

LINC00649 underexpression is an adverse prognostic marker in acute myeloid leukemia

Chao Guo

China-Japan Friendship Hospital

Ya-yue Gao

China-Japan Friendship Hospital

Qian-qian Ju

China-Japan Friendship Hospital

Chun-xia Zhang

China-Japan Friendship Hospital

Ming Gong

China-Japan Friendship Hospital

Zhen-ling Li (✉ lizhenling03@163.com)

<https://orcid.org/0000-0002-6623-1683>

Research article

Keywords: acute myeloid leukemia, long non-coding RNA, survival analysis, gene expression profile

Posted Date: August 27th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-39798/v3>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published on September 3rd, 2020. See the published version at <https://doi.org/10.1186/s12885-020-07331-0>.

Abstract

Background Long noncoding RNAs (lncRNA) play a role in leukemogenesis, maintenance, development, and therapeutic resistance of AML. While few studies have focused on the prognostic significance of LINC00649 in AML, which we aim to investigate in this present study. **Methods** We compared the expression level of LINC00649 between AML patients and healthy controls. The Kaplan-Meier curves of AML patients expressing high versus low level of LINC00649 was performed. The LINC00649 correlated genes/miRNAs/lncRNAs and methylation CpG sites were screened by Pearson correlation analysis with R (version 3.6.0), using TCGA-LAML database. The LINC00649 associated ceRNA network was established using lncBase 2.0 and miRWalk 2.0 online tools, combining results from correlation analysis. Finally, a prediction model was constructed using LASSO-Cox regression. **Results** LINC00649 was underexpressed in bone marrow of AML group than that in healthy control group. The patients of LINC00649-low group have significantly inferior PFS and OS. A total of 154 mRNAs, 31 miRNAs, 28 lncRNAs and 1590 methylated CpG sites were identified to be significantly correlated with LINC00649. Furthermore, the network of ceRNA was established with 6 miRNAs and 122 mRNAs. The Lasso-Cox model fitted OS/PFS to novel prediction models, which integrated clinical factors, ELN risk stratification, mRNA/miRNA expression and methylation profiles. The analysis of time-dependent ROC for our model showed a superior AUC (AUC = 0.916 at 1 year, AUC = 0.916 at 3 years, and AUC = 0.891 at 5 years). **Conclusions** Low expression of LINC00649 is a potential unfavorable prognostic marker for AML patients, which requires the further validation. The analysis by LASSO-COX regression identified a novel comprehensive model with a superior diagnostic utility, which integrated clinical and genetic variables.

Background

AML is one of most common hematological malignancies, which is characterized by unlimited proliferation of clonal myeloid progenitors and impaired production of normal hematopoietic cells[1]. The prognosis of AML is still heterogenous and unsatisfying: the rate of 5-year survival for AML patients is less than 50%[2], 2-year survival rate of elderly patients is only 20%[3]. Several prediction models have been constructed, among which the updated ELN risk stratification is the most widely used in clinical practice [4], recognizing 3 subgroups of patients according to pretreatment molecular mutations and cytogenetics. Whereas the advances in high-throughput methodology produced multidimensional information on genomes, such as noncoding RNA expression, methylation profile etc.

The size of long noncoding RNAs are generally longer than two hundred nucleotides, and do not have the potential of protein-coding. The interactions of lncRNAs and protein-coding genes are diverse and complex, among which the regulatory mechanism of HOTAIR was well studied in AML. HOTAIR exerts a pro-oncogenic effect in AML, which suppresses p15 expression by methylation of p15 promoter mediated by PRC2, and increases of HOXA5 methylation by directly recruiting DNMT3B[5, 6]. The HOTAIRM1 is another well-studied lncRNA in AML, which played a potential oncogenic role by enhancing expression of HOXA1-4 genes[7]. These results suggest HOXA family genes are important targets of AML-related lncRNAs. HOXA family included 11 genes (HOXA1-7, HOXA9-11 and HOXA13), encoding conserve transcription factors in relation with normal hematopoiesis[8-10]. While dysregulated expression of HOXA family genes is associated with oncogenesis[11].

LINC00649, which we focused in the research, was identified as a prognostic marker in prostate and colorectal cancers by previous bioinformatic analysis[12, 13]. Few studies have investigated the prognostic value of LINC00649 in AML[14]. Notably, the expression of LINC00649 was significantly correlated with HOXA family genes, indicated by the results derived from GEPIA[15] 2.0 online tools (<http://gepia2.cancer-pku.cn/>) and our own analysis. These results suggest that LINC00649 may be associated with AML survival through regulating HOXA genes.

lncRNA regulated the expression of target genes mainly by the following mechanisms: epigenetic regulation, directly transcription regulation by lncRNA binding proteins, splicing regulation, sponging target miRNAs to form competing endogenous RNA, post-translation regulation[16]. Differentially methylated positions (DMPs) and differentially methylated regions (DMRs) were identified between LINC00649-high and -low expression groups, to reveal the epigenetic changes related to HOXA gene family. Then the possibly binding protein of LINC00649 was uncovered, and its relationship with HOXA genes were conducted. Moreover, we established ceRNA network using overlapped results of prediction by online databases and correlation analysis regarding LINC00649. We did not only aim to reveal the regulating effect of LINC00649 on HOXA genes, but to uncover the association of AML survival with LINC00649 related epigenetic and genetic changes. Therefore, the LASSO regression analysis was employed to fit the survival data of AML patients into the prediction model. An overall flowchart was shown in Figure 1. Our study presented the rationality to use the expression level of LINC00649 as a prognostic biomarker and established the novel risk model to predict survival of AML cases.

Methods

Data source

The RNAseq data was downloaded from BeatAML [17] (<http://www.vizome.org/aml>) and TCGA database (<https://portal.gdc.cancer.gov/>), as well as the corresponding clinical and genetic information. The transcriptome dataset in the format of RPKM was obtained from BeatAML database. While the raw count dataset of mRNA and lncRNA was originally obtained from TCGA database, and then transformed to TPM format. The last access to the two database is on 2019.12.25. The GEPIA 2.0 online tool [15] (<http://gepia2.cancer-pku.cn>) was used to compare expression level of LINC00649 between diverse cancers and corresponding normal tissues to explore the expression signature in AML. The cohorts from BeatAML database were employed to reveal the different expression level of LINC00649 between AML (n = 474) and healthy controls (n = 33).

Kaplan-Meier analysis of LINC00649 on AML survival

The AML cohorts from TCGA/BeatAML database were classified as LINC00649-high and LINC00649-low groups respectively, using mean TPM/RPKM of LINC00649. The survival data of AML patients, including OS and PFS, was obtained for TCGA and BeatAML database. Kaplan-Meier analysis were conducted using survival data and log rank test was performed, using p value < 0.05 as the cutoff value.

Identification of LINC00649 binding proteins

The catRAPID is an algorithm predicting RNA-protein pairing, by combining hydrogen bond, secondary structure and other inter-molecular factors [18, 19]. The sequence of LINC00649 was downloaded from nucleotide database of PubMed, which was then input into catRAPID omics tools (<http://service.tartagialab.com/>). The results were downloaded with predicted protein name and binding sites of protein/RNA.

Identification of LINC00649 associated protein coding genes/miRNA and lncRNA

The miRNA expression dataset was downloaded from TCGA database, in the format of RPM value. We conducted a correlation analysis between LINC00649 and other protein-coding genes/miRNA/lncRNA by Pearson's method, using R (version 3.6.0) and cor.test function in stats package. The variables with |Pearson's coefficient| > 0.4 and p value < 0.05 are defined as LINC00649 associated genes. Then to access the enrichment on signaling pathways, we conducted the GSEA based on MSigDB database (<http://software.broadinstitute.org/gsea/msigdb>) [20-22] using LINC00649 associated gene set and corresponding Pearson's coefficients. Meanwhile, the ORA was performed based on GO database and KEGG database. The KEGG analysis was conducted by the ClueGO plugin of Cytoscape software (version 3.7.2) and GO analysis by the "topGO" and "REVIGO" package and R (version 3.6.0).

Identification of LINC00649 associated methylation profile

To investigate the methylation signature in relation to LINC00649 expression, we obtained normalized beta value matrix for AML cohort from TCGA database (146 AML patients, Illumina Infinium HumanMethylation450 platform). The correlation analysis was performed between LINC00649 expression and methylation status (beta value) of individual methylation position and methylation regions.

Establishment of LINC00649-centric competitive endogenous RNA (ceRNA) network

The predicted target miRNA set of LINC00649 was obtained by prediction module of IncBase v2 database, which is based on the microRNA/lncRNA target predicting algorithm [23] (http://carolina.imis.athena-innovation.gr/diana_tools/web). The target mRNAs, the 3' UTR of which were predicted to bind the identified miRNAs, were screened by miRWalk 2.0 online tools [22, 24] (<http://zmf.umm.uni-heidelberg.de/>). We defined the LINC00649-centric ceRNAs as intersection between predicted target miRNA set and LINC00649-associated miRNA set generated from previous correlation analysis.

Establishment of prediction model for AML survival

In above analysis, we identified the LINC00649-centric ceRNA network and methylation changes in relevance of LINC00649, which were supposed to be key elements linking to prognosis of AML. Moreover, to improve the prognostic model, we constructed a multidimensional survival analysis, integrating clinical features, expression level of LINC00649 and related mRNAs/miRNAs in ceRNA network, and methylation status of correlated CpG sites. We downloaded expression dataset (RNA-seq data and miRNAseq data), methylation dataset (beta value), from TCGA database (<https://portal.gdc.cancer.gov/>). Because the APL patients, also known as M3 type in FAB classification, received the quite different treatment and have more superior survival than other types of AML, we excluded such patients from AML cohort. The traditional clinical variables were taken into analysis, including age, race, gender, risk stratification based on molecular/cytogenetic signature, and counts of white blood cells. According to ELN2017 recommendations [4], AML patients were classified as 'good', 'intermediate' and 'poor' groups, based on the karyotype and gene variation. Finally, 124 AML patients with the intact data were included in our survival analysis. The PFS and OS were fitted to LASSO-Cox model, establishing a completely novel prognostic model for AML patients. The glmnet package were implemented for LASSO regression analysis, which penalized the variables to eliminate less informative predictors, resulting in more interpretable and simpler models. The final coefficient for each variable included in the model, was the average value of the coefficient estimates obtained for the set of cross-validation evaluations. To compare continuous variables between groups, we conducted the Wilcoxon rank-sum test. The Fisher exact test was employed for testing the correlation between categorical variables between groups.

Results

LINC00649 is under-expressed in bone marrow of AML

Using GEPIA 2.0 database [15], the comparison of LINC00649 expression levels (TPM) between the tumors and normal tissues across multi-cancer types, derived from TCGA and GTEx database, was shown in Figure 2. The expression of LINC00649 in normal hematopoietic cells is the highest among all included cancers and tissues (TPM value = 12.83). Moreover, the expression of LINC00649 is much lower in AML cells in comparison with normal hematopoietic cells (TPM value 2.96 vs 12.83), which is converse in most of other cancer types. While in BeatAML database, AML patients also have a trend of lower expression of LINC00649 than that of healthy controls (p = 0.0567, Figure 3A).

Expression of LINC00649 is correlated with survival of AML patients

The clinical/genetic features of LINC00649-low and LINC00649-high group were described in Table 1. LINC00649-low patients were associated more unfavorable cytogenetic and molecular risk profiles ($p = 0.001$) and higher WBC counts ($p = 0.001$).

The PFS and OS were significantly inferior for LINC00649-low group in comparison with LINC00649-high group (Figure 3C&D). The median OS of LINC00649-high and LINC00649-low groups are 45.797 versus 12.197 months, and p value of log rank test is 0.0202 (Figure 3C). The median PFS of LINC00649-low and LINC00649-high groups are 26.202 versus 13.808 months respectively, and p value of log rank test is 0.119 (Figure 3D). These results are consistent with OS analysis using BeatAML database (Figure 3B).

Prediction of LINC00649 binding proteins

The predicted proteins were listed in Supplementary Table 1. A total of 120 binding sites involving 9 proteins were identified (ELAVL2/TIAL1/PTBP1/CELF1/SRSF9/SRSF2/SRSF3/

ESRP2/RBFOX2). Then correlation analysis of predicted proteins and HOXA genes was conducted using RNA-seq data of TCGA database (Table 2), where the significantly correlated gene pairs were colored in red. The TIAL1, SRSF9, SRSF2, SRSF3 and RBFOX2 were significantly correlated with the expression of HOXA genes ($p < 0.05$).

Identification of LINC00649 associated protein coding genes/miRNAs and lncRNAs

In total, 154 protein-coding genes, 28 lncRNAs and 31 miRNAs were identified to be significantly correlated with LINC00649 expression level (Supplementary Table 2). The expression of 9 HOXA family members (HOXA1/2/3/4/5/6/7/9/10) was negatively correlated with LINC00649 significantly (Figure 4), indicating that LINC00649 involves in downregulation of HOXA genes.

The results of GESA indicated that under-expression of LINC00649 was associated with activation of 19 pathways and suppression of 6 pathways (Figure 5). The activated pathways included oxidative phosphorylation, IL6-JAK-STAT3 signaling, PI3K-Akt-mTOR signaling (Figure 6), angiogenesis, etc. while the suppressed pathways included P53 pathway, Hedgehog signaling, epithelial mesenchymal transition, etc (Figure 6).

For the results of ORA (Figure 7), the following biological processes were identified by GO enrichment analysis: negative regulation of hematopoiesis, DNA-templated transcription, myeloid cell differentiation, etc. The cell components of LINC00649 associated genes were enriched in protein complex involved in cell adhesion, cell periphery, plasma membrane, etc. The molecular function enriched by LINC00649-associated genes were double-stranded DNA binding, transcription regulatory region DNA binding, sequence-specific DNA binding, etc. Based on the KEGG database, LINC00649 associated genes were enriched in PI3K-Akt signaling pathway, Ras signaling pathway, etc. The analysis based on Reactome database indicated that these genes were enriched in Signaling by ERBB2, Signaling by Receptor Tyrosine Kinases, Signaling by VEGF, etc.

Identification of LINC00649 associated methylation prolife

1590 methylated CpG sites were identified to be significantly correlated with LINC00649 expression ($p < 0.05$, $|r| > 0.4$), listed in Supplementary Table 3), among which 7 methylation positions involving with HOXA6/HOXA9/HOXA10 (cg21172377, cg07483304, cg05490659, cg02000808, cg19816811, cg16880946, cg18931036) have prognostic significance for OS, based on TCGA database using MethSurv online tools [25] (<https://biit.cs.ut.ee/methsurv/>). Therefore, we inferred that the similar methylation modulation of HOTAIR on HOXA5, may be related to regulation of HOXA genes by LINC00649[6]. In the following construction of prediction model, the methylation level of 1590 positions were included in initial LASSO analysis.

The competitive endogenous RNA network of LINC00649

6 miRNAs and 122 mRNAs were included in LINC00649 centric ceRNA network (Supplementary Table 4). The miR-10a-3p, miR-500a-5p, miR-500b-5p, miR-532-3p, miR-502-3p and miR-362-5p were both predicted as sponging miRNAs and statistically significantly correlated with LINC00649 expression (Figure8). Notably, these miRNAs are predicted to negatively inhibit the expression of HOX family genes, suggesting LINC00649 may exert biological effect through sponging miRNAs.

Establishment of the prediction model

A LASSO-COX regression analysis was conducted to identify the prediction models fitting AML OS/PFS, which initially included age, ELN2017 risk stratification, expression level of elements in the LINC00649-centric ceRNA network, and methylation status of LINC00649-associated CpG sites. After 1000 times of iteration, the developing process of model screened the optimal risk scores using summation of selected variables weighted by coefficients (Table 3). The 'cutoffROC' package was employed to determine the cut-off value for OS and PFS model respectively, based on the time-dependent ROC curves. Then AML cohort from TCGA database, was divided into high-risk group and low risk group, by the cutoff value.

The AUC at 1/3/5-year AUC of the prediction model are 0.916/0.916/0.891 respectively for OS (Figure 9), and 0.818/0.881/0.89 for PFS (Figure 10). The distribution of risk scores, survival-events plots and the heatmap of variables for individual patients were shown in Figure 11&12 for OS and PFS, respectively. Then Kaplan-Meier plot was employed to elucidate the difference of survival between high risk and low risk group (Figure 13&14). The median OS and PFS of low-risk group were not reached, which were much better than the high-risk group. The results of Kaplan-Meier plots implicated

that the novel prediction model was efficient for selecting AML patients with superior prognosis. The performance of our model is encouraging, while further prospective research is needed to evaluate the diagnostic value of this model more precisely.

Discussion

The long noncoding RNAs have been uncovered to exert an pivotal influence on cell proliferation and apoptosis of AML, the mechanisms of which include altering methylation status of gene promoters[5, 6], recruiting epigenetic complex on gene promoters[26], reshaping chromatin [27, 28], sponging miRNAs to regulate gene expression[29-32], etc. HOTAIR is one of the most studied lncRNAs in AML, which is upregulated in de novo AML patients and predicts an adverse prognosis[33]. HOTAIR locates in HOXC gene cluster on chromosome 12 and exerts biological effect through modulating HOXA family genes. Intriguingly, through analysis of TCGA expression data, we found LINC00649 was also correlated with most of HOXA family genes, which encoded crucial transcription factors in normal hematopoiesis, pathogenesis of AML and resistance to chemotherapy[34-36]. In comparison with healthy controls, AML patients have aberrantly lower LINC00649 expression in our results. Meanwhile, for most of cancers other than AML, expression level of LINC00649 in cancer cells is higher than that of corresponding normal tissues. Furthermore, the survival (OS and PFS), of LINC00649-low group, was significantly worse than that of LINC00649-high group. The unusual expression signature and prognostic value of LINC00649 drove us to explore the possible molecular mechanisms and uncover its biological function.

According to catRAPID algorithm, 9 proteins containing 120 sites were identified to be potentially binding to LINC00649. TIAL1, SRSF9, SRSF2, SRSF3 and RBFOX2 were identified to be associated with HOXA gene expression. TIAL1 is the RNA binding protein, which binds to target sites and splice the pre-mRNA alternatively[37, 38]. SRSF9 involves in constitutive mRNA splicing and can modulate the target of alternative splicing[39]. SRSF9 was reported to be involved in the cell proliferation and apoptosis in bladder and cervical cancer[40, 41], and related to prognostic alternative splicing events of renal clear cell carcinoma[42]. SRSF2 and SRSF3 are also splicing factors, which belong to serine/arginine-rich protein family. Functional mutations of SRSF2 drive the cancer genesis of hematopoietic cells[43]. SRSF3 is a multiple cancer related splicing factor, namely glioblastoma[44], colon cancer[45], oral squamous carcinoma[46], etc. Moreover, the expression of SRSF2/3 is significantly decreased in de novo AML patients in comparison with that of healthy controls. RBFOX2 can bind to 5'-UGCAUGU-3' element of target RNA, exerting alternative splicing. RBFOX2 can modulate erythropoiesis, by promoting alternative selection of exon 16 in protein 4.1R, the product of which is essential for erythrocyte membrane stability[47, 48]. Notably, the expression of RBFOX2 is significantly correlated with all members of HOXA family genes (Supplementary Figure 1), suggesting potential interaction between them. Furthermore, the pancancer-TCGA expression data was download from UCSC database (<https://xenabrowser.net/hub/>), the correlation of RBFOX2 and HOXA genes was analyzed by Pearson's method (Supplementary Figure 2-1/2/3). Notably, the significant association of RBFOX2 and HOXA is a common feature among cancers generated from different tissues. The expression dataset of normal tissue was downloaded from GTEx database (<https://www.gtexportal.org/home/>), similar analysis showed that the correlation is insignificant in normal bone marrow (Supplementary Figure 3-1/2/3), which indicated the relationship was a disease-specific feature for AML instead of normal hematopoiesis. All 5 splicing factors and LINC00649 are potential co-regulators for HOXA genes in AML, which has not been explored before.

Based on the results of GSEA, the upregulation of PI3K-Akt-mTOR signaling, IL6-JAK-STAT3 signaling, oxidative phosphorylation was identified in LINC00649-low group. The activation of PI3K-Akt-mTOR signaling were found in 50% of AML patients[49, 50]. The PI3K-Akt signaling controls leukemic blast cells proliferation and clonogenicity[51, 52]. Aberrantly functional receptor tyrosine kinases drive the activation of PI3K-Akt-mTOR pathway, including IGF1/IGF1R[53, 54], activated FLT3[55] and DEK-NUP14 fusion protein[56]. The inhibitors of PI3K-Akt-mTOR axis have shown preliminary anti-leukemia effects against AML both in vivo and in vitro [57-63]. The IL6-JAK-STAT3 pathway plays a crucial role in oncogenesis of diverse cancers[64]. Constitutive phosphorylation of STAT3 by autocrine secretion of IL6 is revealed in AML cells[65]. Activation of STAT3 is also uncovered in pediatric AML samples, and the small-molecule inhibitor of STAT3 can induce apoptosis and inhibit formation of blast colonies in vitro[66]. The maintenance of leukemia stem cells depends on BCL2 mediated oxidative respiration, instead of glycolysis as in normal hematopoietic cells[67]. The metformin, targeting oxidative phosphorylation (OXPHOS), induces apoptosis of human leukemia cells in an AMPK-independent way[68]. Cytarabine resistant leukemia cells are characterized by activated OXPHOS, with the high level of reactive oxygen species. Additionally, the resistance can be reversed by agents inducing low OXPHOS status[69]. The p53 signaling and Hedgehog signaling were found to be suppressed in GSEA. Non-mutational p53 dysfunction was common in AML and implicated in diverse inactivating mechanisms[70]. Dysregulation and activation of PI3K-Akt-mTOR signaling pathway can activate MDM2 and interact with NF-kappaB signaling pathway, leading to dysfunction of p53[71]. The activation of PI3K pathway was revealed in LINC00649-low group, which may cause the suppression of p53 signaling and inferior survival considering the central role of p53 in the complex network of AML-associated signaling pathway.

The KEGG analysis showed that the LINC00649-associated genes were enriched in AGE-RAGE, PI3K-Akt, Ras and VEGFR signaling pathways. The AGE and RAGE signaling has been studied in AML, which indicated AGE activated MAP kinase, PI3K and JAK/STAT pathway, leading to proliferation of primary AML samples and AML cell lines[72]. Activation of Ras signaling can also promote the dysfunction of p53 by similar mechanism of PI3K-Akt signaling[71]. VEGFR is reported to be overexpressed[73] in AML, which is in accordance with our results. The activated VEGFR signaling promoted the proliferation, survival and resistance to chemotherapy of AML blasts[74]. VEGFR targeting therapy has been developing and showing preliminary benefit for AML in vitro[75-77]. While the Reactome analysis demonstrated other enriched pathways, including signaling by ERBB2 and VEGFR2 mediated cell proliferation. Mudritinib, an ERBB2 inhibitor, was reported to eliminate AML cell both in vivo and in vitro [78]. VEGFR2 is a 'hot' target in AML, and relevant to chemotherapy-sensitivity, pro-survival effect and angiogenesis in bone marrow[79, 80]. VEGFR2-targeting therapy is being developed in preclinical stage[80, 81]. The dysregulation of all above pathways contributed to the difference of survival between LINC00649-high and low groups.

Furthermore, HOXA family genes methylation status was shown to be correlated with LINC00649. The methylation status of seven CpG sites involving with HOXA6/HOXA9/HOXA10 was correlated with expression of LINC00649. Notably, all involved sites were of significance for AML overall survival (Supplementary Figure 4). Considering that lncRNA HOTAIR can modulate the methylation status of HOXA5 by inhibiting DNMT3B[6], our results suggested similar epigenetic mechanism may implicated in the regulation of HOXA genes.

To improve the diagnostic utility, we brought in multi-dimension information to establish a prediction model on AML survival. The traditional prognostic factors (age, gender, ELN2017 risk stratification, etc) and the associated expression data (predicted LINC00649 binding proteins, miRNAs/mRNAs in the ceRNA network), and methylation data (altered methylated CpG sites) were included into the prediction model, by which the OS and PFS data were fitted into using the LASSO regression analysis. A few prediction models, including genetic information of AML patients, have been developed previously, including Clinseq-G[82] (AUC for 3-year OS is 0.730), ELN2017 stratification in the validation cohort [82] (AUC for 3-year OS is 0.65), Li Z et al[82] (AUC for 3-year OS is 0.70), Huang R et al[83] (AUC for 1 year OS is 0.666, AUC for 5 year OS is 0.707), Ha M et al[84] (AUC for 5-year OS is 0.613). The AUC of our prediction models is better than all these models, possibly attributing to the integrated multi-dimension information. On the other hand, the Kaplan-Meier plots supported the risk stratification using our models to divide patients into high-risk and low-risk group, which well recognized the patients with much better prognosis (the median OS of low-risk group has not reached). While due to lack of integrated information in one cohort like TCGA database, which included clinical/RNAseq (protein-coding and noncoding RNA)/miRNAseq/methylation datasets, we can hardly validate this model independently. However, the present work brings clues and insights to further studies, by providing potential biomarker and therapeutic targets.

In our OS-prediction model, novel markers were identified (Table 3). EVPL is a component of the cornified envelope of keratinocytes, the genetic variations of which are associated with several solid cancer types[85-89]. While the association of other protein-coding genes or noncoding RNAs with either hematological or solid malignancy has not been investigated. Among the included methylation positions, individual methylation status of cg27456487 (MPO), cg05140293 (TTLL4), cg10152449 (CHST12), cg22291265 (SHANK1), cg18597188 (XRCC3), cg14533068 (SYNJ2), cg00081084 (TBCB) and cg20386404 (PTPN14) were significantly associated with AML survival, according to MethSurv online tools (<https://biit.cs.ut.ee/methsurv/>) (Supplementary Figure 5). A low expression ratio of MPO has been reported as a deleterious marker for AML, indicating a lower complete remission rate[90] and shorter PFS [91]. In untreated AML patients, hypermethylation status of MPO is detected and correlates with MPO expression, which can be induced by demethylating agents[92]. The alteration of MPO is demonstrated as an indicator for DNA methylation pattern implicating downregulation of DNMT3B[93], our results supported its significance in pathogenesis of AML. Other included methylated sites have not been reported to implicated in AML. In PFS prediction model (Table 3), no genetic variables were described in relation with AML previously. Notably, KIF26A was included in both OS and PFS model, which belongs to kinesin superfamily and is reported as an oncogenic marker for breast cancer[94] and pancreatic ductal carcinoma[95].

Conclusion

To the best of our knowledge, this is the first research to demonstrate the under-expression of LINC00649 is a potential unfavorable prognostic indicator for AML. Additionally, the novel multi-dimensional prediction models were established with superior diagnostic utility. Further studies are needed on the precise molecular mechanisms and validation of data analysis.

Abbreviations

Acute myeloid leukemia (AML)

European LeukemiaNet (ELN)

Long Intergenic Non-Protein Coding RNA 649 (LINC00649)

HOX Transcription Antisense RNA (HOTAIR)

HOXA Transcript Antisense RNA, Myeloid-Specific 1 (HOTAIRM1)

homeobox protein HOX cluster A(HOXA)

Gene Expression Profiling Interactive Analysis (GEPIA)

Differentially methylated positions (DMP)

Differentially methylated regions [DMR]

competing endogenous RNA (ceRNA)

The Cancer Genome Atlas (TCGA)

overall survival (OS)

Reads Per Kilobase Million (RPKM)
Per Kilobase Million (TPM)
overall survival (OS)
progression free survival (PFS)
Kyoto Encyclopedia of Genes and Genomes (KEGG)
Gene ontology (GO)
least absolute shrinkage and selection operator (LASSO)
gene set enrichment analysis (GESA)
over-representation analysis (ORA)
acute promyelocytic leukemia (APL)
Nucleolysin TIAR (TIAL1)
Serine/arginine-rich splicing factor 9 (SRSF9)
Serine/arginine-rich splicing factor 2 (SRSF2)
Serine/arginine-rich splicing factor 3 (SRSF3)
RNA binding protein fox-1 homolog 2 (RBFox2)
receiver operating characteristics (ROC)
area under curve (AUC)
Envoplakin (EVPL)

Declarations

Ethics approval and consent to participate

This study does not involve animal and/or human tissue/individual data/ participants, there is no ethics related issues. No permissions were required to use any repository data involved in the present study.

Consent for publication

Not Applicable.

Availability of data and materials

The data that support the findings of this study are available from BeatAML database (<http://www.vizome.org/aml>) and TCGA database (<https://portal.gdc.cancer.gov/>), which are all publicly available.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

GC, GYY, JQQ, ZCX, GM and LZL participated in the study design. GC performed most of the microarray and RNA-seq analyses. GYY performed cluster and classification analyses. GC drafted the paper. GC, GYY and LZL participated in the final preparation and revision of the paper. All authors read and approved this final manuscript.

Acknowledgements

Not applicable.

Authors' information

Affiliations: Department of Hematology, China-Japan Friendship Hospital, Beijing, 100029, China

Chao Guo¹, Ya-yue Gao¹, Qian-qian Ju¹, Chun-xia Zhang¹, Ming Gong¹, Zhen-ling Li¹

References

1. Barbui T, Thiele J, Gisslinger H, Kvasnicka HM, Vannucchi AM, Guglielmelli P, Orazi A, Tefferi A: **The 2016 WHO classification and diagnostic criteria for myeloproliferative neoplasms: document summary and in-depth discussion.** *Blood Cancer J* 2018, **8**(2):15.
2. Short NJ, Rytting ME, Cortes JE: **Acute myeloid leukaemia.** *Lancet* 2018, **392**(10147):593-606.
3. Riva L, Luzi L, Pelicci PG: **Genomics of acute myeloid leukemia: the next generation.** *Front Oncol* 2012, **2**:40.
4. Dohner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Buchner T, Dombret H, Ebert BL, Fenaux P, Larson RA *et al.*: **Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel.** *Blood* 2017, **129**(4):424-447.
5. Gao S, Zhou B, Li H, Huang X, Wu Y, Xing C, Yu X, Ji Y: **Long noncoding RNA HOTAIR promotes the self-renewal of leukemia stem cells through epigenetic silencing of p15.** *Exp Hematol* 2018, **67**:32-40 e33.
6. Wang SL, Huang Y, Su R, Yu YY: **Silencing long non-coding RNA HOTAIR exerts anti-oncogenic effect on human acute myeloid leukemia via demethylation of HOXA5 by inhibiting Dnmt3b.** *Cancer Cell Int* 2019, **19**:114.
7. Zhang X, Lian Z, Padden C, Gerstein MB, Rozowsky J, Snyder M, Gingeras TR, Kapranov P, Weissman SM, Newburger PE: **A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster.** *Blood* 2009, **113**(11):2526-2534.
8. Lawrence HJ, Sauvageau G, Humphries RK, Largman C: **The role of HOX homeobox genes in normal and leukemic hematopoiesis.** *Stem Cells* 1996, **14**(3):281-291.
9. Lawrence HJ, Christensen J, Fong S, Hu YL, Weissman I, Sauvageau G, Humphries RK, Largman C: **Loss of expression of the Hoxa-9 homeobox gene impairs the proliferation and repopulating ability of hematopoietic stem cells.** *Blood* 2005, **106**(12):3988-3994.
10. Shah N, Sukumar S: **The Hox genes and their roles in oncogenesis.** *Nat Rev Cancer* 2010, **10**(5):361-371.
11. Bhatlekar S, Fields JZ, Boman BM: **HOX genes and their role in the development of human cancers.** *J Mol Med (Berl)* 2014, **92**(8):811-823.
12. Ye G, Guo L, Xing Y, Sun W, Yuan M: **Identification of prognostic biomarkers of prostate cancer with long non-coding RNA-mediated competitive endogenous RNA network.** *Exp Ther Med* 2019, **17**(4):3035-3040.
13. He M, Lin Y, Xu Y: **Identification of prognostic biomarkers in colorectal cancer using a long non-coding RNA-mediated competitive endogenous RNA network.** *Oncol Lett* 2019, **17**(3):2687-2694.
14. Guo C, Ju QQ, Zhang CX, Gong M, Li ZL, Gao YY: **Overexpression of HOXA10 is associated with unfavorable prognosis of acute myeloid leukemia.** *BMC Cancer* 2020, **20**(1):586.
15. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z: **GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses.** *Nucleic Acids Res* 2017, **45**(W1):W98-W102.
16. Gourvest M, Brousset P, Bousquet M: **Long Noncoding RNAs in Acute Myeloid Leukemia: Functional Characterization and Clinical Relevance.** *Cancers (Basel)* 2019, **11**(11).
17. Tyner JW, Tognon CE, Bottomly D, Wilmot B, Kurtz SE, Savage SL, Long N, Schultz AR, Traer E, Abel M *et al.*: **Functional genomic landscape of acute myeloid leukaemia.** *Nature* 2018, **562**(7728):526-531.
18. Livi CM, Klus P, Delli Ponti R, Tartaglia GG: **catRAPID signature: identification of ribonucleoproteins and RNA-binding regions.** *Bioinformatics* 2016, **32**(5):773-775.
19. Bellucci M, Agostini F, Masin M, Tartaglia GG: **Predicting protein associations with long noncoding RNAs.** *Nat Methods* 2011, **8**(6):444-445.
20. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al.*: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102**(43):15545-15550.
21. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E *et al.*: **PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.** *Nat Genet* 2003, **34**(3):267-273.
22. Dweep H, Sticht C, Gretz N: **In-Silico Algorithms for the Screening of Possible microRNA Binding Sites and Their Interactions.** *Curr Genomics* 2013, **14**(2):127-136.
23. Paraskevopoulou MD, Vlachos IS, Karagkouni D, Georgakilas G, Kanellos I, Vergoulis T, Zagganas K, Tsanakas P, Floros E, Dalamagas T *et al.*: **DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts.** *Nucleic Acids Res* 2016, **44**(D1):D231-238.
24. Min H, Yoon S: **Got target? Computational methods for microRNA target prediction and their extension.** *Exp Mol Med* 2010, **42**(4):233-244.
25. Modhukur V, Iljasenko T, Metsalu T, Loka K, Laisk-Podar T, Vilo J: **MethSurv: a web tool to perform multivariable survival analysis using DNA methylation data.** *Epigenomics* 2018, **10**(3):277-288.

26. Zhang X, Weissman SM, Newburger PE: **Long intergenic non-coding RNA HOTAIRM1 regulates cell cycle progression during myeloid maturation in NB4 human promyelocytic leukemia cells.** *RNA Biol* 2014, **11**(6):777-787.
27. Wang H, Li W, Guo R, Sun J, Cui J, Wang G, Hoffman AR, Hu JF: **An intragenic long noncoding RNA interacts epigenetically with the RUNX1 promoter and enhancer chromatin DNA in hematopoietic malignancies.** *Int J Cancer* 2014, **135**(12):2783-2794.
28. Sun J, Li W, Sun Y, Yu D, Wen X, Wang H, Cui J, Wang G, Hoffman AR, Hu JF: **A novel antisense long noncoding RNA within the IGF1R gene locus is imprinted in hematopoietic malignancies.** *Nucleic Acids Res* 2014, **42**(15):9588-9601.
29. Cheng P, Lu P, Guan J, Zhou Y, Zou L, Yi X, Cheng H: **LncRNA KCNQ10T1 controls cell proliferation, differentiation and apoptosis by sponging miR-326 to regulate c-Myc expression in acute myeloid leukemia.** *Neoplasma* 2019.
30. Sun MD, Zheng YQ, Wang LP, Zhao HT, Yang S: **Long noncoding RNA UCA1 promotes cell proliferation, migration and invasion of human leukemia cells via sponging miR-126.** *Eur Rev Med Pharmacol Sci* 2018, **22**(8):2233-2245.
31. Chen L, Wang W, Cao L, Li Z, Wang X: **Long Non-Coding RNA CCAT1 Acts as a Competing Endogenous RNA to Regulate Cell Growth and Differentiation in Acute Myeloid Leukemia.** *Mol Cells* 2016, **39**(4):330-336.
32. Zhang N, Chen Y, Shen Y, Lou S, Deng J: **Comprehensive analysis the potential biomarkers for the high-risk of childhood acute myeloid leukemia based on a competing endogenous RNA network.** *Blood Cells Mol Dis* 2019, **79**:102352.
33. Zhang YY, Huang SH, Zhou HR, Chen CJ, Tian LH, Shen JZ: **Role of HOTAIR in the diagnosis and prognosis of acute leukemia.** *Oncol Rep* 2016, **36**(6):3113-3122.
34. Johansen S, Brenner AK, Bartaula-Brevik S, Reikvam H, Bruserud O: **The Possible Importance of beta3 Integrins for Leukemogenesis and Chemoresistance in Acute Myeloid Leukemia.** *Int J Mol Sci* 2018, **19**(1).
35. Wang H, Lindsey S, Konieczna I, Bei L, Horvath E, Huang W, Saberwal G, Eklund EA: **Constitutively active SHP2 cooperates with HoxA10 overexpression to induce acute myeloid leukemia.** *J Biol Chem* 2009, **284**(4):2549-2567.
36. Shah CA, Bei L, Wang H, Plataniias LC, Eklund EA: **HoxA10 protein regulates transcription of gene encoding fibroblast growth factor 2 (FGF2) in myeloid cells.** *J Biol Chem* 2012, **287**(22):18230-18248.
37. Aznarez I, Barash Y, Shai O, He D, Zielenski J, Tsui LC, Parkinson J, Frey BJ, Rommens JM, Blencowe BJ: **A systematic analysis of intronic sequences downstream of 5' splice sites reveals a widespread role for U-rich motifs and TIA1/TIAL1 proteins in alternative splicing regulation.** *Genome Res* 2008, **18**(8):1247-1258.
38. Le Guiner C, Lejeune F, Galiana D, Kister L, Breathnach R, Stevenin J, Del Gatto-Konczak F: **TIA-1 and TIAR activate splicing of alternative exons with weak 5' splice sites followed by a U-rich stretch on their own pre-mRNAs.** *J Biol Chem* 2001, **276**(44):40638-40646.
39. Stoss O, Schwaiger FW, Cooper TA, Stamm S: **Alternative splicing determines the intracellular localization of the novel nuclear protein Nop30 and its interaction with the splicing factor SRp30c.** *J Biol Chem* 1999, **274**(16):10951-10962.
40. Zhang Q, Lv R, Guo W, Li X: **microRNA-802 inhibits cell proliferation and induces apoptosis in human cervical cancer by targeting serine/arginine-rich splicing factor 9.** *J Cell Biochem* 2019, **120**(6):10370-10379.
41. Yoshino H, Enokida H, Chiyomaru T, Tatarano S, Hidaka H, Yamasaki T, Gotannda T, Tachiwada T, Nohata N, Yamane T *et al.* **Tumor suppressive microRNA-1 mediated novel apoptosis pathways through direct inhibition of splicing factor serine/arginine-rich 9 (SRSF9/SRp30c) in bladder cancer.** *Biochem Biophys Res Commun* 2012, **417**(1):588-593.
42. Chen T, Zheng W, Chen J, Lin S, Zou Z, Li X, Tan Z: **Systematic analysis of survival-associated alternative splicing signatures in clear cell renal cell carcinoma.** *J Cell Biochem* 2019.
43. Liang Y, Tebaldi T, Rejeski K, Joshi P, Stefani G, Taylor A, Song Y, Vasic R, Maziarz J, Balasubramanian K *et al.* **SRSF2 mutations drive oncogenesis by activating a global program of aberrant alternative splicing in hematopoietic cells.** *Leukemia* 2018, **32**(12):2659-2671.
44. Song X, Wan X, Huang T, Zeng C, Sastry N, Wu B, James CD, Horbinski C, Nakano I, Zhang W *et al.* **SRSF3-Regulated RNA Alternative Splicing Promotes Glioblastoma Tumorigenicity by Affecting Multiple Cellular Processes.** *Cancer Res* 2019, **79**(20):5288-5301.
45. Kuranaga Y, Sugito N, Shinohara H, Tsujino T, Taniguchi K, Komura K, Ito Y, Soga T, Akao Y: **SRSF3, a Splicer of the PKM Gene, Regulates Cell Growth and Maintenance of Cancer-Specific Energy Metabolism in Colon Cancer Cells.** *Int J Mol Sci* 2018, **19**(10).
46. Peiqi L, Zhaozhong G, Yaotian Y, Jun J, Jihua G, Rong J: **Expression of SRSF3 is Correlated with Carcinogenesis and Progression of Oral Squamous Cell Carcinoma.** *Int J Med Sci* 2016, **13**(7):533-539.
47. Huang SC, Ou AC, Park J, Yu F, Yu B, Lee A, Yang G, Zhou A, Benz EJ, Jr.: **RBFox2 promotes protein 4.1R exon 16 selection via U1 snRNP recruitment.** *Mol Cell Biol* 2012, **32**(2):513-526.
48. Ponthier JL, Schluepen C, Chen W, Lersch RA, Gee SL, Hou VC, Lo AJ, Short SA, Chasis JA, Winkelmann JC *et al.* **Fox-2 splicing factor binds to a conserved intron motif to promote inclusion of protein 4.1R alternative exon 16.** *J Biol Chem* 2006, **281**(18):12468-12474.
49. Park S, Chapuis N, Tamburini J, Bardet V, Cornillet-Lefebvre P, Willems L, Green A, Mayeux P, Lacombe C, Bouscary D: **Role of the PI3K/AKT and mTOR signaling pathways in acute myeloid leukemia.** *Haematologica* 2010, **95**(5):819-828.
50. Prijic S, Ugrina I, Labar B, Nemet D, Batinic J, Zadro R, Ries S, Gjadrov-Kuvedzic K, Davidovic S, Batinic D: **Prognostic significance of constitutive phosphatidylinositol 3-kinase/Akt and mitogen-activated protein kinase phosphorylation in acute myeloid leukemia.** *Leuk Lymphoma* 2015, **56**(8):2281-2288.

51. Sujobert P, Bardet V, Cornillet-Lefebvre P, Hayflick JS, Prie N, Verdier F, Vanhaesebroeck B, Muller O, Pesce F, Ifrah N *et al*: **Essential role for the p110delta isoform in phosphoinositide 3-kinase activation and cell proliferation in acute myeloid leukemia.** *Blood* 2005, **106**(3):1063-1066.
52. Xu Q, Simpson SE, Scialla TJ, Bagg A, Carroll M: **Survival of acute myeloid leukemia cells requires PI3 kinase activation.** *Blood* 2003, **102**(3):972-980.
53. Tazzari PL, Tabellini G, Bortul R, Papa V, Evangelisti C, Grafone T, Martinelli G, McCubrey JA, Martelli AM: **The insulin-like growth factor-I receptor kinase inhibitor NVP-AEW541 induces apoptosis in acute myeloid leukemia cells exhibiting autocrine insulin-like growth factor-I secretion.** *Leukemia* 2007, **21**(5):886-896.
54. Tamburini J, Chapuis N, Bardet V, Park S, Sujobert P, Willems L, Ifrah N, Dreyfus F, Mayeux P, Lacombe C *et al*: **Mammalian target of rapamycin (mTOR) inhibition activates phosphatidylinositol 3-kinase/Akt by up-regulating insulin-like growth factor-1 receptor signaling in acute myeloid leukemia: rationale for therapeutic inhibition of both pathways.** *Blood* 2008, **111**(1):379-382.
55. Chen W, Drakos E, Grammatikakis I, Schlette EJ, Li J, Leventaki V, Staikou-Drakopoulou E, Patsouris E, Panayiotidis P, Medeiros LJ *et al*: **mTOR signaling is activated by FLT3 kinase and promotes survival of FLT3-mutated acute myeloid leukemia cells.** *Mol Cancer* 2010, **9**:292.
56. Sanden C, Ageberg M, Petersson J, Lennartsson A, Gullberg U: **Forced expression of the DEK-NUP214 fusion protein promotes proliferation dependent on upregulation of mTOR.** *BMC Cancer* 2013, **13**:440.
57. Bertacchini J, Heidari N, Mediani L, Capitani S, Shahjahani M, Ahmadzadeh A, Saki N: **Targeting PI3K/AKT/mTOR network for treatment of leukemia.** *Cell Mol Life Sci* 2015, **72**(12):2337-2347.
58. Su Y, Li X, Ma J, Zhao J, Liu S, Wang G, Edwards H, Taub JW, Lin H, Ge Y: **Targeting PI3K, mTOR, ERK, and Bcl-2 signaling network shows superior antileukemic activity against AML ex vivo.** *Biochem Pharmacol* 2018, **148**:13-26.
59. Stengel C, Jenner E, Meja K, Mayekar S, Khwaja A: **Proliferation of PTEN-deficient haematopoietic tumour cells is not affected by isoform-selective inhibition of p110 PI3-kinase and requires blockade of all class 1 PI3K activity.** *Br J Haematol* 2013, **162**(2):285-289.
60. Willems L, Chapuis N, Puissant A, Maciel TT, Green AS, Jacque N, Vignon C, Park S, Guichard S, Herault O *et al*: **The dual mTORC1 and mTORC2 inhibitor AZD8055 has anti-tumor activity in acute myeloid leukemia.** *Leukemia* 2012, **26**(6):1195-1202.
61. Sen S, Hassane DC, Corbett C, Becker MW, Jordan CT, Guzman ML: **Novel mTOR inhibitory activity of ciclopirox enhances parthenolide antileukemia activity.** *Exp Hematol* 2013, **41**(9):799-807 e794.
62. Bertacchini J, Guida M, Accordi B, Mediani L, Martelli AM, Barozzi P, Petricoin E, 3rd, Liotta L, Milani G, Giordan M *et al*: **Feedbacks and adaptive capabilities of the PI3K/Akt/mTOR axis in acute myeloid leukemia revealed by pathway selective inhibition and phosphoproteome analysis.** *Leukemia* 2014, **28**(11):2197-2205.
63. Kang SA, Pacold ME, Cervantes CL, Lim D, Lou HJ, Ottina K, Gray NS, Turk BE, Yaffe MB, Sabatini DM: **mTORC1 phosphorylation sites encode their sensitivity to starvation and rapamycin.** *Science* 2013, **341**(6144):1236566.
64. Johnson DE, O'Keefe RA, Grandis JR: **Targeting the IL-6/JAK/STAT3 signalling axis in cancer.** *Nat Rev Clin Oncol* 2018, **15**(4):234-248.
65. Schuringa JJ, Wierenga AT, Kruijer W, Vellenga E: **Constitutive Stat3, Tyr705, and Ser727 phosphorylation in acute myeloid leukemia cells caused by the autocrine secretion of interleukin-6.** *Blood* 2000, **95**(12):3765-3770.
66. Redell MS, Ruiz MJ, Alonzo TA, Gerbing RB, Tweardy DJ: **Stat3 signaling in acute myeloid leukemia: ligand-dependent and -independent activation and induction of apoptosis by a novel small-molecule Stat3 inhibitor.** *Blood* 2011, **117**(21):5701-5709.
67. Lagadinou ED, Sach A, Callahan K, Rossi RM, Neering SJ, Minhajuddin M, Ashton JM, Pei S, Grose V, O'Dwyer KM *et al*: **BCL-2 inhibition targets oxidative phosphorylation and selectively eradicates quiescent human leukemia stem cells.** *Cell Stem Cell* 2013, **12**(3):329-341.
68. Scotland S, Saland E, Skuli N, de Toni F, Boutzen H, Micklow E, Senegas I, Peyraud R, Peyriga L, Theodoro F *et al*: **Mitochondrial energetic and AKT status mediate metabolic effects and apoptosis of metformin in human leukemic cells.** *Leukemia* 2013, **27**(11):2129-2138.
69. Farge T, Saland E, de Toni F, Aroua N, Hosseini M, Perry R, Bosc C, Sugita M, Stuani L, Fraisse M *et al*: **Chemotherapy-Resistant Human Acute Myeloid Leukemia Cells Are Not Enriched for Leukemic Stem Cells but Require Oxidative Metabolism.** *Cancer Discov* 2017, **7**(7):716-735.
70. Prokocimer M, Molchadsky A, Rotter V: **Dysfunctional diversity of p53 proteins in adult acute myeloid leukemia: projections on diagnostic workup and therapy.** *Blood* 2017, **130**(6):699-712.
71. Steelman LS, Chappell WH, Abrams SL, Kempf RC, Long J, Laidler P, Mijatovic S, Maksimovic-Ivanic D, Stivala F, Mazzarino MC *et al*: **Roles of the Raf/MEK/ERK and PI3K/PTEN/Akt/mTOR pathways in controlling growth and sensitivity to therapy-implications for cancer and aging.** *Aging (Albany NY)* 2011, **3**(3):192-222.
72. Kim JY, Park HK, Yoon JS, Kim SJ, Kim ES, Ahn KS, Kim DS, Yoon SS, Kim BK, Lee YY: **Advanced glycation end product (AGE)-induced proliferation of HEL cells via receptor for AGE-related signal pathways.** *Int J Oncol* 2008, **33**(3):493-501.
73. Padro T, Bieker R, Ruiz S, Steins M, Retzlaff S, Burger H, Buchner T, Kessler T, Herrera F, Kienast J *et al*: **Overexpression of vascular endothelial growth factor (VEGF) and its cellular receptor KDR (VEGFR-2) in the bone marrow of patients with acute myeloid leukemia.** *Leukemia* 2002, **16**(7):1302-1310.
74. Kampen KR, Ter Elst A, de Bont ES: **Vascular endothelial growth factor signaling in acute myeloid leukemia.** *Cell Mol Life Sci* 2013, **70**(8):1307-1317.

75. Rodriguez-Ariza A, Lopez-Pedraza C, Aranda E, Barbarroja N: **VEGF targeted therapy in acute myeloid leukemia.** *Crit Rev Oncol Hematol* 2011, **80**(2):241-256.
76. Hiramatsu A, Miwa H, Shikami M, Ikai T, Tajima E, Yamamoto H, Imai N, Hattori A, Kyo T, Watarai M *et al*: **Disease-specific expression of VEGF and its receptors in AML cells: possible autocrine pathway of VEGF/type1 receptor of VEGF in t(15;17) AML and VEGF/type2 receptor of VEGF in t(8;21) AML.** *Leuk Lymphoma* 2006, **47**(1):89-95.
77. Ding Y, Zhu X, Li Y, Fei J, Zhang Y: **Targeted inhibition of VEGF-modulated survival and arsenic sensitivity in acute myeloid leukemia (AML).** *Hematology* 2012, **17**(3):157-162.
78. Baccelli I, Gareau Y, Lehnertz B, Gingras S, Spinella JF, Comeau S, Mayotte N, Girard S, Frechette M, Blouin-Chagnon V *et al*: **Mubritinib Targets the Electron Transport Chain Complex I and Reveals the Landscape of OXPHOS Dependency in Acute Myeloid Leukemia.** *Cancer Cell* 2019, **36**(1):84-99 e88.
79. Nobrega-Pereira S, Caiado F, Carvalho T, Matias I, Graca G, Goncalves LG, Silva-Santos B, Norell H, Dias S: **VEGFR2-Mediated Reprogramming of Mitochondrial Metabolism Regulates the Sensitivity of Acute Myeloid Leukemia to Chemotherapy.** *Cancer Res* 2018, **78**(3):731-741.
80. Deng M, Zha J, Zhao H, Jia X, Shi Y, Li Z, Fu G, Yu L, Fang Z, Xu B: **Apatinib exhibits cytotoxicity toward leukemia cells by targeting VEGFR2-mediated prosurvival signaling and angiogenesis.** *Exp Cell Res* 2020, **390**(1):111934.
81. Deng M, Zhao H, Chen Q, Zhao J, Shi Y, Yu L, Fang Z, Xu B: **CS2164 suppresses acute myeloid leukemia cell growth via inhibiting VEGFR2 signaling in preclinical models.** *Eur J Pharmacol* 2019, **853**:193-200.
82. Wang M, Lindberg J, Klevebring D, Nilsson C, Lehmann S, Gronberg H, Rantalainen M: **Development and Validation of a Novel RNA Sequencing-Based Prognostic Score for Acute Myeloid Leukemia.** *J Natl Cancer Inst* 2018, **110**(10):1094-1101.
83. Huang R, Liao X, Li Q: **Identification and validation of potential prognostic gene biomarkers for predicting survival in patients with acute myeloid leukemia.** *Onco Targets Ther* 2017, **10**:5243-5254.
84. Ha M, Kim JY, Han ME, Kim GH, Park SY, Jeong DC, Oh SO, Kim YH: **TMEM18: A Novel Prognostic Marker in Acute Myeloid Leukemia.** *Acta Haematol* 2018, **140**(2):71-76.
85. Sigurdson AJ, Brenner AV, Roach JA, Goudeva L, Muller JA, Nerlich K, Reinert C, Schwab R, Pfeiffer L, Waldenberger M *et al*: **Selected single-nucleotide polymorphisms in FOXE1, SERPINA5, FTO, EVPL, TICAM1 and SCARB1 are associated with papillary and follicular thyroid cancer risk: replication study in a German population.** *Carcinogenesis* 2016, **37**(7):677-684.
86. Ruhrberg C, Williamson JA, Sheer D, Watt FM: **Chromosomal localisation of the human envoplakin gene (EVPL) to the region of the tylosis oesophageal cancer gene (TOCG) on 17q25.** *Genomics* 1996, **37**(3):381-385.
87. Zhang X, Feng H, Li Z, Li D, Liu S, Huang H, Li M: **Application of weighted gene co-expression network analysis to identify key modules and hub genes in oral squamous cell carcinoma tumorigenesis.** *Onco Targets Ther* 2018, **11**:6001-6021.
88. Iwaya T, Maesawa C, Kimura T, Ogasawara S, Ikeda K, Kimura Y, Noda Y, Ishida K, Sato N, Saito K *et al*: **Infrequent mutation of the human envoplakin gene is closely linked to the tylosis oesophageal cancer locus in sporadic oesophageal squamous cell carcinomas.** *Oncol Rep* 2005, **13**(4):703-707.
89. Martins WK, Esteves GH, Almeida OM, Rezze GG, Landman G, Marques SM, Carvalho AF, LF LR, Duprat JP, Stolf BS: **Gene network analyses point to the importance of human tissue kallikreins in melanoma progression.** *BMC Med Genomics* 2011, **4**:76.
90. Matsuo T, Cox C, Bennett JM: **Prognostic significance of myeloperoxidase positivity of blast cells in acute myeloblastic leukemia without maturation (FAB: M1): an ECOG study.** *Hematol Pathol* 1989, **3**(4):153-158.
91. Advani SH, Hegde UP, Iyer RS, Gopal R, Saikia TK, Pai SK, Nair CN, Kurkure PA, Pai VR, Nadkarni KS: **Prognostic significance of myeloperoxidase containing blast cells in acute myelogenous leukaemia.** *Indian J Med Res* 1993, **98**:8-14.
92. Schmelz K, Sattler N, Wagner M, Lubbert M, Dorken B, Tamm I: **Induction of gene expression by 5-Aza-2'-deoxycytidine in acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS) but not epithelial cells by DNA-methylation-dependent and -independent mechanisms.** *Leukemia* 2005, **19**(1):103-111.
93. Itonaga H, Imanishi D, Wong YF, Sato S, Ando K, Sawayama Y, Sasaki D, Tsuruda K, Hasegawa H, Imaizumi Y *et al*: **Expression of myeloperoxidase in acute myeloid leukemia blasts mirrors the distinct DNA methylation pattern involving the downregulation of DNA methyltransferase DNMT3B.** *Leukemia* 2014, **28**(7):1459-1466.
94. Li TF, Zeng HJ, Shan Z, Ye RY, Cheang TY, Zhang YJ, Lu SH, Zhang Q, Shao N, Lin Y: **Overexpression of kinesin superfamily members as prognostic biomarkers of breast cancer.** *Cancer Cell Int* 2020, **20**:123.
95. Wong CH, Li CH, He Q, Chan SL, Tong JH, To KF, Lin LZ, Chen Y: **Ectopic HOTTIP expression induces noncanonical transactivation pathways to promote growth and invasiveness in pancreatic ductal adenocarcinoma.** *Cancer Lett* 2020, **477**:1-9.

Tables

Table 1 The comparison of clinical and genetic features between LINC00649-high and LINC00649-low groups. A total of 145 patients from TCGA and 451 patients from BeatAML database were included. The molecular and cytogenetic risk of LINC00649-low group is more adverse than that of LINC00649-high group (p<0.001).

	TCGA			beatAML		
	LINC00649-low group (n = 72)	LINC00649-high group (n = 73)	p value	LINC00649-low group (n = 225)	LINC00649-high group (n = 226)	p value
Age (year)	55.21±2.011	54.47±2.137	0.81	59.81±1.055	53.62±1.356	0.049
Gender			0.589			0.775
Female	27	31		95	98	
Male	35	31		130	127	
Race			0.838			0.801
White	47	45		39	37	
Other races	15	17		186	188	
Mutation count	9.82±0.624	9.77±0.803	0.962	NA	NA	
Risk stratification of cytogenetics			0.001			NA
Good	0	15		NA	NA	
Intermediate	48	28		NA	NA	
Poor	14	19		NA	NA	
Risk stratification of molecular mutation			0.001			0.002
Good	0	15		99	70	
Intermediate	45	28		44	73	
Poor	17	19		82	82	
WBC	47.498±5.091	22.705±5.040	0.001	34.807±2.979	27.615±3.465	0.112

Table 2 The results of Pearson's correlation analysis between predicted LINC00649 binding proteins and HOXA genes. The significant correlated pairs were colored in red.

	HOXA1	HOXA2	HOXA3	HOXA4	HOXA5	HOXA6	HOXA7	HOXA9	HOXA10	HOXA11	HOXA13
ELAVL2	r = 0.0023,p =0.9779	r = 0.0589,p =0.4726	r = -0.0165,p =0.8407	r = -0.018,p =0.8261	r = -0.0169,p =0.8373	r = -0.0295,p =0.7189	r = -0.026,p =0.7515	r = -0.0456,p =0.5785	r = -0.0193,p =0.8142	r = 0.0708,p =0.388	r = -0.0294,p =0.72
TIAL1	r = 0.035,p =0.67	r = 0.0906,p =0.2684	r = 0.1081,p =0.1864	r = 0.0744,p =0.3636	r = 0.0975,p =0.2338	r = 0.1125,p =0.169	r = 0.1077,p =0.1882	r = 0.1206,p =0.1402	r = 0.0819,p =0.3176	r = 0.0575,p =0.4833	r = 0.1934,p =0.0173
PTBP1	r = -0.1561,p =0.0557	r = -0.1234,p =0.1311	r = -0.0634,p =0.4394	r = -0.0056,p =0.9455	r = -0.0785,p =0.3382	r = -0.0241,p =0.769	r = -0.0101,p =0.9018	r = -0.0512,p =0.5321	r = -0.0716,p =0.3823	r = 0.0553,p =0.5004	r = 0.0394,p =0.6312
CELF1	r = 0.0074,p =0.9284	r = 0.0047,p =0.9547	r = -0.0135,p =0.8698	r = -0.0084,p =0.918	r = -0.0004,p =0.996	r = -0.003,p =0.9711	r = -0.0359,p =0.6614	r = -0.0335,p =0.6829	r = -0.0448,p =0.5849	r = -0.0795,p =0.3318	r = 0.0876,p =0.2851
SRSF9	r = -0.1415,p =0.083	r = -0.0669,p =0.4147	r = -0.0447,p =0.586	r = -0.0475,p =0.5626	r = -0.0477,p =0.5612	r = -0.0192,p =0.8149	r = -0.0139,p =0.8655	r = -0.0278,p =0.7351	r = -0.0258,p =0.7529	r = 0.2046,p =0.0117	r = 0.0108,p =0.8953
SRSF2	r = -0.1219,p =0.1358	r = -0.0305,p =0.7103	r = -0.0416,p =0.6117	r = -0.0385,p =0.6385	r = -0.0363,p =0.6582	r = -0.0396,p =0.6293	r = 0.008,p =0.9223	r = -0.05,p =0.5417	r = -0.0655,p =0.4242	r = 0.0419,p =0.6095	r = 0.1725,p =0.0342
SRSF3	r = -0.0213,p =0.7952	r = 0.0063,p =0.939	r = -0.0804,p =0.3263	r = -0.0784,p =0.3384	r = -0.0536,p =0.5134	r = -0.0683,p =0.4047	r = -0.03,p =0.7143	r = -0.0723,p =0.3774	r = -0.0445,p =0.5872	r = 0.0936,p =0.2531	r = 0.1823,p =0.0251
ESRP2	r = -0.0907,p =0.2682	r = 0.0335,p =0.6833	r = -0.0546,p =0.5055	r = -0.0771,p =0.3467	r = -0.0694,p =0.3971	r = -0.0236,p =0.7734	r = -0.0492,p =0.5484	r = -0.0959,p =0.2415	r = -0.1254,p =0.1248	r = -0.0139,p =0.8659	r = 0.1218,p =0.1364
RBFOX2	r = 0.2132,p =0.0086	r = 0.2744,p =0.0007	r = 0.3405,p =0	r = 0.3295,p =0	r = 0.3332,p =0	r = 0.3681,p =0	r = 0.313,p =0.0001	r = 0.3837,p =0	r = 0.351,p =0	r = 0.3358,p =0	r = 0.2299,p =0.0045

Table 3 The variables and coefficients of prediction models for OS and PFS. The risk score for individual patients was calculated as summation of each variable and corresponding coefficient.

OS		DFS	
Variables	Coefficients	Variables	Coefficients
Age	0.01517872	HSDL1	-0.069116277
Risk_moleuclar	0.08487131	KIF26A	-0.105321193
KIF26A	-0.06888634	ZNF124	-0.027336956
SERINC5	-0.34892134	LPAR5	0.183893293
EVPL	0.01905074	PATE2	-0.056221659
SMAGP	0.04229574	hsa-miR-502-3p	0.141651616
CD320	0.07386408	cg07613391	-0.05203838
hsa-miR-502-3p	0.04133291	cg23495279	0.174167913
cg27456487	0.10692372	cg06637001	-0.143845051
cg15440158	-0.20311719	cg02942845	-0.181993998
cg21760402	-0.32770473	cg00081084	5.458822697
cg00081084	1.0182775	cg10520887	-0.129876909
cg14533068	0.13680058	cg13331200	0.568422555
cg18597188	0.1459364	cg00599124	0.082618413
cg22291265	-0.17893264	cg21347874	-0.358914855
cg13475665	-0.05682765	cg14459021	0.079482707
cg06812991	-0.29442296	cg20386404	-0.013977599
cg02057391	-0.38221762	cg27100436	0.563627584
cg14459021	0.35068618	cg21844856	-1.256126321
cg20386404	-0.09889255		
cg05140293	0.03990318		
cg10152449	-0.73415156		
cg16280141	1.19906562		
cg15275758	-0.65571266		

Supplementary Information

Supplementary table 1 The results of predicted binding proteins of LINC00649 by catRAPID algorithm.

Supplementary table 2 The results of genome-wide expression correlation analysis of LINC00649, for protein-coding genes/miRNAs/lncRNAs respectively (p value < 0.05 and $|r| > 0.4$).

Supplementary table 3 The results of genome-wide methylation correlation analysis of LINC00649, showing the significantly correlated methylated positions.

Supplementary table 4 the identified lncRNA, miRNA and mRNA in the ceRNA network.

Supplementary Figure 1 The results of correlation analysis of RBFOX2 and HOXA genes.

Supplementary Figure 2-1/2/3 pan-cancer correlation analysis of RBFOX2 and HOXA genes. The X axis stands for $-\log_{10}(p \text{ value})$, while the Y axis represent Pearson coefficients. The purple dots in the right upper quadrant represent cancer types, in which the correlation is significant and positive. The result in AML was red dots and annotated by text.

Supplementary Figure 3-1/2/3 pan-tissue correlation analysis of RBFOX2 and HOXA genes. The X axis stands for $-\log_{10}(p \text{ value})$, while the Y axis represent Pearson coefficients. The purple dots in the right upper quadrant represent normal tissue types, in which the correlation is significant and positive. The result in normal bone marrow was red dots and annotated by text.

Figures

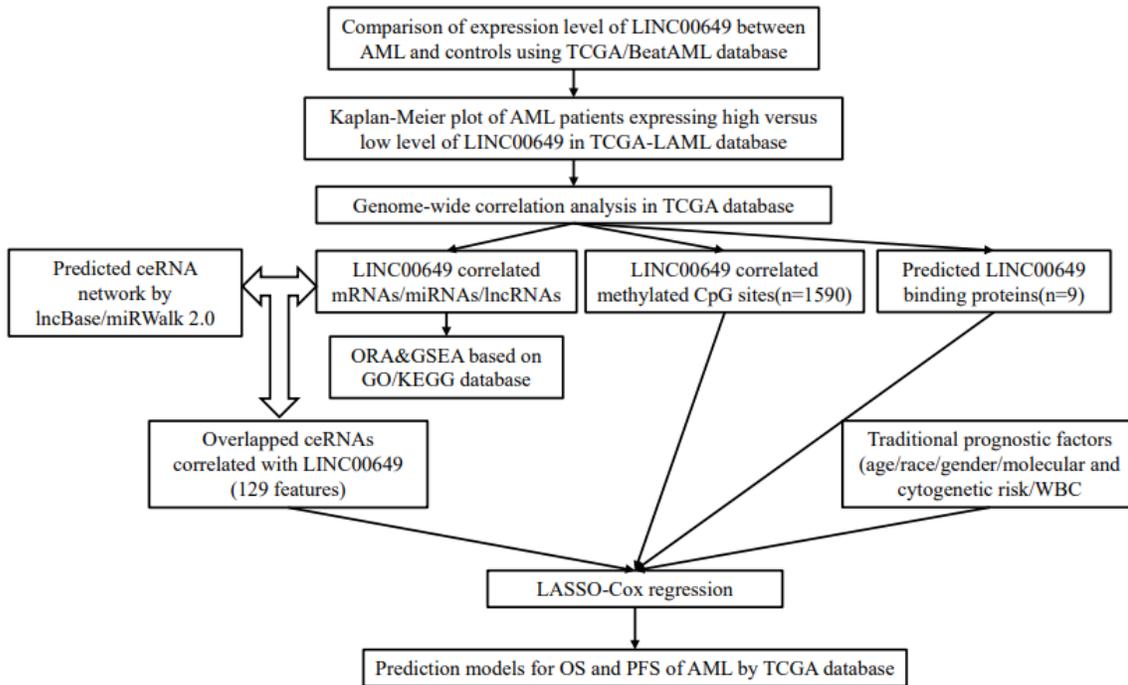


Figure 1

The flow chart of the overall study design. AML, acute myeloid leukemia; TCGA, the Cancer Genome Atlas; OS, overall survival; PFS, progression free survival; LASSO, least absolute shrinkage and selection operator).

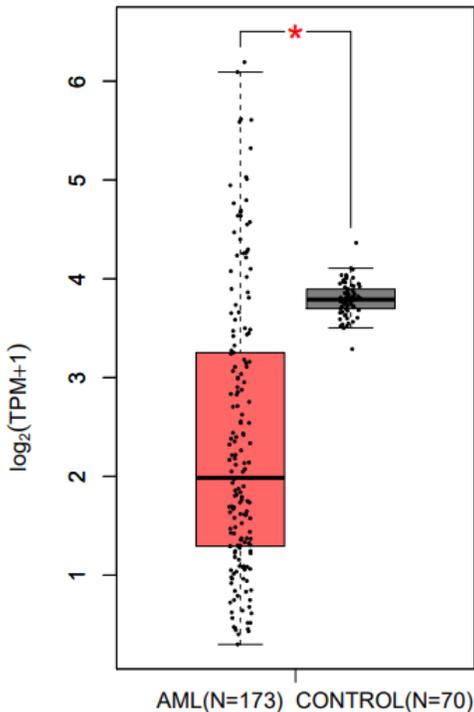


Figure 2

The comparison of expression level for LINC00649 between TCGA-AML cohort (N=173) and GTEx normal bone marrow samples (n=70).

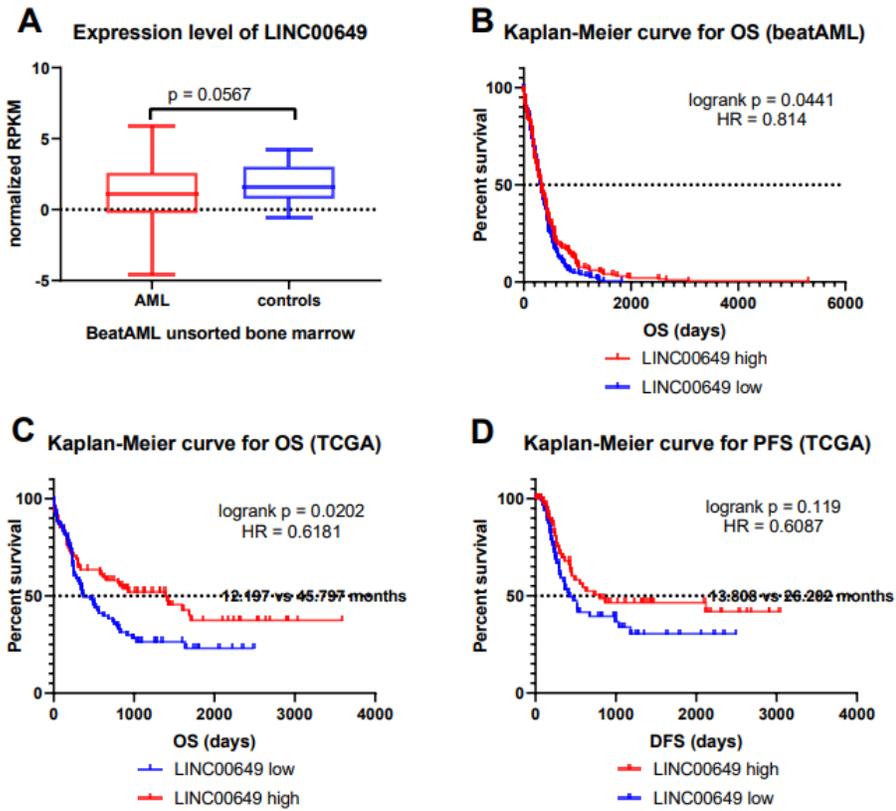


Figure 3

relevance of LINC00649 to AML, expression data of LINC00649 in bone marrow of AML patients versus healthy controls in BeatAML database (A). From TCGA and BeatAML database, the Kaplan-Meier curves of AML patients whose bone marrow cells express high versus low level of LINC00649, for OS (B&C) and PFS (D). The prognosis of LINC00649-low group is more unfavorable than that of LINC00649-high group.

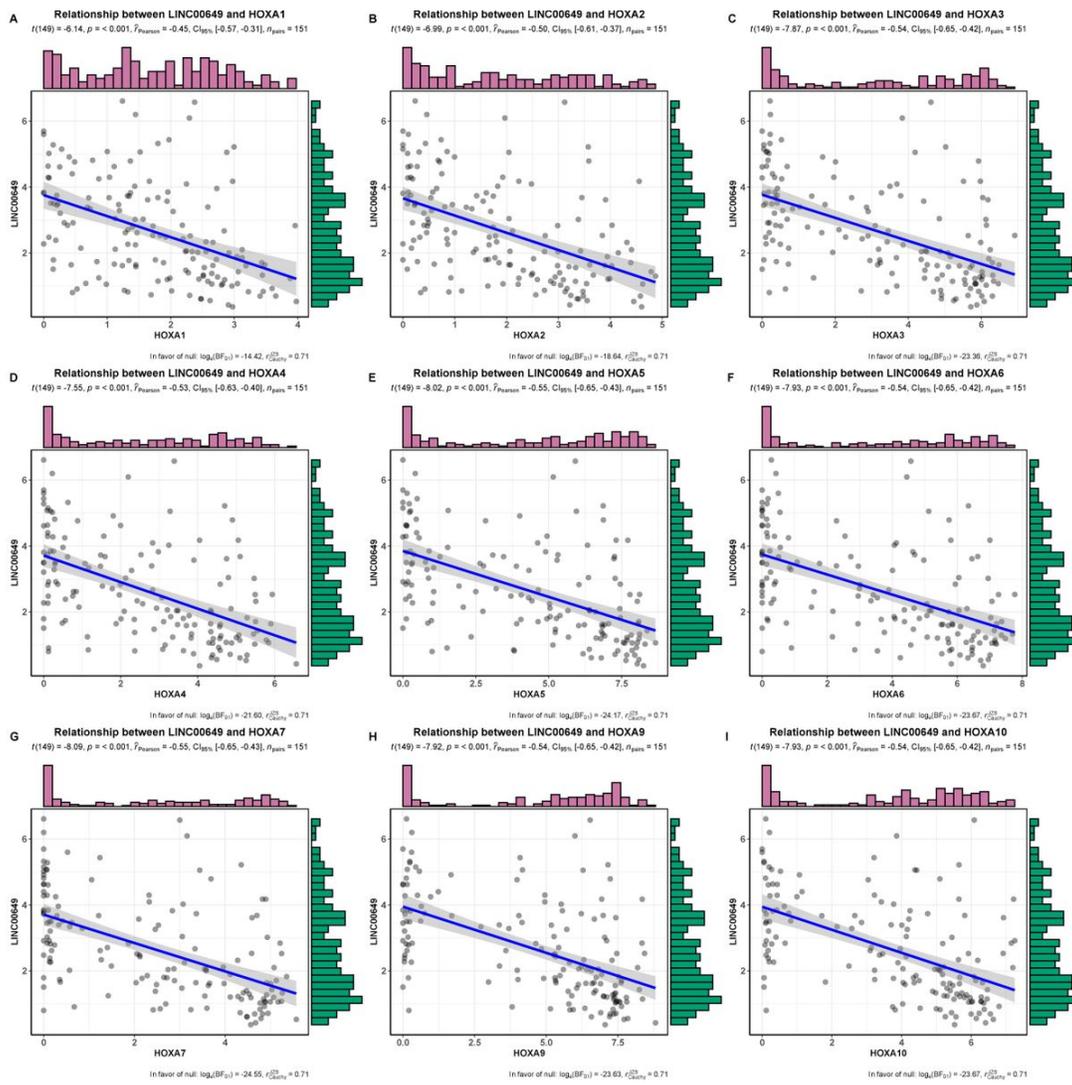


Figure 4

The Pearson correlation and linear regression for LINC00649 and HOXA family genes. The expression is significantly correlated ($p < 0.05$).

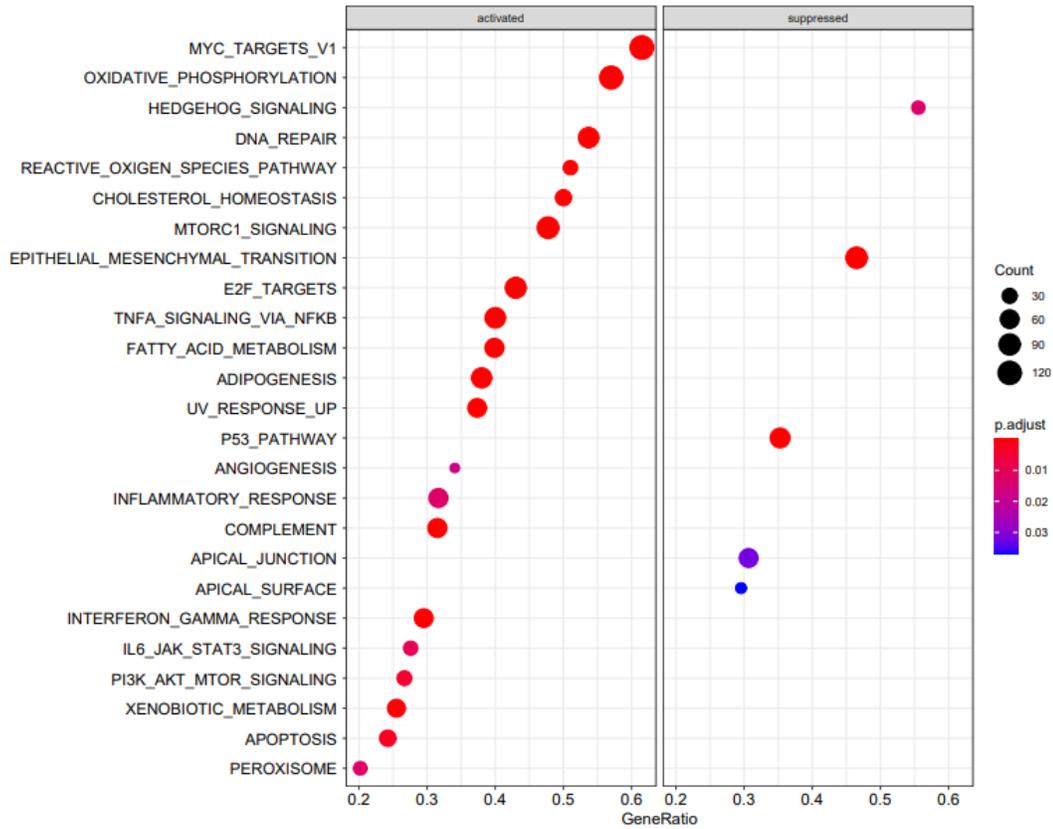


Figure 5
 The dotplot of GESA results for LINC00649 correlated genes. The size of dots stands for gene counts in the specific pathway, and the color represents correlated with adjusted p value. 19 activated and 6 suppressed pathways were identified.

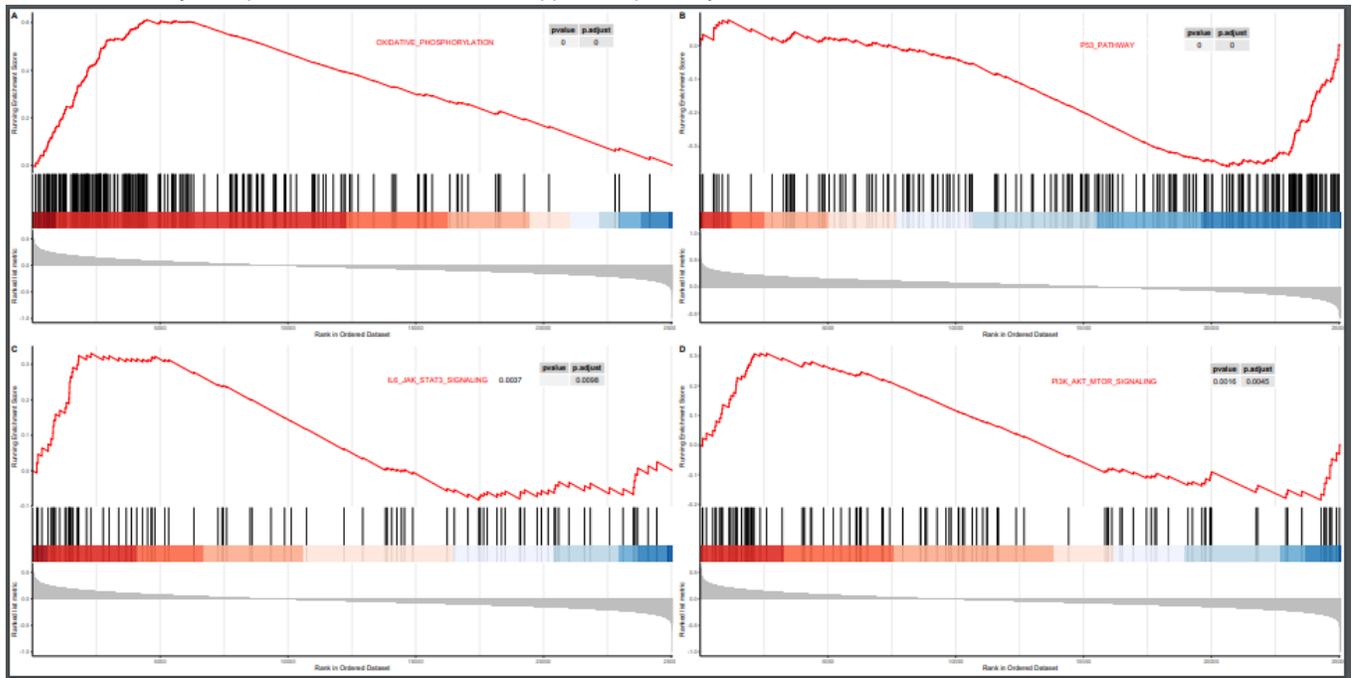


Figure 6
 The running enrichment score curve for oxidative phosphorylation, p53 signaling, IL-6/JAK/STAT3, PI3K/Akt/mTOR.

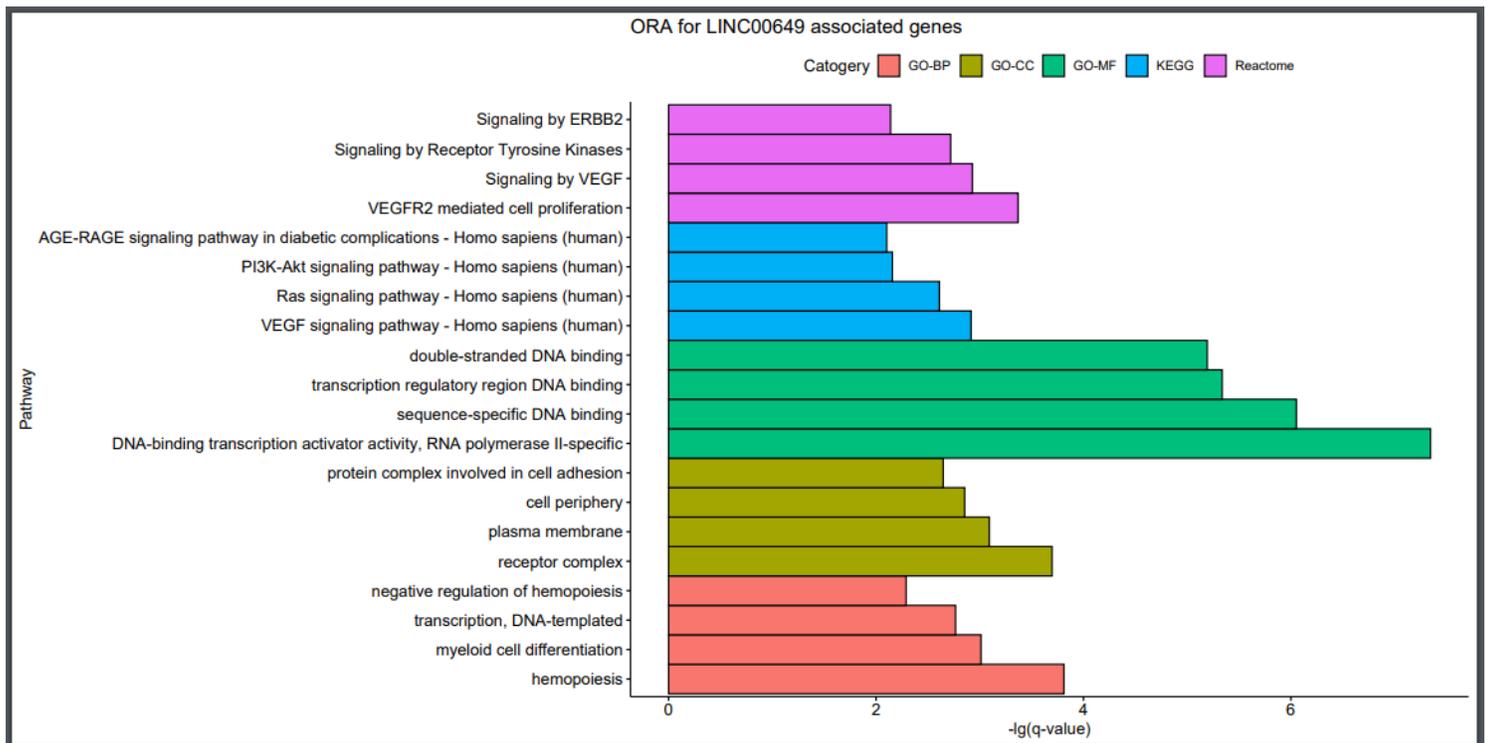


Figure 7

The dotplot of ORA (GO/KEGG/Reactome analysis) results for LINC00649 correlated genes. The size of dots stands for gene counts in the specific pathway, and the color represents correlated with adjusted p value.

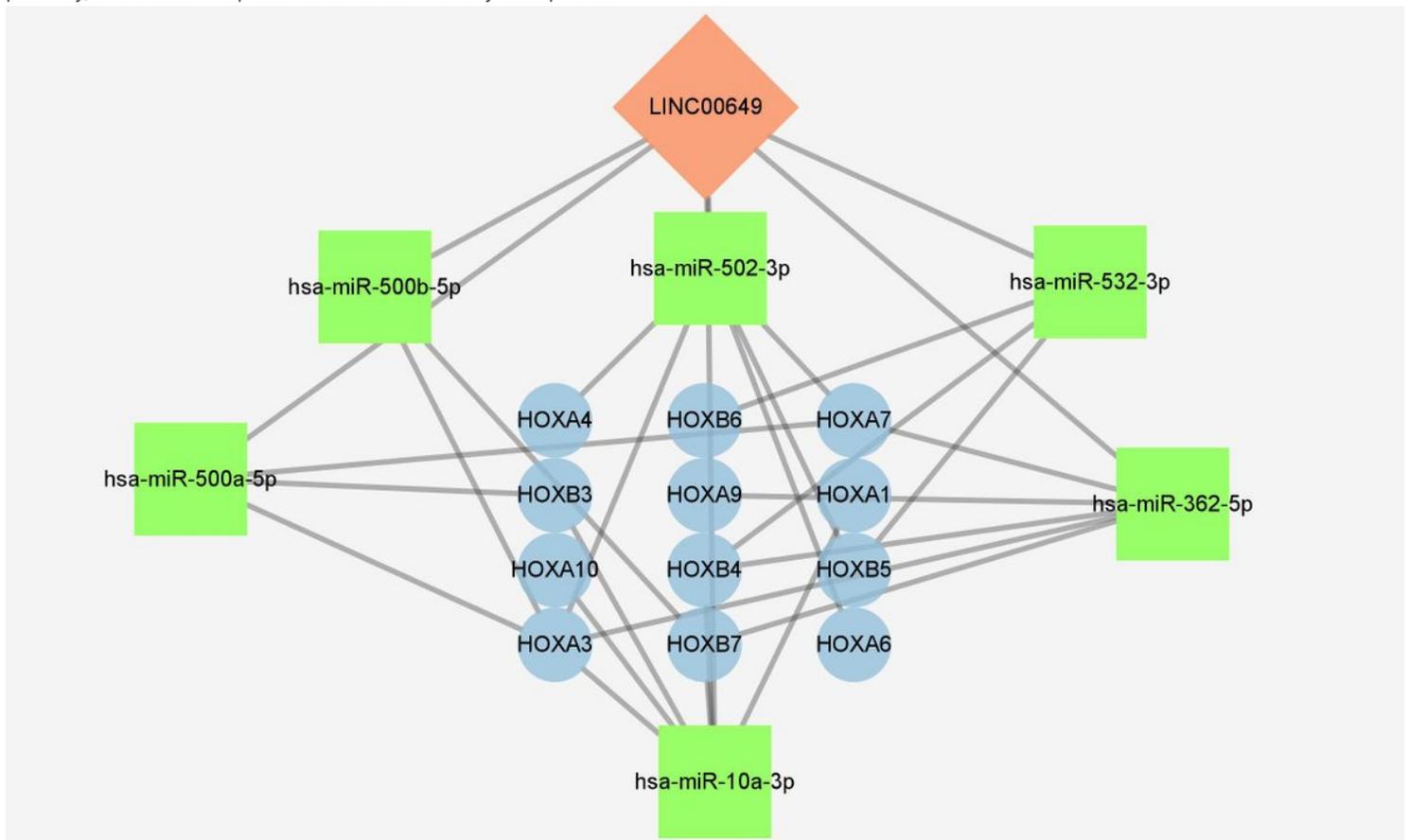


Figure 8

The ceRNA network of LINC00649 related with HOX family genes.

Time dependent ROC curve of riskScore

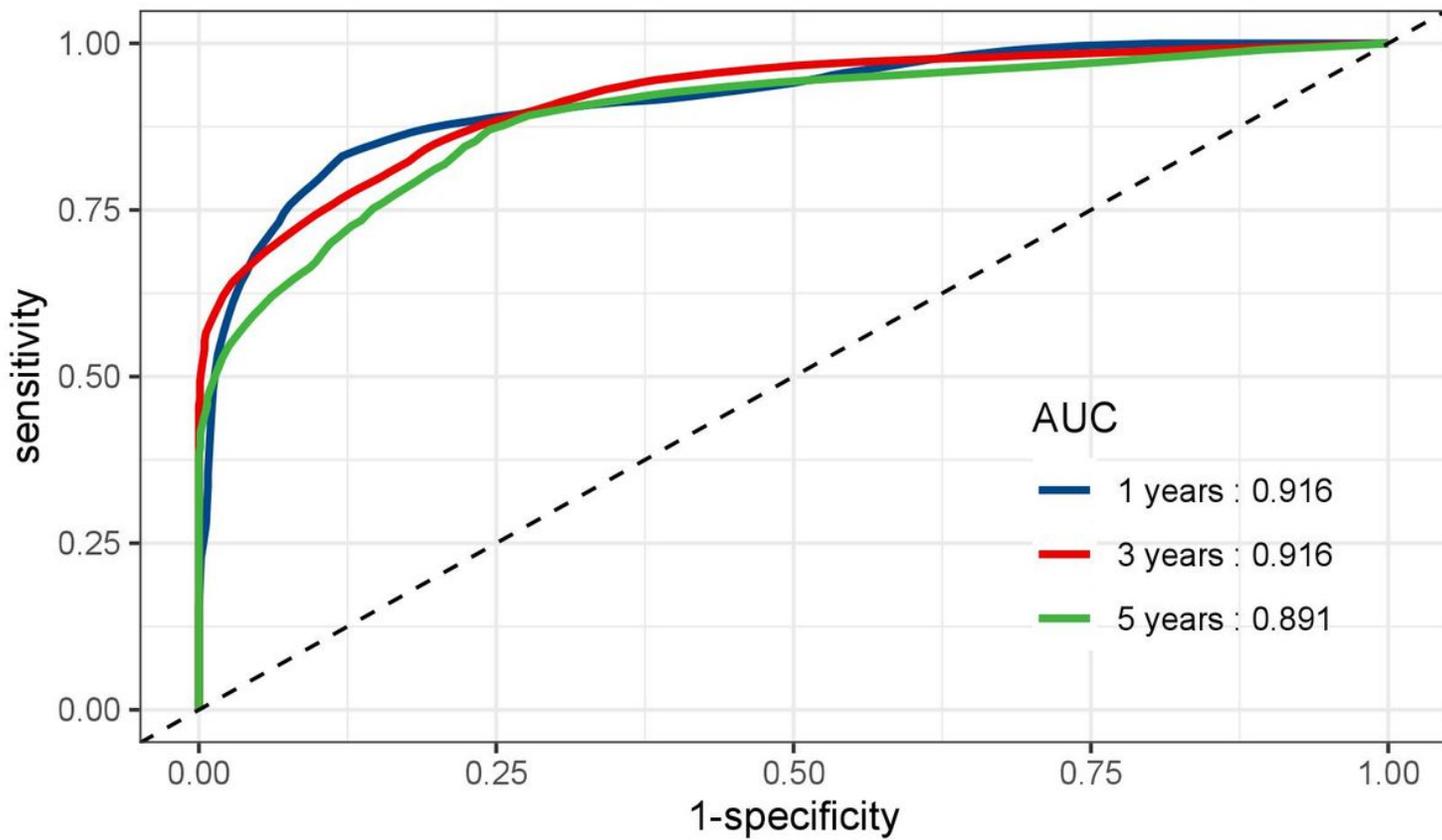


Figure 9

The time-dependent ROC curves of risk score by our prediction model for OS.

Time dependent ROC curve of riskScore

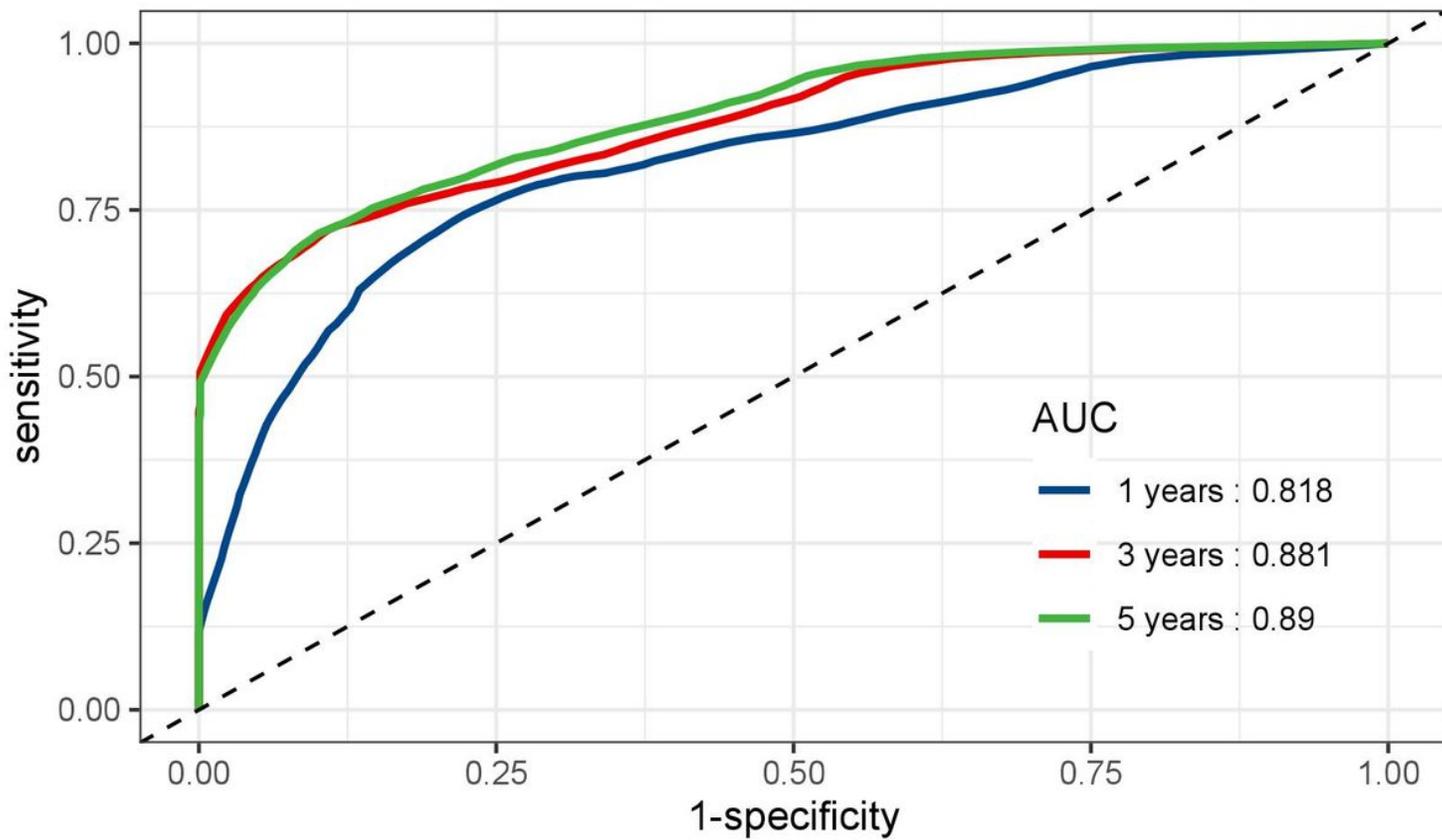


Figure 10

The time-dependent ROC curves of risk score by our prediction model for PFS.

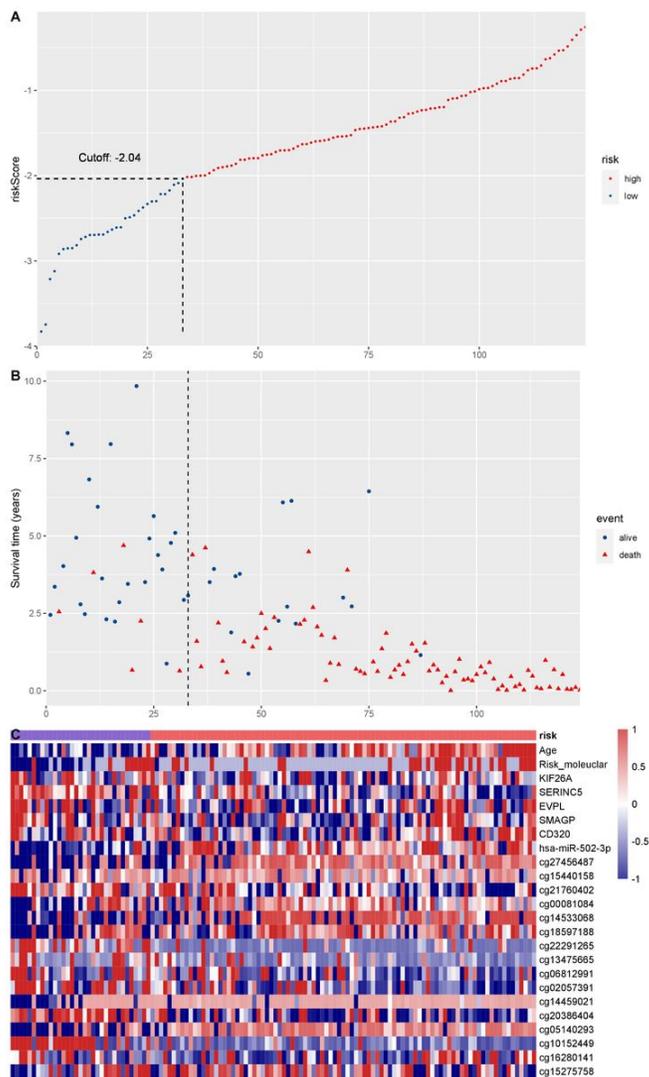


Figure 11

The distribution of risk score (A), patients' survival events (B), and risk to variable heatmap(C) for OS predicting model, the X-axis stands for individual AML patients ordered by ascending risk score.

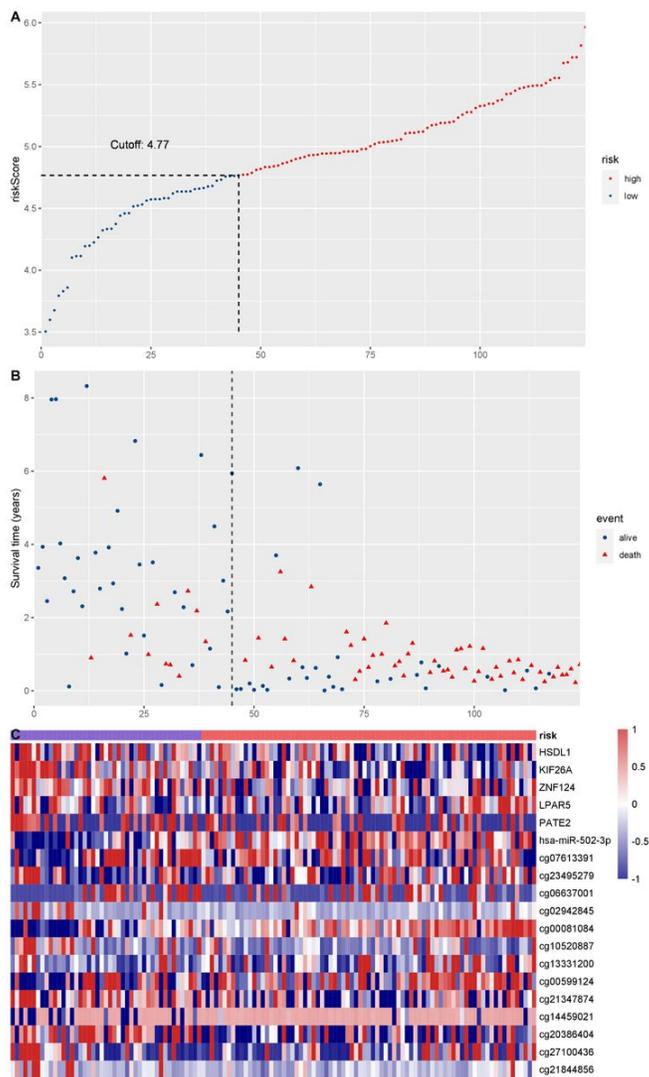


Figure 12

The distribution of risk score (A), patients' survival events (B), and risk to variable heatmap(C) for PFS predicting model, the X-axis stands for individual AML patients ordered by ascending risk score.

Kaplan-Meier curve for OS risk score

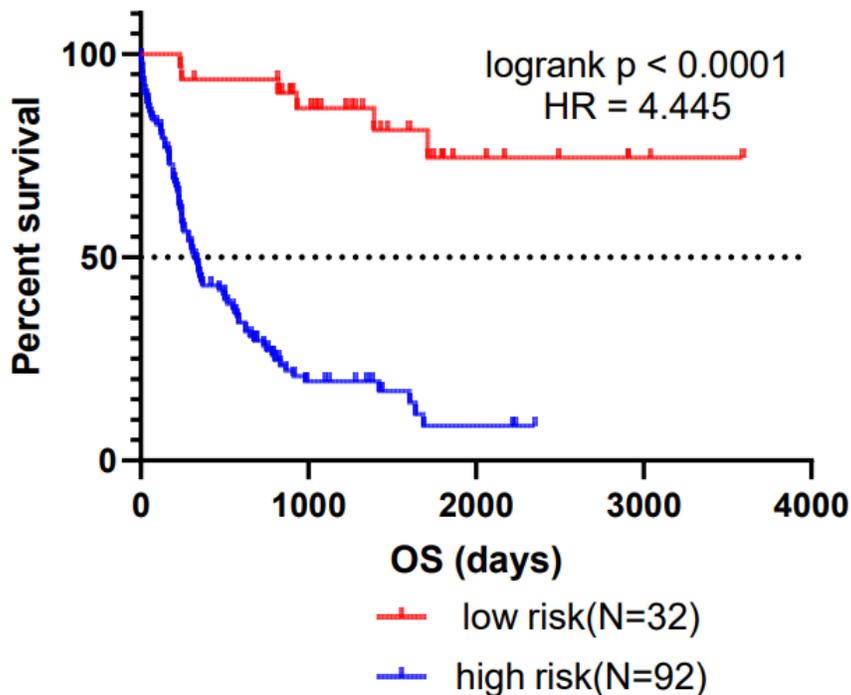


Figure 13

The Kaplan-Meier OS curves of AML patients for low and high-risk group dichotomized by risk score of our prediction model.

Kaplan-Meier curve for PFS risk score

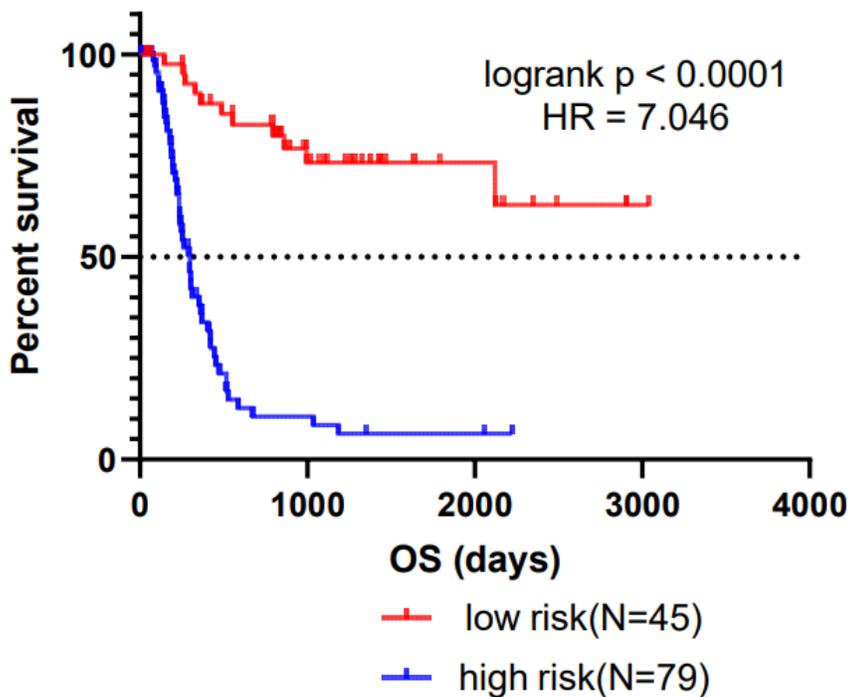


Figure 14

Kaplan-Meier PFS curves of AML patients for low and high-risk group dichotomized by risk score of our prediction model.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigure5.jpg](#)
- [SupplementaryFigure4.jpg](#)
- [SupplementaryFigure33.pdf](#)
- [SupplementaryFigure32.pdf](#)
- [SupplementaryFigure31.pdf](#)
- [SupplementaryFigure23.pdf](#)
- [SupplementaryFigure22.pdf](#)
- [SupplementaryFigure21.pdf](#)
- [SupplementaryFigure1.pdf](#)
- [SupplementaryTable4.xlsx](#)
- [SupplementaryTable3.xls](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable1.xlsx](#)