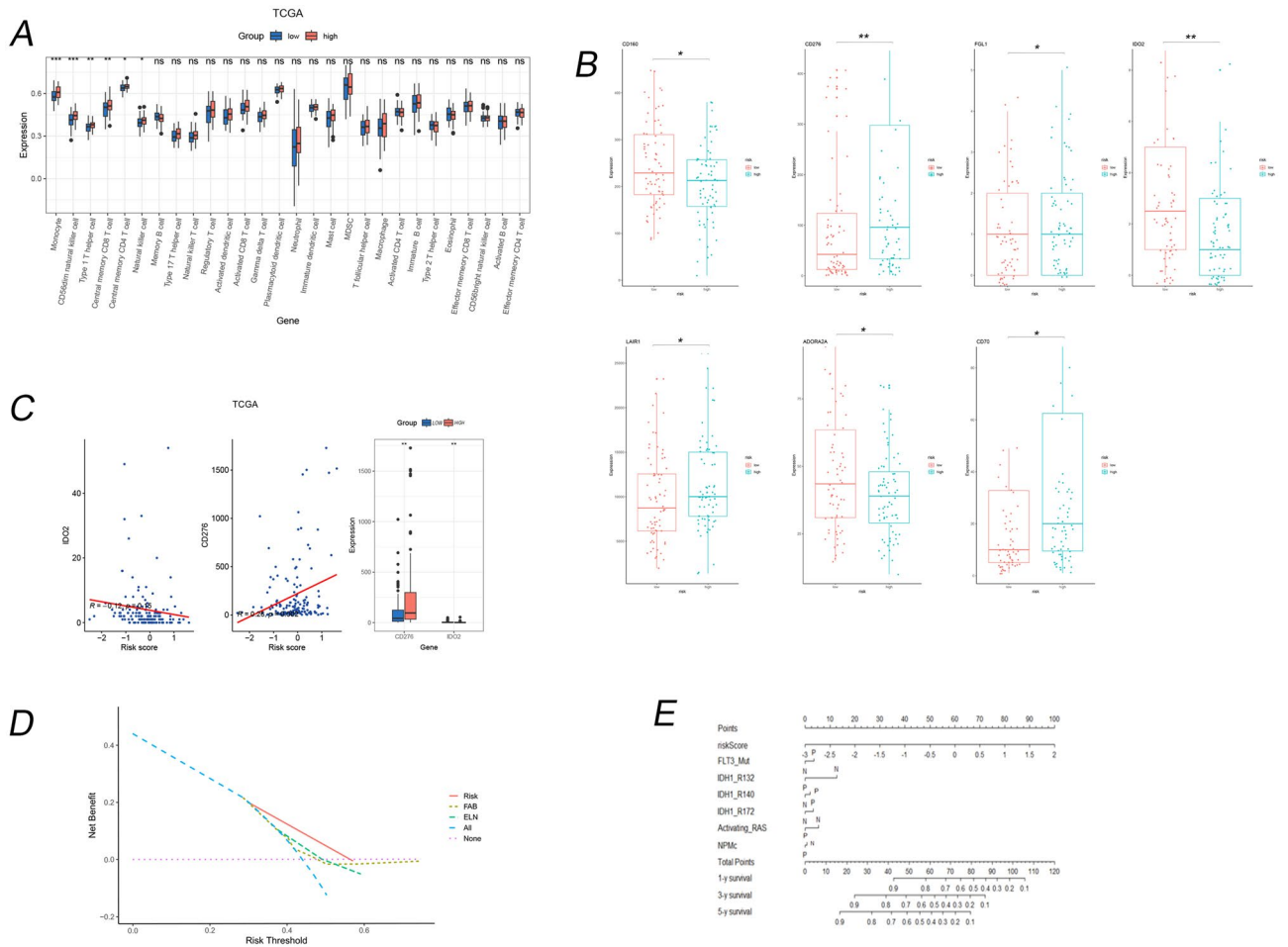


**Figure 5.** Molecular characteristics of different risk score model. (A) The pathway of DEGs based on high vs low risk group. (B) Gene set enrichment analysis (GSEA) of focal adhesion. (C) GSEA of cell adhesion molecules. (D) GSEA of T cell receptor signaling pathway. (E) Oncoplot based on different risk groups. (F) PPI network of up genes in DEGs’ high versus low risk group. (G) PPI network of down genes in DEGs high versus low risk group. Each node represents each protein. The thickness of the line represents the strength of the association between the proteins.

role in the high and low expression groups, our focus shifted towards identifying hub genes associated with the immune microenvironment. By exploring the immune-related gene expression profiles, we aimed to pinpoint genes that potentially exert a significant influence on the TME.

TME is influenced by a multitude of genes, and the use of WGCNA provides a valuable approach for identifying potential immune-related biomarkers<sup>17</sup>. In our study, we employed WGCNA using TCGA datasets to identify hub genes within two immune-related modules. Subsequently, we filtered these genes based on their impact on patient overall survival (OS). By utilizing these identified genes, we developed a risk score model that incorporates their expression levels. The constructed risk score model emerged as a robust prognostic biomarker for AML. It exhibited favorable survival outcomes in the low-risk group and poorer survival outcomes in the high-risk group, as observed across both the TCGA and GEO cohorts. Moreover, the risk score model





**Figure 6.** Immune characteristics of different risk score model. (A) The expression of immune cells in different risk groups. (B) Statistically differentially expressed immune checkpoint (IC) in different risk groups. A total of seven IC differed in expression levels between high and low risk groups. (C) Correlation of *IDO2* and *CD275* with risk model, expression in high and low risk groups respectively. (D) Decision curve analysis (DCA) plot. (E) The nomogram based on risk score model and other prognosis-related genes.

demonstrated notable efficacy in predicting survival rates at 1, 3, and 5 years, with a particular emphasis on the 5-year survival endpoint.

Following the categorization of samples into high and low-risk groups based on their respective risk scores, we proceeded to conduct pathway analysis on the DEGs within each group. Our analysis revealed significant enrichment of pathways related to leukocyte cell–cell adhesion, regulation of T cell activation, and I-κB kinase/NF-κB signaling. Notably, these pathways exhibited substantial enrichment in the high-risk group, indicating their potential involvement in disease progression and prognosis. The role of adhesion is to regulate cellular connections and their interaction with the extracellular matrix, thereby influencing the localization of cells in their respective environments. In AML, the processes of quiescence, migration, and adhesion within the bone marrow (BM) exert significant influence<sup>18</sup>. Adhesion molecules involved in the interaction between HSCs or LSCs and the BM microenvironment have been implicated in LSC survival and resistance to chemotherapy<sup>19,20</sup>. The adhesion-dependent survival of LSCs in AML patients contributes to the development of treatment resistance and reduced overall survival. Therefore, targeting the adhesion process holds potential for therapeutic interventions.

The immune microenvironment of the BM in AML patients exhibits distinct characteristics compared to that of healthy individuals. Notably, activated regulatory T cells (Treg) are present in the immune microenvironment of AML, promoting the emergence of suppressive subsets<sup>21</sup>. High expression of Treg in AML impacts the response to induction chemotherapy and contributes to synergistic and amplified drug resistance<sup>22</sup>. Primary AML CD34(+) cells exhibit detectable NF-kappaB activity<sup>23</sup>. Furthermore, LSCs aberrantly express NF-kappaB, rendering it a potential therapeutic target in AML<sup>23,24</sup>. Interestingly, the stimulation of *TIM-3* by Gal-9 has been found to co-activate signaling pathways for LSC self-renewal, specifically NF-κB and β-catenin<sup>9</sup>. The hub genes associated with *TIM-3* expression in our risk model suggest the involvement of the NF-κB signaling pathway in the AML microenvironment. Furthermore, we performed PPI analyses of DEGs in the high- and low-risk groups and found that FGF13 and BMP4, PPARG in up gene network, NRXN2, NLGN4Y,CNTN1, EGFR,ERBB2,ERBB3 in down gene network.

To further investigate the immunological characteristics associated with the risk groups, we conducted an analysis of gene mutations within each group. Correlation analysis of clinical parameters results showed that the risk score was correlated with patients' gene mutation (*FLT3*, *IDH1*, *NPM1c*). Our findings revealed the presence of the *NPM1* mutation, which is known to confer a favorable prognostic effect in AML, within the high-risk group. Interestingly, we also observed the presence of the *FLT3* mutation within the high-risk group. *FLT3* mutations are the most commonly observed genetic aberrations in AML and are associated with a poor prognosis<sup>25</sup>. This suggests that the poor prognosis observed in the high-risk group cannot be reversed by the presence of the *NPM1* mutation<sup>26,27</sup>. Mutations in *TP53* were found to be more frequent in the high-risk group. AML patients with *TP53* mutations have been associated with poor outcomes, and the specific class-defining mutations have an independent and additive impact on prognosis<sup>27</sup>. *TP53* mutations have been incorporated into the risk stratification guidelines for AML recommended by the 2017 European Leukemia Net (ELN)<sup>28</sup>.

Furthermore, we observed variations in the prevalence of gene mutations between the low-risk and high-risk groups. Specifically, *IDH2* mutations were more prevalent in the low-risk group, suggesting their potential as therapeutic targets to improve overall survival in AML patients within this group. We specifically identified an *IDH2* mutation within the low-risk group. These *IDH* mutations disrupt the differentiation process in AML cells by causing abnormal epigenetic regulation<sup>29</sup>. Encouragingly, substances such as enasidenib, which serve as inhibitors of *IDH2*, have shown promising remission rates and have received approval from regulatory authorities in the United States<sup>30</sup>. *MUC16*, also known as *CA125*, exhibits distinct expression patterns under normal physiological conditions as well as during tumorigenesis. While *MUC16* plays a protective role in healthy physiology, it has been implicated in disease progression and metastasis in various cancers. Its abnormal overexpression makes it an appealing target for diagnostic purposes and immune therapy<sup>31,32</sup>.

TME can provide valuable insights for the development of innovative strategies to treat AML and improve the effectiveness of immunotherapies. We observed differences in the composition of immune cells between the two risk groups. Specifically, the high-risk subgroup exhibited higher levels of monocytes, CD56dim natural killer cells, Type 1T helper cells, central memory CD8 T cells, central memory CD4 T cells, and natural killer (NK) cells. These cell types are involved in the inflammatory response. Chronic inflammation has been associated with the development of premature aging phenotypes and myeloid neoplasms. Soluble cytokines such as tumor necrosis factor (TNF), interferons (IFNs), and interleukin 6 (IL6) have been linked to hematologic neoplasms related to aging, including myeloproliferative neoplasms (MPN), myelodysplastic syndromes, and AML<sup>33–35</sup>. In a recent study, individuals who showed positive clinical responses to immunotherapy involving αCTLA-4 and/or αPD-1 antibodies along with a hypomethylating agent (HMA) demonstrated a significant increase in the population of central memory (CM) CD8+ T cells<sup>36</sup>.

The advent of immune checkpoint inhibitors has revolutionized the treatment of previously untreatable malignancies in cancer immunotherapy. In this study, we aimed to explore the relationship between established immunotherapy prediction biomarkers and the risk group classification. Our findings revealed a significant correlation between the risk group and two immune checkpoints: *IDO2* and *CD276*. *CD276*, also known as *B7-H3*, belongs to the *B7* family of immune checkpoint molecules. High expression of *B7-H3* was found to be associated with poor overall survival in AML patients<sup>37</sup>. Our study supports these observations by showing a positive association between *CD276* expression and the risk score. Additionally, a significant proportion of AML patients exhibit elevated levels of *B7-H3* on leukemic blasts, making it a promising target antigen for the development of CAR-T therapy specifically directed towards AML<sup>38</sup>. Evidence suggests that *IDO2* exhibits lower efficiency in metabolizing l-tryptophan compared to *IDO1*. Instead, its activities are believed to be influenced by interactions with other yet unidentified proteins, which may vary in different inflammatory and neoplastic contexts. Consequently, identifying the interactome and function of *IDO2* in different neoplastic conditions may pave the way for the development of novel treatment approaches<sup>39,40</sup>. *CD276* and *IDO2* exhibit promise as viable targets for immunotherapeutic interventions in AML.

Minimal residual disease (MRD) should indeed be taken into account as a prognostic index. Unfortunately, however, MDR was not found in the corresponding clinical features of the TCGA and several other GEO datasets. However, comparing the blasts of PB and BM in the high- and low-risk groups suggested that their high values were independently associated with the high- and low-groups. Blast of PB and BM may have less effect on high and low risk. Besides, Multivariate survival analysis found that risk prognostic model was the independent prognostic factor for AML patients. To further understand the predictive ability of the model for prognosis, we compare it with the two standard scenarios of ELN recommendations and FAB classification in DCA curves. We can easily find that the risk prognosis model is closer to the curves of these two scenarios, which demonstrates that the risk score model has a better ability to predict prognosis. In addition to the risk score, there are numerous known prognostic factors for genes, so we selected the genes in ELN recommendations for the nomogram construction. This nomogram combined with genes based on ELN recommendations can better predict the prognosis of AML patients.

Risk score group based on the level of *TIM-3* expression is a promising immune-related prognostic model. We found that the high-risk group is involved in the process of cell adhesion and the process of inflammatory response, such as migration, homing, and quiescence of LSCs, which plays an important role in AML. The results implied that high-risk group was characteristics of active tumor progression. The two immune checkpoints: *IDO2* and *B7-H3* would be the potential immune therapy targets. The risk score model demonstrated superior predictive power in AML with high *TIM-3* expression. The utilization of a risk score model exhibits potential in distinguishing the immune and molecular characteristics of patients, thereby enabling the prediction of patient outcomes. Further studies are needed to clarify this assertion.

## Methods

**Patients and datasets.** In order to obtain clinical insights pertaining to AML, a cohort consisting of 140 cases was examined through the utilization of RNA sequencing (RNA-seq) data and gene mutation information. The database of choice for this investigation was the Cancer Genome Atlas (TCGA). Furthermore, to fully validate the reliability of the model, we chose different Gene Expression Omnibus (GEO) databases as external validation (GSE71014, GSE12417, GSE146173). This included samples of cytogenetically normal AML (CN-AML) and secondary AML, treatment-resistant AML, with a view to approximating real-world.

**Functional enrichment analysis.** To identify a collection of differentially expressed genes (DEGs), a rigorous differential expression analysis was performed. Subsequently, the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses<sup>41</sup>, widely recognized methodologies in the field, were conducted. The clusterProfiler package, implemented in the R programming language, served as the tool of choice for these analyses.

**Protein–Protein interaction networks.** We added Protein–Protein Interaction Networks analysis for DEGs in high and low risk groups through STRING websites.

**Evaluation of immune cell fractions.** Immunosuppression within the microenvironment of AML is thought to involve a range of cell types, including myeloid-derived suppressor cells (MDSCs), regulatory T cells (Tregs), natural killer (NK) cells, neutrophils, and various other cells. To investigate the composition of immune cells mentioned above, an analytical approach known as single-sample gene set enrichment analysis (ssGSEA) was employed. This approach utilizes the R package "GSVA" to quantify the extent of immune cell enrichment within the microenvironment, based on gene expression levels obtained from a single tumor sample<sup>42</sup>. Furthermore, the R package "estimate" was utilized to evaluate the ImmuneScore, a measure indicative of immune cell infiltration within the tumor microenvironment.

**Identification of hub genes.** By employing RNA-seq data obtained from AML samples sourced from the TCGA database, we conducted an investigation to identify differentially expressed genes based on the categorization of TIM-3 expression levels into "High" and "Low" groups. The identification of such genes was accomplished using the R package "DESeq" with statistical significance defined as  $P < 0.05$  and  $|\log_2FC| > 2$ <sup>43</sup>.

To explore the relationship between gene expression patterns and Gene Ontology (GO) findings, we employed WGCNA<sup>17</sup>. This analysis focused on identifying hub genes within immune cells and the microenvironment. Initially, a similarity matrix was computed, which was subsequently transformed into an adjacency matrix. To facilitate this transformation, a signed network type was utilized along with a soft threshold of 5. The resulting adjacency matrix was further transformed into a topological matrix utilizing the topological overlap measure (TOM) to quantify the degree of gene linkage. Modules were identified through the implementation of a dynamic pruning tree, which grouped genes based on the 1-TOM distance. Through this process, a total of 11 modules were identified by adjusting the merging threshold function to 0.25.

Following the completion of the WGCNA analysis, we focused on genes belonging to modules that exhibited significant associations with ImmuneScore and ESTIMATEScore. Specifically, our attention was directed towards the "blue" and "turquoise" modules. To construct a network, edges with a weight greater than 0.4 were considered, taking into account the genes from these significantly related modules. Subsequently, we narrowed our analysis to hub genes within these modules that demonstrated significant Kaplan–Meier (K–M) survival analysis results. This rigorous selection process resulted in the identification of 16 immune-related hub genes that displayed significant associations with survival ( $P < 0.05$ , log-rank test) and were thus chosen for further analysis.

**Construction and validation of the prognostic risk-scoring model.** In order to identify hub genes that significantly influenced overall survival (OS), we employed multivariate Cox regression analysis. These hub genes were then utilized to construct a model. To generate the most optimal model, we utilized the R package "My.stepwise" for screening hub genes associated with OS. The resulting model was defined by the following formula:  $(\text{Riskscore} = \sum_{i=1}^N (\text{NCoef}_i \times \text{Exp}_i))$ . To evaluate the predictive capability of the model, Kaplan–Meier survival curves and log-rank tests were performed on both the TCGA and GEO cohorts. Furthermore, time-dependent ROC curve analyses were conducted using the R package "timeROC" to calculate the area under the curve (AUC) and compare the prognostic value among different risk groups.

**Comprehensive analysis of molecular and immune characteristics in different risk groups.** To gain insights into the gene expression profiles of samples with high ( $n = 71$ ) and low ( $n = 70$ ) risk scores, a differential expression analysis was initially conducted on all genes. This analysis utilized the R package "DESeq". Subsequently, gene set enrichment analysis (GSEA) was performed using the R package "clusterProfiler" and the KEGG database to identify signaling pathways associated with the differentially expressed genes. Significance was defined as  $P < 0.05$  and  $\text{FDR} < 0.25$ . Additionally, to delve further into representative gene sets, we conducted single-sample gene set enrichment analysis (ssGSEA) using the R package "GSVA". This analysis provided a deeper understanding of the functional enrichment of the identified gene sets.

To investigate the differences in immune cell composition between the two risk groups, we conducted a comparative analysis of the relative proportions of 22 types of immune cells. This analysis aimed to gain further insights into the immune and molecular functions associated with these risk groups. Additionally, we utilized

the R package “Maftools” to analyze gene mutations between the two groups, providing a comprehensive understanding of the genetic landscape.

In order to assess the prognostic value of the risk group in patients undergoing immunotherapy, we generated a correlation heatmap between immune checkpoint (IC) and risk group samples. This analysis allowed us to evaluate the relationship between the risk group and the expression of immune checkpoint genes, shedding light on potential therapeutic implications.

**Development and assessment of the nomogram.** To develop a personalized prediction model for clinical events, we employed a nomogram. A nomogram is a visual representation that utilizes statistical prediction models to provide a straightforward graphical representation of the probability of specific clinical outcomes. In our study, the creation of the nomogram involved the utilization of R packages such as “survival” and “rms”.

To assess the accuracy of the nomogram in predicting survival rates (1, 3, and 5-year) for patients with AML, we conducted decision curve analysis (DCA). This analysis compared the decision curve of the nomogram with those derived from other prognostic factors, enabling an evaluation of its performance and clinical utility.

**Statistical analysis.** To compare continuous variables between the two groups, an independent t-test was performed. The chi-square test was used for categorical data, where either predicted frequency was less than 5, using Fisher’s exact test. For univariate survival analysis, the Kaplan–Meier (K–M) method was employed. This method enabled the estimation of survival probabilities over time, and the log-rank test was used to assess the significance of differences in survival curves between the two groups. Cox regression was used for univariate and multivariate survival analyses. *P*-values were corrected using the Bonferroni correction method. The significance level for all statistical tests was set at a two-sided *P*-value of less than 0.05, indicating statistical significance.

**Ethics approval.** All data in the database passed ethical review and informed consent before being uploaded to TCGA. Ethics Approval is allowed by Ethics Committee of Shanghai Pudong Hospital. Our research is based on open-source data and is therefore free of ethical issues and other conflicts of interest.

### Data availability

In our study, TCGA (<https://portal.gdc.cancer.gov>) and GEO (<https://www.ncbi.nlm.nih.gov/geo/>) belong to public databases. The samples involved in the database have received ethical approval. Users are free to download relevant data for research and publication of relevant articles. All data in the database passed ethical review and informed consent before being uploaded to TCGA.

Received: 24 June 2023; Accepted: 13 September 2023

Published online: 20 September 2023

### References

- Vetrie, D., Helgason, G. V. & Copland, M. The leukaemia stem cell: Similarities, differences and clinical prospects in CML and AML. *Nat. Rev. Cancer* **20**, 158–173. <https://doi.org/10.1038/s41568-019-0230-9> (2020).
- Estey, E. H. Acute myeloid leukemia: 2019 update on risk-stratification and management. *Am. J. Hematol.* **93**, 1267–1291. <https://doi.org/10.1002/ajh.25214> (2018).
- Jones, C. L. *et al.* Inhibition of amino acid metabolism selectively targets human leukemia stem cells. *Cancer Cell* **35**, 333–335. <https://doi.org/10.1016/j.ccell.2019.01.013> (2019).
- Ho, T. C. *et al.* Evolution of acute myelogenous leukemia stem cell properties after treatment and progression. *Blood* **128**, 1671–1678. <https://doi.org/10.1182/blood-2016-02-695312> (2016).
- Marchand, T. & Pinho, S. Leukemic stem cells: From leukemic niche biology to treatment opportunities. *Front. Immunol.* **12**, 775128. <https://doi.org/10.3389/fimmu.2021.775128> (2021).
- Zhu, C. *et al.* The Tim-3 ligand galectin-9 negatively regulates T helper type 1 immunity. *Nat. Immunol.* **6**, 1245–1252. <https://doi.org/10.1038/ni1271> (2005).
- Sakuishi, K. *et al.* Targeting Tim-3 and PD-1 pathways to reverse T cell exhaustion and restore anti-tumor immunity. *J. Exp. Med.* **207**, 2187–2194. <https://doi.org/10.1084/jem.20100643> (2010).
- Pagliano, O. *et al.* Tim-3 mediates T cell trogocytosis to limit antitumor immunity. *J. Clin. Investig.* <https://doi.org/10.1172/JCI152864> (2022).
- Kikushige, Y. *et al.* A TIM-3/Gal-9 autocrine stimulatory loop drives self-renewal of human myeloid leukemia stem cells and leukemic progression. *Cell Stem Cell* **17**, 341–352. <https://doi.org/10.1016/j.stem.2015.07.011> (2015).
- Kikushige, Y. *et al.* TIM-3 is a promising target to selectively kill acute myeloid leukemia stem cells. *Cell Stem Cell* **7**, 708–717. <https://doi.org/10.1016/j.stem.2010.11.014> (2010).
- Jan, M. *et al.* Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci. Transl. Med.* **4**, 149ra118. <https://doi.org/10.1126/scitranslmed.3004315> (2012).
- Haubner, S. *et al.* Coexpression profile of leukemic stem cell markers for combinatorial targeted therapy in AML. *Leukemia* **33**, 64–74. <https://doi.org/10.1038/s41375-018-0180-3> (2019).
- Wu, Z. *et al.* Upregulation of Tim-3 is associated with poor prognosis in acute myeloid leukemia. *Cancer Med.* **12**, 8956–8969. <https://doi.org/10.1002/cam4.5549> (2023).
- Wang, Z., Chen, J., Wang, M., Zhang, L. & Yu, L. One stone, two birds: The roles of Tim-3 in acute myeloid leukemia. *Front. Immunol.* **12**, 618710. <https://doi.org/10.3389/fimmu.2021.618710> (2021).
- Prada-Arismendy, J., Arroyave, J. C. & Rothlisberger, S. Molecular biomarkers in acute myeloid leukemia. *Blood Rev.* **31**, 63–76. <https://doi.org/10.1016/j.blre.2016.08.005> (2017).
- Zang, K. *et al.* TIM-3 as a prognostic marker and a potential immunotherapy target in human malignant tumors: A meta-analysis and bioinformatics validation. *Front. Oncol.* **11**, 579351. <https://doi.org/10.3389/fonc.2021.579351> (2021).
- Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559. <https://doi.org/10.1186/1471-2105-9-559> (2008).

18. Gruszka, A. M., Valli, D., Restelli, C. & Alcalay, M. Adhesion deregulation in acute myeloid leukaemia. *Cells* <https://doi.org/10.3390/cells8010066> (2019).
19. Ashok, D., Polcik, L., Dannewitz Prosseda, S. & Hartmann, T. N. Insights into bone marrow niche stability: An adhesion and metabolism route. *Front. Cell Dev. Biol.* **9**, 798604. <https://doi.org/10.3389/fcell.2021.798604> (2021).
20. Grenier, J. M. P., Testut, C., Fauriat, C., Mancini, S. J. C. & Aurrand-Lions, M. Adhesion molecules involved in stem cell niche retention during normal haematopoiesis and in acute myeloid leukaemia. *Front. Immunol.* **12**, 756231. <https://doi.org/10.3389/fimmu.2021.756231> (2021).
21. Guo, R. *et al.* Single-cell map of diverse immune phenotypes in the acute myeloid leukemia microenvironment. *Biomark. Res.* **9**, 15. <https://doi.org/10.1186/s40364-021-00265-0> (2021).
22. Szczepanski, M. J. *et al.* Increased frequency and suppression by regulatory T cells in patients with acute myelogenous leukemia. *Clin. Cancer Res.* **15**, 3325–3332. <https://doi.org/10.1158/1078-0432.CCR-08-3010> (2009).
23. Guzman, M. L. *et al.* Nuclear factor-kappaB is constitutively activated in primitive human acute myelogenous leukemia cells. *Blood* **98**, 2301–2307. <https://doi.org/10.1182/blood.v98.8.2301> (2001).
24. Jin, Y. *et al.* Antineoplastic mechanisms of niclosamide in acute myelogenous leukemia stem cells: Inactivation of the NF-kappaB pathway and generation of reactive oxygen species. *Cancer Res.* **70**, 2516–2527. <https://doi.org/10.1158/0008-5472.CAN-09-3950> (2010).
25. Zhao, J. C. *et al.* A review of FLT3 inhibitors in acute myeloid leukemia. *Blood Rev.* **52**, 100905. <https://doi.org/10.1016/j.blre.2021.100905> (2022).
26. Dohner, K. *et al.* Mutant nucleophosmin (NPM1) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: Interaction with other gene mutations. *Blood* **106**, 3740–3746. <https://doi.org/10.1182/blood-2005-05-2164> (2005).
27. Papaemmanuil, E. *et al.* Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med.* **374**, 2209–2221. <https://doi.org/10.1056/NEJMoa1516192> (2016).
28. Dohner, H. *et al.* Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood* **140**, 1345–1377. <https://doi.org/10.1182/blood.2022016867> (2022).
29. McMurry, H., Fletcher, L. & Traer, E. IDH Inhibitors in AML-promise and pitfalls. *Curr. Hematol. Malign. Rep.* **16**, 207–217. <https://doi.org/10.1007/s11899-021-00619-3> (2021).
30. Bose, P., Vachhani, P. & Cortes, J. E. Treatment of relapsed/refractory acute myeloid leukemia. *Curr. Treat. Options Oncol.* **18**, 17. <https://doi.org/10.1007/s11864-017-0456-2> (2017).
31. Felder, M. *et al.* MUC16 (CA125): Tumor biomarker to cancer therapy, a work in progress. *Mol. Cancer* **13**, 129. <https://doi.org/10.1186/1476-4598-13-129> (2014).
32. Aithal, A. *et al.* MUC16 as a novel target for cancer therapy. *Expert Opin. Ther. Targets* **22**, 675–686. <https://doi.org/10.1080/1472822.2018.1498845> (2018).
33. Puissant, A. & Medyouf, H. Walking the tightrope: Balancing delicate inflammation response to eradicate acute myeloid leukemia. *Cancer Discov.* **12**, 1617–1619. <https://doi.org/10.1158/2159-8290.CD-22-0473> (2022).
34. Wu, J. *et al.* A single-cell survey of cellular hierarchy in acute myeloid leukemia. *J. Hematol. Oncol.* **13**, 128. <https://doi.org/10.1186/s13045-020-00941-y> (2020).
35. Mueller, S. N., Gebhardt, T., Carbone, F. R. & Heath, W. R. Memory T cell subsets, migration patterns, and tissue residence. *Annu. Rev. Immunol.* **31**, 137–161. <https://doi.org/10.1146/annurev-immunol-032712-095954> (2013).
36. Lee, S. E. *et al.* Immunologic predictors for clinical responses during immune checkpoint blockade in patients with myelodysplastic syndromes. *Clin. Cancer Res.* <https://doi.org/10.1158/1078-0432.CCR-22-2601> (2023).
37. Wang, P. *et al.* Optimal combination of immune checkpoint and senescence molecule predicts adverse outcomes in patients with acute myeloid leukemia. *Ann. Med.* **55**, 2201507. <https://doi.org/10.1080/07853890.2023.2201507> (2023).
38. Lichtman, E. I. *et al.* Preclinical evaluation of B7–H3-specific chimeric antigen receptor T cells for the treatment of acute myeloid leukemia. *Clin. Cancer Res.* **27**, 3141–3153. <https://doi.org/10.1158/1078-0432.CCR-20-2540> (2021).
39. Mondanelli, G. *et al.* Current challenges for IDO2 as target in cancer immunotherapy. *Front. Immunol.* **12**, 679953. <https://doi.org/10.3389/fimmu.2021.679953> (2021).
40. Platten, M., Nollen, E. A. A., Rohrig, U. F., Fallarino, F. & Opitz, C. A. Tryptophan metabolism as a common therapeutic target in cancer, neurodegeneration and beyond. *Nat. Rev. Drug Discov.* **18**, 379–401. <https://doi.org/10.1038/s41573-019-0016-5> (2019).
41. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287. <https://doi.org/10.1089/omi.2011.0118> (2012).
42. Hanzelmann, S., Castelo, R. & Guinney, J. GSEA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7. <https://doi.org/10.1186/1471-2105-14-7> (2013).
43. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106. <https://doi.org/10.1186/gb-2010-11-10-r106> (2010).

## Author contributions

Each author contributed equally to this article.

## Funding

The manuscript supported by the Project of Key Medical Specialty and Treatment Center of Pudong Hospital of Fudan University. (Grant No. MDT2021-02). Project Funded by the China National Natural Science Foundation (No. 32100628). Declaration of generative AI and AI-assisted technologies in the writing process. During the preparation of this work the authors used ChatGPT in order to check grammar and improve sentence fluency. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-42700-2>.

**Correspondence** and requests for materials should be addressed to N.Z. or Z.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023