frontiers in Oncology

ORIGINAL RESEARCH
published: 19 June 2020
doi: 10.3389/fonc.2020.00829

High Expression of the SH3TC2-DT/SH3TC2 Gene Pair Associated With FLT3 Mutation and Poor Survival in Acute Myeloid Leukemia: An Integrated TCGA Analysis

Pengfei Yu[1,2], *Haifeng Lan*[1], *Xianmin Song*[3*] and *Zengkai Pan*[3,4*]

[1] Department of Hematology, Shanghai East Hospital, Tongji University School of Medicine, Shanghai, China, [2] Hannover Medical School, Institute of Virology, Hanover, Germany, [3] Department of Hematology, Shanghai General Hospital Affiliated to Shanghai Jiao Tong University, Shanghai, China, [4] Department of Hematology, Hemostasis, Oncology, and Stem Cell Transplantation, Hannover Medical School, Hanover, Germany

OPEN ACCESS

Edited by:
Cyrus Khandanpour, University Hospital Münster, Germany

Reviewed by:
Puneet Agarwal, Cincinnati Children's Hospital Medical Center, United States
Gregor Hoermann, Medical University of Vienna, Austria

***Correspondence:**
Zengkai Pan
panzengkai@hotmail.com
Xianmin Song
shongxm@sjtu.edu.cn

Specialty section:
This article was submitted to Hematologic Malignancies, a section of the journal Frontiers in Oncology

Received: *12 February 2020*
Accepted: *28 April 2020*
Published: *19 June 2020*

Citation:
Yu P, Lan H, Song X and Pan Z (2020) High Expression of the SH3TC2-DT/SH3TC2 Gene Pair Associated With FLT3 Mutation and Poor Survival in Acute Myeloid Leukemia: An Integrated TCGA Analysis. Front. Oncol. 10:829. doi: 10.3389/fonc.2020.00829

Fms-like tyrosine kinase 3 (FLT3) mutation is one of the most common mutations in acute myeloid leukemia (AML). However, the effect of FLT3 mutation on survival is currently still controversial and the leukemogenic mechanisms are still under further investigation. The aim of our study is to identify differentially expressed genes (DEGs) in FLT3-mutant AML and to find crucial DEGs whose expression level is related to prognosis for further analysis. By mining the TCGA-LAML dataset, 619 differentially expressed lncRNAs (DElncRNAs) and 1,428 differentially expressed mRNAs (DEmRNAs) were identified between FLT3-mutant and FLT3-wildtype samples. Through weighted gene correlation network analysis (WGCNA) and the following Cox proportional hazards regression analysis, we constructed the prognostic risk models to identify the hub DElncRNAs and DEmRNAs associated with AML prognosis. The presence of both SH3TC2 divergent transcript (SH3TC2-DT) and SH3TC2 in respective prognostic risk models promotes us to further study the significance of this gene pair in AML. SH3TC2-DT and SH3TC2 were identified to be coordinately high expressed in FLT3-mutant AML samples. High expression of this gene pair was associated with poor survival. Using logistic regression analysis, we found that high SH3TC2-DT/SH3TC2 expression was associated with FLT3 mutation, high WBC count, and intermediate cytogenetic and molecular–genetic risk. AML with SH3TC2-DT/SH3TC2 high expression showed enrichment of transcripts associated with stemness, quiescence, and leukemogenesis. Our study suggests that the SH3TC2-DT/SH3TC2 gene pair may be a possible biomarker to further optimize AML prognosis and may function in stemness or quiescence of FLT3-mutant leukemic stem cells (LSCs).

Keywords: acute myeloid leukemia, divergent transcription, the cancer genome atlas, prognostic signature, FLT3 mutation

INTRODUCTION

Acute myeloid leukemia (AML) is a group of clonal diseases that are characterized by heterogeneously genetic or epigenetic mutations. Fms-like tyrosine kinase 3 (FLT3) gene mutation is frequently found in adult AML. Two types of FLT3 activating mutations have been identified: internal tandem duplication (ITD) mutation and tyrosine kinase domain (TKD) point mutation. Twenty to thirty percent of adult AML patients have ITD mutation and 5–10% of them have TKD mutation (1, 2). FLT3 activating mutations were considered to result in constitutive activation of FLT3 and downstream signaling pathways to promote growth and survival of leukemic cells. However, to emphasize, the mechanism of mutant FLT3 activation in leukemogenesis has not been definitely confirmed. In clinical setting, FLT3 mutation has been closely related to leukocytosis and high blast percentage in marrow (3, 4). However, the prognostic significance of FLT3 mutation is still controversial. The prognostic value of FLT3-ITD mutation was reported to be associated with ITD allelic ratio and the presence of NPM1 mutation (1). Although ELN 2017 recommendation indicates that adult patients with concurrent mutant NPM1 and FLT3-ITD of low allelic ratio (<0.5) have favorable outcome (1), another study suggests that these patients have inferior survival compared to other favorable-risk AML patients, and concurrent DNMT3A mutation may serve as a risk modifier (5). The effect of FLT3-TKD mutation on survival is currently still unclear. Recent studies showed that FLT3-TKD mutation was associated with favorable outcome in the presence of NPM1 mutation (6), but associated with poor outcome in the presence of MLL-PTD (7). To further elucidate the potential genes in association with FLT3 mutation and prognosis of AML is meaningful.

Over these years, the genomic and transcriptomic features of AML have been comprehensively characterized by next-generation sequencing. Several publicly available databases from large patient cohorts have been constructed, which provide the opportunity to identify biomarkers in correlation with disease development, evolution, and treatment response. Identification of these potentially critical genes could provide hints for validation and clinical application (3, 4).

In this study, using a series of bioinformatic tools, we identified for the first time that the SH3TC2-DT/SH3TC2 gene pair was highly expressed in FLT3-mutant AML and presented as crucial genes associated with prognosis. The SH3TC2-DT/SH3TC2 gene pair may play a role in stemness or quiescence of FLT3-mutant LSCs. Our study provides the basis to further analyze the function of the SH3TC2-DT/SH3TC2 gene pair in the development and prognosis of FLT3-mutant AML.

MATERIALS AND METHODS

Data Collection and Preprocessing

A workflow chart of this study is shown in **Figure 1**. The data of 151 human AML samples used in this study were downloaded from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/), including the RNA sequencing data derived from the IlluminaHiSeq_RNASeq platform and clinical follow-up data, such as age, blast percentage, and survival time (data status as of Sept. 26, 2018) (3).

Identification of Differentially Expressed lncRNAs and mRNAs

The TCGA-LAML dataset consists of 43 FLT3-mutant AML and 108 FLT3-wildtype AML samples. R package "edgeR" was used to screen for differentially expressed genes (DEGs) between FLT3-mutant and FLT3-wildtype samples. The *q*-values use false discovery rate (FDR) to adjust the statistical significance of multiple hypothesis testing. Fold change (FC) ≥ 2 and adjusted $P < 0.05$ were considered to be statistically significant (8–10). The ensemble ID was switched to gene symbol according to the human genome assembly GRCh38.93. Afterwards, for DElncRNAs and DEmRNAs, volcano maps were depicted by "gplots" package in the R platform.

Functional Enrichment Analysis

R package "clusterProfiler" was used for Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis (11, 12). Enriched pathways with the significant level at $P < 0.05$ were selected. Gene-set enrichment analysis (GSEA) was used to identify the significant gene sets enriched in SH3TC2-DT or SH3TC2 high-expression phenotype (13, 14).

Weighted Gene Co-expression Network Analysis

R package "WGCNA" was used to construct co-expression modules for DEGs (15). Briefly, by "hclust" method to evaluate the expression matrix, the sample of TCGA-AB-2935 was removed from further analysis, because it was considered to be outliers as the cluster height was over 3×10^5 (as shown in **Figure S1**). The other samples of TCGA-LAML were clustered by methods of average linkage and Pearson's correlation. The weighted adjacency matrix between genes *i* and *j* was defined as $aij = |Cij|^{\beta}$ (*aij*: adjacency between gene *i* and gene *j*, *Cij*: the Pearson's correlation, β: soft-power threshold = 4). Afterwards, the clinical traits were accessed.

Three samples (TCGA-AB-2946, TCGA-AB-2895, and TCGA-AB-2810) were excluded from further analyses due to lack of clinical data. The scale independence and mean connectivity were calculated. Then, a topological overlap matrix (TOM) was transformed from the adjacency matrix (16). Finally, according to the TOM-based dissimilarity measure with a minimum size threshold of 30, average linkage hierarchical clustering dendrogram was constructed to classify genes with similar expression profiles into the same modules using the DynamicTreeCut algorithm. To identify the clinical significance of each module, gene significance (GS) was calculated to quantify associations of individual gene with clinical trait. Module significance (MS) was defined as the association between the module eigengenes (MEs) and the gene expression profiles. The different MEs were then correlated to the clinical traits (17, 18).

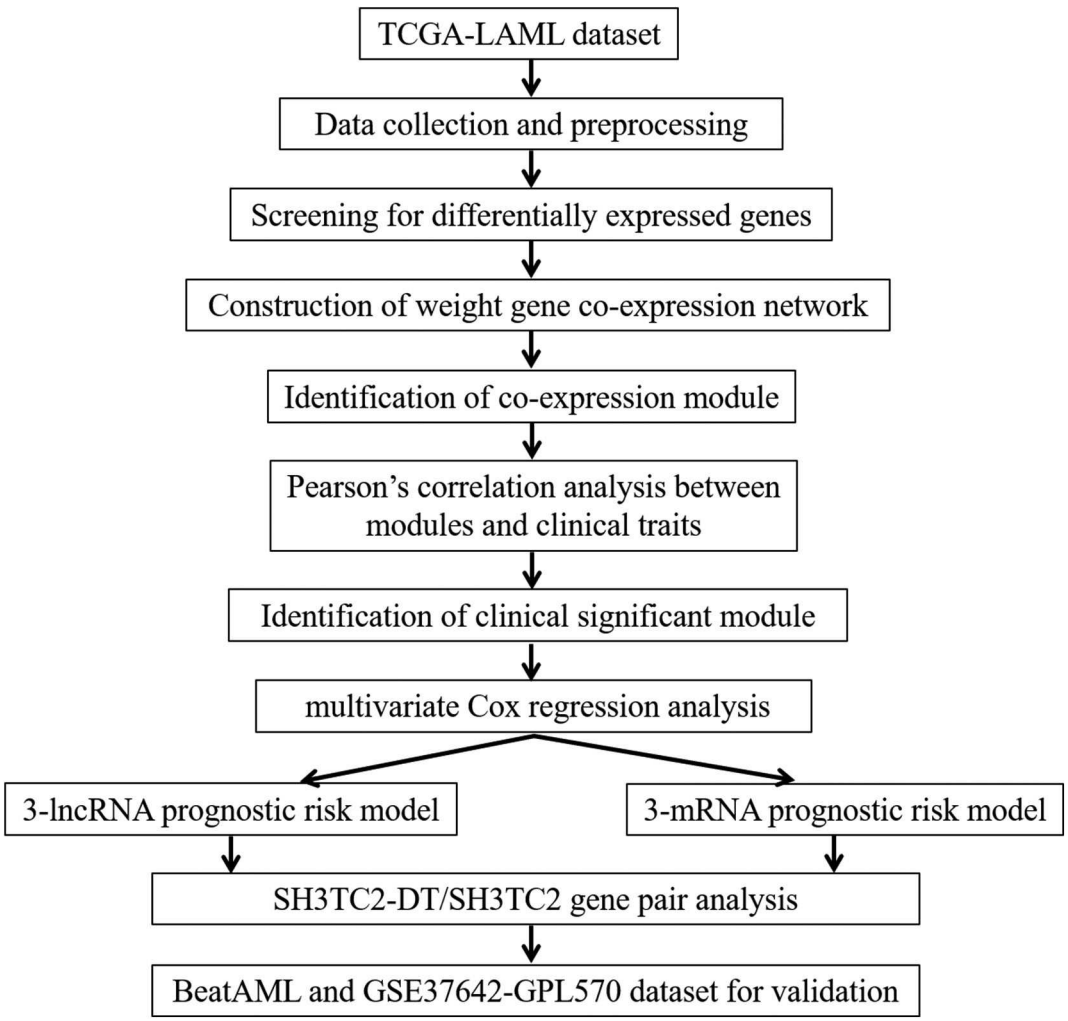

FIGURE 1 | Workflow chart of this study.

Cox Proportional Hazards Regression Analysis

The prognostic significance of each yellow module gene was evaluated by univariate Cox proportional hazards regression. Kaplan–Meier curve was depicted for each gene using R package "survival." Afterwards, multivariate Cox regression analysis was used to construct a 3-lncRNA prognostic risk model from prognosis-related lncRNAs. As to mRNA, least absolute shrinkage and selection operator (LASSO) regression analysis was first performed to select mRNAs in order to enhance the prediction accuracy of the prognostic risk model. Then, multivariate Cox regression analysis was used to construct a 3-mRNA prognostic risk model from the selected mRNAs. The AML samples were separated into high-risk and low-risk groups according to the median of the risk score. The risk score and survival status of each sample were depicted. Heatmap was generated by R package "pheatmap" to show the expression of risk genes in each sample. Kaplan–Meier analysis was conducted to identify the prognostic value of the risk model. The receiver operating characteristic (ROC) curve was used to evaluate the accuracy of the risk model or gene by R package "survivalROC." Nomogram was drawn to predict overall survival (OS) by the results of the multivariate Cox regression analysis.

SH3TC2-DT/SH3TC2 Gene Pair Analysis

For single gene analysis, the Student's *t*-test was used for expression comparison. Survival curve, ROC curve, and multivariate Cox regression analysis were performed as previously described. Logistic regression was used to analyze the association between the SH3TC2-DT/SH3TC2 expression and clinical characteristics.

To predict the potential targets of SH3TC2, we firstly analyzed the DEGs between SH3TC2 high-expression group ($n = 76$) and SH3TC2 low expression group ($n = 75$) by R package "edgeR." Then, DEGs list was annotated by the module "UCSC_TFBS" under the "Protein_Interactions" function of DAVID (19, 20).

Significantly enriched transcription factors (TFs) in DEGs were identified (adjusted $P < 0.05$). The TFs network was visualized by Cytoscape (3.7.1).

The RNA sequencing data and clinical follow-up data of BeatAML dataset were downloaded from Vizome (http://www.vizome.org/) (21) and TCGA (https://portal.gdc.cancer.gov/) to validate the differential expression of SH3TC2-DT/SH3TC2 between FLT3-ITD and FLT3-wildtype AML. DEGs were calculated as previously described. The GSE37642-GPL570 AML dataset (22) was used to validate the association between SH3TC2 expression level and OS. AML survival and microarray data were downloaded from the GEO platform (https://www.ncbi.nlm.nih.gov/geo/). The 136 AML samples were separated into two groups according to the median of SH3TC2 expression level. Kaplan–Meier curve was used to compare the OS between high ($n = 68$) and low ($n = 68$) SH3TC2 expression AML samples.

All statistical tests and graphing were performed by R and GraphPad Prism 7.0. $P < 0.05$ was considered of statistical significance. In the figures, statistical significance was shown as follows: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$.

RESULTS

DEmRNAs and DElncRNAs Between FLT3-Mutant and FLT3-Wildtype AML

To identify the transcriptomic features related to FLT3 mutation, R package "edgeR" was used to perform differential expression analysis in TCGA-LAML dataset. Compared with FLT3-wildtype AML, 619 lncRNAs (113 upregulated and 506 downregulated) and 1,428 mRNAs (194 upregulated and 1,234 downregulated) were significantly differentially expressed in FLT3-mutated AML (FC> 2, FDR< 0.05) (**Figures 2A,B**). KEGG analysis

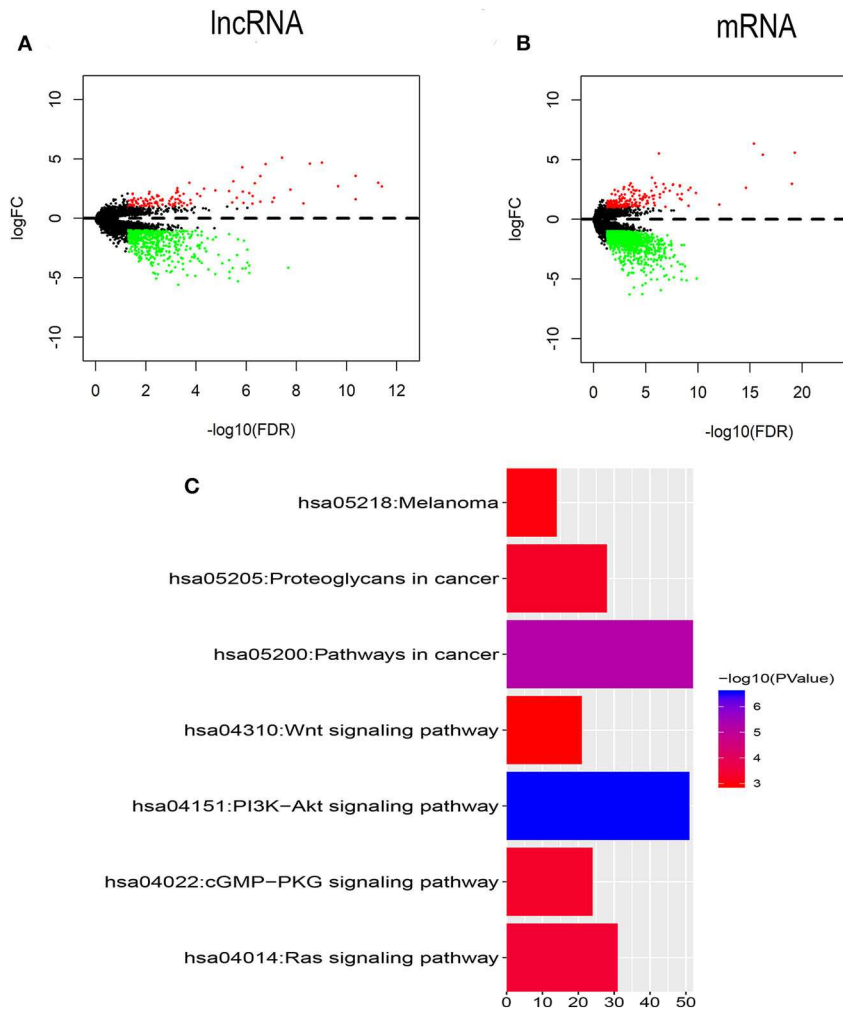

FIGURE 2 | Identification of DElncRNAs and DEmRNAs between FLT3-mutant and FLT3-wildtype AML. **(A,B)** Volcano plots of DElncRNAs **(A)** and DEmRNAs **(B)**. The cutoff values are FC> 2 and FDR< 0.05. Red and green dots indicate upregulated and downregulated RNAs, respectively. Black dots denote no difference in RNA expression. FC, fold change; FDR, false discovery rate; lncRNA, long non-coding RNA; mRNA, messenger RNA; DElncRNAs, differentially expressed lncRNAs; DEmRNAs, differentially expressed mRNAs. **(C)** KEGG pathways in relation with tumorigenesis enriched for DEmRNAs are shown. *X* axis showing the number of DEmRNAs enriched in each pathway. KEGG, Kyoto Encyclopedia of Genes and Genomes.

revealed that DEmRNAs were enriched in pathways closely related to tumorigenesis, such as Wnt signaling pathway, PI3K-Akt signaling pathway, and Ras signaling pathway (**Figure 2C**), suggesting the possible function of FLT3 mutation in AML pathogenesis.

Construction of Weighted Co-expression Network and Identification of Survival-Associated Module

R package "WGCNA" was used to construct co-expression modules of DEGs and to further identify prognosis-related modules. First, the samples of TCGA-LAML were clustered by methods of average linkage and Pearson's correlation (**Figure 3A**). To ensure a scale-free network, $\beta = 4$ (scale free $R^2 = 0.9$) was set as the soft-thresholding parameter (**Figure 3B**). After merging modules with high similarity, a total of 27 modules with the size ranging from 31 to 327 genes were generated by the average linkage hierarchical clustering (**Figure 4A**). The non-co-expressed genes were grouped into "gray" module and removed from further analysis. The heatmap of 400 randomly selected DEGs showed high degree of topological overlap of co-expressed genes in each module (**Figure 4B**). The

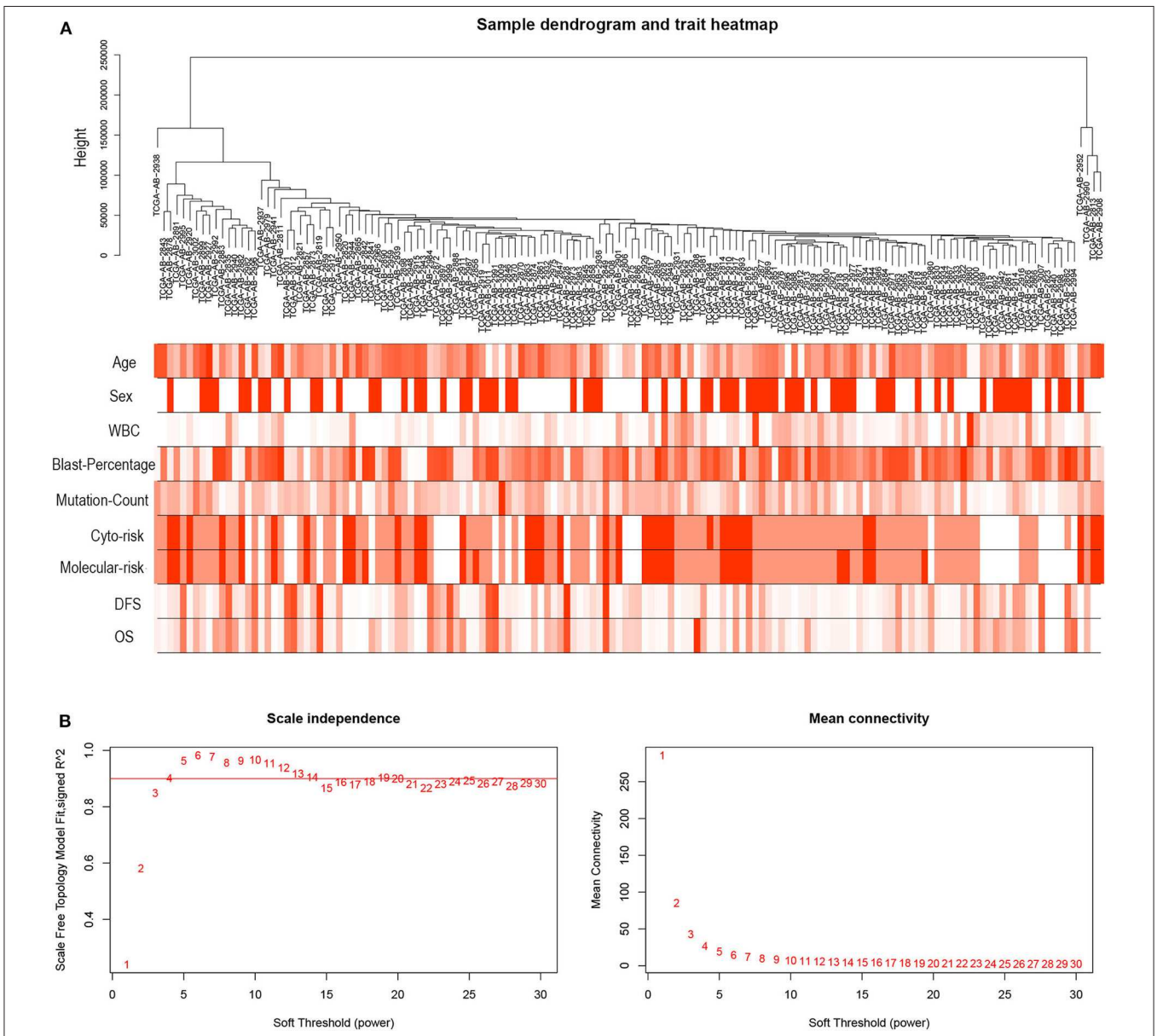

FIGURE 3 | Hierarchical clustering tree and soft-thresholding values estimation. **(A)** Hierarchical clustering tree of the TCGA-LAML samples. Dendrogram tips are denoted with TCGA-LAML sample names. The nine bands below dendrogram indicate clinical traits of each sample. WBC, white blood cell; Cyto-risk, cytogenetic risk; Molecular-risk, molecular genetic risk; DFS, disease-free survival; OS, overall survival. **(B)** Scale independence and mean connectivity across different soft-thresholding values (β).

eigengene adjacency heatmap showed relationships between the 27 co-expression modules (**Figure 4C**). At last, the correlation between these modules and clinical traits was determined (**Figure 4D**). The yellow module was correlated with high white blood cell (WBC) count [Pearson correlation coefficient (PCC) = 0.26, $P = 0.002$] and blast percentage in bone marrow (PCC = 0.2, $P = 0.01$), but not correlated with age, sex, mutation count, cytogenetic risk, or molecular–genetic risk. Of note, this module had the highest association with worse disease-free survival (DFS) (PCC = −0.21, $P = 0.009$)

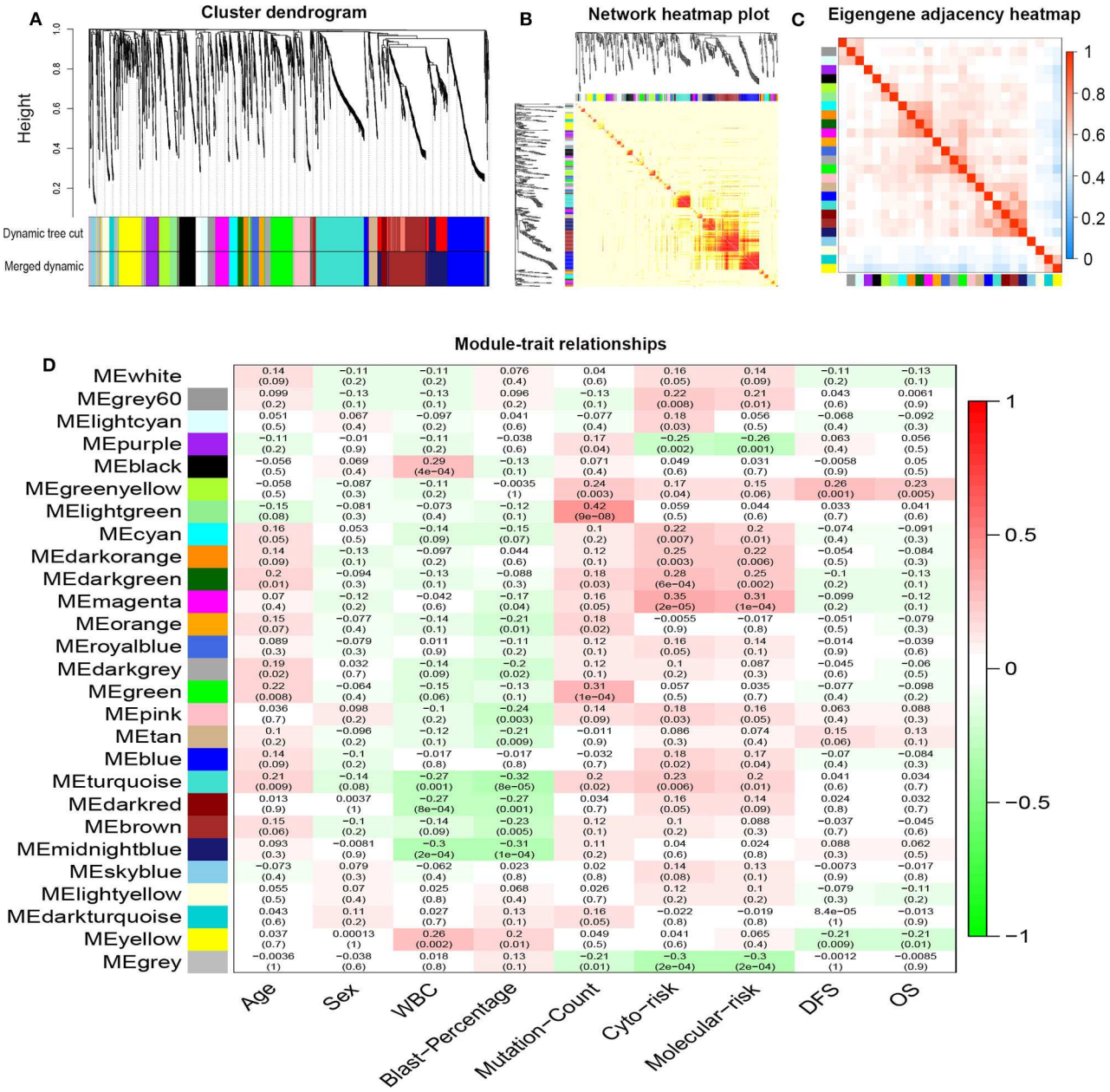

FIGURE 4 | Network construction of co-expressed genes and module–trait relationships. **(A)** Cluster dendrogram and the co-expression network modules produced by average linkage hierarchical clustering of DEGs based on topological overlaps. Each branch within the dendrogram indicates a single gene. Height depicts the Euclidean distance. Each color indicates a single module that contains weighted co-expressed genes. **(B)** Heatmap view of topological overlap. Four hundred randomly selected genes grouped into modules displayed color codes, which are shown beneath the cluster dendrogram. Dark yellow and red represent a high degree of topological overlap. **(C)** Eigengene adjacency heatmap. The heatmap represents the relationship among distinctive co-expression modules. **(D)** Module–trait relationships. Each row represents a color module and each column indicates a clinical trait. Each cell contains R^2-values of Pearson correlations between the modules and clinical features and the corresponding P-value in parentheses. Gradient color of each cell indicates the R^2-values of Pearson correlations (red = 1, green = −1). DEGs, differentially expressed genes; ME, module eigengene.

and OS (PCC = −0.21, $P = 0.01$), and was selected for further analysis.

Prognostic Significance of Each Gene in the Yellow Module

A total of 43 genes in the yellow module were identified to be significantly related to OS by univariate Cox proportional hazards regression. Among them, high expression of 12 lncRNAs (SH3TC2-DT, LINC00982, LINC00899, GRM7-AS1, MIR155HG, AC005392.2, LINC01132, AL133353.1, AF064858.1, LINC01979, AC103702.1, and DLGAP1-AS3) and 31 mRNAs (TMEM273, PRDM16, SH3TC2, LCT, ENPP2, CCDC113, ATRNL1, TRIM16, LDLRAD2, MPZL2, HOXB6, SCHIP1, ARHGEF5, CAMK2A, NLRP2, CCL1, H2AFY2, GLI2, APOL4, HOXA7, PBX3, PDGFD, HOXB2, HOXB5, LCN8, TRIM15, GOLGA8B, CACNG4, FAM47E-STBD1, REN, and FAM47E) were associated with poor OS (**Figure 5**, **Tables S1, S2**). These lncRNAs and mRNAs were then subjected to further construct lncRNA or mRNA prognostic risk model.

Establishment of the lncRNA Prognostic Risk Model

By multivariate Cox proportional hazards regression analysis, we built a 3-lncRNA prognostic risk model to predict OS in AML cases as follows: risk score = (0.006899 × expression level of SH3TC2-DT) + (0.00026 × expression level of AF064858.1) + (0.016446 × expression level of AL133353.1) (**Table 1**). Of note, SH3TC2-DT is the most significant prognosis-associated lncRNA in this model ($P = 0.000234$) (**Table 1**). A total of 148 patients were categorized into high-risk ($N = 74$) and low-risk ($N = 74$) group according to the median of risk score (**Figures 6A–C**). The OS rate of high-risk patients was significantly lower compared with low-risk patients (**Figure 6D**). Multivariate Cox regression analysis revealed that age and the lncRNA risk score were independent prognostic factors affecting OS. The lncRNA risk score had a greater influence on survival than WBC count,

TABLE 1 | lncRNA prognostic risk score model.

lncRNA	Coefficient	HR in OS	*P*-value
SH3TC2-DT	0.006899	1.0069 (1.0032–1.0106)	0.0002
AF064858.1	0.00026	1.0003 (1.0000–1.0005)	0.0386
AL133353.1	0.016446	1.0166 (0.9991–1.0344)	0.0637

HR, hazard ratio; OS, overall survival. Numbers in parentheses show the 95% CI (confidence interval) of HR.

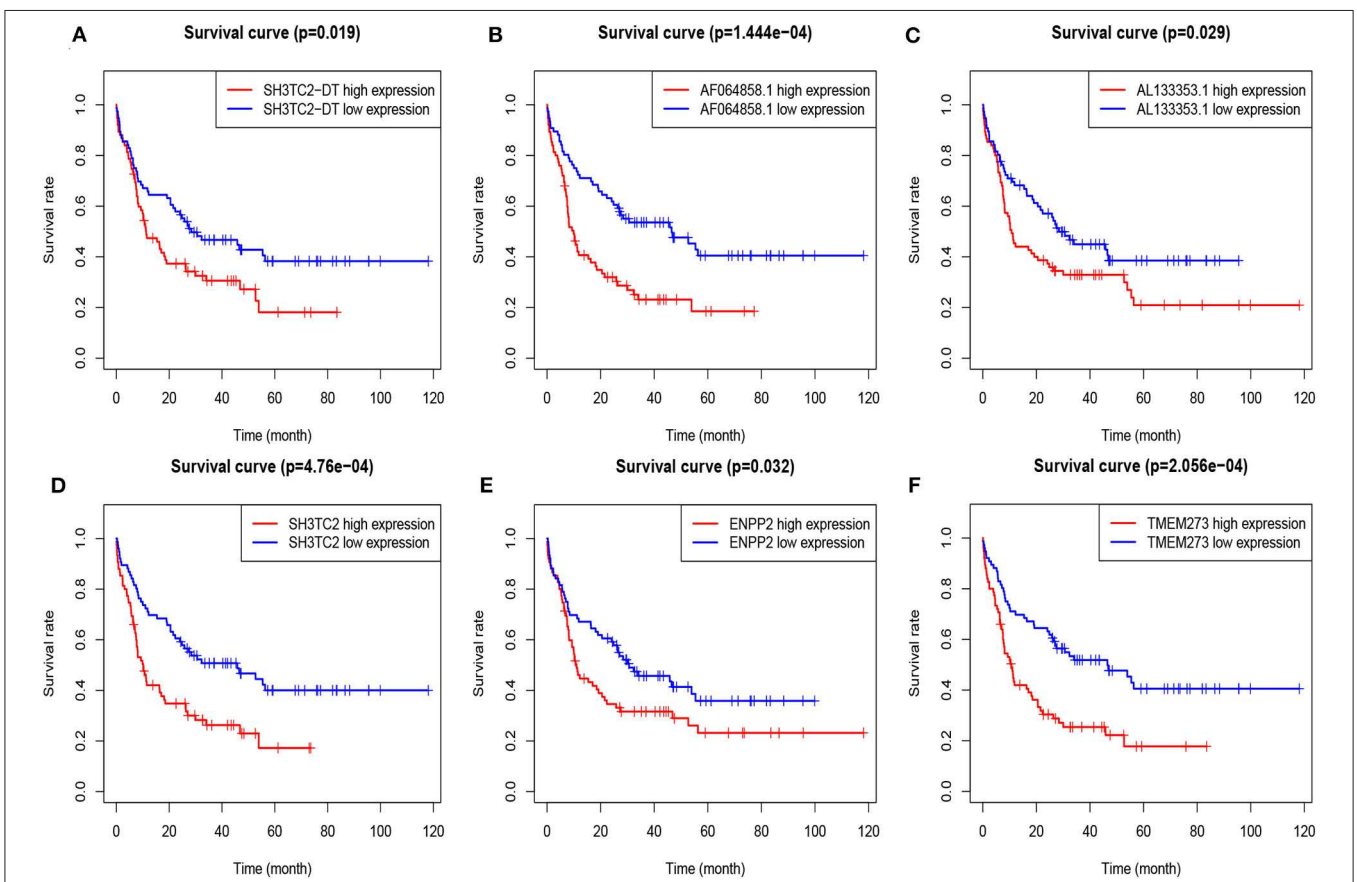

FIGURE 5 | Overall survival analysis of TCGA-LAML cohort based on expression of yellow module genes by Kaplan–Meier plotter. The patients were grouped into high-expression and low-expression group based on median expression of each gene. Survival plots of selected lncRNAs **(A–C)** and mRNAs **(D–F)** are shown.

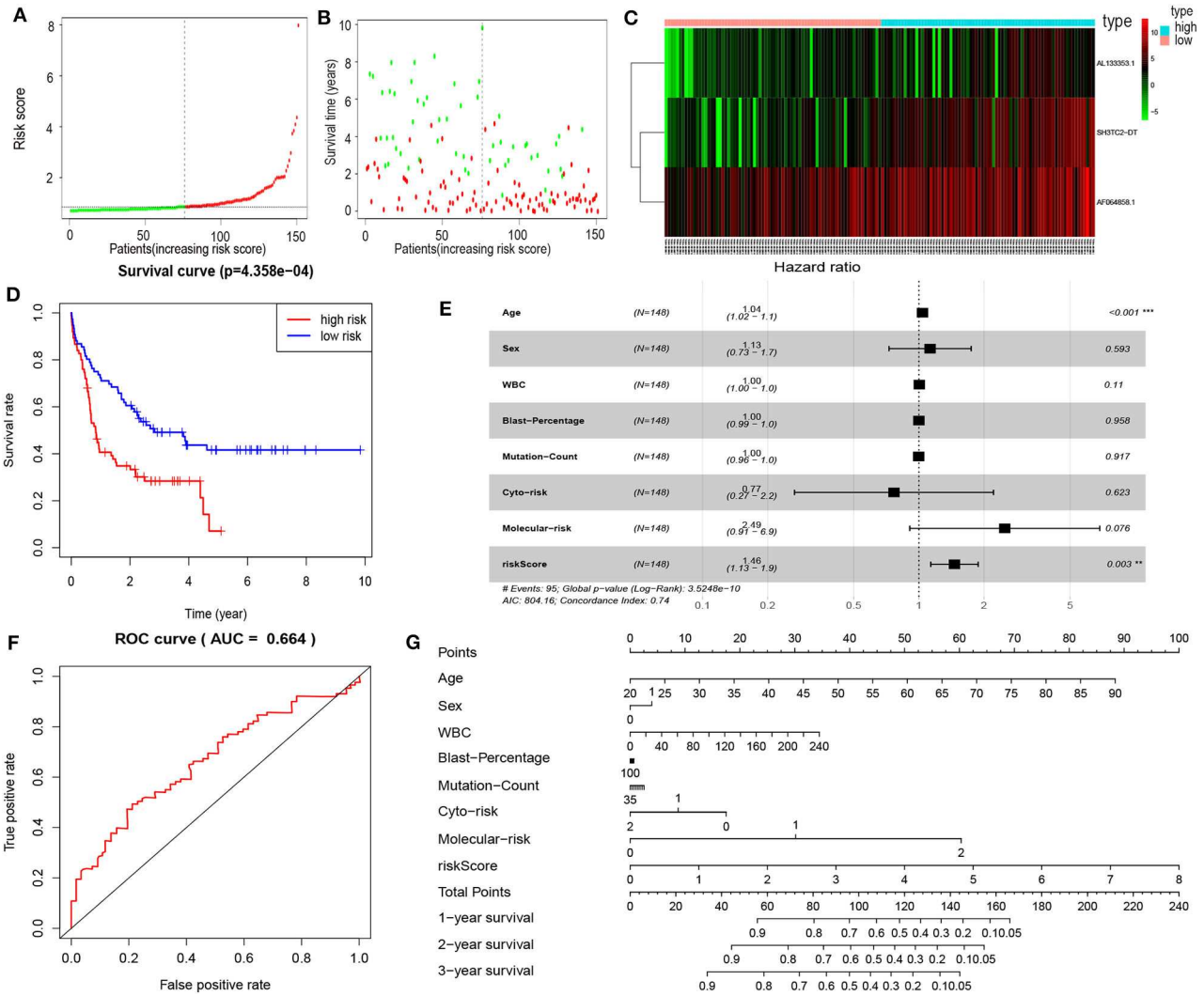

FIGURE 6 | Cox proportional hazards regression analysis of lncRNAs. **(A)** The lncRNA risk score distribution of TCGA-LAML samples. **(B)** Survival status and time of all AML patients across lncRNA risk score. The AML patients were separated into two groups according to the median of risk score. Red dots indicate high-risk samples ($N = 74$). Green dots indicate low-risk samples ($N = 74$). **(C)** Heatmap and hierarchical clustering of lncRNAs from the prognostic risk model were generated. Values are normalized with $\log_{10}$. The right longitudinal axis: the names of lncRNAs; the left longitudinal axis: the clustering information of lncRNAs. Red and green denote upregulated and downregulated lncRNAs, respectively. high, high-risk samples; low, low-risk samples. **(D)** Kaplan–Meier curve to compare OS of high-risk with low-risk samples ($P = 4.358e{-}04$). **(E)** Forest plot of multivariable Cox regression analysis. The squares on the transverse lines show the hazard ratio (HR), and the transverse lines represent 95% confidence interval (CI). **(F)** ROC curves showing predictive value of the prognostic risk model. **(G)** Nomogram of OS prediction in AML. $^{**}P < 0.01$; $^{***}P < 0.001$.

cytogenetic risk, and molecular risk (**Figure 6E**). The area under ROC curve was 0.664, showing a high predictive value of the risk model (**Figure 6F**). Nomogram was drawn to visualize the result of multivariate Cox regression analysis (**Figure 6G**). Furthermore, Kaplan–Meier curves also confirmed that these three lncRNAs were predictive indicators for OS (**Figures 5A–C**).

Establishment of the mRNA Prognostic Risk Model

In order to enhance the prediction accuracy of the prognostic risk model, we firstly performed LASSO regression analysis and selected four mRNAs (SH3TC2, ENPP2, TMEM273, and PRDM16) for further analysis from the 31 mRNAs with prognostic value in yellow module (**Figure S2**). Afterwards, through multivariate Cox proportional hazards regression analysis, we identified a 3-mRNA prognostic risk model to predict OS in AML cases as follows: risk score= (0.000612 × expression level of SH3TC2) + (0.000507 × expression level of ENPP2) + (0.000277 × expression level of TMEM273) (**Table 2**). A total of 148 patients were categorized into high-risk ($N = 74$) and low-risk ($N = 74$) group according to the median of risk score (**Figures 7A–C**). The OS rate of high-risk patients was significantly lower compared with low-risk patients (**Figure 7D**). Multivariate Cox regression analysis revealed that

age, WBC count, molecular risk, and the mRNA risk score were independent prognostic factors affecting OS. The mRNA risk score had a greater influence on survival than WBC count, cytogenetic risk, and molecular risk (**Figure 7E**). The area under ROC curve was 0.744, showing a high predictive value of the risk model (**Figure 7F**). Finally, Nomogram was drawn to visualize the result of multivariate Cox regression analysis (**Figure 7G**). Furthermore, Kaplan–Meier curves also confirmed that these three mRNAs were predictive indicators for OS (**Figures 5D–F**).

TABLE 2 | mRNA prognostic risk score model.

mRNA	Coefficient	HR in OS	*P*-value
SH3TC2	0.000612	1.0006 (1.0002–1.0010)	0.0016
ENPP2	0.000507	1.0005 (1.0002–1.0009)	0.0042
TMEM273	0.000277	1.0003 (1.0001–1.0004)	0.0007

HR, hazard ratio; OS, overall survival. Numbers in parentheses show the 95% CI (confidence interval) of HR.

The SH3TC2-DT/SH3TC2 Gene Pair Is an Independent Prognostic Factor for AML

The presence of SH3TC2-DT and SH3TC2 in respective prognostic risk models emphasized the importance of this divergent lncRNA/mRNA gene pair in prognosis of FLT3-mutant AML (**Tables 1**, **2**), but interestingly, the expression of the

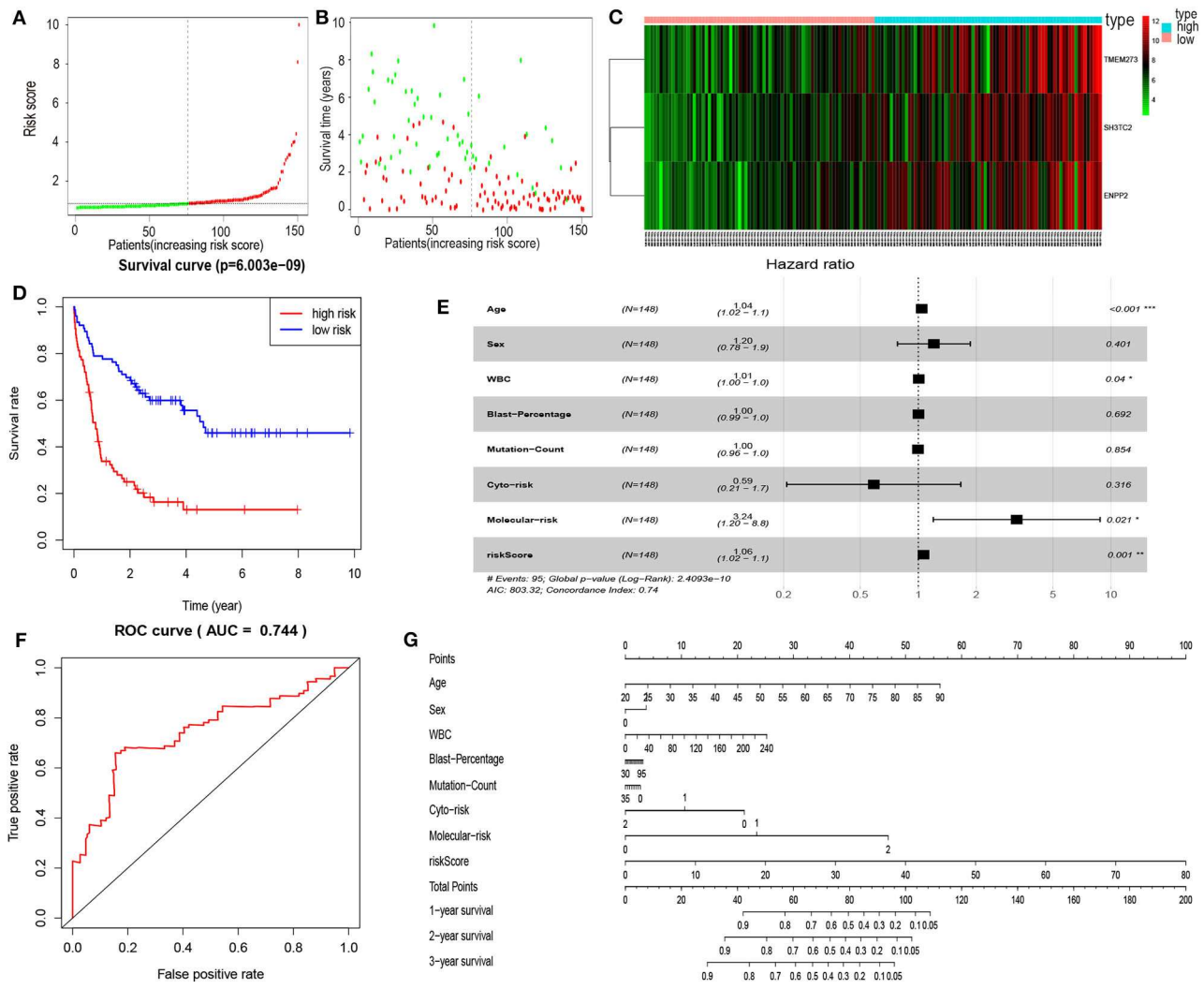

FIGURE 7 | Cox proportional hazards regression analysis of mRNAs. **(A)** The mRNA risk score distribution of TCGA-LAML samples. **(B)** Survival status and time of all AML patients across mRNA risk score. The AML patients were separated into two groups according to the median of risk score. Red dots indicate high-risk samples ($N = 74$). Green dots indicate low-risk samples ($N = 74$). **(C)** Heatmap and hierarchical clustering of the mRNAs in prognostic risk model were generated. Values used are normalized with $\log_{10}$. The right longitudinal axis: the names of mRNAs; the left longitudinal axis: the clustering information of mRNAs. Red and green denote upregulated and downregulated mRNAs, respectively. high, high-risk samples; low, low-risk samples. **(D)** Kaplan–Meier curve to compare the OS of high-risk and low-risk samples ($P = 6.003e{-}09$). **(E)** Forest plot of multivariable Cox regression analysis. The squares on the transverse lines show the HR, and the transverse lines represent 95% CI. **(F)** ROC curves showing predictive value of the prognostic risk model. **(G)** Nomogram of OS prediction in AML. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

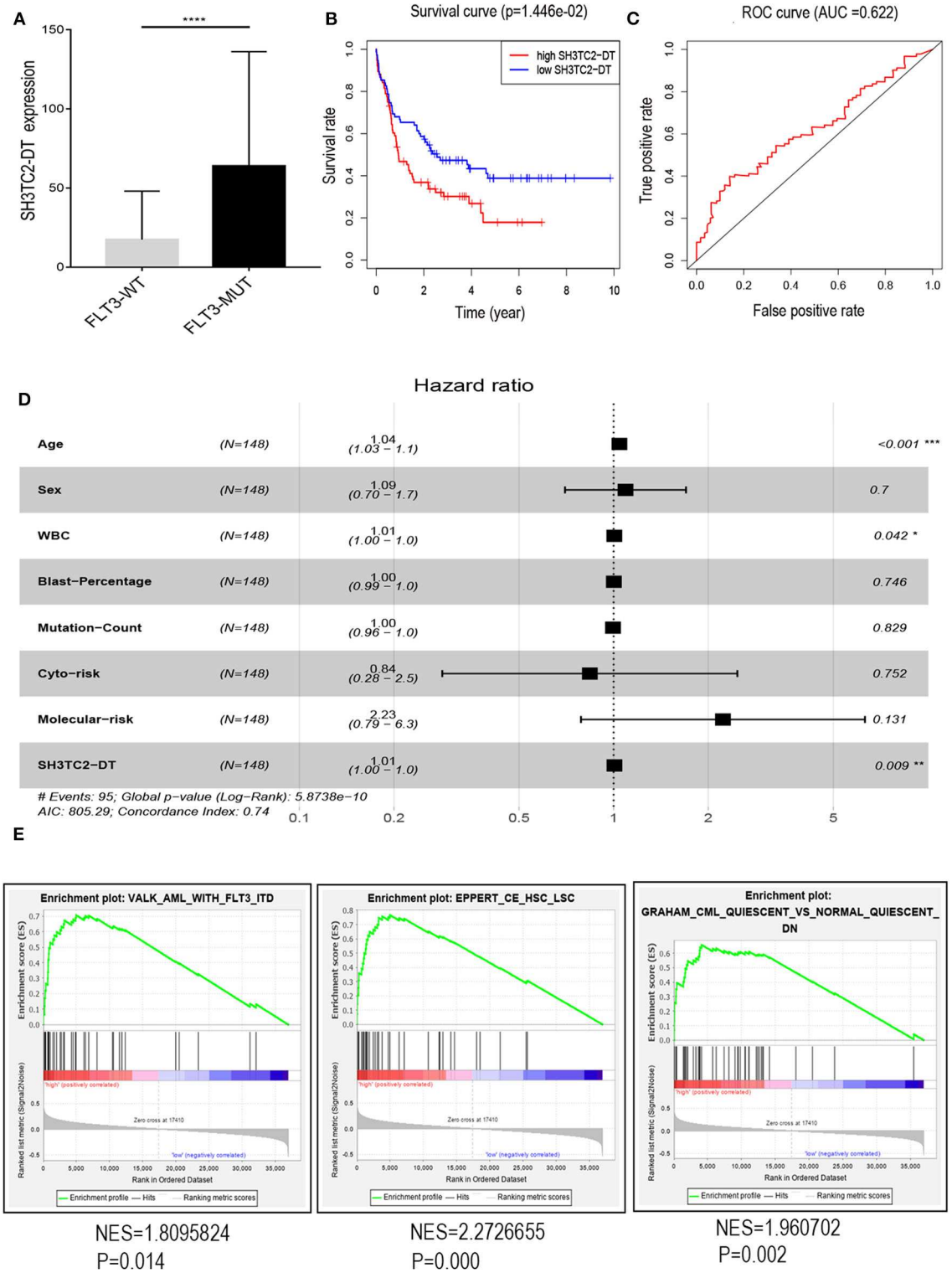

FIGURE 8 | The association of SH3TC2-DT expression with overall survival and gene-set enrichment. **(A)** The normalized RNA-sequencing expression value of SH3TC2-DT in FLT3-WT ($N = 108$) and FLT3-MUT ($N = 43$) AML patients. **(B)** Kaplan–Meier curve to compare OS between high-SH3TC2-DT ($N = 74$) and low-SH3TC2-DT ($N = 74$) samples. The AML patients were separated into two groups according to the median of SH3TC2-DT expression level. **(C)** ROC curves showing predictive value of the SH3TC2-DT expression. **(D)** Forest plot of multivariable Cox regression analysis. **(E)** GSEA identified gene sets enriched in SH3TC2-DT high-expression phenotype. FLT3-WT, FLT3-wildtype; FLT3-MUT, FLT3-mutant. NES, normalized enrichment score. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

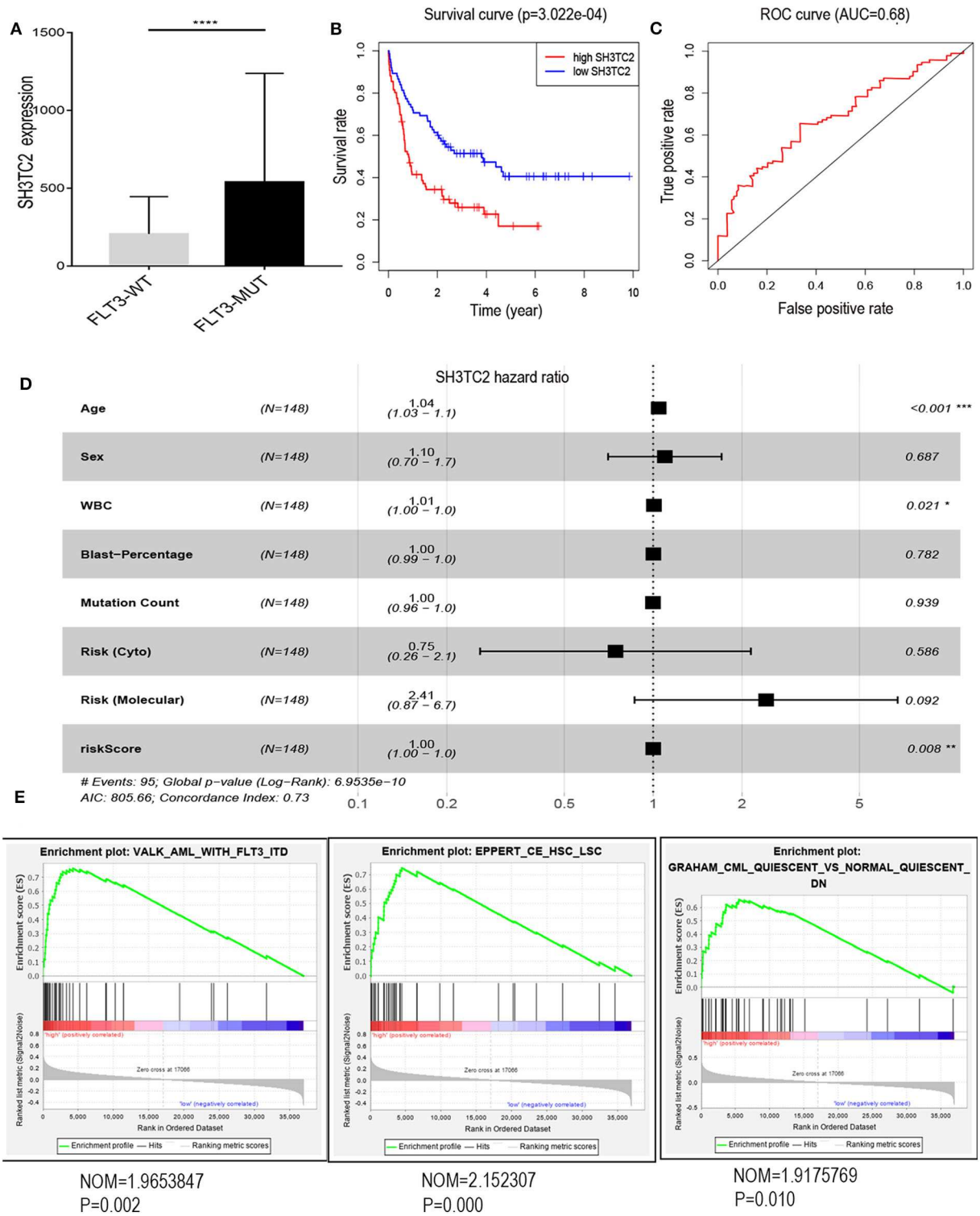

FIGURE 9 | The association of SH3TC2 expression with overall survival and gene-set enrichment. **(A)** The normalized RNA-sequencing expression value of SH3TC2 in FLT3-WT ($N = 108$) and FLT3-MUT ($N = 43$) AML patients. **(B)** Kaplan–Meier curve to compare OS between high-SH3TC2 ($N = 74$) and low-SH3TC2 ($N = 74$) samples. The AML patients were separated into two groups according to the median of SH3TC2 expression level. **(C)** ROC curves showing predictive value of the SH3TC2 expression. **(D)** Forest plot of multivariable Cox regression analysis. **(E)** GSEA identified gene sets enriched in SH3TC2 high expression phenotype. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

SH3TC2-DT/SH3TC2 gene pair showed no significant difference between FLT3-ITD and FLT3-TKD AML samples (**Figure S3**). As the general importance of divergent transcription is poorly understood, we further studied the clinical significance of SH3TC2-DT and SH3TC2 expression in TCGA-LAML cohort. The previous finding that divergent lncRNA/mRNA pairs exhibited coordinate changes in transcription suggests that divergent transcript might regulate gene transcription (23, 24). Our study revealed that SH3TC2-DT and SH3TC2 were also coordinately high expressed in FLT3-mutant AML samples (**Figures 8A**, **9A**), suggesting that SH3TC2-DT might regulate SH3TC2 expression during AML pathogenesis. High expression of SH3TC2-DT or SH3TC2 was associated with poor OS (**Figures 8B**, **9B**). The area under ROC curve was 0.622 for SH3TC2-DT and 0.68 for SH3TC2, both showing a high predictive value (**Figures 8C**, **9C**). Multivariate Cox regression analyses revealed that both SH3TC2-DT and SH3TC2 expression were independent prognostic factors (**Figures 8D**, **9D**). Furthermore, we used logistic regression analysis to relate SH3TC2-DT/SH3TC2 gene pair with clinical features and revealed that both high expression of SH3TC2-DT and SH3TC2 were associated with higher WBC count, intermediate cytogenetic and molecular–genetic risk, and FLT3 mutation. The SH3TC2 high expression was also associated with older age (**Tables 3**, **4**). GSEA showed that gene sets of AML with FLT3-ITD, HSC (hematopoietic stem cell)/LSC and CML (chronic myelocytic leukemia) quiescence were enriched in both SH3TC2-DT and SH3TC2 high-expression phenotype (**Figures 8E**, **9E**). We found that TFs associated with stemness (e.g., MEF2, AP1, SOX9, GATA3, and GFI1) or leukemogenesis (CEBPA and BACH1) were significantly enriched for DEGs between SH3TC2 high-expression and SH3TC2 low-expression group, suggesting that these TFs may be potential targets of SH3TC2 in AML (**Figure 10**).

To validate our findings, we analyzed the BeatAML (Vizome) dataset and found that both SH3TC2-DT and SH3TC2 were significantly highly expressed in FLT3-mutant AML (**Figure S4**). Furthermore, another validation cohort GSE37642-GPL570 also showed that high expression of SH3TC2 was associated with poor OS in AML ($P = 2.04e{-}02$) (**Figure S5**).

DISCUSSION

Although FLT3 mutation is commonly found in adult AML, its prognostic significance is still controversial (1). The mechanism of mutant FLT3 activation in leukemogenesis has not been definitely confirmed. Further elucidating the potential genes that are associated with FLT3 mutation and prognosis of AML is of great significance. Mining transcriptomic data from TCGA dataset could help to find prognostic factors that possibly participate in cancer development or evolution.

In the present study, using the TCGA-LAML dataset, we identified the DElncRNAs and DEmRNAs between FLT3-mutant and FLT3-wildtype AML. Functional enrichment analysis revealed that DEmRNAs were enriched in Wnt signaling pathway, PI3K–Akt signaling pathway, and Ras signaling pathway. Wnt signaling pathway was reported to cooperate with FLT3-ITD in leukemic signal transduction (25), PI3K-Akt and Ras were considered as downstream molecular pathways of FLT3 in leukemogenesis (26). Our result confirmed the function of FLT3 mutation in AML pathogenesis as previously reported. Afterwards, we constructed co-expression modules and related them to clinical traits through WGCNA. Using multivariate Cox regression analyses, we constructed prognostic risk models of lncRNA and mRNA to identify hub DEGs associated with AML

TABLE 3 | Association between SH3TC2-DT expression[a] and clinical characteristics (logistic regression).

Clinical characteristics	Total	OR in SH3TC2-DT expression	*P*-value
Age (continuous)	148	1.01 (0.99–1.03)	0.218
Sex (male vs. female)	148	1.18 (0.62–2.26)	0.620
WBC (continuous)	148	1.01 (1.00–1.02)	0.042*
Blast percentage (continuous)	148	1.00 (0.98–1.012)	0.892
Mutation counts (continuous)	148	1.01 (0.96–1.07)	0.684
Cyto-risk (intermediate vs. good)	112	4.40 (1.82–11.63)	0.002**
Molecular risk (intermediate vs. good)	108	4.35 (1.79–11.53)	0.002**
Therapy (auto-SCT or chemo vs. allo-SCT)	148	0.57 (0.29–1.10)	0.097
NPM1 (WT vs. MUT)	148	0.55 (0.25–1.18)	0.128
DNMT3A (WT vs. MUT)	148	0.46 (0.20–1.00)	0.054
FLT3 (WT vs. MUT)	148	0.28 (0.13–0.60)	0.001**

[a]*Categorical dependent variable, higher or less than the median expression level. OR, odds ratio; SCT, stem cell transplantation; WT, wildtype; MUT, mutant. *P < 0.05; **P < 0.01.*

TABLE 4 | Association between SH3TC2 expression[a] and clinical characteristics (logistic regression).

Clinical characteristics	Total	OR in SH3TC2 expression	*P*-value
Age (continuous)	148	1.02 (1.00–1.05)	0.024*
Sex (male vs. female)	148	1.06 (0.55–2.02)	0.869
WBC (continuous)	148	1.01 (1.00–1.02)	0.011*
Blast percentage (continuous)	148	0.85 (0.44–1.62)	0.620
Mutation counts (continuous)	148	0.98 (0.92–1.04)	0.487
Cyto-risk (intermediate vs. good)	112	3.56 (1.49–9.05)	0.005**
Molecular risk (intermediate vs. good)	108	3.71 (1.56–9.48)	0.004**
Therapy (auto-SCT or chemo vs. allo-SCT)	148	0.89 (0.46–1.72)	0.739
NPM1 (WT vs. MUT)	148	0.55 (0.25–1.18)	0.128
DNMT3A (WT vs. MUT)	148	0.54 (0.24–1.17)	0.121
FLT3 (WT vs. MUT)	148	0.17 (0.07–0.38)	3.550e-05***

[a]*Categorical dependent variable, higher or less than the median expression level. OR, odds ratio; SCT, stem cell transplantation; WT, wildtype; MUT, mutant. *P < 0.05; **P < 0.01; ***P < 0.001.*

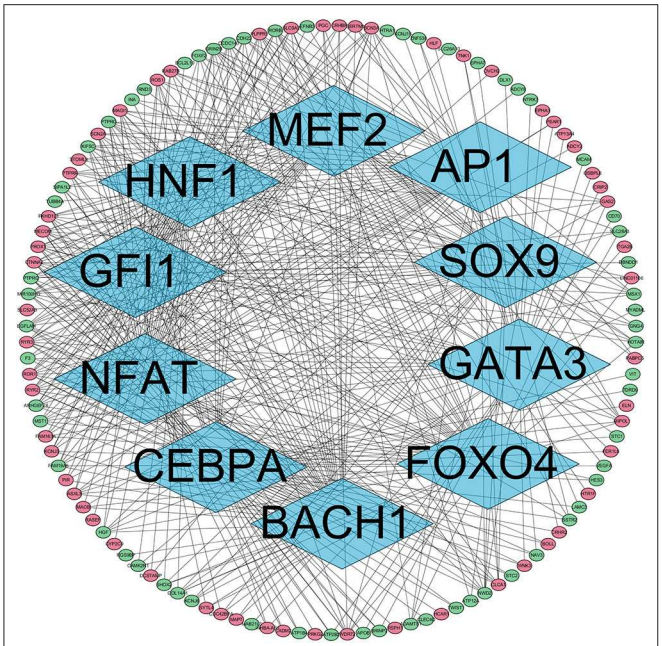

FIGURE 10 | The TFs regulatory network enriched for DEGs between SH3TC2 high-expression AML and SH3TC2 low-expression AML. Diamonds, TFs; ellipses, genes regulated by TFs; red, high expressed genes; green, low expressed genes. TFs associated with stemness and leukemogenesis were selected to shown.

prognosis. The presence of both SH3TC2-DT and SH3TC2 in respective prognostic risk models promotes us to further analyze this divergent lncRNA/mRNA pair in AML dataset.

Divergent transcription is a common phenomenon that mammalian promoters initiate transcription on both sides with opposite directions. However, the biological function and importance are still poorly understood. The findings that divergent transcription is found in most actively transcribed genes and divergent lncRNA/mRNA pairs exhibit coordinated changes in transcription suggests that it might regulate gene transcription (23, 24). Our study revealed that SH3TC2-DT and SH3TC2 were also coordinately highly expressed in FLT3-mutant AML samples, suggesting that divergent transcription might regulate SH3TC2 expression to play a role in AML pathogenesis.

SH3TC2 is displayed as one of the most prevalent genes and prognostic factors across multiple tumor types (27, 28). In neuroblastoma, lower SH3TC2 expression level was associated with MYCN amplification and poor survival (29). In diffuse large B cell lymphoma, SH3TC2 was identified as a signature gene of the poor prognostic $CD5^+$ activated B-cell like (ABC) subtype (30). However, the prognostic value of SH3TC2 expression in AML has not been well-understood. In our study, we found that high expression of SH3TC2-DT and SH3TC2 were both associated with poor OS, FLT3 mutation, high WBC count, and intermediate cytogenetic and molecular–genetic risk in AML. These results suggest the association of the SH3TC2-DT/SH3TC2 gene pair with the proliferation function of FLT3 mutation. The association with high WBC count and intermediate genetic risk contributes to explanation of high SH3TC2-DT/SH3TC2 expression in relation to poor OS.

Although SH3TC2 expression is prevalent in multiple types of tumors, the function of SH3TC2 in cancer development or therapeutic resistance is less understood. The adjacent gene miR-584 is located in the first intron of the SH3TC2 gene and was reported to function as a tumor suppressor gene for glioma (31) and clear cell renal cell carcinoma (32), but to induce migration in breast cancer through TGF-β (33). We found that the phenotype of SH3TC2-DT/SH3TC2 high expression was enriched in HSC/LSC and CML quiescence gene sets by GSEA. TFs associated with stemness or leukemogenesis may be potential targets of SH3TC2. Our study suggests that the SH3TC2-DT/SH3TC2 gene pair may be associated with the stemness or quiescence of FLT3-mutant LSCs. Further study to elucidate the pathologic function of SH3TC2-DT/SH3TC2 in FLT3-mutant AML is probably valuable.

In the validating BeatAML dataset, the SH3TC2-DT/SH3TC2 gene pair was also highly expressed in FLT3-mutant AML. However, we did not find the correlation between high expression of this gene pair and OS. AML samples of this dataset are heterogeneous, including both *de novo* and secondary, both chemotherapy-treated and palliative therapy-treated, which might lead to the unexpected result. Thus, we turned to the GEO dataset GSE37642-GPL570 and found that high expression of SH3TC2 was associated with poor OS. Overall, SH3TC2-DT/SH3TC2 expression may be a possible biomarker to further optimize the prognosis of FLT3-mutant AML, but larger AML cohorts for further study are needed.

In summary, through integrated bioinformatic analyses, we identified hub lncRNAs and mRNAs associated with FLT3 mutation and AML prognosis. Among them, high expression of the SH3TC2-DT/SH3TC2 gene pair was found in FLT3-mutant AML and associated with poor prognosis, high WBC count, and intermediate genetic risk. The correlation of SH3TC2-DT/SH3TC2 expression with stemness, quiescence, and leukemogenesis suggests a possible role for this divergent gene pair in FLT3-mutant LSCs. Overall, our findings would help to better understand the possibly leukemogenic mechanisms of FLT3-mutant AML and to find possible candidate genes for prognostic and therapeutic usage.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in The Cancer Genome Atlas database (https://portal.gdc.cancer.gov/), Vizome (http://www.vizome.org/), and GEO (GSE37642-GPL570, https://www.ncbi.nlm.nih.gov/geo).

AUTHOR CONTRIBUTIONS

PY performed research, collected, analyzed, interpreted data, and wrote the manuscript. HL interpreted data and provided support. XS interpreted data, provided support, and edited the manuscript. ZP conceived the concept, designed the studies, performed research, collected, analyzed, interpreted data, wrote

the paper, and took responsibility for the whole manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by Key Disciplines Group Construction Project of Pudong Health Bureau of Shanghai (Grant No. PWZxq2017-13), Three-year development project from Shanghai Shenkang Hospital Development Center (16CR1010A), and National Natural Science Foundation of China (NSFC 81370601).

ACKNOWLEDGMENTS

The results of our study are based on data from TCGA (https://portal.gdc.cancer.gov/) and GEO platform (https://www.ncbi.nlm.nih.gov/geo/).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00829/full#supplementary-material

REFERENCES

1. Dohner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Ebert BL, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel hartmut. *Blood*. (2017) 129:424–48. doi: 10.1182/blood-2016-08-733196
2. Pan Z, Yang M, Huang K, Büsche G, Glage S. Flow cytometric characterization of acute leukemia reveals a distinctive "blast gate" of murine T-lymphoblastic leukemia / lymphoma. *Oncotarget*. (2018) 9:2320–8. doi: 10.18632/oncotarget.23410
3. The Cancer Genome Atlas Research Network, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson A et al. Genomic and epigenomic landscapes of adult *de novo* acute myeloid leukemia. *N Engl J Med*. (2013) 368:2059–74. doi: 10.1056/NEJMoa1301689
4. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med*. (2016) 374:2209–21. doi: 10.1056/NEJMoa1516192
5. Sander A, Zimmermann M, Dworzak M, Bullinger L, Ehrich M, Dna Q, et al. The impact of age, NPM1 mut, and FLT3 ITD allelic ratio in patients with acute myeloid leukemia. *Blood*. (2018) 131:1148–53. doi: 10.1182/blood-2017-09-807438
6. Boddu P, Kantarjian H, Borthakur G, Kadia T, Daver N, Pierce S, et al. Co-occurrence of FLT3-TKD and NPM1 mutations defines a highly favorable prognostic AML group. *Blood Adv*. (2017) 1:1546–50. doi: 10.1182/bloodadvances.2017009019
7. Bacher U, Haferlach C, Kern W, Haferlach T, Schnittger S. Prognostic relevance of FLT3-TKD mutations in AML: the combination matters–an analysis of 3082 patients. *Blood*. (2008) 111:2527–37. doi: 10.1182/blood-2007-05-091215
8. Lai Y. A statistical method for the conservative adjustment of false discovery rate (q-value). *BMC Bioinformatics*. (2017) 18:69. doi: 10.1186/s12859-017-1474-6
9. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. (2010) 26:139–40. doi: 10.1093/bioinformatics/btp616
10. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. (2012) 40:4288–97. doi: 10.1093/nar/gks042
11. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J Integr Biol*. (2012) 16:284–7. doi: 10.1089/omi.2011.0118
12. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res*. (2010) 38:D355–60. doi: 10.1093/nar/gkp896
13. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
14. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. (2003) 34:267–73. doi: 10.1038/ng1180
15. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. (2008) 9:559. doi: 10.1186/1471-2105-9-559
16. Iancu OD, Colville A, Oberbeck D, Darakjian P, McWeeney SK, Hitzemann R. Cosplicing network analysis of mammalian brain RNA-Seq data utilizing WGCNA and mantel correlations. *Front Genet*. (2015) 6:174. doi: 10.3389/fgene.2015.00174
17. Fuller TF, Ghazalpour A, Aten JE, Drake TA, Lusis AJ, Horvath S. Weighted gene coexpression network analysis strategies applied to mouse weight. *Mamm Genome*. (2007) 18:463–72. doi: 10.1007/s00335-007-9043-3
18. Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, et al. Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet*. (2006) 2:e130. doi: 10.1371/journal.pgen.0020130
19. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. (2009) 37:1–13. doi: 10.1093/nar/gkn923
20. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. (2009) 4:44–57. doi: 10.1038/nprot.2008.211
21. Tyner JW, Tognon CE, Bottomly D, Wilmot B, Kurtz SE, Savage SL, et al. Functional genomic landscape of acute myeloid leukaemia. *Nature*. (2018) 562:526–31. doi: 10.1038/s41586-018-0623-z
22. Janke H, Pastore F, Schumacher D, Herold T, Hopfner K-P, Schneider S, et al. Activating FLT3 mutants show distinct gain-of-function phenotypes *in vitro* and a characteristic signaling pathway profile associated with prognosis in acute myeloid leukemia. *PLoS ONE*. (2014) 9:e89560. doi: 10.1371/journal.pone.0089560
23. Seila AC, Core LJ, Lis JT, Sharp PA. Divergent transcription: a new feature of active promoters. *Cell Cycle*. (2009) 8:2557–64. doi: 10.4161/cc.8.16.9305
24. Sigova AA, Mullen AC, Molinie B, Gupta S, Orlando DA, Guenther MG, et al. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc Natl Acad Sci USA*. (2013) 110:2876–81. doi: 10.1073/pnas.1221904110
25. Tickenbrock L, Schwa J, Wiedehage M, Choudhary C, Brandts C, Berdel WE, et al. Flt3 tandem duplication mutations cooperate with Wnt signaling in leukemic signal transduction. *Blood*. (2014) 105:3699–707. doi: 10.1182/blood-2004-07-2924
26. Takahashi S. Downstream molecular pathways of FLT3 in the pathogenesis of acute myeloid leukemia: biology and therapeutic implications. *J Hematol Oncol*. (2011) 4:13. doi: 10.1186/1756-8722-4-13
27. Doecke JD, Wang Y, Baggerly K. Co-localized genomic regulation of miRNA and mRNA via DNA methylation affects survival in multiple tumor types. *Cancer Genet*. (2016) 209:463–73. doi: 10.1016/j.cancergen.2016.09.001
28. Abdelfattah N, Rajamanickam S, Panneerdoss S, Timilsina S, Yadav P, Onyeagucha BC, et al. MiR-584-5p potentiates vincristine and radiation response by inducing spindle defects and DNA damage in medulloblastoma. *Nature Commun*. (2018) 9:4541. doi: 10.1038/s41467-018-06808-8
29. Xiang X, Mei H, Qu H, Zhao X, Li D, Song H, et al. miRNA-584-5p exerts tumor suppressive functions in human neuroblastoma through repressing transcription of matrix metalloproteinase 14. *Biochim Biophys Acta*. (2015) 1852:1743–54. doi: 10.1016/j.bbadis.2015.06.002

30. Miyazaki K, Yamaguchi M, Imai H, Kobayashi K, Tamaru S, Kobayashi T, et al. Gene expression profiling of diffuse large B-cell lymphomas supervised by CD5 expression. *Int J Hematol*. (2015) 102:188–94. doi: 10.1007/s12185-015-1812-2
31. Wang X-P, Deng X-L, Li L-Y. MicroRNA-584 functions as a tumor suppressor and targets PTTG1IP in glioma. *Int J Clin Exp Pathol*. (2014) 7:8573–82.
32. Ueno K, Hirata H, Shahryari V, Chen Y, Zaman MS, Singh K, et al. Tumour suppressor microRNA-584 directly targets oncogene Rock-1 and decreases invasion ability in human clear cell renal cell carcinoma. *Br J Cancer*. (2011) 104:308–15. doi: 10.1038/sj.bjc.6606028
33. Fils-Aimé N, Dai M, Guo J, El-Mousawi M, Kahramangil B, Neel J-C, et al. MicroRNA-584 and the protein phosphatase and actin regulator 1 (PHACTR1), a new signaling route through which transforming growth factor-β Mediates the migration and actin dynamics of breast cancer cells. *J Biol Chem*. (2013) 288:11807–23. doi: 10.1074/jbc.M112.430934

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