Research Paper

Identification of hub genes associated with adult acute myeloid leukemia progression through weighted gene co-expression network analysis

Ruiqi Zhu[1,*], Wenyi Lin[1,*], Liang Tang[1], Yu Hu[1]

[1]Institute of Hematology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China
*Equal contribution

Correspondence to: Yu Hu, Liang Tang; **email:** dr_huyu@126.com, https://orcid.org/0000-0002-2815-4568; lancet.tang@qq.com, https://orcid.org/0000-0002-3253-0220
Keywords: AML, hub genes, WGCNA
Received: June 24, 2020 **Accepted:** December 18, 2020 **Published:** February 11, 2021

ABSTRACT

Acute myeloid leukemia (AML) is a malignancy of hematopoietic stem cells. Although many candidate genes such as *CEBPA*, *FLT3*, *IDH1*, and *IDH2* have been associated with AML initiation and prognosis, the molecular mechanisms underlying this disease remain unclear. In this study, we used a systemic co-expression analysis method, namely weighted gene co-expression network analysis (WGCNA), to identify new candidate genes associated with adult AML progression and prognosis. We identified around 5,138 differentially expressed genes (DEGs) between AML samples (from The Cancer Genome Atlas database) and normal control samples (from the Genotype-Tissue Expression database). WGCNA identified nine co-expression modules with significant differences based on the DEGs. Among modules, the turquoise and blue ones were the most relevant to AML (P-value: turquoise 0, blue 4.64E-77). GO term and KEGG pathway analyses revealed that pathways that are commonly dysregulated in AML were all enriched in the blue and turquoise modules. A total of 15 hub genes were identified to be crucial for AML progression. PIVOT analysis revealed non-coding RNAs, transcriptional factors, and drugs associated with the hub genes. Finally, survival analysis revealed that one of the hub genes, *CEACAM5*, was significantly associated with AML prognosis and could serve as a potential target for AML treatment.

INTRODUCTION

Acute myeloid leukemia (AML) is a hematopoietic malignancy with high heterogeneity. The incidence of AML is approximately 4.3 per 100,000 per year in the United States according to the Surveillance, Epidemiology, and End Results database. The mortality of patients with AML is still relatively high, despite the application of advanced therapy methods such as intensive chemotherapy, bone marrow transplantation, and targeted therapy. In the past decades, through novel high-throughput sequencing techniques, multiple somatic mutations have been identified as associated with AML initiation or prognosis. For example, *DNMT3A*, *ASXL1*, *IDH1*, and *IDH2* are frequently mutated in patients with AML. These gene mutations are considered to be acquired early in AML. In contrast, mutations in *FLT3*, *RAS*, and *NPM1* are regarded as secondary events in leukemogenesis. Nevertheless, these findings cannot fully explain why and how AML occurs. Hence, bioinformatic analysis is also used to decipher the molecular mechanisms of leukemogenesis and AML progression, which could provide new therapeutic targets for AML treatment.

Various bio-informatic methods have been used to uncover the mechanisms underlying AML progression or to construct prognosis signatures for patient risk stratification [1–3]. Previously, we had also constructed a 4-microRNA signature to predict the prognosis of pediatric and adolescent AML [4]. Weighted gene co-expression network analysis (WGCNA) is a type of bio-informatic method applied to discover the relationship between genes and phenotypes. It provides a comprehensive method to determine the key regulators, crucial pathways, and potential drug targets associated with AML. In this study, we aimed to identify differentially expressed genes (DEGs) between adult patients with AML and healthy control subjects by using information from The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) database, respectively. Our purpose was to provide a new perspective on genes associated with AML that could be used as potential diagnostic and therapeutic targets.

RESULTS

Differentially expressed genes between AML and normal control samples

After screening and filtration of RNA-seq data in TCGA [5] and GTEx [6] databases, a total of 19,148 genes were obtained, of which 5,166 DEGs were identified between AML and whole-blood normal control samples. Fifty-nine additional genes associated with AML from the NCBI GENE and OMIM [7] databases were also included for further co-expression network construction. Subsequently, a total of 5,225 genes were put forth for the WGCNA.

Co-expression network construction

By carrying out the WGCNA package in R Bioconductor, we constructed a co-expression network for AML. In order to ensure average connectivity and high independence, we screened the power value for the modules, which ranged from 1 to 30. The power value in this study was set at 28 when the scale free R^2 reached 0.9 at this moment of time (Figure 1). Nine modules were identified, and the number of genes in each module was as follows: 41 in the black, 568 in the blue, 134 in the brown, 66 in the green, 3111 in the gray, 26 in the pink, 45 in the red, 1017 in the turquoise, and 130 in the yellow module. Detailed gene information for each module is provided in Supplementary Table 1. The cluster tree is shown in Figure 2A. Among modules, the turquoise and blue ones were the most relevant for AML (P = 4.64E-77 and 0, respectively; Figure 2B). Network heat-map was used to depict the correlation of genes in and among modules. The depth of the red color correlated with the strength of the relationship between the pairs of modules. As illustrated in Figure 2C, genes within the same module strongly correlated with each other, while genes in different modules were almost independent of each other. This indicated that the modules had great scale independence.

A heat-map was drawn for the modules in 115 AML and 755 control samples. As shown in Figure 2D, genes in the blue module were highly expressed in AML samples, while those in the turquoise module were expressed at lower levels compared to the respective levels in the control samples, suggesting that these two modules were strongly associated with the AML phenotype.

Next, we performed gene significance (GS) and module membership (MM) analyses for genes in the blue and turquoise modules to examine whether they were well correlated with the AML phenotype. As shown in Figure 2E, 2F, genes in both these modules showed strong correlations with the AML phenotype, with P-values < 1e-200.

Hub genes were considered as genes with connectivity > 5 in both blue and turquoise modules. The 15 hub genes that were identified as associated with AML progression in the present study were: *BDP1*, *RFX7*, *LARP4*, *TCEGR1*, *MPHOSPH9*, *CCDC18*, *PDS5A*, *FANCL*, and *ICE2* in the turquoise module; and *SERPINB7*, *CEACAM5*, *MUC2*, *RHOV*, *ALDH3A1*, and *CBLC* in the blue module. Visualization of the co-expression network for the turquoise and blue modules are depicted in Figure 3.

GO and KEGG analysis for the turquoise and blue modules

In order to obtain a comprehensive understanding of the biological functions of genes in the turquoise and blue modules, we carried out GO and KEGG pathway enrichment analyses. According to GO analysis, genes in the turquoise module were most enriched in DNA replication, chromosome segregation, microtubule cytoskeleton organization, and nuclear division (Figure 4B). In contrast, genes in the blue module were enriched in neutrophil activation, neutrophil degranulation, and neutrophil mediated immunity (Figure 4A). According to KEGG analysis, genes in the turquoise module were most enriched in pathways associated with herpes simplex virus 1 infection, cell cycle, Fanconi anemia pathway, homologous recombination, and DNA replication (Figure 4D), while those in the blue module were enriched in pathways associated with Salmonella infection, phagosome, tuberculosis, and chemokine signaling (Figure 4C). The results of genes in black and red modules were shown in Supplementary Tables 3, 4.

PIVOT analysis for identifying non-coding RNAs, transcription factors, and drugs associated with AML in the turquoise and blue modules

Transcription factors are a variety of proteins participating in the initiation of transcription. Mutations or functional dysregulation of transcription factors may result in the transformation of hematopoietic precursors into leukemic stem cells. For example, CCAAT/enhancer binding protein alpha (CEBPA) deficiency inhibits the differentiation of myeloid cells both *in vitro* and *in vivo* [8]. A total of 9,395 transcription factor pairs were enrolled in the present study. PIVOT analysis revealed eight transcription factors (CHD8, CTBP1, E2F1, E2F4, E4F1, TP53, TP53BP1, and ZNF143) in the turquoise module ($P < 0.01$) and four (CEBPA, DEDD, IRF8, and SPI1) in the blue module as significantly correlated with AML (Figure 5A, 5B).

Non-coding RNAs are important post-transcriptional regulators. Their dysregulation has been tightly correlated to AML progression in a post-transcriptional level. For example, transcriptional activation of some long non-coding RNAs, such as GAS6-AS2, has been reported to lead to chemotherapy resistance in AML [9]. In the present study, 40 and 25 non-coding RNAs in the turquoise and blue module, respectively, were highly associated with AML (Supplementary Table 2 and Figure 5A, 5B). We further investigated drugs associated with hub genes. After screening the DrugBank database, PIVOT analysis identified nine drugs in the blue module [(2S)-2-(3-bromophenyl)-3-(5-chloro-2-hydroxyphenyl)-1,3-thiazolidin-4-one, artenimol, capecitabine, dasatinib, dextromethorphan, interferon-gamma-1b, morniflumate, N-[2-(2-methyl-1H-indol-3-yl)ethyl]thiophene-2-carboxamide, and sargramostim] and three drugs (caffeine, methionine, and thimerosal) in the turquoise module as significantly associated with AML ($P < 0.01$ for both).

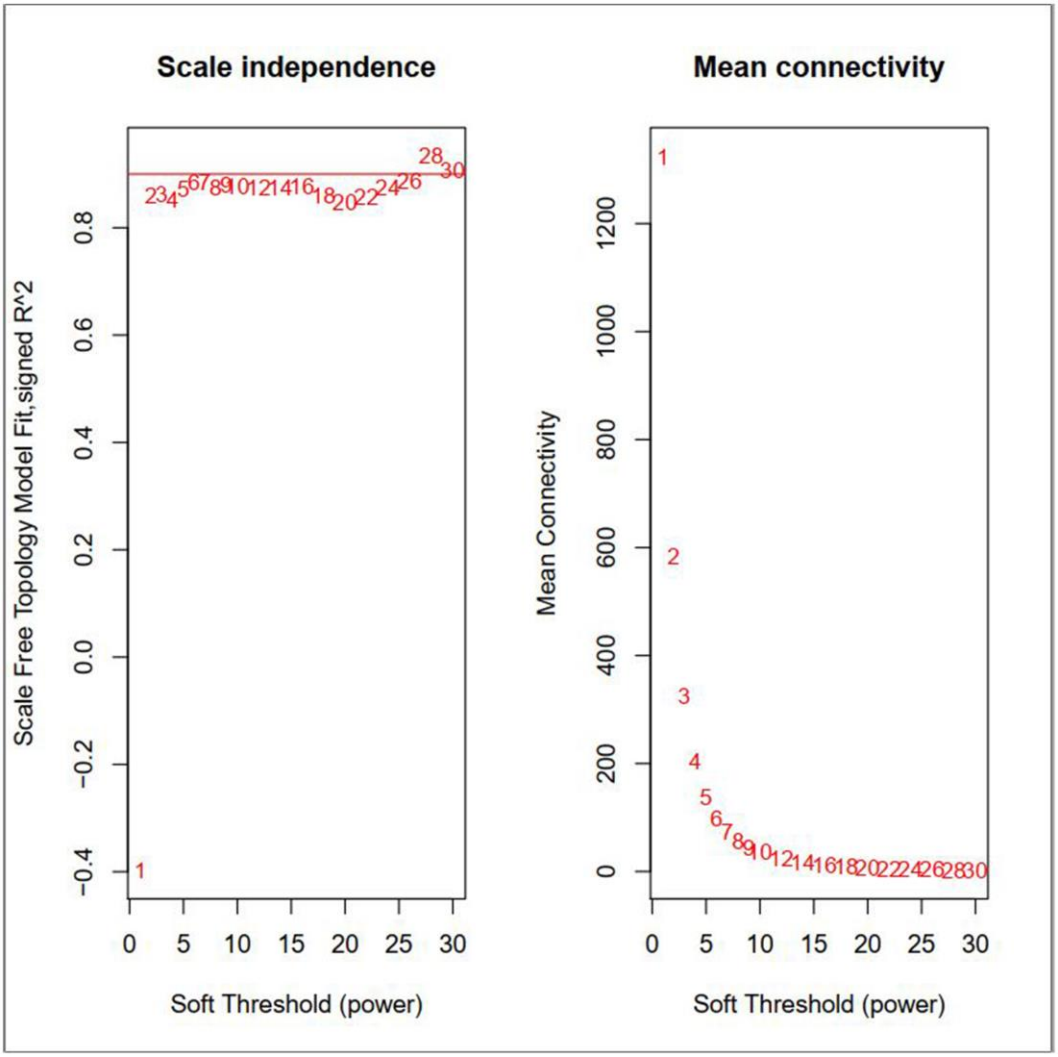

Figure 1. Effects of different soft-threshold values for adult AML co-expression network.

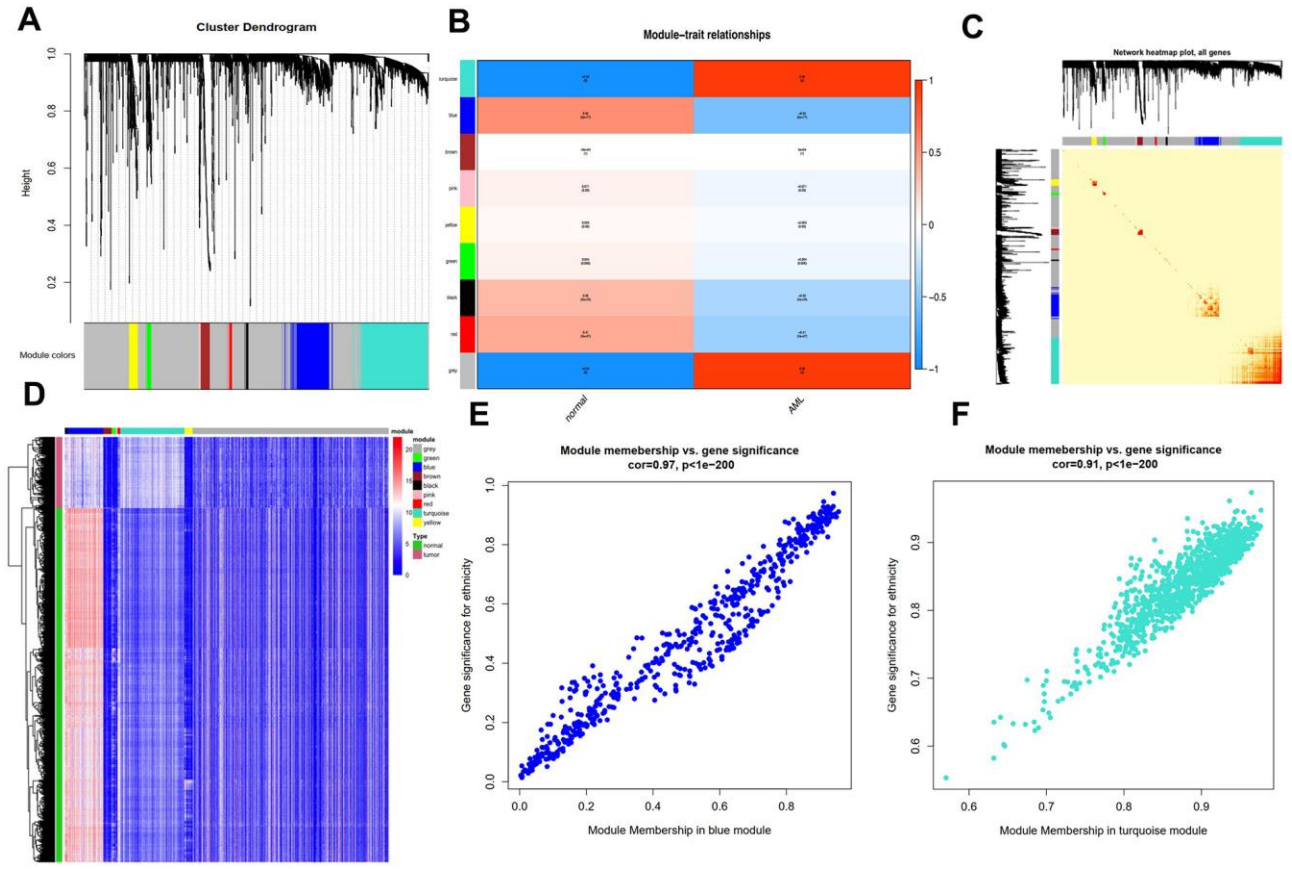

Figure 2. Characterization of adult AML co-expression network. (**A**) Cluster dendrogram and colored display of the network. (**B**) Relationship between modules and phenotypes. Turquoise and blue module are the top two modules related with AML phenotype according to P-value. (**C**) Network heatmap plot for genes in the modules in hierarchical clustering dendrograms. The deeper the red, the more correlated between the genes. (**D**) Gene expression differences between AML samples and normal control in modules. (**E**) Gene Significance (GS) and Module Membership (MM) analyses in turquoise and (**F**) blue module.

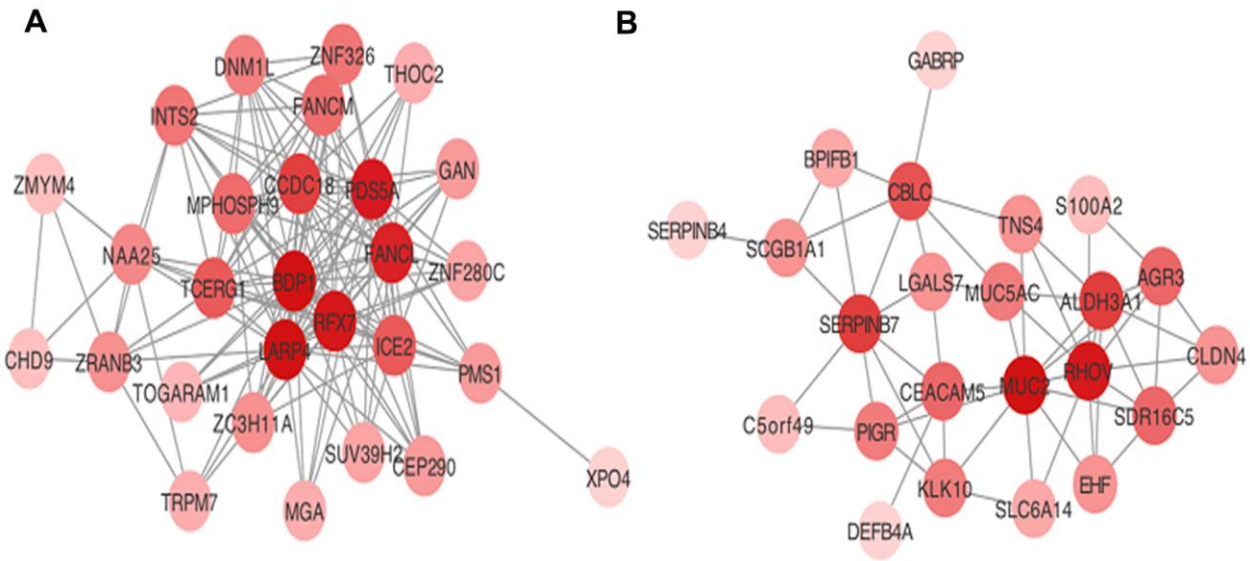

Figure 3. Visualization of hub genes network in (**A**) turquoise and (**B**) blue modules.

Prognostic significance of hub genes in blue and turquoise modules

The prognostic value of the 15 identified hub genes was assessed using OS and expression data in TCGA database. We found that high expression of *CEACAM5* was significantly associated with worse OS for patients with AML (Figure 6). Other hub genes were not statistically significantly associated with the survival of these patients.

DISCUSSION

The molecular mechanisms underlying AML progression and initiation are still not well understood, although advanced progress has been made in the past decades. With the development of high-throughput sequencing technologies, a large amount of genomic data can be acquired from patient samples. In this study, by using published adult AML RNA-seq data from TCGA-AML project and the GTEx database, we constructed a co-expression network by the WGCNA method. A total of nine independent modules were identified, and 15 hub genes were selected from two modules mostly associated with the AML phenotype. Among all hub genes, only *CEACAM5* was significantly associated with the OS of patients with AML, indicating its role as biomarker for AML prognosis and treatment.

WGCNA is a bio-informatic method used to identify clusters of biologically relevant genes associated with a particular disease [10]. It has been widely used to

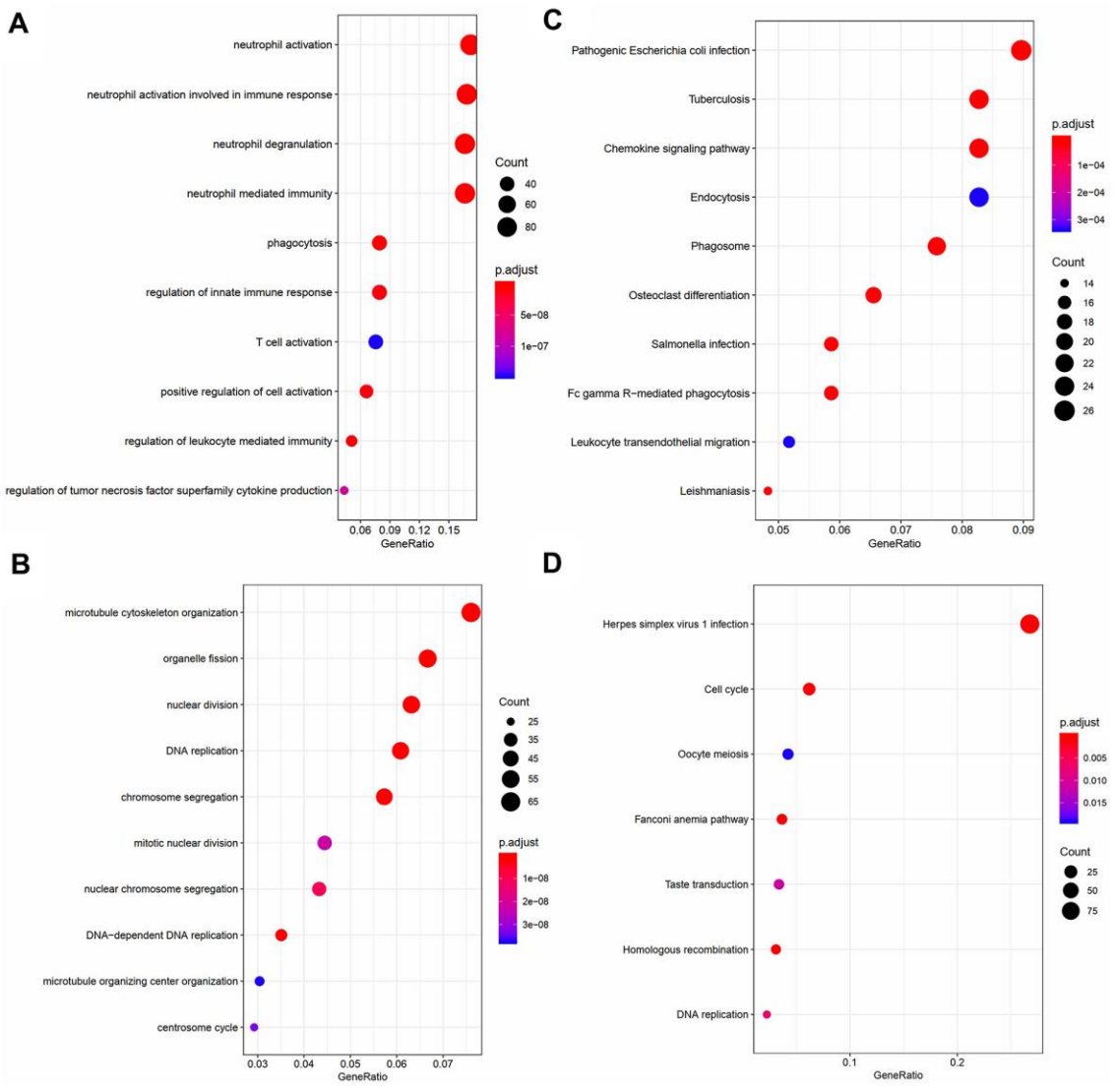

Figure 4. Visualization of part of GO bio-functional analysis results in (**A**) blue and (**C**) turquoise module. Visualization of part of KEGG analysis results in (**B**) blue and (**D**) turquoise module.

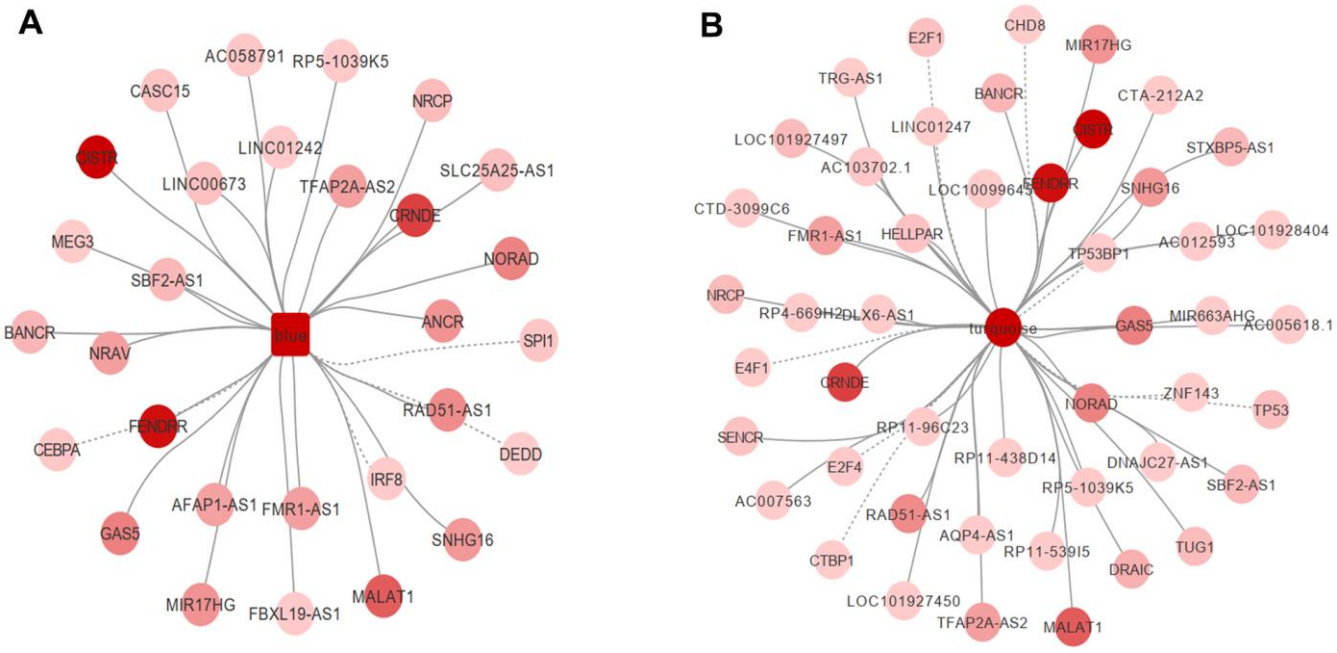

Figure 5. PIVOT analysis revealed TFs and ncRNAs associated with (**A**) blue and (**B**) turquoise module.

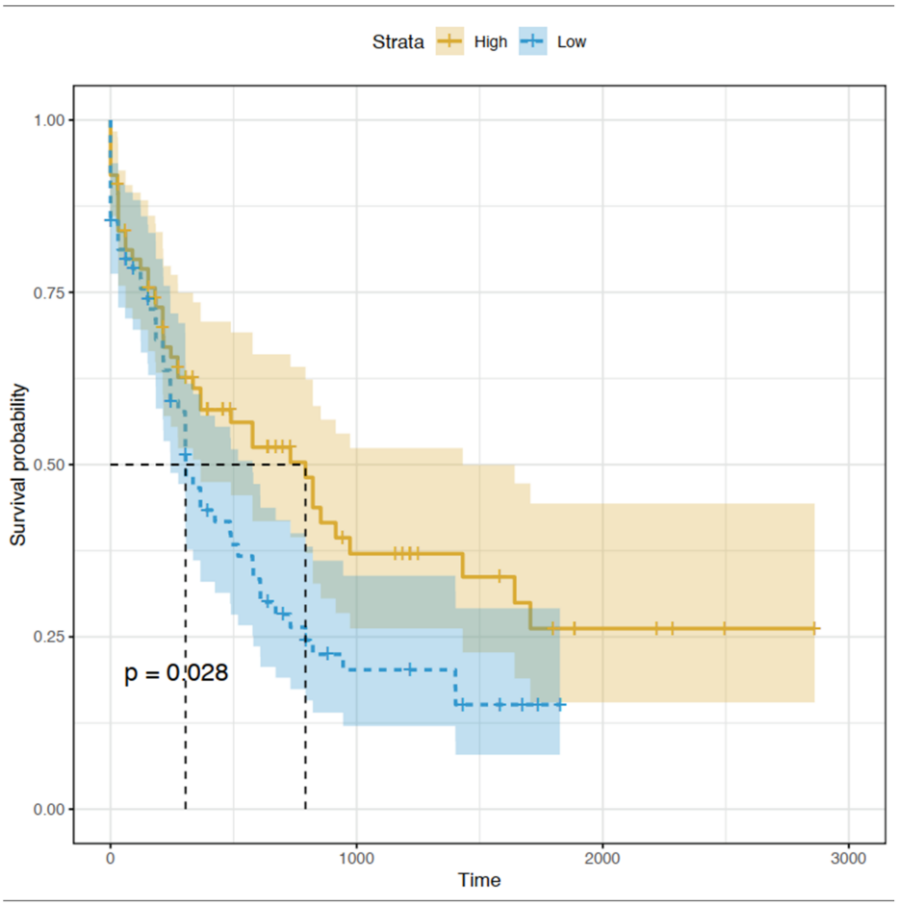

Figure 6. Kaplan-Meier curve for CEACAM5 low and high expression patients in TCGA database.

reveal the molecular mechanisms underlying cancer progression and initiation. Bao et al. reported four hub genes (*FOXC1*, *BCL11A*, *FAM171A1*, and *RGMA*) as positively correlated with the triple-negative breast cancer subtype using the WGCNA method [11]. Chen et al. conducted WGCNA and identified the long non-coding RNA LOC646762 as a biomarker for the prognosis of AML in adult patients [12]. According to WGCNA, biologically relevant genes are classified into the same module. In the present study, the turquoise and blue modules were the ones mostly associated with the AML phenotype; hence, genes in these two modules were used to perform GO and KEGG analyses. For the turquoise module, the identified biological processes, i.e., DNA replication [13], chromosome abnormalities [14], and microtubule cytoskeleton organization [15], have been reported to have a great impact on AML progression and drug resistance. For the blue module, the identified biological processes, i.e., neutrophil activation, neutrophil degranulation [16], and neutrophil mediated immunity [17], are all AML oncogene-related processes. Dysregulation of these processes would lead to cell differentiation disorder and ultimately result in AML. Regarding our results of the KEGG pathway analysis, we found that genes in the turquoise and blue modules were highly enriched in cell cycle, homologous recombination [18], DNA replication [19], chemokine signaling, and leukocyte transendothelial migration pathways, which were highly correlated with the AML progression and initiation.

Regarding the identified hub genes, the functions of *BDP1*, *RFX7*, *LARP4*, *TCEGR1*, *MPHOSPH9*, *CCDC18*, *ICE2*, *SERPINB7*, *RHOV*, and *CBLC* are understudied in leukemia. It was reported that the deletion of *RFX7*, encoding for a transcriptional factor, decreases natural killer (NK) cell maintenance and immunity [20]. The dysfunction of NK cells is common in human tumors, and activation of these cells has become a promising strategy to prevent relapse and induce remission when treating AML [21]. *MPHOSPH9* was reported as a susceptible locus for multiple sclerosis [22]. *CDCC18* was reported as a susceptibility gene for familial colorectal cancer. *PDS5A* is a cell cycle-related gene and precocious dissociation of PDS5A is a translocation partner of MLL in AML [23]. FANCL is a family member of DNA repair molecules and is frequently mutated in myelodysplastic syndrome (MDS), a pre-malignant hematopoietic disease; patients with MDS have increased risk of AML [24]. Moreover, hypermethylation of the *FANCL* promoter region was also suggested to be associated with sporadic acute leukemia [25]. *MUC2* belongs to the mucin family, is located on 11p15.5, and is associated with childhood AML [26]. *CEACAM5* has been widely studied in various kinds of cancers, including breast cancer [27], colorectal cancer [28, 29], pancreatic cancer [30], gastric cancer [31], and carcinoma of the tongue [32]. It is an adhesion molecule and its aberrant expression is always associated with tumor metastasis and poor prognosis. However, its function in AML has not been studied yet. In the present study, we reveal that CEACAM5 is not only one of the hub genes in AML but is also associated with unfavorable prognosis. ALDH3A1 was reported to be important in metabolizing reactive aldehydes and reactive oxygen species in hematopoietic stem cells. Its loss was associated with drug resistance and poor prognosis in AML [33].

Further, we investigated transcription factors and non-coding RNAs associated with AML. CEBPA mutations [34, 35] and GAS6-AS2 dysregulation [9] were previously reported to be strongly associated with AML prognosis and chemotherapy resistance. These results indicate that the hub genes identified in the present study are highly associated with the AML phenotype and could be used as potential therapeutic targets and biomarkers for AML.

MATERIALS AND METHODS

TCGA and GTEx datasets

RNA-seq data of 151 samples from adult patients with AML were downloaded from TCGA (https://portal.gdc.cancer.gov/) database. The age of the patients ranged from 18 to 88 years, and 45.0% of the patients were females. Control data from whole-blood samples of healthy subjects were obtained from GTEx (https://www.gtexportal.org/home/) database. A total of 755 control samples were included in the present study. For each sample, probe data less than 25% were excluded from the study. After annotating a probe name to each gene symbol according to the annotation files, RNA-seq data from different sources were unified by realigning raw reads, removing degraded samples, and performing batch effect correction to correct non-biological variation. Batch effect correction was performed using the SVAseq R package. AML-related genes in the NCBI GENE and OMIM databases were also downloaded for co-expression network construction.

Identification of DEGs and co-expression network construction

The DEGs between AML and control samples were identified using the edgeR package in R Bioconductor. Genes with log fold change $|\log FC| > 2$ and P-value < 0.05 were considered as DEGs. The co-expression network was subsequently constructed based on the DEGs using the WGCNA package. The normalized count was used when WGCNA was conducted and was standardized

by the TMM method of the R package edgeR. The parameters used for WGCNA were as follows: minModuleSize = 25, mergeCutHeight = 0.25, corType = "pearson". To identify which co-expression module had the highest relevance to the clinical phenotype, we applied the module-trait association method. Genes were clustered, and a heat-map was drawn to illustrate the association between modules and phenotype. All analyses were conducted using R Bioconductor.

Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) enrichment analysis for co-expression modules

To get a comprehensive understanding of the function of genes associated with AML, a functional annotation was carried out for the two most relevant modules using Database for Annotation, Visualization, and Integrated Discovery (DAVID) (https://david.ncifcrf.gov/). GO and KEGG analyses were accordingly applied. Results with P-value < 0.05 were considered as significant terms and pathways, respectively.

PIVOT analysis for identifying transcription factors, non-coding RNAs, and drugs associated with modules

To investigate the gene transcription and post-transcriptional regulations in the co-expression modules, we carried out PIVOT analysis. Non-coding RNA-gene interactions and transcription factor target data were downloaded from RAID and TRRUST databases, respectively. P-value < 0.01 indicated a significant interaction between the PIVOT regulator and the module. R Bioconductor was used for predicting the target non-coding RNAs/transcriptions factors associated with the modules. Module-related drugs were also screened by the same method. Drugs associated with AML were extracted from the DRUGBANK Database.

Survival analysis

By considering patients with available gene expression data and clinical data, a total of 132 samples were enrolled in the survival analysis. Hub genes identified from the two most relevant modules were selected for the survival analysis. Survival data including the overall survival (OS) time and the living status were downloaded from TCGA database. Survival analysis was performed using the Kaplan-Meier method using R Bioconductor. Adjusted P values <0.05 were considered statistically significant.

AUTHOR CONTRIBUTIONS

RZ contributed to the conception of the study and performed the data analysis. WL contributed significantly to the data analysis. RZ and WL prepared the manuscript. YH and LT supervised the study, revised the manuscript and contributed to the conception of the study.

CONFLICTS OF INTEREST

The authors declare there are no conflicts of interest.

FUNDING

Yu Hu: National Natural Science Fund, No. 31620103909.
Liang Tang: Chen-Guang Project of Wuhan Science and Technology Bureau, 2017050304010276.

REFERENCES

1 Duployez N, Marceau-Renaut A, Villenet C, Petit A, Rousseau A, Ng SW, Paquet A, Gonzales F, Barthélémy A, Leprêtre F, Pottier N, Nelken B, Michel G, et al. The stem cell-associated gene expression signature allows risk stratification in pediatric acute myeloid leukemia. Leukemia. 2019; 33:348–57. https://doi.org/10.1038/s41375-018-0227-5 PMID:30089916

2 Beck D, Thoms JA, Palu C, Herold T, Shah A, Olivier J, Boelen L, Huang Y, Chacon D, Brown A, Babic M, Hahn C, Perugini M, et al. A four-gene LincRNA expression signature predicts risk in multiple cohorts of acute myeloid leukemia patients. Leukemia. 2018; 32:263–72. https://doi.org/10.1038/leu.2017.210 PMID:28674423

3 Bullinger L, Döhner K, Kranz R, Stirner C, Fröhling S, Scholl C, Kim YH, Schlenk RF, Tibshirani R, Döhner H, Pollack JR. An FLT3 gene-expression signature predicts clinical outcome in normal karyotype AML. Blood. 2008; 111:4490–95. https://doi.org/10.1182/blood-2007-09-115055 PMID:18309032

4 Zhu R, Lin W, Zhao W, Fan F, Tang L, Hu Y. A 4-microRNA signature for survival prognosis in pediatric and adolescent acute myeloid leukemia. J Cell Biochem. 2019; 120:3958–68. https://doi.org/10.1002/jcb.27679 PMID:30242879

5 Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn). 2015; 19:A68–77. https://doi.org/10.5114/wo.2014.47136 PMID:25691825

6 Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B, Moser M, Karasik E, et al, and GTEx Consortium. The genotype-tissue expression (GTEx) project. Nat Genet. 2013; 45:580–85.

https://doi.org/10.1038/ng.2653
PMID:23715323

7. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 2005; 33:D514–7. https://doi.org/10.1093/nar/gki033 PMID:15608251

8 Pabst T, Mueller BU. Transcriptional dysregulation during myeloid transformation in AML. Oncogene. 2007; 26:6829–37. https://doi.org/10.1038/sj.onc.1210765 PMID:17934489

9 Bester AC, Lee JD, Chavez A, Lee YR, Nachmani D, Vora S, Victor J, Sauvageau M, Monteleone E, Rinn JL, Provero P, Church GM, Clohessy JG, Pandolfi PP. An integrated genome-wide CRISPRa approach to functionalize lncRNAs in drug resistance. Cell. 2018; 173:649–64.e20. https://doi.org/10.1016/j.cell.2018.03.052 PMID:29677511

10 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559. https://doi.org/10.1186/1471-2105-9-559 PMID:19114008

11 Bao C, Lu Y, Chen J, Chen D, Lou W, Ding B, Xu L, Fan W. Exploring specific prognostic biomarkers in triple-negative breast cancer. Cell Death Dis. 2019; 10:807. https://doi.org/10.1038/s41419-019-2043-x PMID:31649243

12 Chen CT, Wang PP, Mo WJ, Zhang YP, Zhou W, Deng TF, Zhou M, Chen XW, Wang SQ, Wang CX. Expression profile analysis of prognostic long non-coding RNA in adult acute myeloid leukemia by weighted gene co-expression network analysis (WGCNA). J Cancer. 2019; 10:4707–18. https://doi.org/10.7150/jca.31234 PMID:31528236

13 Ziv O, Zeisel A, Mirlas-Neisberg N, Swain U, Nevo R, Ben-Chetrit N, Martelli MP, Rossi R, Schiesser S, Canman CE, Carell T, Geacintov NE, Falini B, et al. Identification of novel DNA-damage tolerance genes reveals regulation of translesion DNA synthesis by nucleophosmin. Nat Commun. 2014; 5:5437. https://doi.org/10.1038/ncomms6437 PMID:25421715

14 Sashida G, Ito Y, Nakajima A, Kawakubo K, Kuriyama Y, Yagasaki F, Bessho M, Ohyashiki K. Multiple myeloma with monosomy 13 developed in trisomy 13 acute myelocytic leukemia: numerical chromosome abnormality during chromosomal segregation process. Cancer Genet Cytogenet. 2003; 141:154–56. https://doi.org/10.1016/s0165-4608(02)00792-6 PMID:12606134

15 Pera B, Calvo-Vidal MN, Ambati S, Jordi M, Kahn A, Díaz JF, Fang W, Altmann KH, Cerchietti L, Moore MA. High affinity and covalent-binding microtubule stabilizing agents show activity in chemotherapy-resistant acute myeloid leukemia cells. Cancer Lett. 2015; 368:97–104. https://doi.org/10.1016/j.canlet.2015.07.038 PMID:26277539

16 Zheng J, Song Y, Li Z, Tang A, Fei Y, He W. The implication of lncRNA expression pattern and potential function of lncRNA RP4-576H24.2 in acute myeloid leukemia. Cancer Med. 2019; 8:7143–60. https://doi.org/10.1002/cam4.2518 PMID:31568697

17 Yang TH, St John LS, Garber HR, Kerros C, Ruisaard KE, Clise-Dwyer K, Alatrash G, Ma Q, Molldrem JJ. Membrane-associated proteinase 3 on granulocytes and acute myeloid leukemia inhibits T cell proliferation. J Immunol. 2018; 201:1389–99. https://doi.org/10.4049/jimmunol.1800324 PMID:30021768

18 Gaymes TJ, Mohamedali A, Eiliazadeh AL, Darling D, Mufti GJ. FLT3 and JAK2 mutations in acute myeloid leukemia promote interchromosomal homologous recombination and the potential for copy neutral loss of heterozygosity. Cancer Res. 2017; 77:1697–708. https://doi.org/10.1158/0008-5472.CAN-16-1678 PMID:28108507

19 David L, Fernandez-Vidal A, Bertoli S, Grgurevic S, Lepage B, Deshaies D, Prade N, Cartel M, Larrue C, Sarry JE, Delabesse E, Cazaux C, Didier C, et al. CHK1 as a therapeutic target to bypass chemoresistance in AML. Sci Signal. 2016; 9:ra90. https://doi.org/10.1126/scisignal.aac9704 PMID:27625304

20 Castro W, Chelbi ST, Niogret C, Ramon-Barros C, Welten SP, Osterheld K, Wang H, Rota G, Morgado L, Vivier E, Raeber ME, Boyman O, Delorenzi M, et al. The transcription factor Rfx7 limits metabolism of NK cells and promotes their maintenance and immunity. Nat Immunol. 2018; 19:809–20. https://doi.org/10.1038/s41590-018-0144-9 PMID:29967452

21 Cooley S, Parham P, Miller JS. Strategies to activate NK cells to prevent relapse and induce remission following hematopoietic stem cell transplantation. Blood. 2018; 131:1053–62. https://doi.org/10.1182/blood-2017-08-752170 PMID:29358179

22. International Multiple Sclerosis Genetics Conssortium (IMSGC). IL12A, MPHOSPH9/CDK2AP1 and RGS1 are

novel multiple sclerosis susceptibility loci. Genes Immun. 2010; 11:397–405. https://doi.org/10.1038/gene.2010.28 PMID:20555355

23 Put N, Van Roosbroeck K, Vande Broek I, Michaux L, Vandenberghe P. PDS5A, a novel translocation partner of MLL in acute myeloid leukemia. Leuk Res. 2012; 36:e87–89. https://doi.org/10.1016/j.leukres.2011.12.006 PMID:22230297

24 Ganguly BB, Kadam NN. Mutations of myelodysplastic syndromes (MDS): an update. Mutat Res Rev Mutat Res. 2016; 769:47–62. https://doi.org/10.1016/j.mrrev.2016.04.009 PMID:27543316

25 Hess CJ, Ameziane N, Schuurhuis GJ, Errami A, Denkers F, Kaspers GJ, Cloos J, Joenje H, Reinhardt D, Ossenkoppele GJ, Zwaan CM, Waisfisz Q. Hypermethylation of the FANCC and FANCL promoter regions in sporadic acute leukaemia. Cell Oncol. 2008; 30:299–306. https://doi.org/10.3233/clo-2008-0426 PMID:18607065

26 Paris MJ, Williams BR. Characterization of a 500-kb contig spanning the region between c-Ha-Ras and MUC2 on chromosome 11p15.5. Genomics. 2000; 69:196–202. https://doi.org/10.1006/geno.2000.6339 PMID:11031102

27 Wang XM, Zhang Z, Pan LH, Cao XC, Xiao C. KRT19 and CEACAM5 mRNA-marked circulated tumor cells indicate unfavorable prognosis of breast cancer patients. Breast Cancer Res Treat. 2019; 174:375–85. https://doi.org/10.1007/s10549-018-05069-9 PMID:30535933

28 Li Q, Li Y, Li J, Ma Y, Dai W, Mo S, Xu Y, Li X, Cai S. FBW7 suppresses metastasis of colorectal cancer by inhibiting HIF1α/CEACAM5 functional axis. Int J Biol Sci. 2018; 14:726–35. https://doi.org/10.7150/ijbs.24505 PMID:29910683

29 Saeland E, Belo AI, Mongera S, van Die I, Meijer GA, van Kooyk Y. Differential glycosylation of MUC1 and CEACAM5 between normal mucosa and tumour tissue of colon cancer patients. Int J Cancer. 2012; 131:117–28. https://doi.org/10.1002/ijc.26354 PMID:21823122

30 Govindan SV, Cardillo TM, Moon SJ, Hansen HJ, Goldenberg DM. CEACAM5-targeted therapy of human colonic and pancreatic cancer xenografts with potent labetuzumab-SN-38 immunoconjugates. Clin Cancer Res. 2009; 15:6052–61. https://doi.org/10.1158/1078-0432.CCR-09-0586 PMID:19789330

31 Blumenthal RD, Hansen HJ, Goldenberg DM. Inhibition of adhesion, invasion, and metastasis by antibodies targeting CEACAM6 (NCA-90) and CEACAM5 (Carcinoembryonic Antigen). Cancer Res. 2005; 65:8809–17. https://doi.org/10.1158/0008-5472.CAN-05-0420 PMID:16204051

32 Zhang F, Jin Liu X, Qu X, Sheng Hu Z, Yang YM, Ma L, Liu P, Shi P, Cai Wei F. Osteopontin's colocalization with the adhesion molecule CEACAM5 in cytoplasm of carcinoma of tongue and its correlation with the invasion of that diease. Cancer Cell Int. 2012; 12:33. https://doi.org/10.1186/1475-2867-12-33 PMID:22738781

33 Gasparetto M, Smith CA. ALDHs in normal and malignant hematopoietic cells: potential new avenues for treatment of AML and other blood cancers. Chem Biol Interact. 2017; 276:46–51. https://doi.org/10.1016/j.cbi.2017.06.020 PMID:28645468

34 Green CL, Koo KK, Hills RK, Burnett AK, Linch DC, Gale RE. Prognostic significance of CEBPA mutations in a large cohort of younger adult patients with acute myeloid leukemia: impact of double CEBPA mutations and the interaction with FLT3 and NPM1 mutations. J Clin Oncol. 2010; 28:2739–47. https://doi.org/10.1200/JCO.2009.26.2501 PMID:20439648

35. Pabst T, Mueller BU, Zhang P, Radomska HS, Narravula S, Schnittger S, Behre G, Hiddemann W, Tenen DG. Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. Nat Genet. 2001; 27:263–70. https://doi.org/10.1038/85820 PMID:11242107

SUPPLEMENTARY MATERIALS

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1, 3, 4.

Supplementary Table 1. Detailed gene information for each module.

Supplementary Table 2. TFs and ncRNAs related to blue and turquoise modules.

module	ncRNA	connection	pvalue
blue	AC058791	56	0.000861
blue	AFAP1-AS1	590	5.31E-06
blue	ANCR	795	1.31E-07
blue	BANCR	311	0.003088
blue	CASC15	120	0.000202
blue	CISTR	2883	2.80E-18
blue	CRNDE	1984	6.75E-14
blue	FBXL19-AS1	30	0.007103
blue	FENDRR	2672	1.39E-28
blue	FMR1-AS1	619	3.66E-06
blue	GAS5	1067	0.000113
blue	LINC00673	157	0.005009
blue	LINC01242	24	0.003754
blue	MALAT1	1569	1.04E-10
blue	MEG3	12	0.000459
blue	MIR17HG	800	4.69E-07
blue	NORAD	1028	0.001439
blue	NRAV	622	1.91E-09
blue	NRCP	221	4.35E-05
blue	RAD51-AS1	905	7.89E-08
blue	RP5-1039K5	27	0.005269
blue	SBF2-AS1	275	0.000336
blue	SLC25A25-AS1	169	0.00741
blue	SNHG16	723	0.003953
blue	TFAP2A-AS2	620	0.00071
turquoise	AC005618.1	4	0.004255
turquoise	AC007563	5	0.006965
turquoise	AC012593	4	0.004255
turquoise	AQP4-AS1	5	0.006965
turquoise	BANCR	311	0.000969
turquoise	CISTR	2883	3.52E-14
turquoise	CRNDE	1984	6.87E-25
turquoise	CTA-212A2	30	0.001183
turquoise	CTD-3099C6	4	0.004255
turquoise	DLX6-AS1	14	0.005798
turquoise	DNAJC27-AS1	4	0.004255
turquoise	DRAIC	386	5.07E-11
turquoise	FENDRR	2672	2.57E-18
turquoise	FMR1-AS1	619	5.27E-16
turquoise	GAS5	1067	5.20E-09

turquoise	HELLPAR	138	0.000106
turquoise	LINC01247	3	0.002167
turquoise	LOC100996455	3	0.002167
turquoise	LOC101927450	3	0.002167
turquoise	LOC101927497	237	1.12E-13
turquoise	LOC101928404	3	0.002167
turquoise	MALAT1	1569	3.33E-35
turquoise	MIR17HG	800	4.93E-05
turquoise	MIR663AHG	15	0.007103
turquoise	NORAD	1028	1.83E-05
turquoise	NRCP	221	0.000128
turquoise	RAD51-AS1	905	2.39E-12
turquoise	AC103702.1	65	0.001888
turquoise	RP11-438D14	2	0.000736
turquoise	RP11-539I5	2	0.000736
turquoise	RP11-96C23	3	0.002167
turquoise	RP4-669H2	3	0.002167
turquoise	RP5-1039K5	27	0.000717
turquoise	SBF2-AS1	275	2.84E-09
turquoise	SENCR	123	0.006505
turquoise	SNHG16	723	4.00E-11
turquoise	STXBP5-AS1	284	1.27E-05
turquoise	TFAP2A-AS2	620	1.50E-23
turquoise	TRG-AS1	2	0.000736
turquoise	TUG1	228	8.76E-22

Supplementary Table 3. The results of GO and KEGG biofunctional analysis in black module.

Supplementary Table 4. The results of GO and KEGG biofunctional analysis in red module.