

Identification of the hub genes and pathways involved in acute myeloid leukemia using bioinformatics analysis

Youping Tan, MD, Liling Zheng, MD, Yuanyuan Du, MD, Qi Zhong, MD, Yangmin Zhu, MD, PhD, Zhi Liu, MD, Shuang Liu, PhD, Qing Zhang, MD, PhD*

Abstract

Background: We identified the hub genes and pathways dysregulated in acute myeloid leukemia and the potential molecular mechanisms involved.

Methods: We downloaded the GSE15061 gene expression dataset from the Gene Expression Omnibus database and used weighted gene co-expression network analysis to identify hub genes. Differential expression of the genes was evaluated using the limma package in R software. Subsequently, we built a protein–protein interaction network followed by functional enrichment analysis. Then, the prognostic significance of gene expression was explored in terms of overall survival. Finally, transcription factor–mRNA (ribonucleic acid) and microRNA–mRNA interaction analysis was also explored.

Results: We identified 100 differentially expressed hub genes. Functional enrichment analysis indicated that the genes were principally involved in immune system regulation, host defense, and negative regulation of apoptosis and myeloid cell differentiation. We identified 4 hub genes, the expression of which was significantly correlated with overall survival. Finally, 26 key regulators for hub genes and 38 microRNA–mRNA interactions were identified.

Conclusion: We performed a comprehensive bioinformatics analysis of hub genes potentially involved in acute myeloid leukemia development. Further molecular biological experiments are required to confirm the roles played by these genes.

Abbreviations: AML = acute myeloid leukemia, BM = bone marrow, DEGs = differentially expressed genes, GEP = gene expression profiling, OS = overall survival, LGA = Leukemia Gene Atlas, PB = peripheral blood, PPI = protein–protein interaction, RNA = ribonucleic acid, TCGA = Cancer Genome Atlas, TF = transcription factor, WGCNA = weighted gene co-expression network analysis.

Keywords: acute myeloid leukaemia, bioinformatics, gene expression profiling, hub genes and pathways, weighted gene co-expression network analysis

Editor: Gunjan Arora.

This work was supported by the Foundation of Guangdong Second Provincial General Hospital (2017-001) (Qing Zhang), Doctoral workstation foundation of Guangdong Second Provincial General hospital (2019BSGZ008) (Shuang Liu), Guangzhou Science and technology plan project (202002030404) (Shuang Liu), Guangdong Medical Scientific Research (B2020092) (Zhi Liu).

The authors have no conflicts of interest to disclose.

Supplemental Digital Content is available for this article.

The datasets generated during and/or analyzed during the current study are publicly available.

Department of Hematology, Guangdong Second Provincial General Hospital, Guangzhou, Guangdong Province, P. R. China.

* Correspondence: Qing Zhang, Department of Hematology, Guangdong Second Provincial General Hospital, Guangzhou, Guangdong Province, 510220, P. R. China (e-mail: gdqingzhang@163.com).

Copyright © 2020 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Tan Y, Zheng L, Du Y, Zhong Q, Zhu Y, Liu Z, Liu S, Zhang Q. Identification of the hub genes and pathways involved in acute myeloid leukemia using bioinformatics analysis. *Medicine* 2020;99:35(e22047).

Received: 20 April 2020 / Received in final form: 4 July 2020 / Accepted: 4 August 2020

<http://dx.doi.org/10.1097/MD.00000000000022047>

1. Introduction

Acute myeloid leukemia (AML) is a heterogeneous hematological malignancy characterized by the clonal expansion of myeloid blasts in peripheral blood (PB), bone marrow (BM), and/or other tissues. AML is the most common form of adult acute leukemia, with an annual incidence of 2 to 4 cases per 100,000 adults. The median age at diagnosis is 64 years. As populations age, the incidence of AML increases.^[1] Despite advances in our understanding of AML molecular heterogeneity and pathogenesis, the standard therapy has not greatly changed over the past few decades. The 5-year median overall survival (OS) is roughly 40% in patients younger than 60 years, but the disease is more serious in older individuals, with a 5-year survival rate of only 10% to 20% in those older than 60 years.^[2,3] Hence, it is essential to identify biomarkers predicting AML progression and prognosis to facilitate the development of new therapeutic approaches.

High-throughput platforms (such as microarrays) allowing analysis of gene expression can be used to explore the complex biological network involved in AML development. Many microarray studies on AML have appeared,^[4–6] substantially unraveling AML pathogenesis and identifying new molecular markers. Microarray-mediated gene expression profiling (GEP) can be used to diagnose leukemia with high accuracy.^[4] Gene

expression signatures associated with distinct clinical subtypes of leukemia have been identified by microarray studies.^[5] Moreover, GEP analysis is also useful for predicting AML prognosis.^[6]

In recent years, many GEP microarrays have been used to identify differentially expressed genes (DEGs) in AML; a few such genes are promising biomarkers in terms of AML diagnosis and prognosis.^[4] However, most previous studies were focused on DEG screening, not on the interactions among genes. Weighted gene co-expression network analysis (WGCNA) can be used to identify correlations among DEGs using microarray or RNA (ribonucleic acid) sequencing data. WGCNA identifies modules of highly correlated genes and links the modules to certain clinical phenotypes. Correlation networks facilitate network-based gene screening of candidate biomarkers or therapeutic targets.^[7]

In the present study, we compared the GEPs of primary AML and non-leukemia BM samples; we collected data from the National Center of Biotechnology Information Gene Expression Omnibus database. The WGCNA algorithm was used to construct a scale-free weighted genetic interaction network featuring specific gene modules that play common biological roles during AML development. Furthermore, we identified certain hub genes as potential candidate biomarkers or novel molecular therapeutic targets for AML.

2. Methods

2.1. GEP

This study as approved by the Institutional Review Board of the Guangdong Second Provincial General Hospital. The GSE15061 dataset (Platforms GPL570) was downloaded from the Gene Expression Omnibus database. GSE15061 data were derived using Affymetrix Human Genome U133 Plus 2.0 GeneChips (Thermo Fisher Scientific, Waltham, MA, USA). The dataset is part of the MILE Study (Microarray Innovations in Leukemia) program, headed by the European Leukemia Network (ELN), contains information on 164 myelodysplastic syndrome, 202 AML, and 69 non-leukemia BM samples. The AML samples include different subtypes, such as AML with t(15; 17), t(8; 21), inv(16), or t(11q23)/MLL. We analyzed only the AML and non-leukemia samples.

2.2. Data preprocessing and probe re-annotation

All probes and samples were first checked to ensure that no data were missing. The data were then quantile-normalized and log-transformed, and the probe sequences were mapped to the RefSeq transcript database (www.ncbi.nlm.nih.gov/refseq/) using seq-Map software (Wong Lab, Stanford University, Stanford, CA, USA).^[8] Only probes exhibiting perfect matches to RefSeq transcripts were retained, and these probes were then annotated at the gene level.

2.3. Selection of genes exhibiting the most variation

We ranked the probes according to the coefficient of variation (CV) of their expression profiles and selected the top 5000 genes. The CV is commonly used to measure relative variability among subjects of interest. The CV is simply the standard deviation divided by the mean. We selected probes in descending order of CV, while ignoring repeat genes, until the number of selected genes attained reached 5000.

2.4. Co-Expression network construction and module detection

The R package WGCNA was used to build a weighted gene co-expression network for the gene expression matrix, as described elsewhere.^[7] First, we cleaned up the data by removing obvious outliers as well as genes and samples with excessive missing entries. We next constructed pairwise Pearson correlation matrices. The weighted gene co-expression network was constructed using the β function; this is a soft thresholding parameter that retains strong correlations between genes but removes weak correlations. The adjacency data were transformed into a topological overlap matrix (TOM) measuring not only between-gene correlations but also the extent of their shared correlations across the weighted network. We next performed average linkage hierarchical clustering using the TOM-based dissimilarity measure, with a minimum module size of 30 and a medium sensitivity of 2 (the other parameters were set to the default values). A dynamic TreeCut algorithm was used to cut the hierarchical clustering dendrogram and the branches were defined as modules. We associated modules with clinical traits and calculated the associated gene significance (GS) and module membership (MM) values. Finally, by setting minimum thresholds for these 2 parameters, we selected the hub genes with the highest GSs and MMs for each module.

2.5. DEG screening

The R package limma was used to identify DEGs between AML and non-leukemia samples. Screening was performed using an adjusted P -value $< .05$ and a \log_2 -fold change (FC) ≥ 2 . The DEG screening results were validated using the Cancer Genome Atlas (TCGA)-AML gene expression profiles identified by GEPIA (<http://gepia.cancer-pku.cn/>) at a P -value $< .01$ and a \log_2 FC value ≥ 1 . GEPIA is a newly developed interactive web server used to analyze RNA expression data from TCGA and Genotype-Tissue Expression projects, using a standard processing pipeline.

2.6. Functional annotation and protein–protein interaction (PPI) network construction

We used the ClueGO plugin in Cytoscape to perform gene ontology (GO) (<http://www.geneontology.org/>), Kyoto Encyclopedia of Genes and Genomes (KEGG) (<http://www.kegg.jp/kegg/>), and REACTOME (<https://reactome.org/>) pathway enrichment analyses of the selected genes. PPI information was also assessed employing the GeneMANIA plugin of Cytoscape. The plugin MCODE was used to screen the PPI network identified by Cytoscape using a degree cutoff of 2, a node score cutoff of 0.2, a K -core of 2, and a maximum depth of 100.

2.7. Survival analysis

The association between hub DEGs and OS was examined using GEPIA. OS curves obtained using the Kaplan–Meier method were generated by reference to the median gene expression levels. The data were validated employing the Leukemia Gene Atlas (LGA) (<http://www.leukemia-gene-atlas.org/>) consisting of AML gene expression profiles created by Verhaak et al.^[9] We used the 50% quantiles of gene expression as cutoffs for the Kaplan–Meier curves generated by the LGA public platform supporting analysis of molecular leukemia data.

2.8. Transcription factor (TF)-mRNA and microRNA-mRNA interaction analysis

We used TRRUST (<https://www.grnpedia.org/trrust/>), a manually curated database of human and mouse transcriptional regulatory networks for TF-mRNA interaction analysis within hub DEGs. We also used miRWalk to perform microRNA-mRNA interaction analysis. MiRWalk (<http://mirwalk.umm.uni-heidelberg.de/>) is an online resource for prediction of microRNA binding sites, which stores predicted data obtained with a machine learning algorithm including experimentally verified microRNA-target interactions. We chose intersection of 2

database Targetscan (http://www.targetscan.org/vert_72/) and miRDB (<http://mirdb.org/index.html>) to present the results.

3. Results

3.1. Data preprocessing and probe re-annotation

Data analysis pipeline is shown in Figure 1. All sample and probe data were complete. The dataset contained 54,675 probes, and after re-annotation 41,373 remained, which mapped to 20,604 unique genes. We selected the top 5000 genes according to CV.

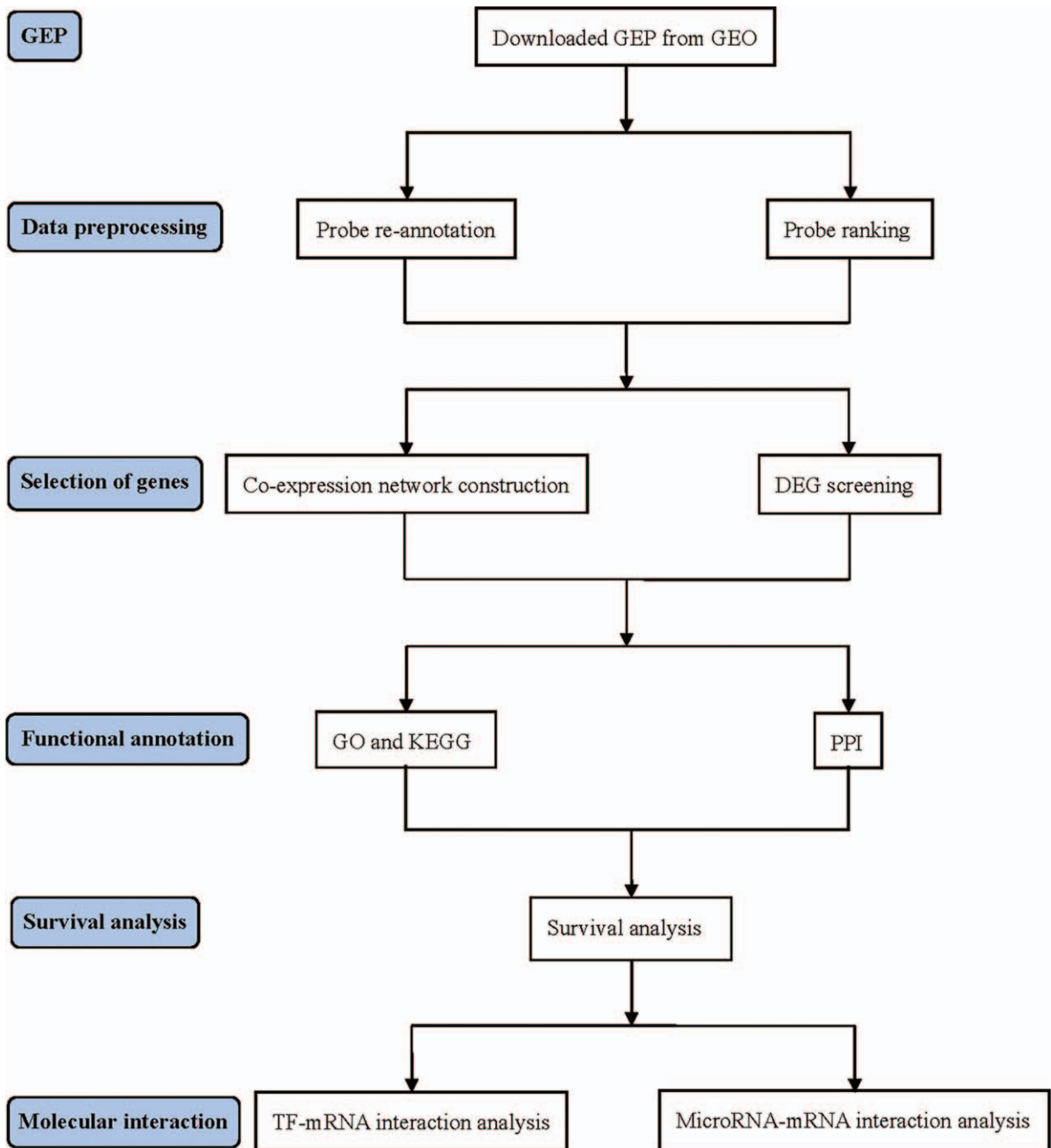


Figure 1. Flow chart diagram of data analysis.

Table 1**Functional enrichment analysis of 100 differentially expressed hub genes with representative pathways for each group.**

Pathway ID	Term	Pathway P-value	Pathway group	Associated genes
GO:0006968	Cellular defense response	3.41E-04	Group 0	C5AR1, CLEC5A, CXCR2, GNLY
GO:0032677	Regulation of interleukin-8 production	6.45E-03	Group 1	BPI, FCN1, PRG3
GO:0045638	Negative regulation of myeloid cell differentiation	7.01E-05	Group 2	CEACAM1, HOXA5, HOXA7, HOXA9, MEIS1
GO:0048704	Embryonic skeletal system morphogenesis	1.26E-04	Group 3	HOXA3, HOXA5, HOXA7, HOXB6, MYCN
KEGG:04640	Hematopoietic cell lineage regulation	1.39E-04	Group 4	CD24, CD8A, FLT3, IL7R, MME
KEGG:05202	Transcriptional dysregulation in cancer	4.95E-06	Group 5	CCNA1, CEBPE, FLT3, HOXA10, HOXA9, JUP, MEIS1, MYCN
KEGG:05150	<i>Staphylococcus aureus</i> infection	3.99E-07	Group 6	C5AR1, CFD, FCAR, FCGR3A, FPR1, FPR2
GO:0002456	T cell-mediated immunity	9.24E-05	Group 7	ARG1, CD8A, CEACAM1, IL18RAP, IL7R
GO:0051668	Localization within membranes	2.96E-04	Group 8	ATP1B1, CD24, CD8A, NRXN2, SHANK3
GO:0042119	Neutrophil activation	1.29E-29	Group 9	ALOX5, ANXA3, ARG1, BPI, C5AR1, CDA, CEACAM1, CEACAM21, CFD, CKAP4, CLEC5A, CRISP3, CXCR2, CYBB, EGFL7, FCAR, FCN1, FPR1, FPR2, HK3, HP, IL18RAP, JUP, MCEMP1, MGAM, MME, MMP25, PADI2, PRG3, QPCT, RETN, S100A12, S100P, SIGLEC5, TCN1

3.2. Co-expression network construction and module detection

The 5000 genes mentioned above were used in subsequent analyses. Seven outliers were excluded. We chose a soft threshold β of 6 when constructing the co-expression network. TOM-based hierarchical clustering yielded 14 gene modules, and module-trait associations and their P -values were evaluated. We selected modules with absolute correlations $> .5$, thereby obtaining 4 modules (blue, 430 genes, $P = 3e-12$, correlation = -0.64 ; brown, 386 genes, $P = 6e-56$, correlation = 0.78 ; green/yellow, 113 genes, $P = 3e-23$, correlation = -0.56 ; yellow, 248 genes, $P = 1e-27$, correlation = 0.6). We set minimum GS and MM thresholds and selected those hub genes with the highest GS and MM values in each module. Initially, we set a minimum GS of 0.3 and a minimum MM of 0.6. Because of the relatively small number of genes obtained in the green/yellow module, we reduced the GS and MM thresholds for this module to 0.2 and 0.4, respectively. The selected hub genes were further analyzed (blue, 151; brown, 131; green/yellow, 94; yellow, 119; Supplementary Tables S1–S4, <http://links.lww.com/MD/E809>).

3.3. Identification of DEGs

Using the limma package of R software, and employing $P < .05$ and $\log_2 FC \geq 2$ as cut-offs, we identified 197 DEGs, of which 63 were upregulated and 134 downregulated (Supplementary Table 5, <http://links.lww.com/MD/E809> and Fig. S1, <http://links.lww.com/MD/E808>). Using cutoffs of $P < .01$ and $\log_2 FC \geq 1$, 133 DEGs were validated by GEPIA using the RNA sequencing data from TCGA-AML.

3.4. Functional enrichment analysis

To obtain further insights into the biological relevance of hub DEGs (Intersection of hub genes and DEGs, $n = 100$) associated with AML, we performed GO, KEGG, and REACTOME pathway enrichment analyses using the ClueGO plugin of Cytoscape. As shown in Supplementary Figure S2, <http://links.lww.com/MD/E808> and Table S6, <http://links.lww.com/MD/E809>, the following 28 pathways were significantly enriched

among the hub DEGs ($P < .01$): neutrophil activation, cellular defense, regulation of interleukin-8 production, negative regulation of myeloid cell differentiation, embryonic skeletal system morphogenesis, hematopoietic cell lineage regulation, transcriptional dysregulation in cancer, *Staphylococcus aureus* infection, T cell-mediated immunity, and postsynaptic density organization (Table 1).

3.5. PPI network analysis

GeneMANIA showed that the PPI network of the hub DEGs featured 117 nodes and 1464 edges. The MCODE plugin of Cytoscape revealed 4 clusters (using the default settings). The top cluster, featuring 38 nodes and 611 edges, was selected for further

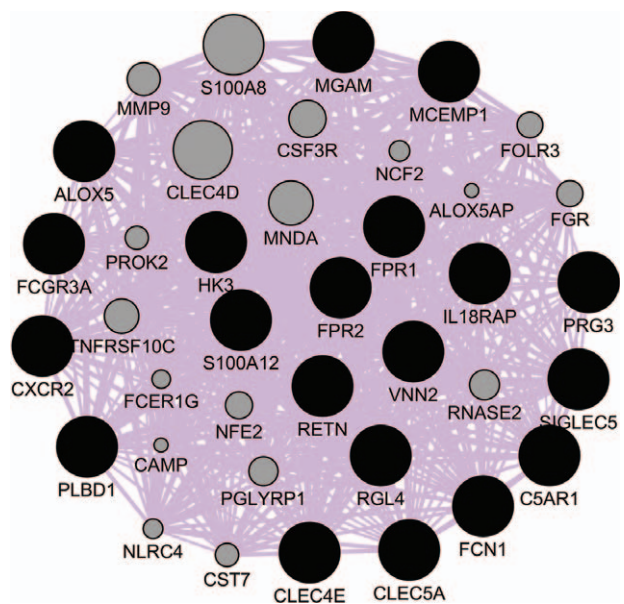


Figure 2. The top cluster of the protein–protein interaction (PPI) network involving the differentially expressed hub genes. Gray circles indicate the top 18 genes.

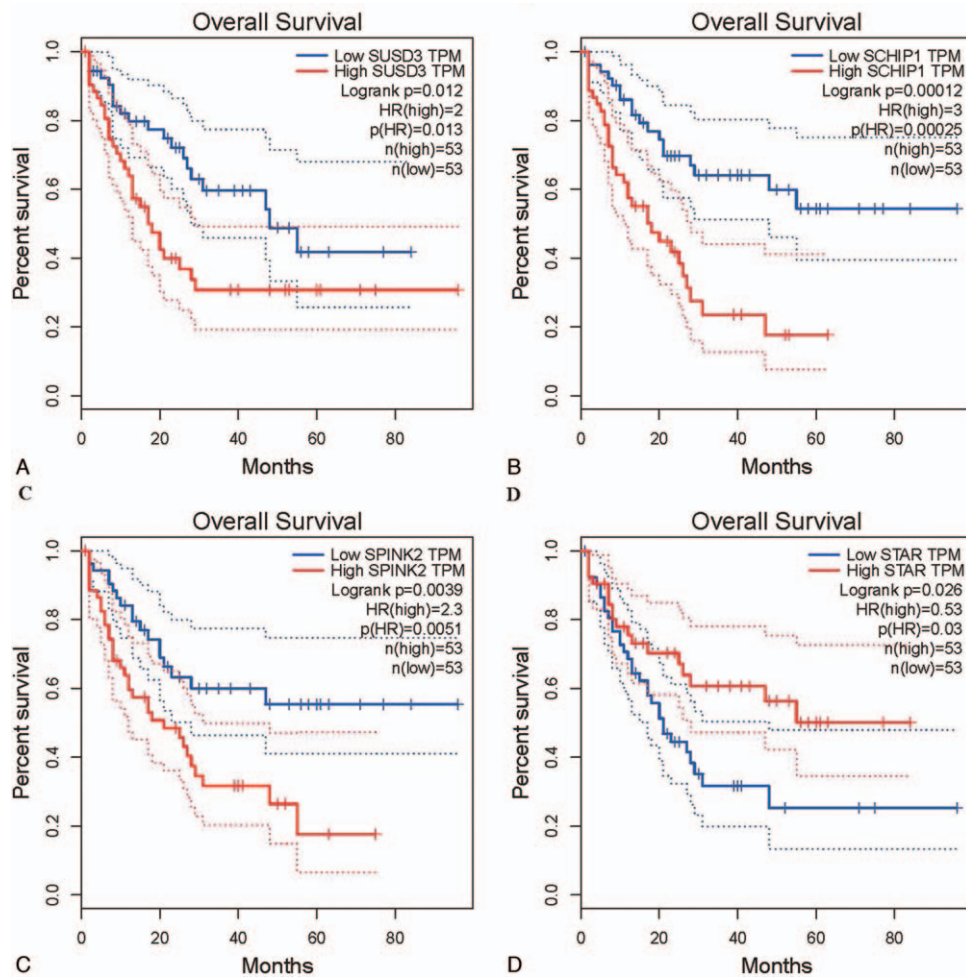


Figure 3. High SUSD3, SCHIP1, and SPINK2 expression and low STAR expression were correlated with poor survival as revealed.

analysis (Fig. 2). Gene functional enrichment analysis revealed significant enrichment of the following 34 pathways ($P < .01$): neutrophil activation, icosanoid biosynthesis, neutrophil chemotaxis, cellular defense, mast cell activation, positive regulation of cytokine secretion, regulation of interleukin-8 production, superoxide anion generation, response to fungi, complement receptor-mediated signaling, Dectin-2 family activity, and negative regulation of myeloid cell apoptosis (Supplementary Fig. S3, <http://links.lww.com/MD/E808> and Table S7, <http://links.lww.com/MD/E809>).

3.6. Survival analysis

The associations among the hub DEGs in the module that were most associated with AML and OS were examined using GEPIA to data-mine TCGA-AML information. Log-rank OS curves showed that FLT3, CEACAM21, SUSD3, RETN, MME, FGF13, HOXA9, HOXB6, FAM30A, CFD, GNLY, CLEC5A, S100P, PRG3, SCHIP1, SPINK2, STAR, CCNA1, and CRNDE expression levels were significantly associated with 100-month OS (Supplementary Fig. S4, <http://links.lww.com/MD/E809>). These results were validated using LGA; high SUSD3, SCHIP1, and SPINK2, and low STAR expression were all significantly

associated with poor 200-month OS ($P = .0329, .0046, .0206,$ and $.0032$, respectively) (Figs. 3–4).

3.7. TF-mRNA and microRNA-mRNA interaction analysis

Using the TRRUST, we identified 26 key regulators for our hub DEGs (Fig. 5, Supplementary Table S8, <http://links.lww.com/MD/E809>). Of note, we found 6 key TFs (CEBPA, CEBPB, NR4A1, CREM, SP1 and JUN) for gene STAR, the expression of which was significantly correlated with OS. 49 microRNA-mRNA interactions were identified by miRWalk analysis. However, we chose 38 microRNA-mRNA interactions from the intersection of 2 database Targetscan and miRDB to present the results (Fig. 6), 4 genes (COL24A1, TRIM71, HOXA9, and MYCN) were regulated by more than 4 microRNAs.

4. Discussion

AML is a complex heterogeneous malignancy caused by malignant transformation of hematopoietic stem cells and accumulation of immature myeloid progenitors in the PB and BM. Over the past few decades, AML treatments have remained largely unchanged; chemotherapy, and stem cell transplantation

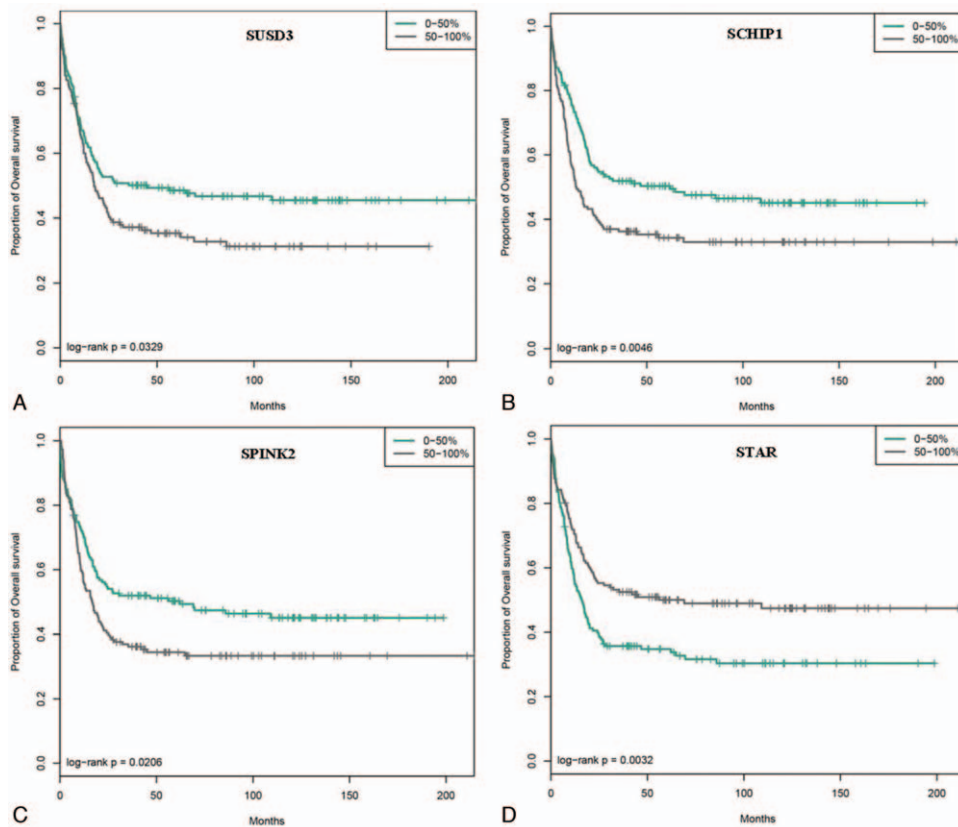


Figure 4. Survival afforded by acute myeloid leukemia hub genes as revealed by the Leukemia Gene Atlas. High SUSD3, SCHIP1, and SPINK2 expression and low STAR expression were correlated with poor.

are commonly employed.^[10] In recent years, considerable progress has been made in our understanding of disease pathogenesis and identification of biomarkers used to diagnose AML and aid treatment. Survival outcomes have improved, especially in younger patients. However, AML has a very poor prognosis in older patients, most of whom die from disease relapse. Therefore, the mechanisms of leukemogenesis and progression require further attention.

GEP uses cDNA microarrays or oligonucleotide probes to measure the levels of many mRNA transcripts simultaneously,^[11] allowing abnormalities and variations to be assessed on a genome-wide basis. In recent years, many studies have explored AML gene expression using microarrays,^[4-6,12] identifying new clinical subgroups^[5] and providing valuable diagnostic and prognostic information.^[4,6,12] Previously, AML progression-related genes were identified using basic methods such as DEM screening,^[4] WGCNA is a new algorithm identifying correlation patterns among genes across multiple microarray samples, creating modules of highly co-expressed genes and relating these modules to clinical phenotypes.^[7]

Here, we identified 100 hub DEGs in AML and non-leukemia BM samples. Functional enrichment indicated that the hub genes were significantly enriched in 28 pathways divided into 10 groups, including regulation of the immune system, defense, and negative regulation of apoptosis and myeloid cell differentiation. Many of these pathways have already been reported to be active in AML. We found that as AML developed, negative regulation of myeloid cell apoptosis, and differentiation triggered accumu-

lation of immature myeloid progenitors in PB and BM, accompanied by immune system activation and dysregulation. We constructed a hub gene PPI network and identified the genes with significant interactions; the most significant cluster was evaluated further. We found 4 hub genes, the expression of which was correlated significantly with OS.

Of these hub DEGs, some have been reported previously to be expressed in AML patients. FLT3 is a receptor tyrosine kinase expressed by immature hematopoietic cells and is involved in the normal development of stem cells and the immune system. FLT3 mutations have been detected in many AML patients with poor prognoses.^[13] CEBPE is a member of the CCAAT/enhancer binding protein family that plays important roles in normal myelopoiesis. Loss of CEBPE activity may play a role in AML pathogenesis. Increased CEBPE activity suppresses the leukemia phenotype of acute promyelocytic leukemia, suggesting that targeted modulation of CEBPE activity may constitute a new form of acute promyelocytic leukemia therapy.^[14] HOX gene dysregulation is a common feature of AML. Dysregulation of HOXA9 or HOXA10 triggers transformation to leukemia by altering hematopoietic stem cell self-renewal and differentiation properties. HOX overexpression is a poor prognostic marker in AML patients.^[15,16] S100A8, a member of the damage-associated molecular pattern family, is differentially expressed in many cell types but abundant in myeloid cells and involved in the progression of various cancers including leukemia. The S100A8 expression level is correlated with poor AML outcomes. S100A8 also plays an important role in drug-resistance

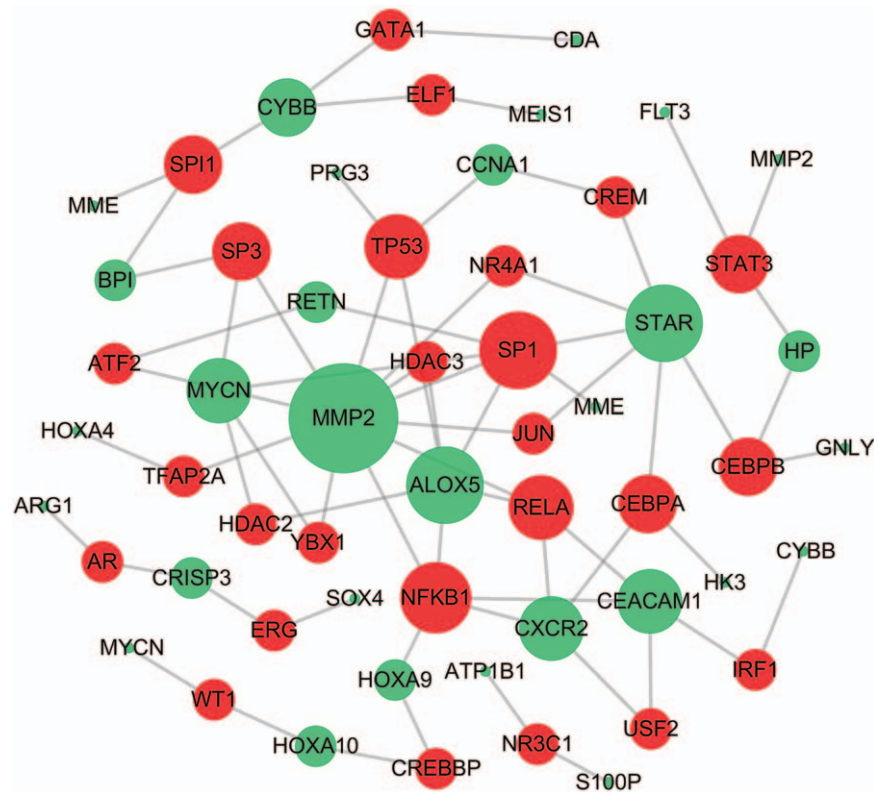


Figure 5. 26 transcription factors identified by TRRUST for hub differentially expressed genes.

development in leukemia cells by promoting autophagy and inhibition of apoptosis.^[17,18] CCNA1 is a cell cycle protein overexpressed in many myeloid leukemia cell lines and in AML patients; its overexpression plays a role in the growth and apoptotic suppression of leukemia cells via repression of WT1 synthesis.^[19] Other hub genes (ATP1B1, CDA, MEIS1, MN1, HK3, CAMP, SIGLEC5, MNDA, FCER1G, ALOX5, and CSF3R) have been associated with AML previously.

Of the various hub genes, some are also active in non-AML hematological or solid tumors. CLEC5A is a type II membrane protein critical for myeloid differentiation. CLEC5A dysregulation plays a role in the pathogenesis of myelodysplastic syndromes.^[20] NFE2 is overexpressed in most patients with myeloproliferative neoplasms, and its expression is regulated by epigenetic mechanisms that are perturbed in myeloproliferative neoplasms.^[21] TAL1 (SCL/TAL1, T-cell acute leukemia protein 1) is a TF involved in hematopoiesis and leukemogenesis. PADI4 is a promoter-dependent epigenetic cofactor of Tal1, which creates important epigenetic histone markers, and is a potential target of molecular therapy in some leukemias.^[22] MMP9 is a proteolytic enzyme that degrades components of the extracellular matrix and basement membrane; its overexpression is correlated with poor survival in ovarian cancer patients and promotes cancer cell invasion.^[23] C5AR1 (the receptor for the complement anaphylatoxin C5a) is a potent immune system mediator that plays a major role in malignant growth and dissemination of non-small cell lung cancer cells. Disruption of C5AR1 signaling in lung cancer cells abrogates tumor-associated osteoclastogenic activity, in turn impairing osseous colonization.^[24]

Of the 4 hub genes significantly correlated with OS in AML, SUSD3 is a cell-surface protein with unknown function that may be involved in the growth and development of breast cancer via the estrogen receptor pathway. Some evidence indicates that the gene acts downstream of the estrogen receptor and that estrogen enhances its expression. Lack of SUSD3 expression in breast cancer tissue was the most significant predictor of non-responsiveness to an aromatase inhibitor.^[25] Hippo signaling is an evolutionarily conserved pathway modulating organ growth, and SCHIP1 is a newly discovered member of the Hippo pathway. Loss of the SCHIP1 protein triggers overproliferation via upregulation of genes targeted by Yorkie (YAP in mammals), indicating that this protein serves as a tumor suppressor regulating the YAP oncogene.^[26] SPINK2 is a member of the family of Kazal-type serine proteinase inhibitors that inhibit trypsin/acrosin; SPINK2 is synthesized principally in the testes and seminal vesicles, promoting fertility. Dysregulation of SPINK2 is closely associated with cancers such as lymphomas; high levels of SPINK2 transcription in patients with primary cutaneous follicular center cell lymphomas are associated with improved prognosis and reduced mortality.^[27] Recently, SPINK2 expression was shown to predict poor outcomes in pediatric AML patients.^[28] Steroid hormones are synthesized by steroidogenic cells in various tissues and influence many developmental and physiological processes. STAR mediates the rate-limiting step in steroid biosynthesis, and a great deal of evidence supports the crucial role played by STAR in the regulation of steroid biosynthesis. Appropriate regulation of steroid hormone expression is essential for appropriate biological functioning, aiding

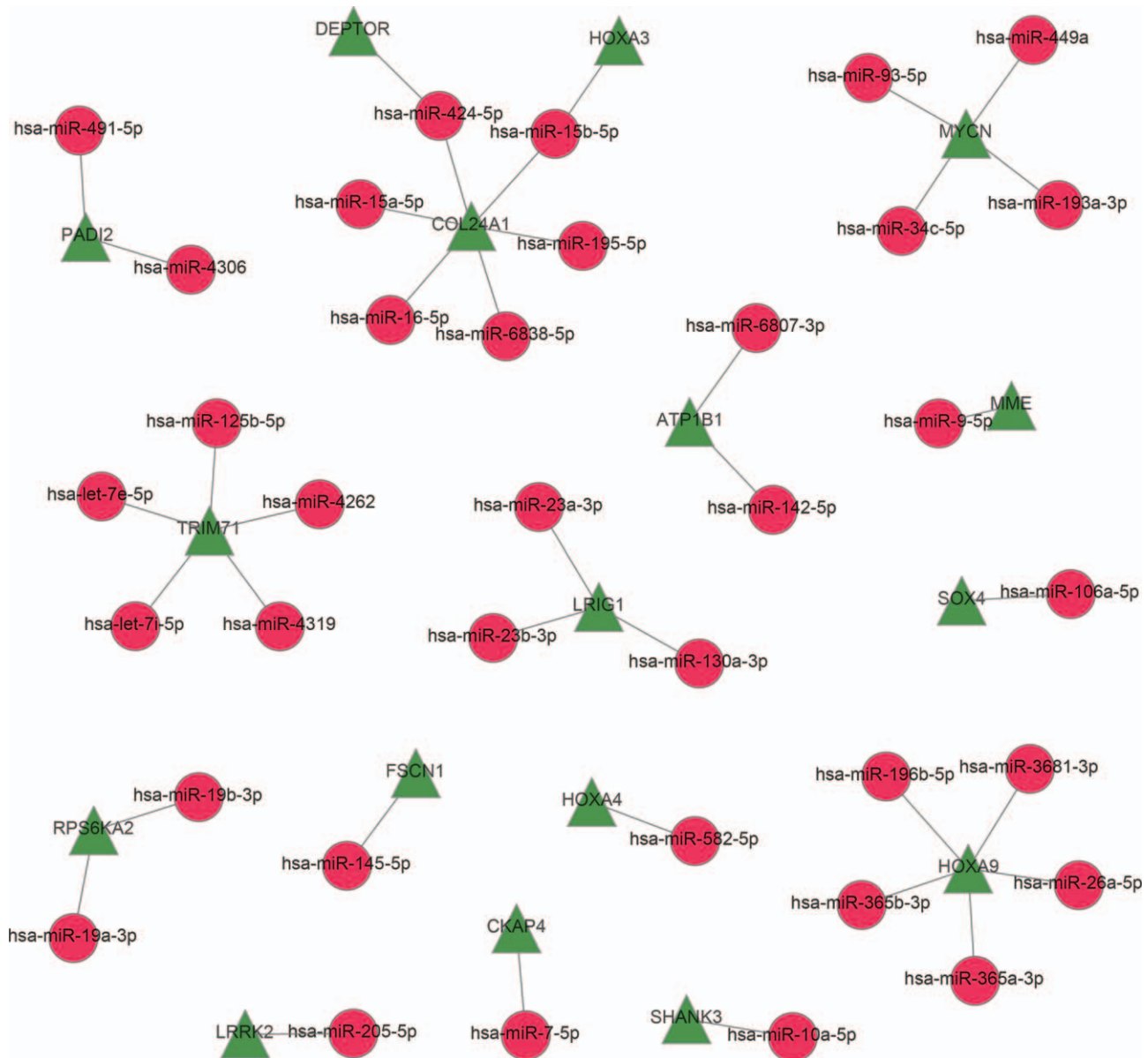


Figure 6. 38 micro ribonucleic acid -mRNA interactions were identified by miRWalk analysis based on Targetscan and miRDB database.

geriatric populations to live longer and healthier.^[29] However, whether such genes play roles in AML pathogenesis requires further investigation.

5. Conclusion

We performed a comprehensive bioinformatics analysis of hub genes that may be involved in AML development. The genes and signaling pathways identified may be of clinical utility. However, further *in vitro* and *in vivo* molecular biological experiments are required to confirm the functions of the identified genes.

Acknowledgments

The authors would like to thank other colleagues of the department of hematology for supporting this project.

Author contributions

ZYM designed the study. TYP, ZLL, and DYY downloaded and analyzed the data. TYP and ZYM interpreted the data and drafted the manuscript. LZ, LS, and ZQ consulted with others and helped to revise the manuscript.

References

- [1] Rodriguez-Abreu D, Bordoni A, Zucca E. Epidemiology of hematological malignancies. *Ann Oncol* 2007;Suppl 1:i3–8.
- [2] Dohner H, Estey EH, Amadori S, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* 2010;115:453–74.
- [3] Dohner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. *N Engl J Med* 2015;373:1136–52.

- [4] Haferlach T, Kohlmann A, Wiczorek L, et al. Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the International Microarray Innovations in Leukemia Study Group. *J Clin Oncol* 2010;28:2529–37.
- [5] Bullinger L, Dohner K, Bair E, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* 2004;350:1605–16.
- [6] Valk PJ, Verhaak RG, Beijen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* 2004;350:1617–28.
- [7] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559–72.
- [8] Jiang H, Wong WH. SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics* 2008;24:2395–6.
- [9] Verhaak RG, Wouters BJ, Erpelinck CA, et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica* 2009;94:131–4.
- [10] Burnett A, Wetzler M, Löwenberg B. Therapeutic advances in acute myeloid leukemia. *J Clin Oncol* 2011;29:487–94.
- [11] de Jonge HJ, Huls G, de Bont ES. Gene expression profiling in acute myeloid leukaemia. *Neth J Med* 2011;69:167–76.
- [12] Wilson CS, Davidson GS, Martin SB, et al. Gene expression profiling of adult acute myeloid leukemia identifies novel biologic clusters for risk classification and outcome prediction. *Blood* 2006;108:685–96.
- [13] Gilliland DG, Griffin JD. The roles of FLT3 in hematopoiesis and leukemia. *Blood* 2002;100:1532–42.
- [14] Truong BT, Lee YJ, Lodie TA, et al. CCAAT/Enhancer binding proteins repress the leukemic phenotype of acute myeloid leukemia. *Blood* 2003;101:1141–8.
- [15] Drabkin HA, Parsy C, Ferguson K, et al. Quantitative HOX expression in chromosomally defined subsets of acute myelogenous leukemia. *Leukemia* 2002;16:186–95.
- [16] Rice KL, Licht JD. HOX deregulation in acute myeloid leukemia. *J Clin Invest* 2007;117:865–8.
- [17] Yang XY, Zhang MY, Zhou Q, et al. High expression of S100A8 gene is associated with drug resistance to etoposide and poor prognosis in acute myeloid leukemia through influencing the apoptosis pathway. *Oncotargets Ther* 2016;9:4887–99.
- [18] Yang M, Zeng P, Kang R, et al. S100A8 contributes to drug resistance by promoting autophagy in leukemia cells. *PLoS One* 2014;9:e97242–53.
- [19] Krug U, Yasmeen A, Beger C, et al. Cyclin A1 regulates WT1 expression in acute myeloid leukemia cells. *Int J Oncol* 2009;34:129–36.
- [20] Inoue D, Kitaura J, Togami K, et al. Myelodysplastic syndromes are induced by histone methylation-altering ASXL1 mutations. *J Clin Invest* 2013;123:4627–40.
- [21] Peeken JC, Jutzi JS, Wehrle J, et al. Epigenetic regulation of NFE2 overexpression in myeloproliferative neoplasms. *Blood* 2018;131:2065–73.
- [22] Kolodziej S, Kuvardina ON, Oellerich T, et al. PADI4 acts as a coactivator of Tal1 by counteracting repressive histone arginine methylation. *Nat Commun* 2014;5:3995–4007.
- [23] Hu X, Li D, Zhang W, et al. Matrix metalloproteinase-9 expression correlates with prognosis and involved in ovarian cancer cell invasion. *Arch Gynecol Obstet* 2012;286:1537–43.
- [24] Ajona D, Zanduetta C, Corrales L, et al. Blockade of the complement C5a/c5ar1 axis impairs lung cancer bone metastasis by cxcl16-mediated effects. *Am J Respir Crit Care Med* 2018;197:1164–76.
- [25] Yu Z, Jiang E, Wang X, et al. Sushi domain-containing protein 3: a potential target for breast cancer. *Cell Biochem Biophys* 2015;72:321–4.
- [26] Chung HL, Choi KW. Schip1, a new upstream regulator of Hippo signaling. *Cell Cycle* 2016;15:2097–8.
- [27] Chen T, Lee TR, Liang WG, et al. Identification of trypsin-inhibitory site and structure determination of human SPINK2 serine proteinase inhibitor. *Proteins* 2009;77:209–19.
- [28] Lamba JK, Cao X, Raimondi SC, et al. Integrated epigenetic and genetic analysis identifies markers of prognostic significance in pediatric acute myeloid leukemia. *Oncotarget* 2018;9:26711–23.
- [29] Manna PR, Stetson CL, Slominski AT, et al. Role of the steroidogenic acute regulatory protein in health and disease. *Endocrine* 2016;51:7–21.