

Expression profile analysis of long non-coding RNA in acute myeloid leukemia by microarray and bioinformatics

Yuandong Feng¹ | Ying Shen¹ | Hongli Chen¹ | Xiaman Wang¹ | Ru Zhang¹ | Yue Peng¹ | Xiaoru Lei² | Tian Liu¹ | Jing Liu¹ | Liufang Gu¹ | Fangxia Wang¹ | Yun Yang¹ | Ju Bai¹ | Jianli Wang¹ | Wanhong Zhao¹ | Aili He¹ 

¹Department of Hematology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

²Department of Hematology, The Xi'an Central Hospital, Xi'an, China

Correspondence

Aili He, Department of Hematology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China.
Email: heaili_2008@163.com

Funding information

Science and Technology Coordinating Innovative Engineering Projects of Shaanxi Province (Grant/Award Number: no. 2014KTCL03-01).

Long non-coding RNAs (lncRNAs) are transcripts longer than 200 nt that are involved in tumorigenesis and play a key role in cancer progression. To determine whether lncRNAs are involved in acute myeloid leukemia (AML), we analyzed the expression profile of lncRNAs and mRNAs in AML. Five pairs of AML patients and iron deficiency anemia (IDA) controls were screened by microarray. Through co-expression analysis, differently expressed transcripts were divided into modules, and lncRNAs were functionally annotated. We further analyzed the clinical significance of crucial lncRNAs from modules in public data. Finally, the expression of three lncRNAs, RP11-222K16.2, AC092580.4, and RP11-305O.6, were validated in newly diagnosed AML, AML relapse, and IDA patient groups by quantitative RT-PCR, which may be associated with AML patients' overall survival. Further analysis showed that RP11-222K16.2 might affect the differentiation of natural killer cells, and promote the immunized evasion of AML by regulating Eomesodermin expression. Analysis of this study revealed that dysregulated lncRNAs and mRNAs in AML vs IDA controls could affect the immune system and hematopoietic cell differentiation. The biological functions of those lncRNAs need to be further validated.

KEYWORDS

acute myeloid leukemia, bioinformatics, long non-coding RNA, microarray, transcriptome network

1 | INTRODUCTION

Acute myeloid leukemia (AML) is the most common form of hematological malignant tumor that threatens human health. In the last decade, the rapid evolution in the detection of molecular abnormalities has brought more and more precise prediction of prognosis and diagnosis, which can efficiently guide post-remission therapy and personal treatment.¹ The development of high-throughput screening technologies (HTS), such as next-generation sequencing and cDNA microarray, has identified several new mutations (including

DNMT3A, IDH1, IDH2, and TET2) in AML.^{2,3} However, the vast majority of transcripts that were detected by HTS do not appear to be protein-coding genes. This phenomenon is notable in non-coding RNAs, which have been vividly described as the "dark matter" of the genome.⁴ In some diseases there are no point mutations in the protein-coding genes. The pathogenic mechanisms may be multifactorial and likely to involve genetic elements additional to small non-coding RNAs, as seen in 13q14.3 of solid tumors and hematopoietic malignancies. Long non-coding RNAs (lncRNAs) are probably the chief culprits.⁵

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2017 The Authors. *Cancer Science* published by John Wiley & Sons Australia, Ltd on behalf of Japanese Cancer Association.

In contrast to small non-coding RNAs, what we know about lncRNAs is confined to mRNA-like transcripts lacking significant ORFs and over 200 nt in length. Recent studies indicate that lncRNAs may play a key role in cancer pathways.⁶ In AML, lncRNAs are important fractions of the biomarkers detected by microarray. Its expression may discriminate acute leukemia molecular subtypes, which may provide a more precise tool to categorize leukemia and stratify patients.^{7,8} Various lncRNAs were reported to be implicated in AML. For example, lncRNA CCAT1 can regulate microRNA-155 that can target for c-Myc, which in turn can activate CCAT1.⁹ The study was just a glance of the complex world constructed by mRNA, microRNA, and lncRNA. Furthermore, Garzon et al built a prognostic lncRNA score system for older patients (>60 years) with cytogenetically normal AML.¹⁰ It is anticipated that lncRNAs will be used in clinical diagnosis and treatment after large-scale clinical trials and functional studies are completed in the near future.

This study analyzed the expression profiles of lncRNAs and mRNAs in AML vs iron deficiency anemia (IDA) controls, with a focus on lncRNAs playing a big part in AML in the modularization process. In particular, using public databases, we identified the clinical significance of lncRNAs, and validated their expression in AML patients by quantitative real-time PCR (qRT-PCR).

2 | MATERIALS AND METHODS

2.1 | Patients and samples

Bone marrow specimens were obtained from AML patients at the Department of Hematology, Second Affiliated Hospital of Xi'an Jiaotong University, and Xi'an Jiaotong University (both Xi'an, China) in 2016. This study was approved by the Medical Ethics Committee of the Second Affiliated Hospital of Xi'an Jiaotong University (#2015182), and written informed consent was obtained from all parents or guardians. Diagnosis of AML was made in accordance with the revised French-American-British classification. One hundred and fifty-one patients included in The Cancer Genome Atlas (TCGA), investigated using RNA sequencing technology, were analyzed to evaluate prognostic values of lncRNAs.¹¹ The ID numbers of the patients are shown in Table S1. To weaken the variation between samples as much as possible, bone marrow samples from IDA were used as controls, for the deficiency of bone marrow samples from normal donors in clinic. Bone marrow mononuclear cells were isolated using lymphocyte separation liquid within 8 hours after bone marrow samples were harvested and subjected for isolation of total cellular RNA, then stored at -80°C . Detailed information of all cases in the study is summarized in Table S2.

2.2 | Profiling of lncRNA expression

Arraystar Human lncRNA Array version 4.0 was used to profile the expression of lncRNAs, which was performed by KangChen Bio-tech (Shanghai, China). Briefly, RNA samples from bone marrow

mononuclear cells were purified to remove rRNA and were amplified and transcribed into fluorescent cRNAs along the entire length of the transcripts without 3' bias. Then, each cRNA was hybridized to the Arraystar Human lncRNA Array. The array was designed for the global expression profiling of human lncRNA and protein-coding mRNA transcripts, which can detect a total of 40 173 lncRNAs in two tiered compilations: gold standard lncRNAs for 7506 well-annotated functionally studied and experimentally supported full-length lncRNAs, and reliable lncRNAs for 32 667 high confidence lncRNAs as the comprehensive collection. The lncRNAs were carefully constructed using the most highly respected public transcriptome databases (e.g. RefSeq, UCSC known genes, and Ensembl), as well as landmark publications. The array also includes an entire collection of 20 730 protein-coding mRNAs further supported by the UniProt catalog. Data were deposited in the Gene Expression Omnibus database (www.ncbi.nlm.nih.gov/geo, accession no. GSE103828).

2.3 | Mapping and identification of differentially expressed genes

We applied Agilent GeneSpring GX version 12.1 to screen out the differentially expressed genes using the following criteria: (i) fold change >2 for upregulation or downregulation; (ii) P -value $<.05$; and (iii) false discovery rate (FDR) <0.05 .

2.4 | Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway analyses

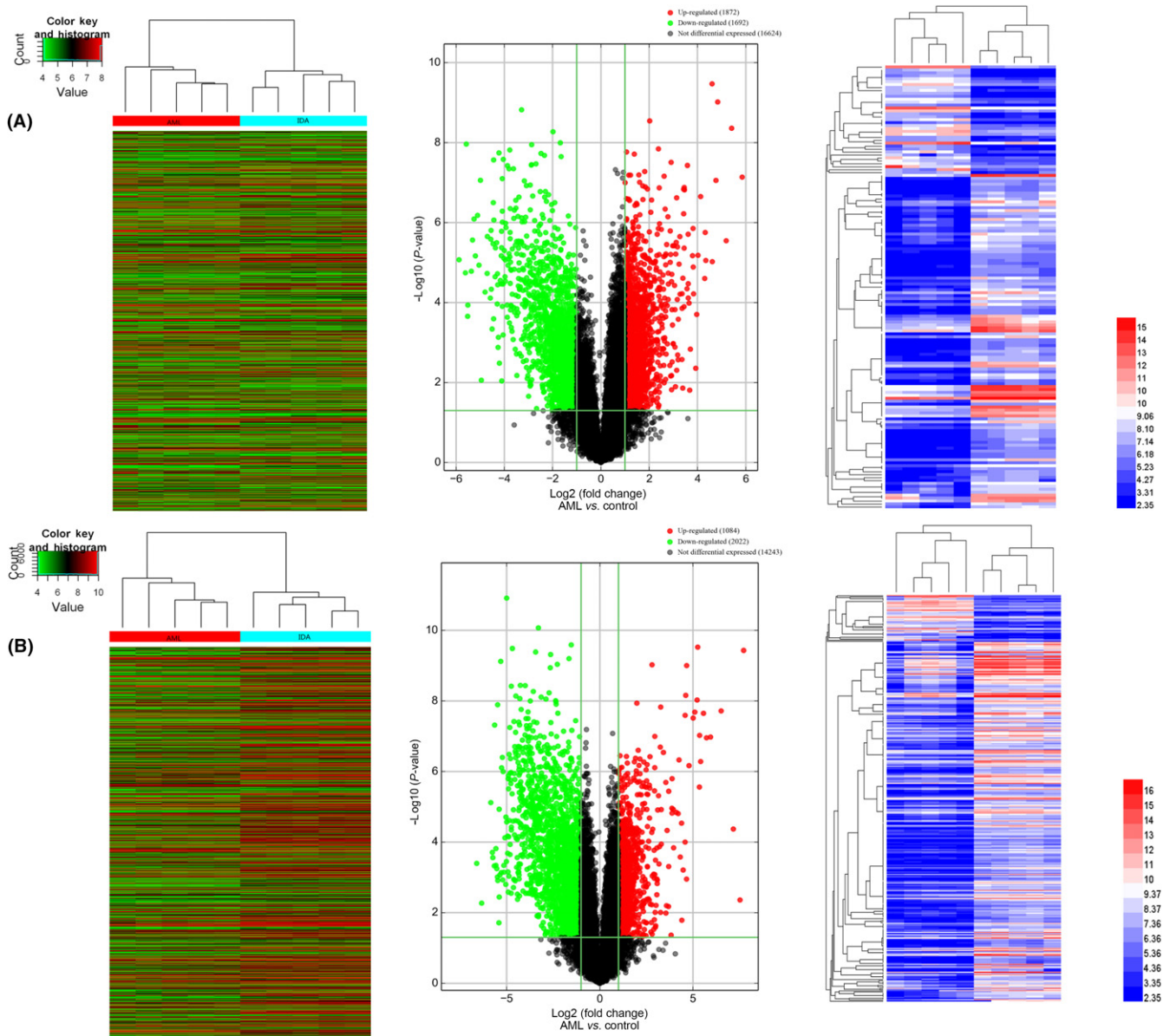
Gene Ontology (GO) analysis was undertaken to facilitate the understanding of the unique biological significance of the genes in the distinctive or representative profiles of the differentially expressed genes.¹² Pathway analysis of differentially expressed genes was carried out to find out the important pathways, based on the latest Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The significant GO terms and pathways were identified by Fisher's exact test, and FDR was utilized to correct the P -values.

2.5 | Correlation and coexpression analysis

The coexpression analysis was based on weighted correlation network analysis (WGCNA), a systems biology method for constructing relationship patterns.¹³ Compared to general methods, such as Pearson's correlation coefficient, WGCNA uses the soft threshold, which can provide more extensive and exact correlation between transcripts. Differentially expressed lncRNAs and mRNAs with fold change ≥ 4 , $P <.05$, and FDR <0.05 were analyzed. The value of parameter soft threshold ≥ 0.98 and P -value $<.05$ was recommended for the coexpression analysis. k -Core scoring was used to determine core transcripts of coexpression networks. A higher k -core score means a more central location of a transcript within a network.¹⁴ The soft threshold was adjusted to 0.8 to obtain the lncRNA coexpressed mRNA cluster for further functional analysis of lncRNAs.

TABLE 1 Real-time quantitative PCR primer sequences used in this study

Primer name	Forward (5'-3')	Reverse (5'-3')
β-actin (H)	GTGGCCGAGGACTTTGATTG	CCTGTAACAACGCATCTCATATT
RP11222k16.2	CTAAACTTTTGGAGTCGCTGTG	CCATTCCGCTGGACACTTAT
AC092580.4	GACCAAAGAGAAACAAGAAAAGC	CAAGAGAGACAGATCGTCCAGC
linc00944	CCCGGAAACATCATCTCATT	GAGTTACAGGGACCGAAGCA
linc00899	CCCAACAGGAAGGTCTGGT	TCAGTGCTGGTCATTCTTG
RP1-109B7.5	CACTCGACACAGGACAGCAG	CAGCTTAACTCCTCCCATGC
RP11-305O6.3	TGCTTAACCTCCCTCAGTG	GTGAGGAACGAGGAGGAGTG

**FIGURE 1** Volcano plots and heat map showing expression profiles of long non-coding RNAs (lncRNAs) (A) and mRNAs (B) in acute myeloid leukemia. Left panels, plots are based on the expression values of all lncRNAs and mRNAs detected by microarray. Middle and right panels, maps showing significantly changed lncRNAs and mRNAs with fold change ≥ 2.0 and ≥ 10.0 respectively ($P < .05$; false discovery rate < 0.05)

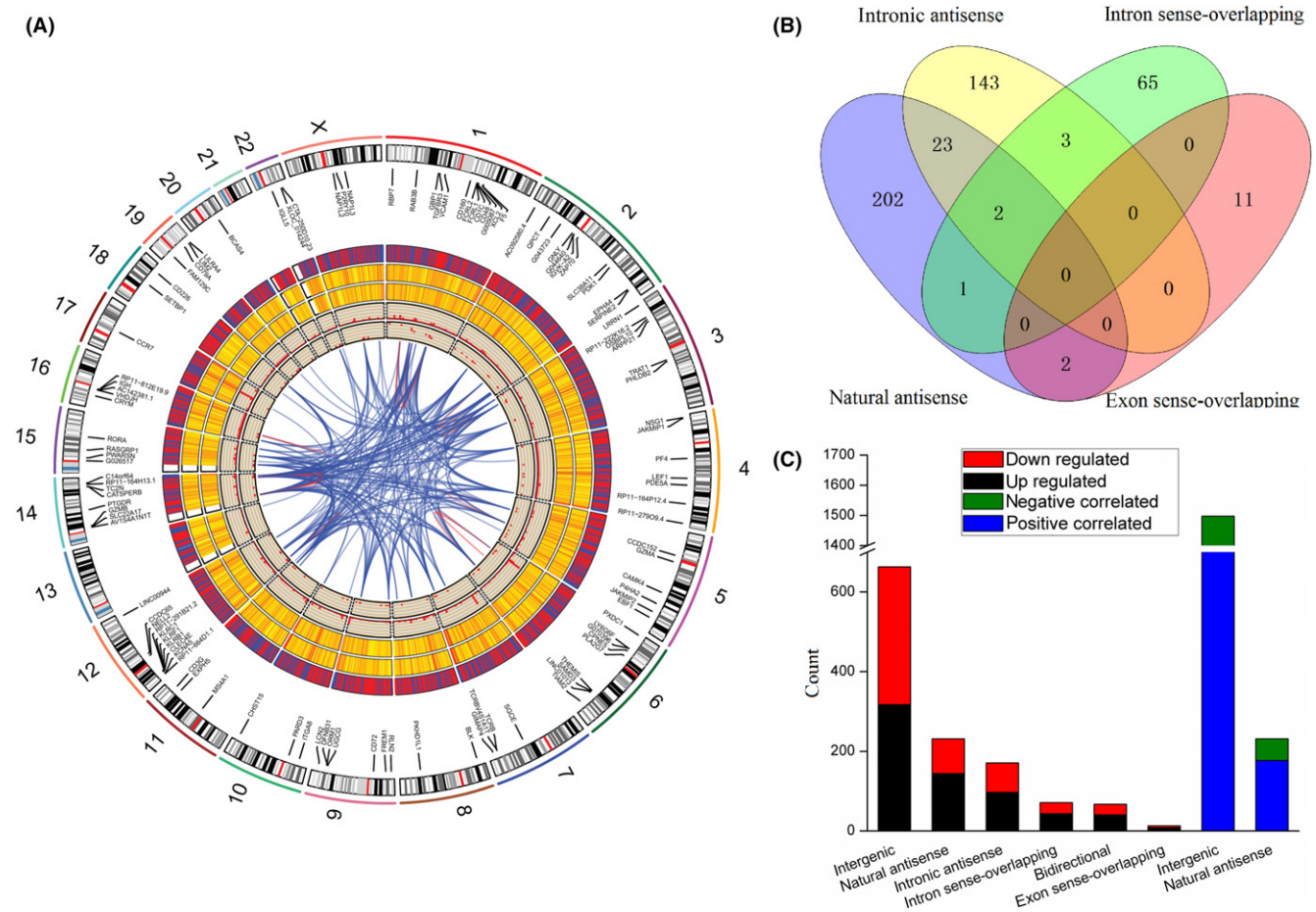


FIGURE 2 Identification of differentially expressed long non-coding RNAs (lncRNAs) in acute myeloid leukemia (AML). A, Circos plot showing lncRNAs and mRNAs on human chromosomes. From the outside in, the first layer of the Circos plot is a chromosome map of the human genome, black and white bars are chromosome cytobands, and red bars represent centromeres. The transcripts, of which the k -score in the coexpression network is >7 , are labeled in the second circle. All differentially expressed lncRNAs and mRNAs are marked in red and blue, respectively, in the third layer. The fourth and fifth layers represent the mean expression values of significantly differentially changed lncRNAs and mRNAs in AML and Control groups. The sixth circle shows the fold change of all differentially expressed lncRNAs and mRNAs with fold change ≥ 2.0 , $P < .05$, and false discovery rate < 0.05 . The innermost circle indicates the k -score of the labeled transcripts. The network in the center of the plot represents the core network; red lines indicate the linked transcripts in the same chromosome, blue in different chromosomes. B, Types and counts of differently regulated lncRNAs classified into six categories according to the genomic loci of their neighboring genes. The two correlation types of intergenic and natural antisense lncRNAs with their associated genes are also shown. C, Venn diagram presents overlapping relationships, and the numbers indicate lncRNA counts

2.6 | Gene Ontology annotations of lncRNA-correlated mRNAs

The lncRNA coexpressed mRNAs, calculated by WGCNA, were analyzed by DAVID tools for GO analysis. Fisher's exact test was applied to identify the significant GO terms, and FDR was utilized to correct the P -values.

2.7 | Validation of lncRNAs by qRT-PCR

Expression of six lncRNAs was validated by qRT-PCR. The cDNA was synthesized by reverse transcription using a PrimeScript RT reagent kit with random primers according to the manufacturer's protocols (TaKaRa). Then, qRT-PCR was carried out using SYBR Premix Ex Taq

II (Tli RNaseH Plus; TaKaRa). Primers for RP11-222k16.2, AC092580.4, LINC00944, LINC00899, RP1-109B7.5, RP11-305O.6, and β -actin were synthesized by Invitrogen (Shanghai, China). All qRT-PCR primer sequences are shown in Table 1.

2.8 | Statistical analysis

Statistically significant differences between groups were estimated by the Mann-Whitney U -test for the expression of lncRNAs using SPSS 23.0 (IBM, SPSS, Chicago, IL, USA); $P < .05$ was considered statistically significant. Correlation of transcripts were evaluated using Pearson's correlation. The Kaplan-Meier survival curves were used to show the differences in patients' overall survival (OS) between the high expression group and low expression group, and the statistical significance

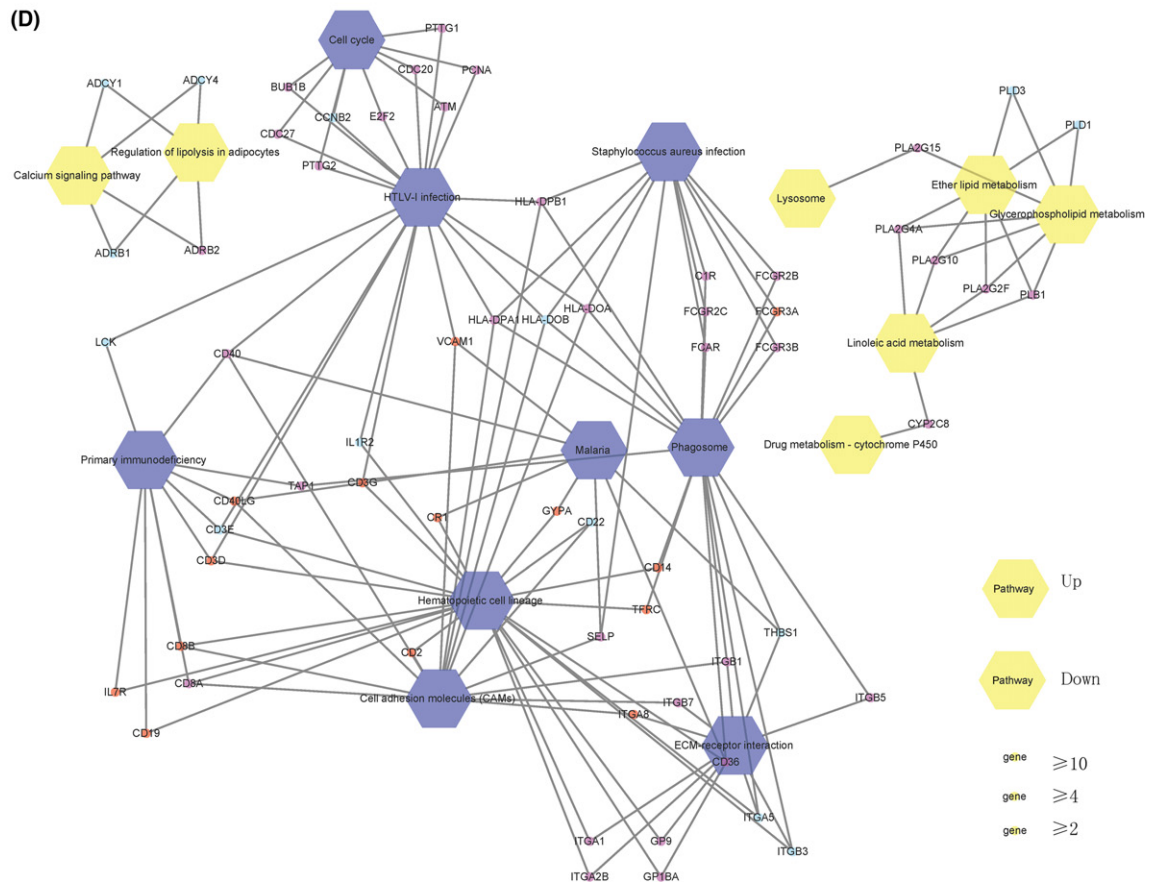
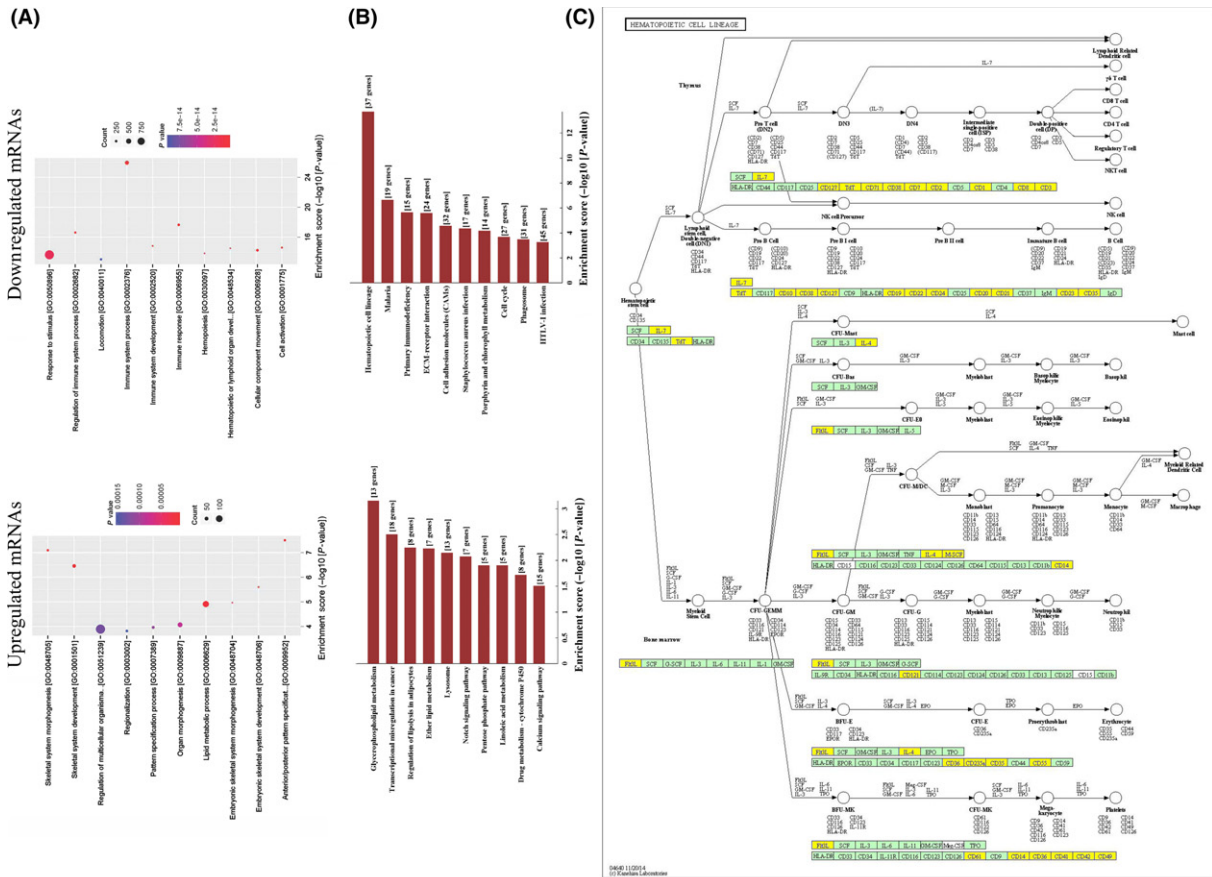


FIGURE 3 Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses of mRNAs in acute myeloid leukemia. A, GO annotations of up- and down regulated mRNAs with top 10 enrichment scores of biological processes. B, KEGG pathway enrichment analysis of up- and downregulated mRNAs with top 10 enrichment scores. C, KEGG pathway annotations of the hematopoietic cell lineage pathway. Yellow marked nodes are associated with downregulated genes; green nodes have no significance. D, Interaction and overlapping of associated molecules among significant pathways. Hexagons represent the most significantly enriched pathways; ellipses indicate mRNAs that act as the link hinge between the pathways. Different colors of mRNA nodes represent the absolute value of fold change

was obtained using the two-sided log-rank test. Cox regression was used to analyze the significance of lncRNAs for OS more deeply.

3 | RESULTS

3.1 | Differentially expressed lncRNAs and mRNAs in AML

Volcano plots were used for assessing gene expression variation between AML and IDA patient groups. In total, 3564 lncRNAs displayed differential expression in AML, including 1872 upregulated lncRNAs and 1692 downregulated lncRNAs. Of 3106 mRNAs that showed differential expression, 1084 were upregulated and 2022 were downregulated. Among them, 37 lncRNAs and 42 mRNAs were significantly upregulated, and 112 lncRNAs and 317 mRNAs were significantly downregulated >10-fold in AML. Hierarchical clustering analysis showed systematic variations in the expression of lncRNAs and mRNAs among samples. The data suggested that the expression of lncRNAs and mRNAs in AML differ from those in IDA controls (Figure 1).

These lncRNAs and mRNAs are widely distributed in all chromosomes covering chromosome X and Y. The transcripts located in chromosome Y are excluded to eliminate gender's effect (Figure 2A). The well-annotated lncRNAs (totally 1216) were classified into six categories: 14.1% were intronic antisense, 5.8% were intron sense-overlapping, 5.5% were bidirectional, and 1.1% were exon sense-overlapping. There are overlaps between these four categories (Figure 2B). Intergenic and natural antisense lncRNAs constitute the largest number in all differentially expressed lncRNAs, and comprised 54.5% and 19.0%, respectively, in this study. We also noted that, among the 1498 and 231 pair relationships, 57.9% of intergenic lncRNAs and 76.6% of natural antisense were positively correlated with their neighboring genes (Figure 2C).

3.2 | Functional analysis of differentially expressed genes

Until now, the functions of most lncRNAs have not been well annotated. Therefore, by analyzing differentially expressed mRNAs, we can forecast the role that lncRNAs play in AML. The GO and KEGG pathway analyses of differentially expressed mRNAs could provide a clue about the AML disease process. We utilized all differentially expressed mRNAs for the GO analysis and found that the most enriched GO targeted by upregulated and downregulated transcripts were involved in anterior/posterior pattern specification, immune system processes, and immune response (Figure 3A). In the KEGG

pathway analysis, the down- and upregulated mRNAs were found to be mostly enriched in hematopoietic cell lineage and glycerophospholipid metabolism, respectively (Figure 3B). Many genes involved in hematopoietic cell differentiation were dysregulated (Figure 3C). A pathway network was constructed using 20 of the most significantly enriched pathways to illustrate the critical pathways in the process of AML. The hematopoietic cell lineage pathway and cell adhesion molecules pathway were considered to be the most central functions in the net because the exchanges with other pathways strongly depended on their existence (Figure 3D).

To further investigate the function of genes at the protein level, and to reveal the core mRNAs in the cellular process of AML, a biological database, Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) was used to filter functional genes, thus providing an intuitive network for annotating structural and functional properties of proteins.¹⁵ The highest confidence score (0.9) was adopted to evaluate the protein interactions for the differentially expressed genes. The network contains 1000 nodes and 636 edges (Figure 4A). The *k*-score was used to assess the importance of genes, and the 30 highest *k*-score genes constitute two important subnetworks (Figure 4B). The networks are enriched in the G-protein-coupled receptor signaling pathway and T-cell receptor signaling pathway (Figure 4C).

3.3 | Long non-coding RNA/mRNA coexpression network in AML

Transcriptome regulation involves a huge network, among which many transcripts form a complex web to function. Coexpression networks facilitate the intricate network based on gene screening methods that can be used to identify candidate biomarkers or therapeutic targets. The coexpression network comprised 676 network nodes and 1283 connections, and several prominent subnetworks were formed. The most crucial subnetwork was constructed by the transcripts with a high *k*-score, which would be the core regulatory modules of the entire coexpression network (Figure 5). What is more, the transcripts in this subnetwork are widely distributed in all chromosomes, indicating the widely interconnected regulation network between lncRNAs and mRNAs (Figure 2A). This subnetwork includes four lncRNAs, RP11-222K16.2, G005087, G044640, and ANKRD36BP2, constituting probably the core of the network.

3.4 | Coexpression and GO annotations to predict the probable functions of lncRNAs

It is well known that lncRNAs could regulate the expression of neighboring coding genes by cis-pattern, and affect distant genes

by trans-pattern, even in different chromosomes.^{16,17} We constructed mRNA functional modules to clarify the biological role of differently expressed lncRNAs. The mRNAs coexpressed with

lncRNA were annotated by GO terms. The enriched GO terms ($P < .05$) could reflect the function of lncRNAs. The analysis flow chart of RP11-222K16.2 is shown as an example (Figure 6A,B). To

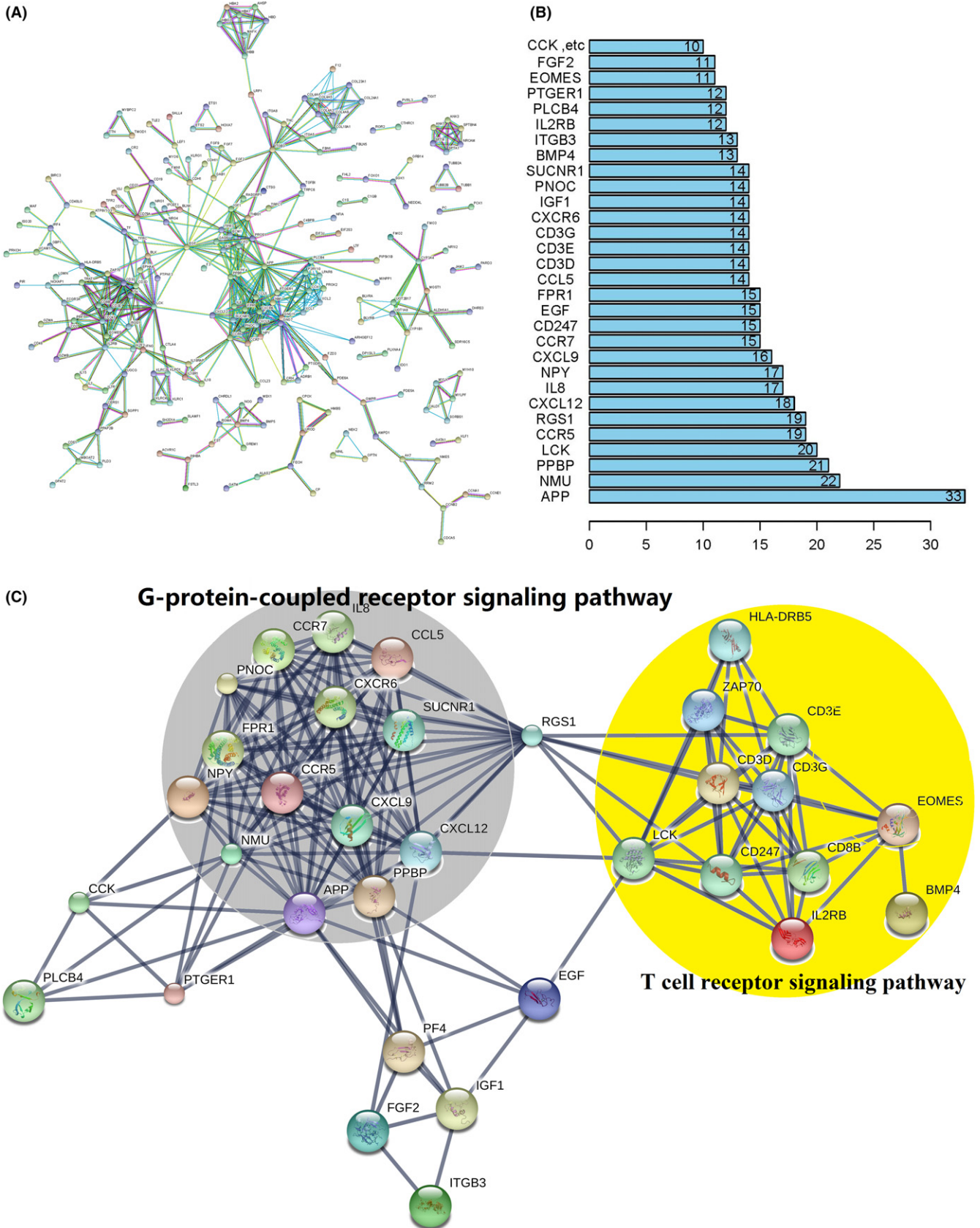


FIGURE 4 Protein-protein interaction networks by Search Tool for the Retrieval of Interacting Genes/Proteins (STRING). A, STRING software constructed the differentially expressed mRNAs (fold change ≥ 2 , $P < .05$, and false discovery rate [FDR] < 0.05) network based on protein-protein interactions. A confidence score that calculated for all protein interactions based on experimentally and computationally interaction was set as the highest (>0.9). B, Thirty top k -score genes involved in the network. C, Thirty top k -score genes constitute the two core subnetworks. Yellow, those enriched in T-cell receptor signaling pathway (FDR = $1.20E-07$); gray, enriched in the G-protein-coupled receptor signaling pathway (FDR = $1.97E-07$)

confirm the correlation, the multi-experiment matrix was applied to obtain more correlated mRNAs, if the lncRNAs were included in the database.¹⁸ Then, by their connections with mRNAs, lncRNAs were functionally annotated. We analyzed all the lncRNAs with a fold change >10 and constructed three separate networks to show the connections of lncRNAs with GO annotations. The core biological process GO terms of each separate system were: GO: 0002250, adaptive immune response; GO: 0009952, anterior/posterior pattern formation; and GO: 0015671, oxygen transport. Referring to the results of GO annotations for differentially

expressed mRNAs, the core GO terms for lncRNAs correctly reflect the biological process of AML.

3.5 | Survival analysis using TCGA for lncRNAs

After the systemic analysis of lncRNA and mRNA expression profiling, we obtained the core lncRNAs from the coexpression network and the GO annotations enrichment map. Through analysis of the TCGA dataset, which provides extensive genetic studies of human gene expression and specific disease associations, we found several

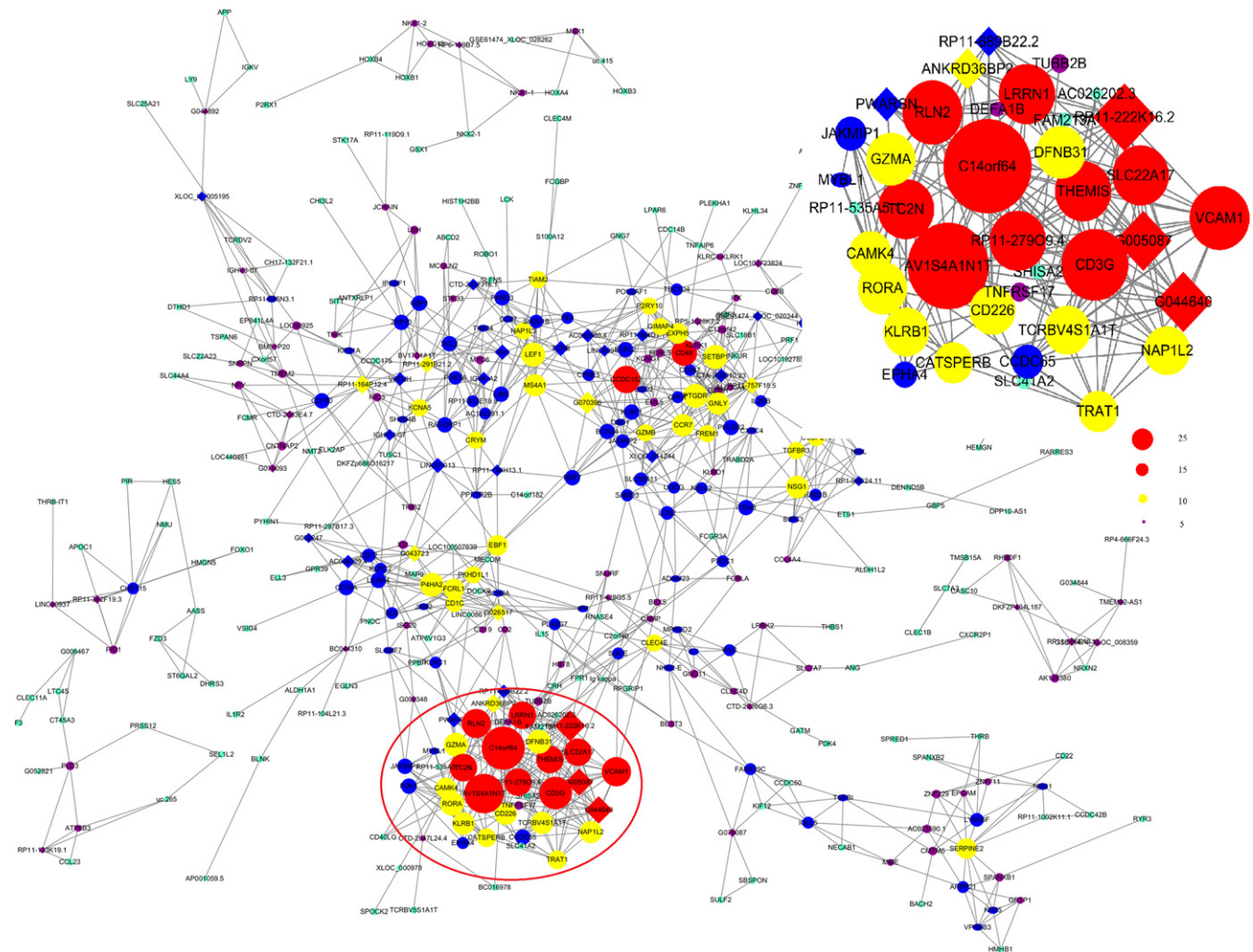


FIGURE 5 Coexpression networks constructed by weighted correlation network analysis. Circle, mRNA; diamond, long non-coding RNA; line, correlative relationship. Different sizes and colors representing the corresponding k -core scoring; we highlighted the highest k -core subnetwork

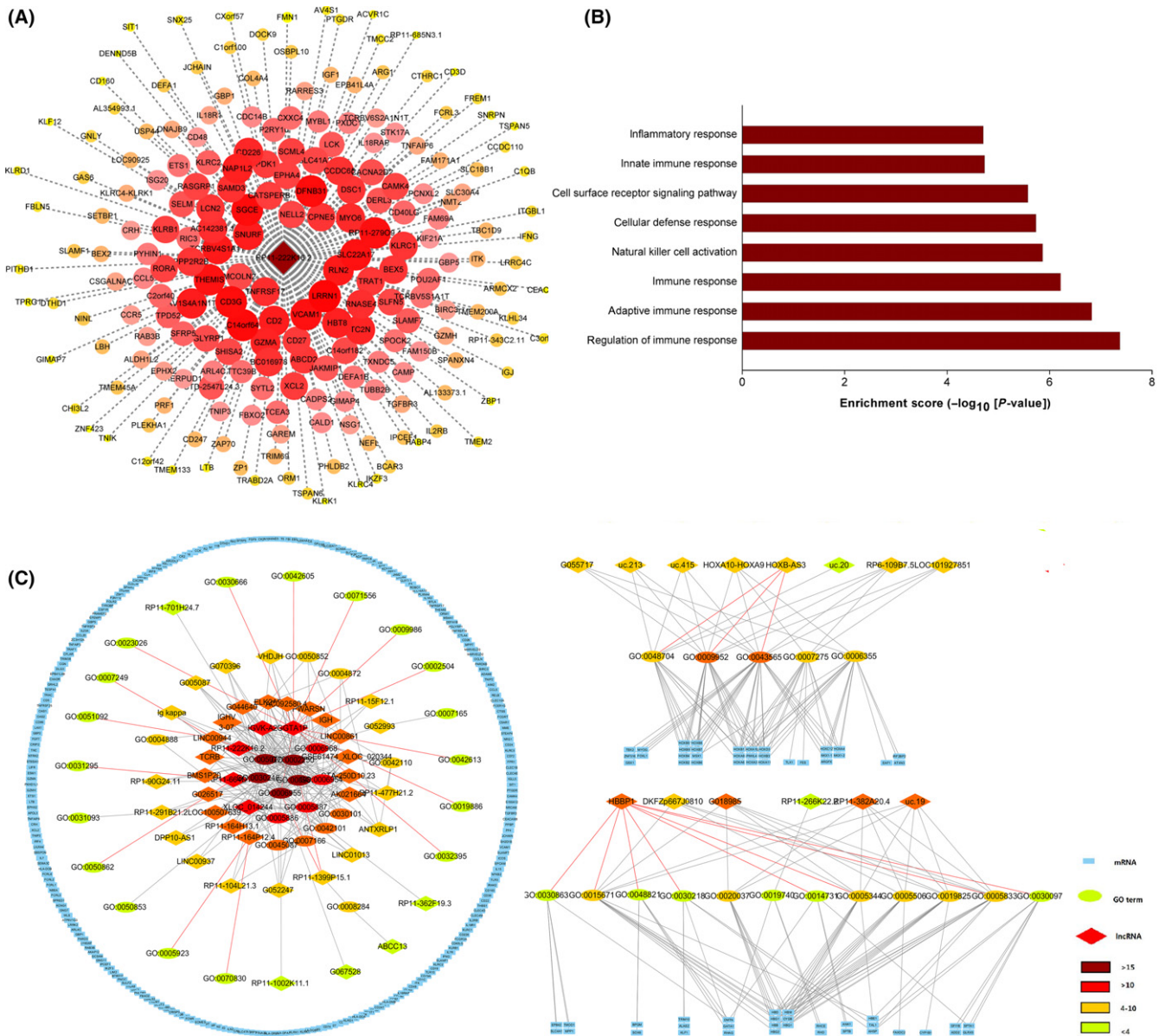


FIGURE 6 Gene Ontology (GO) annotations for long non-coding RNAs (lncRNAs) in acute myeloid leukemia. A, mRNAs that coexpressed with lncRNA RP11-222K16.2. Weighted correlation network analysis based on Pearson's correlation was used to estimate the correlation coefficient between the lncRNA and coding genes, and the soft threshold was set at 0.8. B, GO annotations enriched in the lncRNA RP11-222K16.2 coexpressed mRNA cluster (P -value < .05, false discovery rate [FDR] < 0.05). C, GO annotations enrichment maps for lncRNAs with fold change >10, P -value < 0.05, FDR < 0.05, and GO terms with P -value < 0.05. Different colors represent how many other nodes that the lncRNA/GO term is linked with. Red lines represent the relationship is confirmed by the multi-experiment matrix database

lncRNAs that are involved the prognosis of AML patients. Of 151 patients in the AML dataset, 130 had full transcript sequencing and survival time data. Hundred and forty seven lncRNAs in our microarray had fold change >10 or belonged to the core of the coexpression network, and 44 of these 147 lncRNAs had the expression profiling in the TCGA dataset. Based on the median value of lncRNA expression, patients were divided into high expression and low expression groups for each lncRNA. The Kaplan-Meier survival curves of 130 patients with AML showed that expression of eight lncRNAs was correlated with OS (Figure 7A). Among the eight lncRNAs, lncRNA RP11-222k16.2 is the core of the coexpression network, and

lncRNA HOXB-AS3 has been reported to be upregulated in *NPM1*-mutated AML.¹⁹ In order to find independent factors for patients' OS, a multivariate regression analysis was further carried out on the expression levels of eight lncRNAs with OS, and other individual clinical features were also considered. The results showed RP11-222k16.2, LINC00899 and RP11-3050.6 were significant to independently predict patients' OS ($P < .05$). The significant factors are summarized in Table 2. All 130 patients were assigned to a high-risk group or a low-risk group using the median risk score as the cut-off point. The Kaplan-Meier analysis showed that there are significant differences in patients' OS between high-risk and low-risk

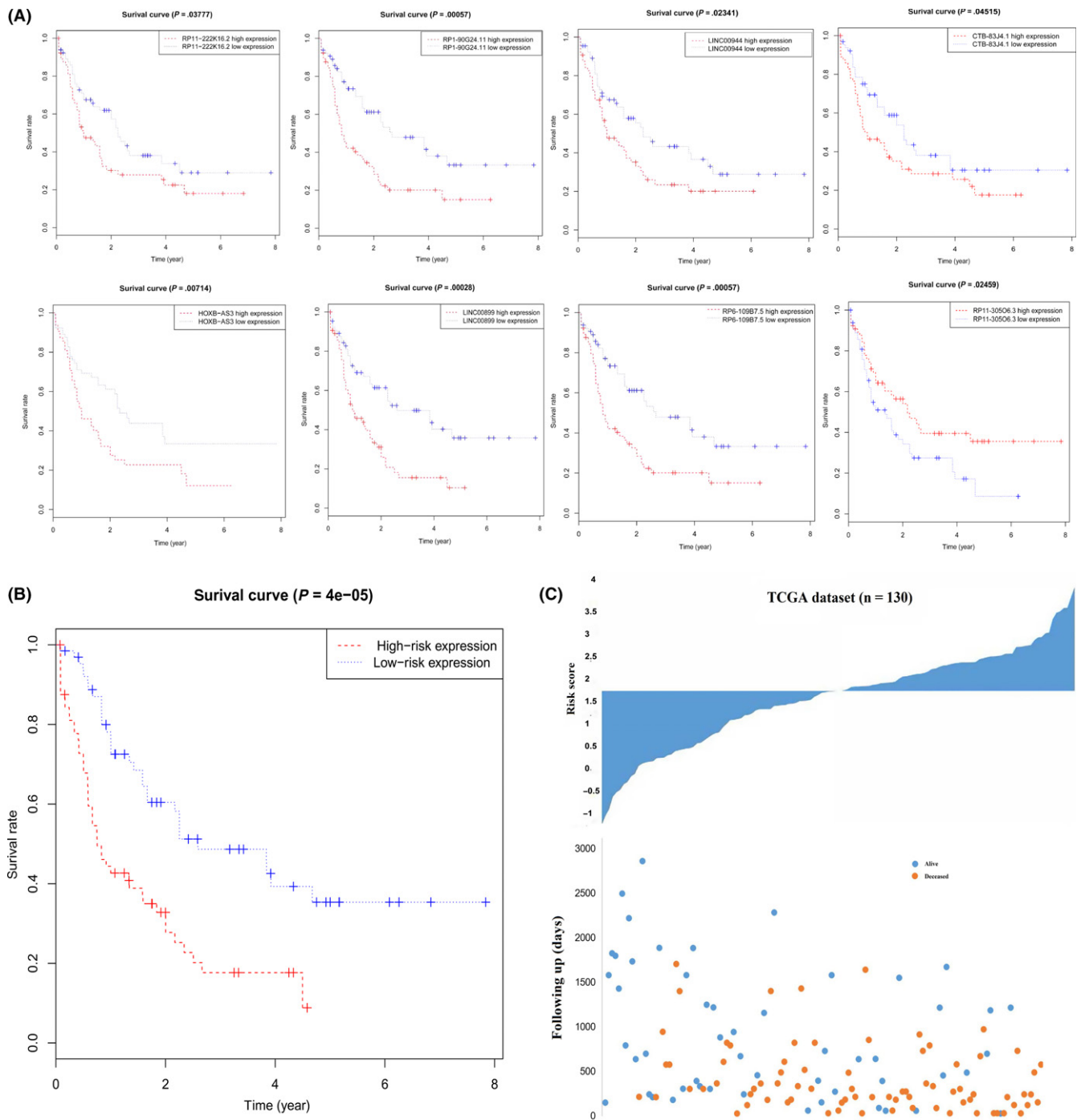


FIGURE 7 Expression of eight long non-coding RNAs (from The Cancer Genome Atlas [TCGA]) correlates with survival in acute myeloid leukemia. A, Kaplan-Meier survival curves of 130 patients with acute myeloid leukemia. The median value of expression was set as the cut-point. B, Kaplan-Meier analysis for overall survival of patients with high-risk or low-risk scores in the TCGA dataset. C, Risk score distribution and survival status of the risk score model in 130 patients of TCGA dataset

groups ($P < .001$, log-rank test; Figure 7B). Patients in the high-risk group had significantly shorter mean OS (569.7 days) than those in the low-risk group (1448.5 days). In conclusion, the three lncRNAs not only expressed abnormally in AML, but also had important clinical significance. Further functional studies for these lncRNAs would be valuable.

3.6 | Validation of dysregulated lncRNAs in AML vs IDA controls

To validate the microarray data, we used qRT-PCR to detect lncRNA expression. RP11-222k16.2 and AC092580.4 that belonged to the core of the coexpression network and had the highest GO

TABLE 2 Multivariable Cox regression analysis of eight long non-coding RNAs and other individual clinical features

	Hazard ratio	Coefficient	Wald	P-value
Variables in the equation				
RP11-222K16.2	0.562	-0.577	3.915	.048
LINC00899	1.811	0.594	4.973	.026
RP11-3050.6	0.464	-0.767	7.487	.006
Age	1.027	0.027	10.230	.001
Cytogenetics risk category	1.632	0.490	5.620	.018
White blood cell	1.007	0.007	4.164	.041
Variables not in the equation				
LINC00944	—	—	0.971	.324
RP1-90G24.11	—	—	0.294	.587
CTB-83J4.1	—	—	0.387	.534
HOXB-AS3	—	—	0.121	.728
RP6-109B7.5	—	—	0.954	.329
Gender	—	—	1.526	.217
Bone blast cell	—	—	0.361	.548
Blast cell	—	—	2.830	.092

The backward step (likelihood ratio) method was used. Variables are summarized. —, not available.

annotation enrichment (Figures 5 and 6C) were confirmed in bone marrow from 82 AML patients (51 newly diagnosed, 8 relapses, and 23 complete response) and 17 IDA patients. LINC00944, LINC00899, RP6-109B7.5, and RP11-3050.6 were detected in 22 newly diagnosed AML and 10 IDA cases. Through our confirmation, lncRNA RP11-222K16.2, AC092580.4, LINC00944, and RP11-3050.6 had different expressions in AML and IDA, but LINC00944 lost significance. Furthermore, RP11-222K16.2 and AC092580.4 had validated difference in newly diagnosed and recurrent AML between complete response patients (Figure 8).

3.7 | Long non-coding RNA RP11-222K16.2 involved in the immune system through cis-regulated Eomes

As mentioned, RP11-222K16.2 was the core lncRNA in the coexpression network. The GO annotations showed that it was involved in the immune system. What is more, its expression level was correlated with AML patients' OS. To understand how lncRNA RP11-222k16.2 functions in AML, we had the deeper insight into the AML transcripts' sequencing data from the TCGA. We calculated the Spearman's correlation of lncRNA RP11-222k16.2 with approximately 37 100 transcripts in 151 AML patients. Eomes is the mRNA with the highest correlation with lncRNA RP11-222K16.2, and this gene is located downstream of the lncRNA within 10K (Figure 9A,B); pan-cancer data from the chip-base database strengthened the correlation (Figure 9C).²⁰ Based on Flank 10K theory,²¹ we supposed that this lncRNA might regulate Eomes expression by cis-pattern.

For further study, 506 mRNAs, whose Spearman's correlation with the lncRNA RP11-222K16.2 were >0.4, were analyzed by Gene Set Enrichment Analysis. An enrichment map was constructed using gene sets with *P*-value <.05 (Figure 9D,E).²² The gene sets involved in immune system development, regulation of leukocyte differentiation, regulation of cell-cell adherence, and KEGG pathways in cancer were significantly downregulated in the AML group. The top-scoring gene in the immune system development category was Eomes. This gene was differentially expressed in our microarray data, and the analysis by string software showed that it was the core mRNA in AML (Figure 4B). Eomes is a master regulator of CD8 T-cell function, and a transcription factor that is critical for terminal natural killer (NK) cell differentiation.^{23,24} A recent study noted the NK cells from leukemic mice and humans with AML showed lower level of Eomes, which at least partly led to the blocking of NK cell differentiation, and then it enabled AML to evade mature NK cell surveillance.²⁵ Based on these results, we predicted that, through the lncRNA-mediated dysregulation of Eomes, blocking of NK cell differentiation occurs, promoting the immunized evasion of AML. Additional studies are underway to probe into this additional mechanism of immune escape in cancer.

4 | DISCUSSION

It is well known that the mutation of genes and chromosomes contribute to the pathogenesis of leukemia.²⁵ However, lncRNAs, the rising stars in biology, have just begun to be understood, and the majority of them have not yet been researched. To provide some insights into the biological functions of lncRNAs in the pathogenesis of AML, we undertook a comprehensive analysis of lncRNA and mRNA profiling data from AML and IDA patients, together with data from a public database. We identified the core lncRNAs and their functional annotations, and validated their expression by qRT-PCR. Overall, our work uncovered an interlaced transcripts network that is involved in AML development, in which lncRNAs play an indispensable role.

We explored the expression patterns of transcripts between AML patients and IDA controls' bone marrow. Microarray data identified vast lncRNAs and mRNAs, supporting an extensive involvement of lncRNAs in AML. There were two important concepts that ran throughout our studies to handle the mass data. First, we simplified the complex transcript network by modularization. The GO and KEGG pathway analyses divided mRNAs into several functional modules, which are related to immunity, hematopoiesis, and cancer, indicating the validity of the microarray. STRING and WGCNA were used to construct coexpressed networks from the public and our microarray data, respectively, and then the networks were facilitated into several subnetworks through the *k*-score method. In the same way, lncRNAs were attributed their correlated functional mRNA modules through the coexpressed network and GO annotations. Overall, modularization contributes to simplifying the intricate network into modules, which were like "big genes". Second, the AML dataset from

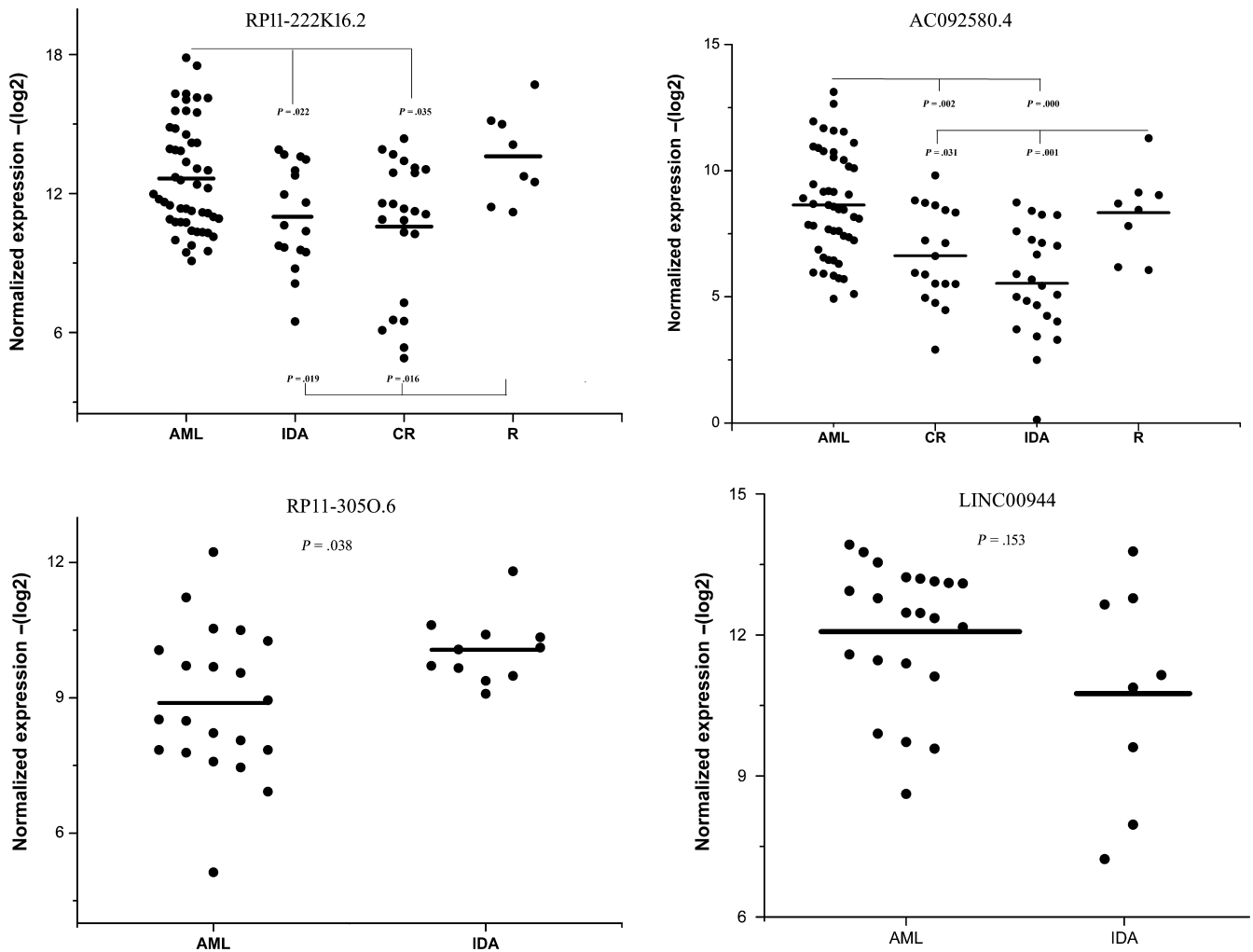


FIGURE 8 Expression of RP11-222K16.2, AC092580.4, RP11-3050.6, and LINC00944 in samples from patients with newly diagnosed acute myeloid leukemia (AML), disease relapse (R), or complete response (CR), and iron deficiency anemia (IDA) controls. RP11-222K16.2 and AC092580.4 were significantly downregulated in newly diagnosed and relapsed AML. RP11-3050.6 was significantly upregulated in newly diagnosed AML. LINC00944 was downregulated in newly diagnosed AML, but lost its significance

TCGA was applied to analyze the clinical significance of lncRNAs from the modules. Eight lncRNAs might have prognostic application in AML. Among them, RP11-222K16.2 gained our attention as it belonged to the most important subnetwork, and its expression is associated with patients' OS. Further analysis of public data showed that the lncRNA may regulate *Eomes* to block NK cell differentiation, leading to the immunized evasion of AML. Finally, three lncRNAs, RP11-222K16.2, AC092580.4, and RP11-3050.6, were confirmed as significantly differentially expressed in AML patients and IDA controls by qRT-PCR.

Our work clearly indicated an important role for lncRNAs in AML. However, many lncRNAs were excluded as they failed to be allocated to functional modules and have not been included in public data. It was difficult to originally understand the functions and targets of these lncRNAs, which may also play a key role in AML. In addition, our analysis showed that lncRNA RP11-222K16.2 was highly correlated with *Eomes*, its neighboring gene. The lncRNA may be directly

or indirectly correlated with *Eomes* and there may be additional transcripts involved in the lncRNA-associated biological process. The lncRNA's biological functions need to be validated further.

The last decades witnessed the discovery of biological functions for non-coding RNA, which triggered the recognition that RNA is not only a simple hinge of the central dogma but also directly takes part in the regulation of biological networks.^{6,26} With the development of next-generation sequencing, especially in terms of depth and scale, significant data has been accumulated. We have to recognize the system is so complex that it is beyond the initial recognition. Fortunately, the progress of methodology simplifies the networks. Through modularization, thousands of transcripts can be facilitated into several "big genes," and then the core lncRNAs of each module can be researched in details. Moreover, the active application of accumulated public data will help us to make the functions of lncRNAs more clear. Hopefully, this study can provide a reference for the broad analysis of HTS data.

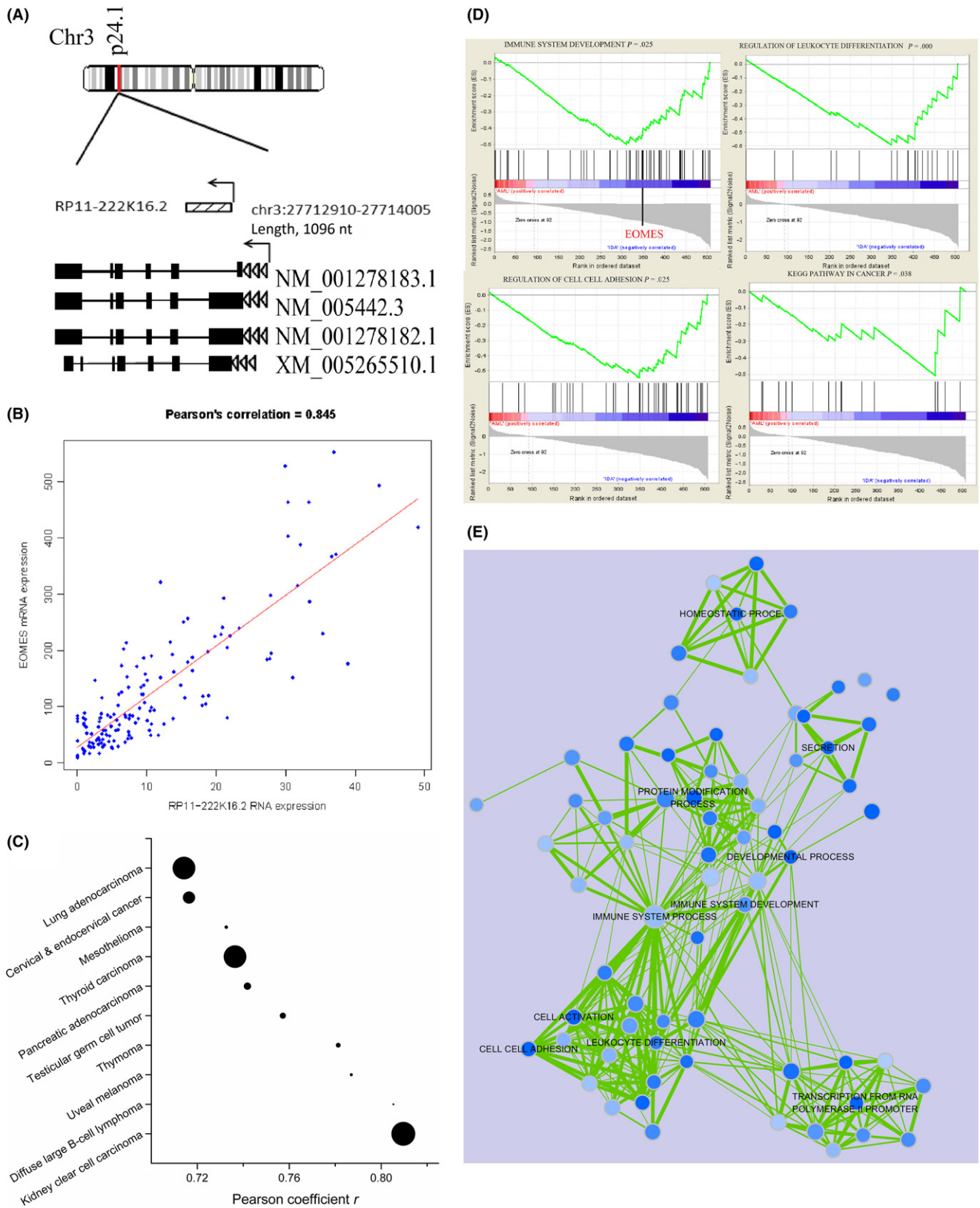


FIGURE 9 Functional analysis of long non-coding RNA (lncRNA) RP11-222K16.2 in acute myeloid leukemia (AML). A, Schematic representation of the composition of the *Eomes* gene and the RP11-222K16.2 loci on human chromosome 3p24.1. B, Pearson's correlation of the lncRNA RP11-222k16.2 with *Eomes* in the AML dataset from The Cancer Genome Atlas (TCGA). C, Pearson's correlation of the lncRNA RP11-222k16.2 with *Eomes* in 10 cancers from TCGA. Data were acquired from the CHIP-BASE database. The size of the points represents the number of patients included in the dataset. D, E, Gene Set Enrichment Analysis (D) and map (E) for all protein-coding genes with Spearman correlations >0.4 with lncRNA RP11-222K16.2

ACKNOWLEDGMENTS

This study was supported by grants from Science and Technology Coordinating Innovative Engineering Projects of Shaanxi Province (Grant No. 2014KTCL03-01).

CONFLICT OF INTEREST

The authors have no conflict of interest.

ORCID

Aili He  <http://orcid.org/0000-0003-2521-3806>

REFERENCES

- Weinberg OK, Sohani AR, Bhargava P, Nardi V. Diagnostic work-up of acute myeloid leukemia. *Am J Hematol*. 2017;92:317-321.
- Cagnetta A, Adamia S, Acharya C, et al. Role of genotype-based approach in the clinical management of adult acute myeloid leukemia with normal cytogenetics. *Leuk Res*. 2014;38:649-659.
- Ilyas AM, Ahmad S, Faheem M, et al. Next generation sequencing of acute myeloid leukemia: influencing prognosis. *BMC Genom*. 2015;16 (Suppl 1):S5.
- ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007;447:799-816.
- Garding A, Bhattacharya N, Claus R, et al. Epigenetic upregulation of lncRNAs at 13q14.3 in leukemia is linked to the lncRNA downregulation of a gene cluster that targets NF- κ B. *PLoS Genet*. 2013;9:e1003373.
- Schmitt AM, Chang HY. Long noncoding RNAs in cancer pathways. *Cancer Cell*. 2016;29:452-463.
- Melo CP, Campos CB, Rodrigues Jde O, et al. Long non-coding RNAs: biomarkers for acute leukaemia subtypes. *Br J Haematol*. 2016;173:318-320.
- Fernando TR, Rodriguez-Malave NI, Waters EV, et al. lncRNA expression discriminates karyotype and predicts survival in B-lymphoblastic leukemia. *Mol Cancer Res*. 2015;13:839-851.
- 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;39:330-336.
- Garzon R, Volinia S, Papaioannou D, et al. Expression and prognostic impact of lncRNAs in acute myeloid leukemia. *Proc Natl Acad Sci USA*. 2014;111:18679-18684.
- Cancer Genome Atlas Research Network, Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013;368:2059.
- Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44-57.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
- Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL. Hierarchical organization of modularity in metabolic networks. *Science*. 2002;297:1551-1555.
- Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43(Database issue):447-452.
- Bonasio R, Shiekhhattar R. Regulation of transcription by long non-coding RNAs. *Annu Rev Genet*. 2014;48:433-455.
- Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*. 2012;81:145-166.
- Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*. 2012;28:573.
- Verhaak RG, Goudswaard CS, van Putten W, et al. Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance. *Blood*. 2005;106:3747-3754.
- Yang JH, Li JH, Jiang S, Zhou H, Qu LH. ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Res*. 2013;41 (Database issue):D177.
- Jia H, Osak M, Bogu GK, Stanton LW, Johnson R, Lipovich L. Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA*. 2010;16:1478-1487.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102:15545-15550.
- Dausy C, Faure F, Mayol K, et al. T-bet and Eomes instruct the development of two distinct natural killer cell lineages in the liver and in the bone marrow. *J Exp Med*. 2014;211:563.
- Pradier A, Simonetta F, Waldvogel S, Bosshard C, Tiercy JM, Roosnek E. Modulation of T-bet and Eomes during Maturation of Peripheral Blood NK Cells Does Not Depend on Licensing/Educating KIR. *Front Immunol*. 2016;7:299. <https://doi.org/10.3389/fimmu.2016.00299>.
- de Jonge HJ, Huls G, de Bont ES. Gene expression profiling in acute myeloid leukaemia. *Neth J Med*. 2011;69:167-176.
- Gibb EA, Brown CJ, Lam WL. The functional role of long non-coding RNA in human carcinomas. *Mol Cancer*. 2011;10:38.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Feng Y, Shen Y, Chen H, et al. Expression profile analysis of long non-coding RNA in acute myeloid leukemia by microarray and bioinformatics. *Cancer Sci*. 2018;109:340-353. <https://doi.org/10.1111/cas.13465>