

# Identification of acute myeloid leukemia-driven genes by multi-dimensional approach

**Jimo Jian**

Department of Hematology, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Wenhua West Road, Jinan, Shandong, 250012 <https://orcid.org/0000-0002-8471-0039>

**Hongyuan Hao**

Department of Hematology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, 758 Hefei Road, Qingdao, Shandong, 266035

**Qian Liu**

Department of Pain, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Wenhua West Road, Jinan, Shandong, 250012

**Wenhao Dong**

Department of Hematology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, 758 Hefei Road, Qingdao, Shandong, 266035

**Lu Li**

Department of Hematology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, 758 Hefei Road, Qingdao, Shandong, 266035

**Yalu Liu**

Department of Hematology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, 758 Hefei Road, Qingdao, Shandong, 266035

**Gege Feng**

Department of Hematology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, 758 Hefei Road, Qingdao, Shandong, 266035

**Juan Wang**

Department of Hematology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, 758 Hefei Road, Qingdao, Shandong, 266035

**Sheng Gao**

Department of Hematology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, 758 Hefei Road, Qingdao, Shandong, 266035

**Lai Chen**

Integrated Chinese & Western Medicine Oncology Research Center, Jiangxi University of Traditional Chinese Medicine, Nanchang, Jiangxi 330004

**Fei Lu** (✉ [lufeisdu2@163.com](mailto:lufeisdu2@163.com))

Department of Hematology, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Wenhua West Road, Jinan, Shandong, 250012 <https://orcid.org/0000-0002-3656-6482>

## Chenglu Yuan

Department of Hematology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University,  
758 Hefei Road, Qingdao, Shandong, 266035


---

## Research

**Keywords:** Acute myeloid leukemia(AML), WGCNA, TCGA, biomark

**Posted Date:** May 28th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-29046/v1>

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

---

---

# Identification of Acute myeloid leukemia-driven genes by multi-dimensional approach

Jimo Jian, Hongyuan Hao, Qian Liu, Wenhao Dong, Lu Li, Yalu Liu, Gege Feng, Juan Wang, Sheng Gao, Lai Chen, Fei Lu\*, Chenglu Yuan \*

Department of Hematology, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Wenhua West Road, Jinan, Shandong, 250012, China

Jimo Jian and Fei Lu

Department of Hematology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, 758 Hefei Road, Qingdao, Shandong, 266035, China

Hongyuan Hao, Wenhao Dong, Lu Li, Yalu Liu, Gege Feng, Juan Wang and Sheng Gao

Department of Pain, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Wenhua West Road, Jinan, Shandong, 250012, China

Qian Liu

Integrated Chinese & Western Medicine Oncology Research Center, Jiangxi University of Traditional Chinese Medicine, Nanchang, Jiangxi 330004, China

Lai Chen

Correspondence to Chenglu Yuan [18660229187@163.com](mailto:18660229187@163.com) or Fei Lu [lufeisdu2@163.com](mailto:lufeisdu2@163.com).

## ABSTRACT

### Background

Acute myeloid leukemia (AML), which is characterized by the uncontrolled proliferation of myeloid leukemia cells in the bone marrow and other hematopoietic tissues, and is highly heterogeneous. While with the progress of sequencing technology, understanding of the AML-related biomarkers are still incomplete. The purpose of this study is to identify potential biomarkers for diagnosis of AML.

### Methods

Based on WGCNA analysis of gene mutation expression, methylation level distribution, mRNA expression and AML-related genes in public databases, were employed for investigating potential biomarkers for prognosis of AML.

### Results

This study screened a total of 6,383 genes by analyzing various changes in 103 acute myeloid leukemia (AML) samples, including gene mutation expression, methylation level distribution, mRNA expression, and AML-related genes in public databases. Moreover, seven AML-related co-expression modules were mined by WGCNA analysis and twelve biomarkers associated with the AML prognosis were identified from each top 10 genes of the seven co-expression modules.

---

The AML samples were then classified into two subgroups, the prognosis of which is significantly different, based on the expression of these twelve genes. The differentially expressed genes are mainly involved in glucose metabolism, glutathione biosynthesis, small G protein-mediated signal transduction, and the Rap1 signaling pathway.

## **Conclusions**

With the utilization of WGCNA mining, seven gene co-expression modules were identified from TCGA database and there are unreported genes that may as potential driven genes of AML and may be the direction to identify the possible molecular signatures to predict survival of AML patients and help guide experiments for potential clinical drug targets.

## **Keywords**

Acute myeloid leukemia(AML), WGCNA, TCGA, biomark

## **1. BACKGROUND**

Leukemia is a malignant clonal hematopoietic stem cell disease, characterized by the uncontrolled proliferation of leukemia cells that accumulate in the bone marrow and other hematopoietic tissues, thereby inhibiting the normal hematopoietic function and infiltrating tissues such as lymph nodes, liver and spleen. [1]. Based on the pathogenic cellular origin, the acute leukemia can be divided into acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML). AML is the most common adult type, it is characterized by uncontrolled proliferation of primary and immature myeloid cells in bone marrow and peripheral blood, and is highly heterogeneous.

Hematopoietic transformation leads to changes in many key genes during the proliferation and differentiation of bone marrow [2]. Important regulatory gene changes may be associated with the pathogenesis of AML, which are potential target for AML treatment [3]. At present, the French-American-British classification is still widely used clinically, which classifies AML into eight subgroups (M0-M7) according to the degree of differentiation and morphological characteristics of myeloid progenitor cells [4, 5]. AML is characterized by genetic heterogeneity, with different clinical behaviors observed even in the same subgroup [6]. With the progress, the complete response (CR) rate and disease-free survival (DFS) rate have slightly improved in AML patients, but the prognosis remains poor, especially in refractory or recurrent AML[7]. The 5-year survival rate of total AML patients is 21.4%, with median survival time of 5-10 months, however, it remarkably reduces to 1% if the patients older than 75 years. The CR rate after initial standard induction therapy is about 60-70% in AML patients younger than 60 years. About 30-40% of patients eventually relapse, and their average overall survival (OS) is approximately 3.8 months[8]. The recurrence rate is significantly increased in AML patients with abnormal chromosomes or genes, and the prognosis for those patients is particularly poor. Moreover, chemotherapy resistance and minimal residual disease after remission also affect the prognosis of AML patients[9].

Currently, AML treatments include chemotherapy, hematopoietic cell transplantation and target therapy, the induction therapy is mainly based on cytarabine combined with anthracyclines,

---

and the consolidation treatment is mainly based on high-dose cytosine combined with hematopoietic cell transplantation. Although extensive studies of AML in the past forty years, the outcome of patients has not improved yet[10]. In recent years, with the progress of sequencing technology, we have a deeper understanding of the gene mutations in AML, and have found a variety of AML-related genes[11, 12]. It has been reported that 20% and 30% patients have Fms-like tyrosine kinase-1(FLT3) and isocitrate dehydrogenase 1/2( IDH1/2) mutations, respectively[13, 14].

The first-generation FLT3 inhibitors include sunitinib (SU11248), midostaurin(PKC412), lestaurtin(CEP701) and tandutinib. In a phase I trial, the old AML patients with FLT3 mutation have achieved high CR rate by sunitinib treatment[15], another four AML patients with FLT3/ITD mutation have achieved pathologic CR or pathologic partial response (PR) by sunitinib monotherapy[16]. However, the first-generation of FLT3 inhibitors is not perfect, such as high recurrence rate, high adverse reactions and poor tolerance, which have limited their application [17-19]. The second-generation FLT3 inhibitors include quizartinib, sorafenib and midostaurin, although the efficacy is significantly improved, the high drug resistance cannot be ignored. Moreover, a study showed that the AML patients had 50% probability of recurrence with 3 months after remission. Therefore, the efficacy of FLT3 inhibitor monotherapy in AML patients with FLT3/ITD mutation is not satisfactory, as well as in recurrent or refractory AML(R/R AML) patients[20]. IDH1/2 mutation sites include IDH1-R132, IDH2-R140 and IDH2-R172[21]. However, current clinical trials suggested unsatisfied effect of IDH1/2 inhibitors in AML. The outcomes of phase I/II clinical trial showed 41.6% and 40.3% objective remission rate (ORR) in relapsed or refractory AML patients with IDH1/2 mutations by Ivosidenib(AG-120) and Enasidenib(AG221), respectively[22, 23]. Among them, 20.2% patients achieved CR, the median duration of remission was 3.7 months, the adverse reactions included hyperbilirubinemia (8%) and IDH-inhibitor-associated differentiation syndrome (7%).

The above studies have suggested that small molecule inhibitors had great therapeutic potential in treatment of AML, it is characterized by strong specificity, low toxicity and good tolerance. However, so far, the studies of AML-related genes are insufficient. The patients with FLT3 and IDH1/2 mutations accounts for 50% of total AML population, and not all patients response positively to these treatments, in another word, a small number of patients benefit from FLT3 and IDH1/2 inhibitors. Therefore, further investigation of AML pathogenesis is urgently needed in order to find more potential target genes, providing the theoretical basis for development of new drugs.

According to cytogenetic abnormalities, the prognosis of AML patient is generally classified as favorable, intermediate or unfavorable[24]. However, approximately 40% of AML patients have no identifiable cytogenetic abnormalities detected by molecular cytogenetics and fluorescence in situ hybridization (FISH) methods [25], they are classified as intermedium risk, but the overall survival of these patients varies widely. AML is a highly heterogeneous disease, investigation of prognostic value of biomarkers is one of the research hotspots. Based on the analysis of single-omics data, researchers identified numerous AML-related factors. As a complex regulatory system, disease occurrence usually involves genetic variation, epigenetic changes, and abnormal gene expression. Therefore, it is very important and meaningful to screen prognostic biomarkers of AML through integrated analysis of multi-omics data.

---

## 2. MATERIALS AND METHODS

### 1. TCGA data download

AML mutation data was downloaded via the R package TCGA biolinks [26] <https://bioconductor.org/packages/release/bioc/html/TCGAbiolinks.html>. The AML data was downloaded from the website <http://firebrowse.org/>, including SNP6 CopyNumber segment data and methylation chip data (platform for Illumina 450K chip). The mRNA and miRNA expression profile data as well as sample clinical data were downloaded from the TCGA website <https://portal.gdc.cancer.gov/>. In total, 103 patients' data from five data sets (mutation, CNV, methylation, mRNA, miRNA) were analyzed.

### 2. Healthy blood sample expression profile data download

All normal tissue sample gene expression profiles were taken from 407 whole blood samples from the GTEx release version V7 <https://gtexportal.org/home/datasets>.

### 3. Mutation (SNV) analysis

#### 3.1 Mutation gene analysis

MutSigCV was used to analyze high-frequency mutation gene, with most mutations occurring at the conserved sites. The analysis was performed by the corresponding MutSigCV module in the online analysis tool GenePattern [27] <https://cloud.genepattern.org/gp/pages/index.jsf> developed by the Broad Institute.

#### 3.2 Analysis of mutation signature

The following six basic mutation types appear in cancer: C>A, C>G, C>T, T>A, T>C, T>G. To investigate the different types of mutations, the researchers proposed the concept of mutational signatures [28], which is based on six basic types of mutations, followed by one base upstream and one base downstream of the mutation. There are four cases of A, T, C, and G for each base, so a total of  $4 \times 6 \times 4 = 96$  mutation types may be obtained. The frequency of these 96 mutation types in different cancers is different. The mutational signatures of AML can be obtained after non-negative matrix decomposition of the frequencies of 96 mutation types in each sample. Herein, the R package maftools [29] <https://bioconductor.org/packages/release/bioc/html/maftools.html> and SomaticSignatures [30] <https://bioconductor.org/packages/release/bioc/html/SomaticSignatures.html> were used to perform mutation signature analysis and the samples were then unsupervised hierarchically clustered based on the contribution of each label to observe the clinical features of the subset of samples with different mutation characteristics.

### 4. Somatic cell copy number variation (CNV) analysis

The GISTIC method was used to detect common CNV regions in all samples, including chromosome arm horizontal CNV and the smallest common region between samples. The GISTIC method parameters were set as:  $Q \leq 0.05$  as the change significance criterion; when determining the peak interval, the confidence level was 0.95; when analyzing the chromosome arm horizontal

---

variation, the region was greater than the chromosome arm length 0.98 as the standard. The analysis was performed using the corresponding GISTIC module in GenePattern.

#### 5. Difference analysis

Data correction, standardization and differential gene analysis were performed on the expression profiles of AML samples and healthy whole blood samples using DESeq[31] in the R algorithm.

#### 6. Heterogeneity analysis

The mean and standard deviation of each methylation site or miRNA expression value in the methylation expression profile and the miRNA expression profile were calculated, followed by calculation of the CV value to screen for sites or miRNAs with high heterogeneity. The calculation method of the CV value is:  $CV = \text{Standard deviation} / \text{Average}$ .

#### 7. Identification of candidate gene sets

The differentially expressed genes and the AML-related genes identified from the GENE database, the OMIM database and the KEGG database were jointly examined to screen at least one gene for the study. The gene annotated in the high heterogeneity methylation site analysis and the significant mutation gene obtained by MutSigCV analysis were further added to expand the scope of the study, and finally obtain the candidate gene. The expression profile data of the candidate gene set in AML samples in TCGA was selected for the next step of weight co-expression network analysis.

#### 8. WGCNA analysis

WGCNA [32] is a systematic biology method that uses gene expression data to construct a scale-free network. First, the similarity matrix is constructed using gene expression data, that is, the absolute value of the Pearson correlation coefficient between the two genes( $i$  and  $j$ ) is calculated using Eq. (1), where  $x_i$  and  $y_j$  represent the expression values of the  $i$ -th and  $j$ -th genes, respectively:

$$S_{ij} = \left| \frac{1 + \text{cor}(x_i + y_i)}{2} \right| \quad (1)$$

Then, Eq. (2) is used to transform the gene expression similarity matrix into adjacency matrix, where  $\beta$  is the soft threshold, which is actually the Pearson correlation coefficient  $\beta$  power of each gene pair:

$$a_{ij} = \left| \frac{1 + \text{cor}(x_i + y_i)}{2} \right|^\beta \quad (2)$$

This step can strengthen the strong correlation and weaken the weak correlation from the index level. The next step is to convert the adjacency matrix into a topological matrix using Eq. (3). The topological overlap measure (TOM) is used to describe the degree of association between genes:

$$TOM = \frac{\sum_{u \neq i, j} a_{i u} a_{u j} + a_{i j}}{\min(\sum_u a_{i u} + \sum_u a_{j u}) - 1} \quad (3)$$

1-TOM indicates the degree of dissimilarity between gene i and gene j. Hierarchical clustering of genes using 1-TOM as a distance is performed, then using the method of dynamic cut tree for module identification.

Using the WGCNA package of the R algorithm, a weighted co-expression network was constructed on the expression profile data of the candidate gene set obtained in the previous step. The modules related to AML survival time were screened, and the inter-subnets were constructed for the co-expressed gene sets in the module, then candidate core genes of AML were screened according to the network node degree.

### 9. Cluster analysis

All cluster heat maps were completed using the R package pheatmap, and the clustering method selects ward.D.

### 10. Identification of miRNAs that regulate the module

The miRNA-mRNA interaction relationship was included in the starBase v2.0 database <http://starbase.sysu.edu.cn/targetSite.php>. The database was used to screen the interaction between differential miRNAs and module genes, and the interaction network between differential miRNAs and module genes was constructed to observe differential miRNAs and module core genes regulating the relationship. The miRNAs with significant regulatory effects on the modules were then screened by hypergeometric analysis based on the interaction between the miRNA and the module gene.

### 11. Survival analysis

Based on the clinical data compiled by TCGA, Kaplan-Meier analysis of candidate core genes of AML was performed to identify genes that had a significant impact on patient prognosis ( $P < 0.05$ ).

### 12. Function and pathway enrichment analysis

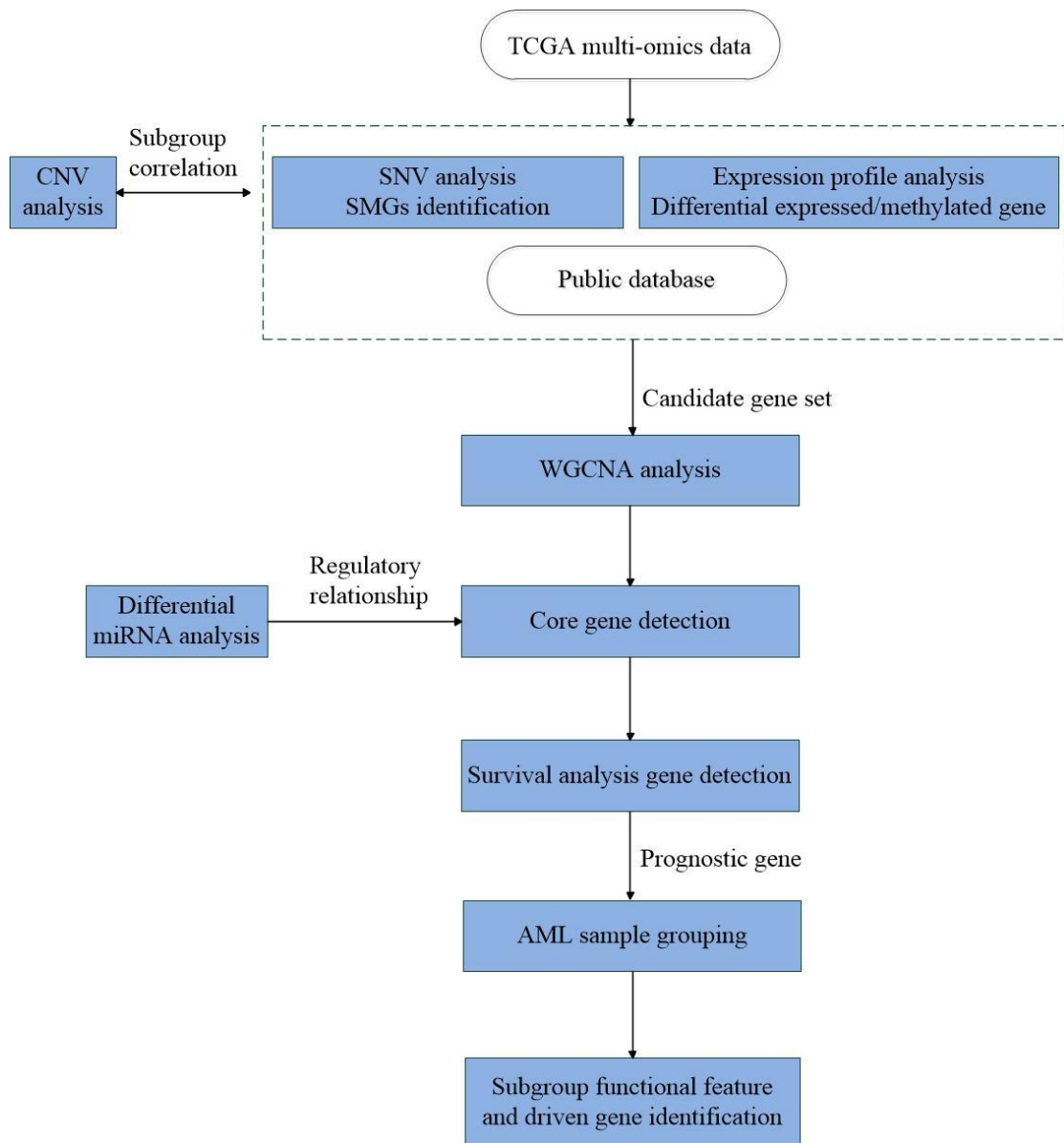
GO and KEGG enrichment analysis was performed on the gene set of interest using the R package clusterProfiler [33] <https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>.

## 3. RESULTS

### 1. Sample and clinical data

The study included a total of 103 patients with clinical information (Supplementary Table 1) and multiomics analysis data (including mutation results, CNV results and methylation expression profile data, mRNA expression profile data and miRNA expression profile data). The mRNA expression profile data of 407 healthy whole blood samples was also downloaded from the GTEx database, and the flowchart was shown in Figure 1.





**Figure1: Workflow**

## 2. Variant genes and mutation characteristics

### 2.1 Identification of significant variant genes

Numerous somatic mutations occur during the development of cancer, and mutations in a few genes, so called ‘driving genes’ directly promote tumorigenesis and development. Driving gene prediction was performed using MutSigCV for mutation data from 103 tumor samples. At the significance threshold  $q < 0.05$ , a total of 378 candidate driving genes were screened, including DNMT3B, NPM1 and other genes with a high proportion of mutations in AML (see Supplementary Table 2).

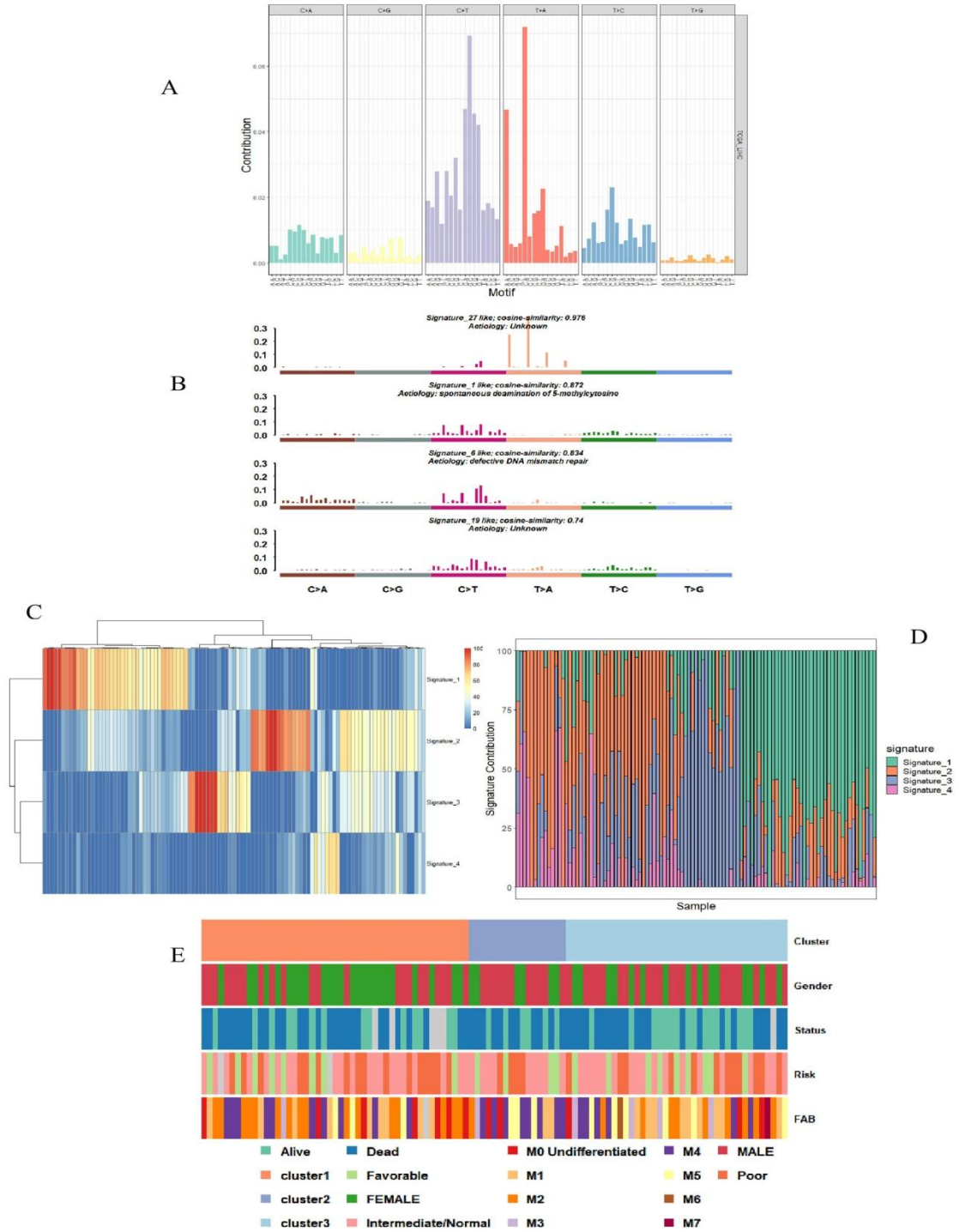
---

## 2.2 Mutation characteristics

We studied the point mutation information of 103 samples. According to the mutation site of each sample, we considered the bases 1 bp upstream and downstream of the mutation site and divided the mutations into 96 types according to the above classification method. The frequency distribution in the 103 tumor samples was counted and the distribution is shown in Figure 2A.

The frequencies of the 96 mutation types appearing in each sample were further subjected to non-negative matrix decomposition to extract the existing mutation characteristics. The four mutation characteristics extracted are shown in Figure 2B, which are highly similar to signature27, signature1, signature6 and signature19 in the cosmic database. Signature6 is associated with a DNA mismatch repair system, and signature1 is associated with spontaneous deamination of 5-methylcytosine.

The distribution of each mutation feature in different samples (see Supplementary Table 3 for analysis results) was then performed for unsupervised hierarchical clustering (see Figure 2C). In general, the three subgroups in the AML sample have different mutation characteristics, with the ratio of signature 1 being the highest in one subgroup, and signature 3 being the highest in the other subgroup (see Figure 2D). The sample classification was performed from the mutation characteristics alone. The FAB classification results were significantly different in the distribution of M2-M5 ( $p$ -value = 0.04715) in the subgroups. The distribution of clinical features is shown in Figure 2E.



**Figure2: (A) Mutant mapping.** The abscissa indicates the type of 96 mutations, and the ordinate indicates the proportion of the corresponding mutation. **(B)AML mutation feature distribution.** The abscissa indicates the type of 96 mutations, and the ordinate indicates the proportion of the corresponding mutation. **(C) Mutation feature proportional cluster analysis.** **(D) Distribution of four mutation characteristics in the AML population.** The abscissa represents the sample, and the ordinate represents the contribution ratio of each signature in the sample. **(E) Distribution of the clinical features in the three subgroups.**

---

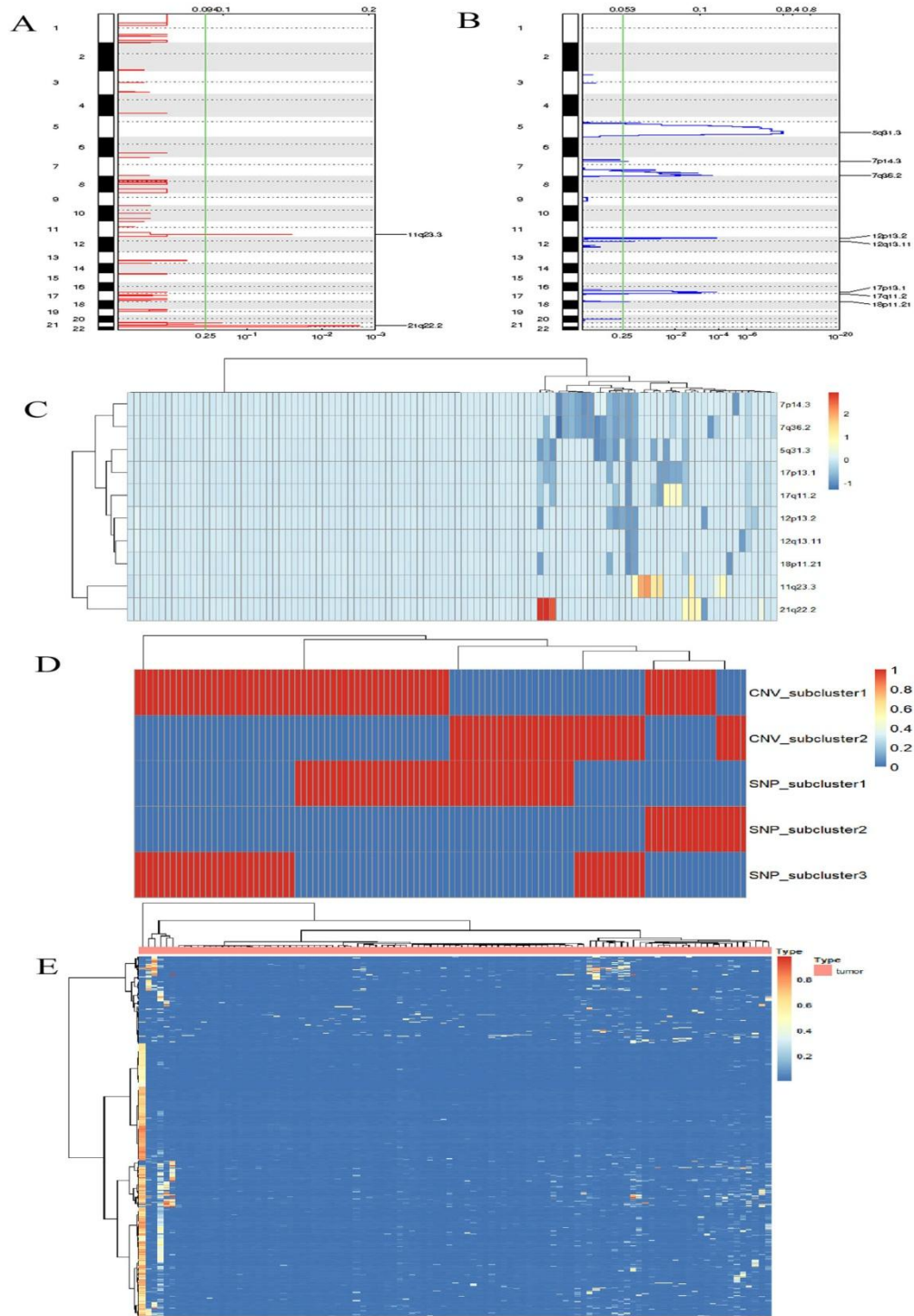
### 3. Somatic cell CNV

The GISTIC analysis of 103 samples detected a total of two copies and eight copy number deletions of concentrated copy number variants (MCRs) (see Supplementary Table 4; Figure 3A-3B). Among them, the most significant areas of amplification were 11q23.3, 21q22.2 etc.; the most significant of the missing areas were 5q31.3, 12p13.2 and so on. The unsupervised hierarchical clustering for the CNV distribution of these samples can divide AML samples into two sub-categories (see Supplementary Table 5 and Figure 3C for classification results), and one of them has significant CNV changes.

According to the classification of SNP and CNV data, the classification results of each sample are shown in following Figure 3D, revealing that the classification effects of different platforms have certain consistency. The classification results are shown in Supplementary Table 6.

### 4. DNA methylation

CV values were calculated for each methylation site of the 103 cancer samples. A total of 1,441 highly heterogeneous methylation sites were screened with CV values greater than  $>2.5$ , 779 genes were annotated, and 143 genes were differentially expressed, including SMAD1, SLC1A2 and the like. The heterogeneity analysis results are shown in Supplementary Table 7. The distribution of these heterogeneous methylation sites is shown in Figure 3E.



**Figure3: (A) CNV amplification distribution. (B) CNV missing distribution. (C) CNV unsupervised hierarchical clustering. (D) Comparison of different platform classification results. The red color indicates that the sample belongs to the subgroup corresponding to the horizontal axis, and the blue color indicates that the sample is not in the corresponding subgroup of the horizontal axis. (E) High heterogeneity methylation site hierarchical clustering.**

---

## 5. mRNA expression profile

### 5.1 Differential expression gene analysis

Differential gene analysis was performed on the mRNA expression profiles of 103 cancer samples and 407 control samples. When the significance threshold was 0.05,  $|\log FC| > 6$ , 4734 differentially expressed genes were screened and the results are shown in Supplementary Table 8.

### 5.2 Differential gene expression and mutation relationship

The effect of each mutation site on the gene is different, so the mutations of each gene in all samples and their effects on the genes were counts and categorized as Frame\_Shift\_Del, Frame\_Shift\_Ins, In\_Frame\_Del, Missense\_Mutation, Nonsense\_Mutation, Nonstop\_Mutation, Splice\_Site, Translation\_Start\_Site. Statistical analysis using t-tests showed that there were significant differences in the distribution of differential and non-differential genes in Splice\_Site ( $p$ -value = 0.01), indicating that SNV has a certain effect on gene expression. The results of the mutation classification are shown in Supplementary Table 9, and the statistical results are shown in Supplementary Table 10.

### 5.3 Candidate gene set

The differentially expressed genes and the GENE database and OMIM database, as well as the AML-related genes identified in the KEGG database were jointly examined (genetic results are shown in Supplementary Table 11). The relationship between the genes included in the database and the differential genes is shown in Figure 4A, of which only 4,504 genes were identified in differential analysis, which may be associated with the onset of AML.

In order to further expand the scope of investigation, 779 genes annotated in the analysis of highly heterogeneous methylation sites and 378 significant mutations obtained by MutSigCV analysis were added to obtain a candidate gene set of 6,383 genes (see Supplementary Table 12). The AML sample expression profile data of the 6,383 genes in TCGA was then used for the next weight co-expression network analysis.

### 5.4 Weighted co-expression network construction of candidate gene sets

The construction of a weighted co-expression network for candidate gene sets was performed using R's WGCNA software package. The co-expression network conforms to the scale-free network, that is, the logarithm  $\log(k)$  of the node with the connection degree  $k$  is negatively correlated with the logarithm  $\log(P(k))$  of the probability of the node, and the correlation coefficient is greater than 0.8. To ensure that the network is a scale-free network, the optimal  $\beta=10$  was used (see Figure 4B). The average-linkage hierarchical clustering method was used to cluster genes (see Figure 4C), obtaining fourteen modules (see Supplementary Table 13).

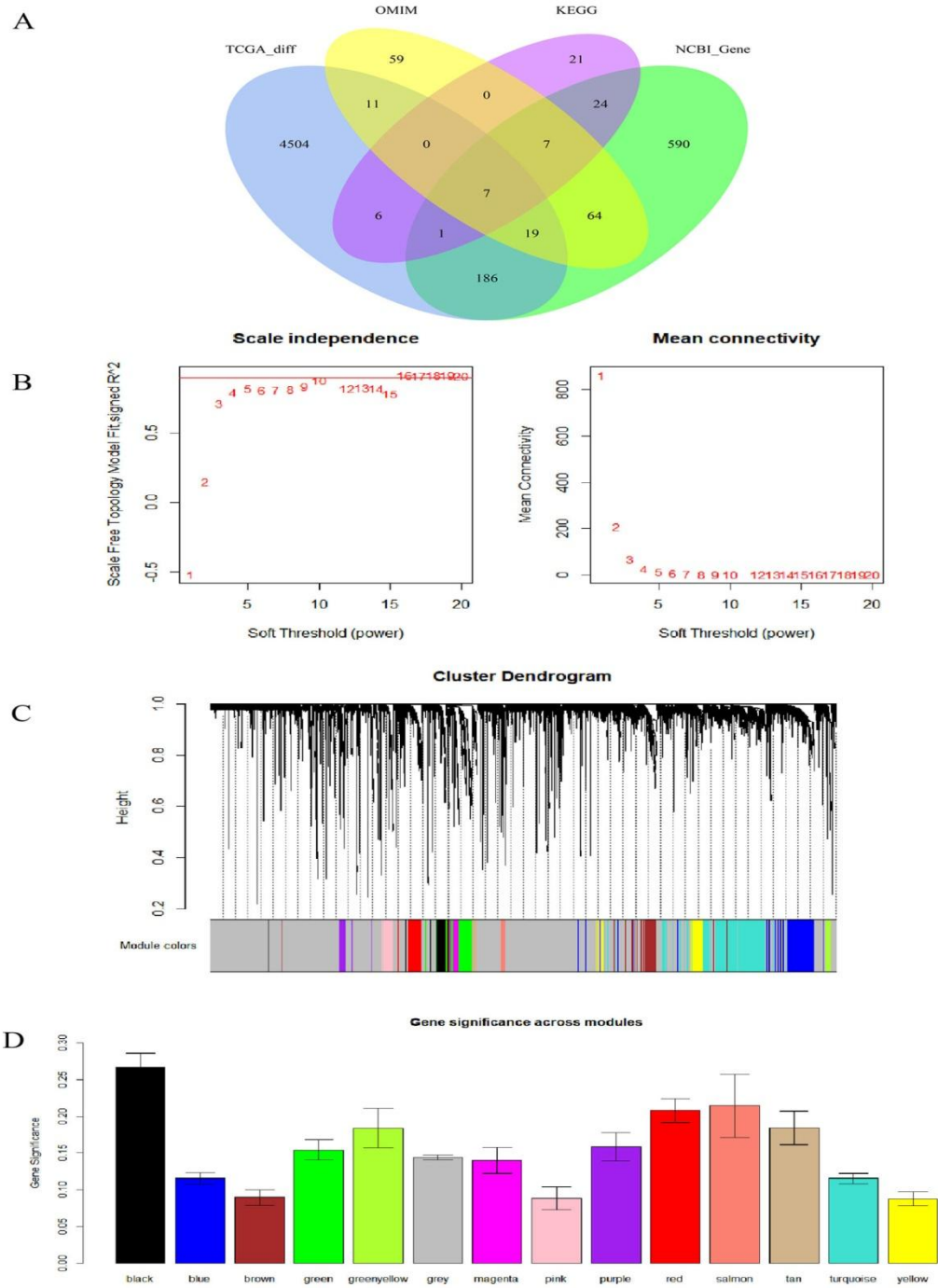
To determine the association of gene modules with prognostic risk (Favorable, Intermediate/Normal, Poor), we calculated the Pearson correlation coefficient (higher for the module) and the significance  $P$  for the corresponding correlation for each module and sample prognosis risk, respectively. The results of gene module characteristics and phenotypic correlation coefficients are shown in Table 1. Subsequently, the gene significance (GS) value of each gene module was calculated, the larger the GS value, the more relevant the module is to the prognosis. Through the correlation analysis and GS value calculation, the last seven modules related to AML

were selected, which are tan, red, salmon, black, green, purple, greenyellow (see Figure4D and Supplementary Table 14 ).

**Table 1. Gene module characteristics and phenotypic correlation results.**

module	Cor	Pvalue
black	3.81E-01	7.06E-05
salmon	2.88E-01	3.16E-03
red	2.83E-01	3.81E-03
grey	2.45E-01	1.27E-02
greenyellow	-2.34E-01	1.75E-02
tan	2.33E-01	1.81E-02
purple	2.09E-01	3.40E-02
green	1.95E-01	4.84E-02
magenta	1.72E-01	8.15E-02
pink	8.22E-02	4.09E-01
turquoise	6.98E-02	4.84E-01
blue	-1.17E-02	9.06E-01
brown	-7.99E-03	9.36E-01
yellow	5.55E-03	9.56E-01

Based on the expression relationship of the genes in the seven co-expression modules analyzed above, we constructed a co-expression network for each module separately (see Supplementary Table 15 for the analysis results), then counted the degree of each network node, the degree of node indicates that there are connections with several genes in the network, the higher the degree, the more important the node is in the network (see Supplementary Table16). A small number of genes in the co-expression network have obvious concentration. After the expression changes, the expression of these genes will interact with other nodes adjacent to them to produce co-expression, which will affect downstream biological functions. Therefore, these have higher node degrees and the gene is likely to be a key gene in the development of AML. Finally, we chose the node top10 genes as the key gene in the co-expression network, which is the key gene related to AML.



**Figure 4: (A) AML-related genes from different platforms (B) Network soft threshold  $\beta$  determination. (C) Co-expression module. (D) GS mean distribution of different gene modules.**



---

## 6. miRNA expression profiling

### 6.1 Heterogeneous miRNA analysis

CV values were calculated from miRNA expression profiles of 103 cancer samples. A total of 568 highly heterogeneous miRNAs were screened at  $CV > 2.5$  (see Supplementary Table 17).

### 6.2 miRNA-mRNA analysis

The miRNA-mRNA interaction pairs of the number of supporting experiments  $\geq 3$  and number of cancer types  $\geq 1$  from the starBase v2.0 database <http://starbase.sysu.edu.cn/starbase2/mirMrna.php> were screened. Then, the hypergeometric test method was used to identify the miRNAs that regulate the AML-related modules, revealing that there were 72 miRNAs that have significant interaction with the above four modules (see Supplementary Table 18).

Regarding the regulatory relationship between highly heterogeneous miRNAs and the core genes of the modules and modules, a total of ten highly heterogeneous miRNAs interacted with the above seven modules (see the Supplementary Table 19), of which two were high. Heterogeneous miRNAs regulate two core genes of the module, which are miR-346 and ID4, miR-449a and FIP1L1-PDGFR $\alpha$ , respectively.

## 7. Survival analysis

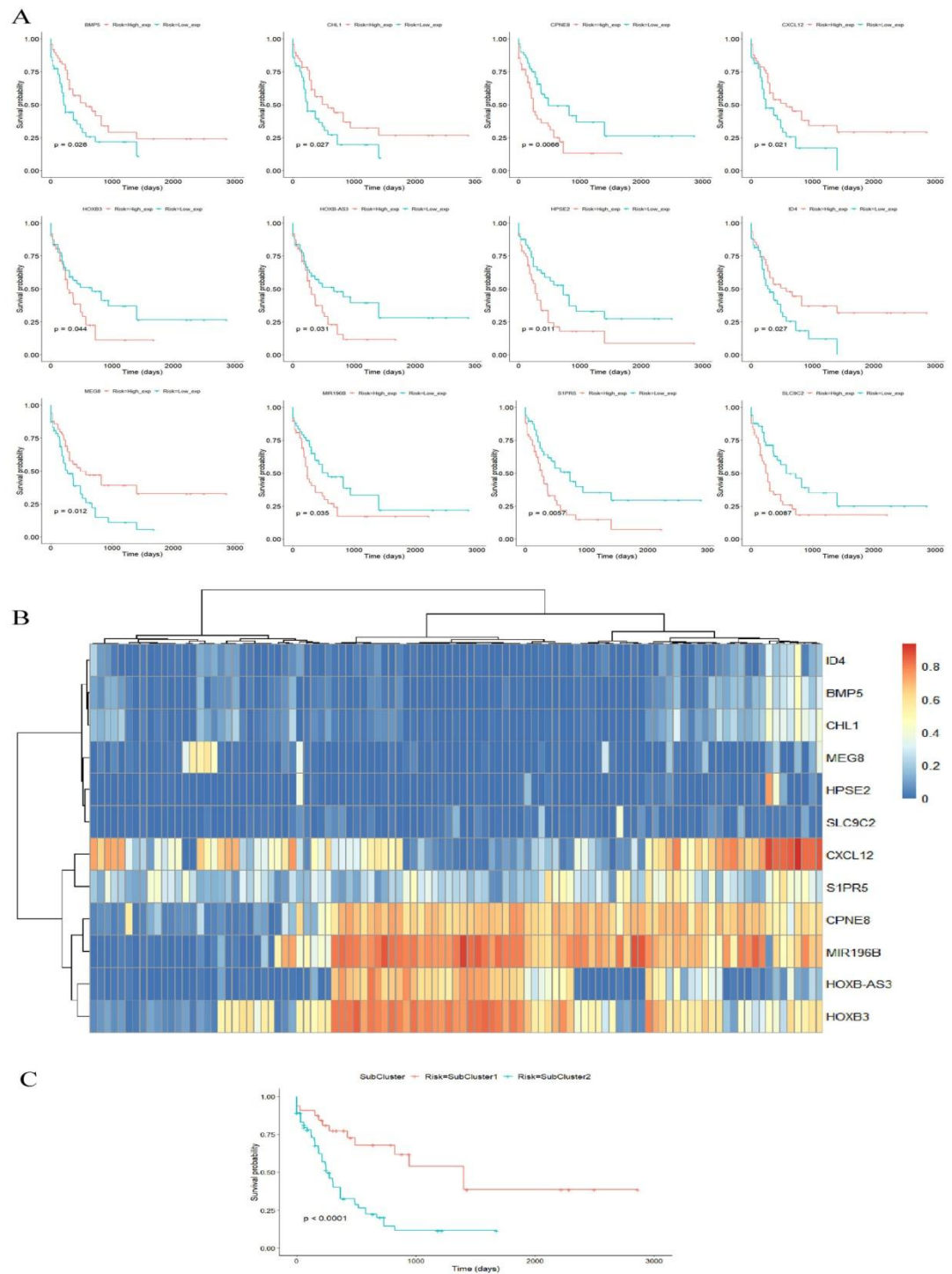
To determine whether the 7 key genes associated with AML (7 modules of the top 10 gene) were significantly associated with prognosis, we combined the clinical data of 103 samples in the TCGA to determine the expression of these genes in the sample. Kaplan-Meier survival analysis was performed, when the significance threshold was set to 0.05, a total of twelve module core genes were screened for prognosis (survival curves see Figure 5A and Supplementary Table 20 for results). Survival analysis was also performed on the previously analyzed genes that are both significant mutations and seven modules combined with their mutations in the sample, when the significance threshold was set to 0.05, no significant mutations and prognosis were related.

## 8. Driving gene-mediated AML molecular typing

Unsupervised clustering of cancer samples was performed using the expression of twelve prognostic-related genes screened in the previous step in 103 AML samples (see Figure 5B and Supplementary Table 21). As shown, all AML samples can be divided into two categories (see Figure 5C and Supplementary Table 22), and the prognosis of the two subgroups is significantly different.

To further analyze the differences between these key genes in different subpopulations, the mean expression of each gene in each sample type was used as the expression value of the gene in this category. After t-test analysis, these twelve genes were associated with prognosis, of which seven (HOXB-AS3, HOXB3, SLC9C2, CPNE8, MEG8, S1PR5, MIR196B) were differentially expressed in the two subgroups. The distribution of seven genes see Figure 6A and the analysis results in Supplementary Table 23.

According to the drug-target interaction relationship included in the drugbank database, the seven genes were separately labeled with drugs, revealing that one gene (S1PR5) is a target gene of one drug (Fingolimod) (see Supplementary Table 24).

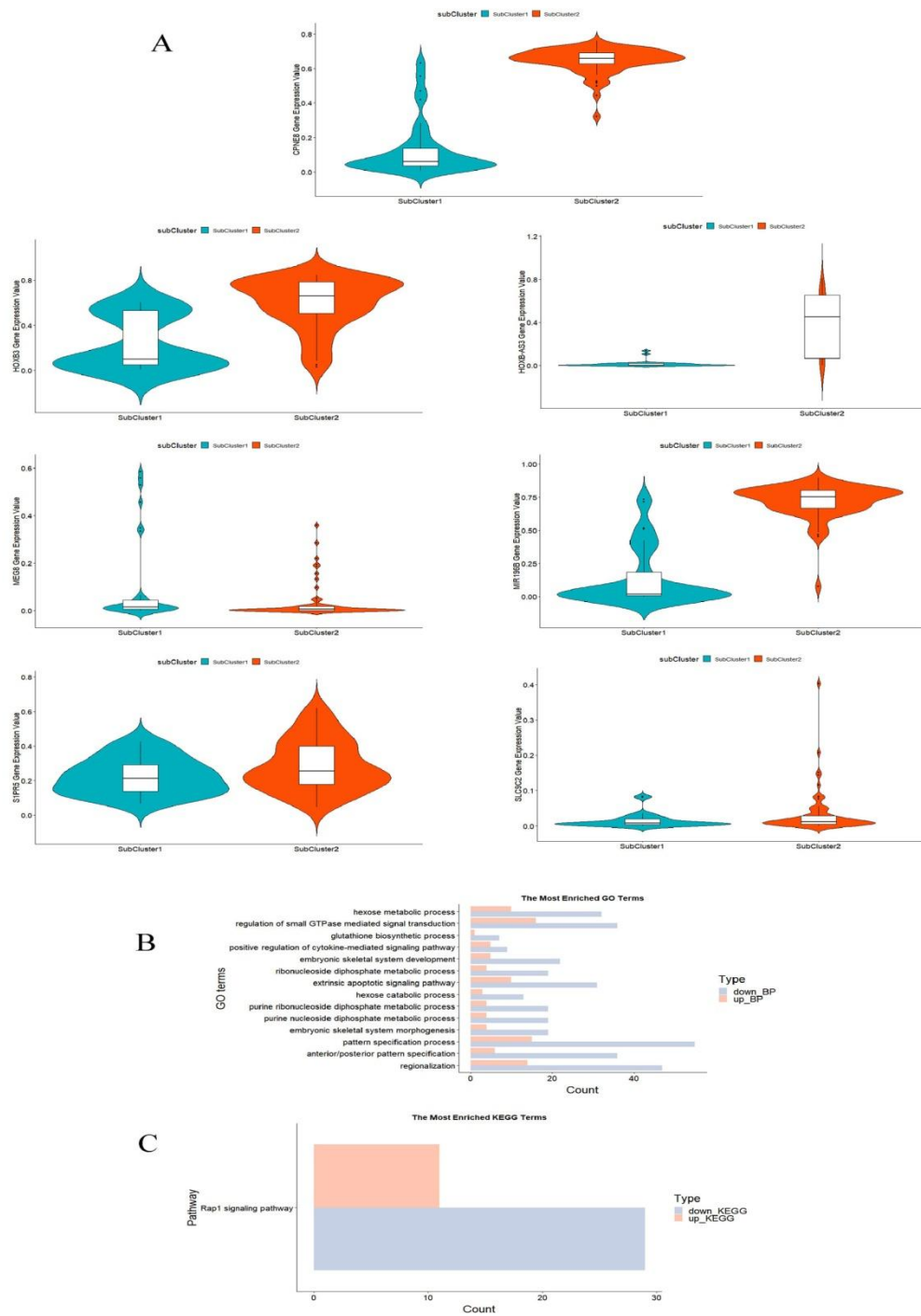


**Figure 5 (A) Survival curve of 12 module core gene expression. (B) Prognosis-related gene expression in AML samples. (C) Differences in the prognosis of the two subgroups.**

---

#### 9. AML subgroup functional differences

Differential genetic analysis was performed on the two types of AML samples obtained in the previous step. When the significance threshold was 0.05, 2293 differentially expressed mRNAs were screened out (see Supplementary Table 25 for the analysis results). Functional and pathway enrichment analysis of these genes (GO and KEGG enrichment analysis see Figure 6B and Figure 6C; see Supplementary Table 26), showed that these genes are mainly involved in glucose metabolism, glutathione biosynthesis, small G protein-mediated signal transduction, and the Rap1 signaling pathway.



**Figure 6 (A)** Seven genes (HOXB-AS3, HOXB3, SLC9C2, CPNE8, MEG8, S1PR5, MIR196B) differentially expressed in two subgroups. **(B)**Differential gene GO enrichment analysis. **(C)** Differential gene KEGG enrichment analysis.

---

#### 4. DISCUSSION

We applied WGCNA to analyse public dataset comprising 103 AML patients to identify the genes associated with prognosis indicators. In this study, seven distinct gene modules and twelve biomarkers associated with the AML prognosis from 6383 genes that satisfied standard of co-expression analysis were identified. The 103 AML samples could be classified into two subgroups which is significantly different in prognosis.

To further analyze the differences between these key genes in different subpopulations, the mean expression of each gene in each sample type was used as the expression value of the gene in this category. After t-test analysis, these twelve genes were associated with prognosis, of which seven (HOXB-AS3, HOXB3, SLC9C2, CPNE8, MEG8, S1PR5, MIR196B) were differentially expressed in the subgroups. Some studies have shown that several of these seven genes are associated with the development or prognosis of AML. It has been previously reported that HOX genes have an important role in the development of AML malignant phenotype and inhibition of HOX/PBX dimer formation leads to necroptosis in AML [34]. In addition, down regulation of HOXB-AS3 can inhibit cell proliferation. In AML, patients with high HOXB-AS3 expression have shorter survival than patients with low HOXB-AS3 expression [35]. More than 30% of AML patients have a type III receptor tyrosine kinase FLT3 mutation, and HOXB2 and HOXB3 are novel regulators of oncogenic FLT3-ITD-driven AML [36]. Hsa-miR-196b has been shown to directly target HOXA9 / MEIS1 oncogenes and FAS tumor suppressors in MLL rearranged leukemia [37]. Recent analysis showed that HOXB3 is an important homeobox protein which contributes to leukemogenesis by a novel *miR-375*-HOXB3-CDCA3/DNMT3B regulatory circuitry in AML [38]. In addition, SLC9C2, CPNE8, MEG8, S1PR5 have not been found to be associated with the prognosis of AML. This indicates that among the seven genes we analyzed, there are unreported genes that can be used as potential driven genes for AML.

Our study identified the miRNAs that regulate the AML-related modules, heterogeneous miRNAs regulate two core genes of the module, which are miR-346 and ID4, miR-449a and FIP1L1-PDGFR $\alpha$ . It has been reported that FIP1L1-PDGFR $\alpha$  is a kinase translocation which could be cancer driver, resulting in eosinophilias and leukemias [39]. However, the role of miR-449a in leukemogenesis has yet to be elucidated. Previous research has shown that as an independent prognostic biomarker in AML, aberrant miR-335/ID4 expression contributed to AML development by activation of PI3K/Akt signaling pathway [40]. It would be worth exploring the interaction and biological functions of miR-346/ID4 and miR-449a/FIP1L1-PDGFR $\alpha$  in AML. More work will be necessary to explore these possibilities.

According to the drug-target interaction relationship included in the drugbank database, the seven genes were separately labeled with drugs, revealing that one gene (S1PR5) is a target gene of one drug (Fingolimod). Current treatment options for AML include protein phosphatase 2A (PP2A) activators [41]. An earlier meta analysis indicate that with caution of infections or cardiac toxicity, fingolimod could be a potential treatment for elderly, low tumour burden AML patients in combination with hydroxyurea or subcutaneous cytarabine [42]. Fingolimod (FTY720, Gilenya; Novartis, Basel, Switzerland) is a known PP2A activator that inhibits multiple signaling pathways like Ras/ERK, PI3K/Akt, JNK, JAK/STAT and Wnt/ $\beta$ -catenin pathways [43] and is currently used in patients with multiple sclerosis [44]. As a tumour suppressor [45], fingolimod demonstrated inhibition of tumor proliferation and in-vivo leukaemogenesis [46]. However, how

---

fingolimod activates PP2A remains unclear, it has been reported that fingolimod might interacts with PP2A inhibitory proteins such as SET oncogene directly[47]. Our results may suggest a possible drug-target interaction.

## 5. CONCLUSION

With the utilization of WGCNA mining, seven gene co-expression modules were identified from TCGA database and there are unreported seven genes( HOXB-AS3, HOXB3, SLC9C2, CPNE8, MEG8, S1PR5, MIR196B) that could be used as potential driven genes for AML. Indeed, in vivo/vitro experiments and multicenter randomized controlled clinical trials should be performed to evaluate the possible application of molecular signatures to predict survival of AML patients and further research of these hub genes may contribute to molecular targeted therapy of AML and help guide experiments for potential clinical drug targets.

## Abbreviations

AML: Acute myeloid leukemia

ALL: Acute lymphoblastic leukemia

CR: Complete response

DFS: Disease-free survival

OS: Overall survival

FLT3: Fms-like tyrosine kinase-1

IDH1/2: Isocitrate dehydrogenase 1/2

PR: Partial response

R/R AML: Recurrent or refractory acute myeloid leukemia

ORR: Objective remission rate

FISH: Fluorescence in situ hybridization

TCGA: The cancer genome atlas

GEO: Gene expression omnibus

CNV: Copy number variation

SNV: Single nucleotide variants

TOM: Topological overlap measure

WGCNA: Weighted gene co-expression network

GO: Gene ontology

KEGG: Kyoto encyclopedia of genes and genomes

## Declarations:

•Ethics approval and consent to participate  
Not required.

•Consent for publication

That the article is original, has not already been published in a journal, and is not currently under consideration by another journal.

---

- Availability of data and materials

The data that support the findings of this study are openly available in TCGA <https://bioconductor.org/packages/release/bioc/html/TCGAbiolinks.html> and the GTEx release version V7 <https://gtexportal.org/home/datasets>.

- Competing interests

The authors have no competing interests to declare.

- Funding

This study was partially supported by the Youth Scientific Foundation of Qilu Hospital (Qingdao), Shandong University (No.QDKY2015QN10).

The award of a China Scholarship Council visiting scholar grant to Jimo Jian (No.201906225042) supported the development of this paper.

- Authors' contributions

Jimo Jian, Fei Lu and Chenglu Yuan conceived and designed this study. Hongyuan Hao, Qian Liu , Wenhao Dong and Lu Li carried out the analysis procedure, Jimo Jian, Yalu Liu and Gege Feng analyzed the results, Sheng Gao, Lai Chen and Juan Wang contributed analysis tools, Chenglu Yuan and Fei Lu participated in the manuscript writing. All authors read and approved the final manuscript.

- Acknowledgments

None

- Author's information:

Jimo Jian, Hongyuan Hao, Qian Liu , Wenhao Dong, Lu Li, Yalu Liu, Gege Feng, Juan Wang , Sheng Gao, Lai Chen, Fei Lu\*, Chenglu Yuan \*.

- Corresponding author

Correspondence to Chenglu Yuan or Fei Lu.

- Author's Affiliations:

Department of Hematology, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Wenhua West Road, Jinan, Shandong, 250012, China

Jimo Jian and Fei Lu

Department of Hematology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, 758 Hefei Road, Qingdao, Shandong, 266035, China

Hongyuan Hao, Wenhao Dong, Lu Li, Yalu Liu, Gege Feng, Juan Wang and Sheng Gao

Department of Pain, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Wenhua West Road, Jinan, Shandong, 250012, China

Qian Liu

Integrated Chinese & Western Medicine Oncology Research Center, Jiangxi University of Traditional Chinese Medicine, Nanchang, Jiangxi 330004, China

Lai Chen

---

## References:

1. Döhner H, Weisdorf DJ, and Bloomfield CD. Acute myeloid leukemia. *N Engl J Med.* 2015; 373: 1136–1152.
2. Siegel RL, Miller KD, and Jemal A. Cancer statistics, 2018. *CA Cancer J. Clin.* 2018; 68:7–30.
3. Hartmann L, and Metzeler KH. Clonal hematopoiesis and preleukemia—Genetics, biology, and clinical implications. *Genes Chromosomes Canc.* 2019; 58:828-838.
4. Bennett JM, Catovsky D, Daniel M T, Flandrin G, Galton DA, Gralnick HR, and Sultan C. Proposals for the classification of the acute leukaemias. *Brit. J. Haematol.* 1976; 33:451–458..
5. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR, and Sultan C. Proposed revised criteria for the classification of acute myeloid leukemia. A report of the French-American-British Cooperative Group. *Ann. Med.* 1985; 103: 626-629.
6. Li S, Mason C, and Melnick A. Genetic and epigenetic heterogeneity in acute myeloid leukemia. *Curr. Opin. Genet. Dev.* 2016; 36:100-106.
7. Shipley J, Butera JN. Acute myelogenous leukemia. *Experimental Hematology.* 2009; 37(6): 649-58.
8. Dohner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood.* 2017;129(4):424-47.
9. Löwenberg B. Treatment of the elderly patient with acute myeloid leukaemia. *Baillière's Clinical Haematology.* 1996;9(1):147-59.
10. Garcia JS, Stone R. The Development of FLT3 Inhibitors in Acute Myeloid Leukemia. *Hematology-oncology Clinics of North America.* 2017;31(4):663-80.
11. Ley TJ, Miller CA, Ding L, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England Journal of Medicine.* 2013;368(22):2059-74.
12. Grimwade D, Mrozek K. Diagnostic and Prognostic Value of Cytogenetics in Acute Myeloid Leukemia. *Hematology-oncology Clinics of North America.* 2011;25(6):1135-61.
13. Levis M. FLT3 mutations in acute myeloid leukemia: what is the best approach in 2013? *2013; 2013(1):220.*
14. Im AP, Sehgal AR, Carroll MP, et al. DNMT3A and IDH mutations in acute myeloid leukemia and other myeloid malignancies: associations with prognosis and potential treatment strategies. *Leukemia.* 2014;28(9):1774-83.
15. Fiedler W, Kayser S, Kebenko M, et al. Sunitinib and Intensive Chemotherapy in Patients with Acute Myeloid Leukemia and Activating FLT3 Mutations: Results of the AMLSG 10-07 Study (ClinicalTrials.gov No. NCT00783653). *Blood.* 2012;120(21):1483-.
16. Walter F, Hubert S, Hartmut D. A phase 1 study of SU11248 in the treatment of patients with refractory or resistant acute myeloid leukemia (AML) or not amenable to conventional therapy for the disease. *2005;105(3):986-93.*
17. Walker AR, Rami SK, Jainulabdeen I, et al. Phase I study of cladribine, cytarabine (Ara-C), granulocyte colony stimulating factor (G-CSF) (CLAG Regimen) and simultaneous escalating doses of imatinib mesylate (Gleevec) in relapsed/refractory AML. *2008;32(12):1830-6.*
18. Smith BD, Levis M, Beran M, et al. Single-agent CEP-701, a novel FLT3 inhibitor, shows biologic and clinical activity in patients with relapsed or refractory acute myeloid leukemia. *2004;103(10):3669-76.*
19. Heidary DK, Huang G, Boucher D, et al. VX-322: A Novel Dual Receptor Tyrosine Kinase Inhibitor for the Treatment of Acute Myelogenous Leukemia. *Journal of Medicinal Chemistry.*



- 
- 2012;55(2):725-34.
20. D RFCJEJ. Phase I/II Study of Combination Therapy With Sorafenib, Idarubicin, and Cytarabine in Younger Patients With Acute Myeloid Leukemia. *Journal of Clinical Oncology*. 2010;28(11):1856-62.
  21. Abbas S, Lugthart S, Kavelaars FG, et al. Acquired mutations in the genes encoding IDH1 and IDH2 both are recurrent aberrations in acute myeloid leukemia: prevalence and prognostic value. *Blood*. 2010;116(12):2122-6.
  22. Dinardo CD, Stein EM, De Botton S, et al. Durable remissions with ivosidenib in IDH1-mutated relapsed or refractory AML. *The New England Journal of Medicine*. 2018;378(25):2386-98.
  23. Stein EM, Dinardo CD, Pollyea DA, et al. Enasidenib in mutant IDH2 relapsed or refractory acute myeloid leukemia. *Blood*. 2017;130(6):722-31.
  24. Kuykendall A, Duployez N, Boissel N, Lancet JE, and Welch JS. Acute myeloid leukemia: the Good, the Bad, and the ugly. *Am Soc Clin Oncol Educ Book*. 2018; 555-573.
  25. Handschuh L. Not Only Mutations Matter: Molecular Picture of Acute Myeloid Leukemia Emerging from Transcriptome Studies. *J. Oncol*. 2019; 2019:1-36.
  26. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, and Noushmehr H. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016; 44:1-11.
  27. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, and Mesirov J P. GenePattern 2.0. *Nat. Gent*. 2006; 38:500-501.
  28. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale A, Boyault S, Burkhardt B, Butler AP, et al. Signatures of mutational processes in human cancer. *Nature*. 2013; 500:415-421.
  29. Mayakonda A, Lin D, Assenov Y, Plass C, and Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018; 28:1747-1756.
  30. Gehring JS, Fischer B, Lawrence M, and Huber W. SomaticSignatures: inferring mutational signatures from single-nucleotide variants. *Bioinformatics*. 2015; 31:3673–3675.
  31. Anders S, and Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010; 11:1-12.
  32. Zhang B, and Horvath S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol*. 2005; 4:1-45.
  33. Yu G, Wang L, Han Y, and He Q. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012; 16:284-287.
  34. Alharbi RA, Pandha HS, Simpson GR, Pettengell R, Poterlowicz K, Thompson A, Harrington K, El-Tanani M, and Morgan R. Inhibition of HOX/PBX dimer formation leads to necroptosis in acute myeloid leukemia cells. *Oncotarget* 2017; 8: 89566-89579.
  35. Huang H, Chen F, Chou W, Hou H, Ko B, Lin C, Tang J, Li C, Yao M, Tsay W, Hsu S, Wu S, Chen C, et al. Long non-coding RNA HOXB-AS3 promotes myeloid cell proliferation and its higher expression is an adverse prognostic marker in patients with acute myeloid leukemia and myelodysplastic syndrome. *BMC Cancer* 2019;19:1-14.
  36. Lindblad O, Chougule R, Moharram SA, Kabir NN, Sun J, Kazi JU, and Rönstrand L. The role of HOXB2 and HOXB3 in acute myeloid leukemia. *Biochem. Bioph. Res. Co*. 2015; 467: 742-747.

- 
37. Li Z, Huang H, Chen P, He M, Li Y, Arnovitz S, Jiang X, He C, Hyjek E, Zhang J, Zhang Z, Elkahloun A, Cao D, et al. miR-196b directly targets both HOXA9/MEIS1 oncogenes and FAS tumour suppressor in MLL-rearranged leukaemia. *Nat. Commun.* 2012; 3:1-13.
  38. Wong KK, Lawrie CH, and Green TM. Oncogenic Roles and Inhibitors of DNMT1, DNMT3A, and DNMT3B in Acute Myeloid Leukaemia. *Biomarker Insights.* 2019; 14:1-12.
  39. Cicens J, Zalyte E, Bairoch A, and Gaudet P. Kinases and Cancer. *Cancers.* 2018; 10:1-7.
  40. Zhou J, Li X, Zhang T, Xu Z, Zhang Z, Gu Y, Wen X, Zhang W, Ji R, Deng Z, Lin J, and Qian J. MicroRNA-335/ID4 dysregulation predicts clinical outcome and facilitates leukemogenesis by activating PI3K/Akt signaling pathway in acute myeloid leukemia. *Aging.* 2019; 11: 3376—3391.
  41. Perrotti D, and Neviani P. Protein phosphatase 2A: a target for anticancer therapy. *Lancet Oncol.* 2013;14: e229-e238.
  42. Enjeti AK, Angel D' C, Melville K, Verrills NM, and Rowlings P. A systematic evaluation of the safety and toxicity of fingolimod for its potential use in the treatment of acute myeloid leukaemia. *Anti-Cancer Drugs.* 2016; 27:560-568.
  43. Arnold HK, and Sears RC. Protein phosphatase 2A regulatory subunit B56alpha associates with c-myc and negatively regulates c-myc accumulation. *Mol. Cell. Biol.* 2006; 26: 2832-2844.
  44. Yiu EM, and Banwell B. Update on emerging therapies for multiple sclerosis. *Expert Rev. Neurother.* 2010;10: 1259-1262.
  45. Janssens V, Goris J, and Hoof CV. PP2A: the expected tumor suppressor. *Curr. Opin. Genet. Dev.* 2005; 15: 34-41.
  46. Roberts KG, Smith AM, McDougall F, Carpenter H, Horan M, Neviani P, Powell JA, Thomas D, Guthridge MA, Perrotti D, Sim ATR, Ashman LK, and Verrills NM. Essential requirement for PP2A inhibition by the oncogenic receptor c-KIT suggests PP2A reactivation as a strategy to treat c-KIT+ cancers. *Cancer Res.* 2010; 70: 5438-5447.
  47. Saddoughi SA, Gencer S, Peterson YK, Ward KE, Mukhopadhyay A, Oaks J, Bielawski J, Szulc ZM, Thomas RJ, Selvam SP, Senkal CE, Garrett-Mayer E, De Palma RM, et al. Sphingosine analogue drug FTY720 targets I2PP2A/SET and mediates lung tumour suppression via activation of PP2A-RIPK1-dependent necroptosis. *EMBO Mol. Med.* 2013; 5: 105-121.

Figures

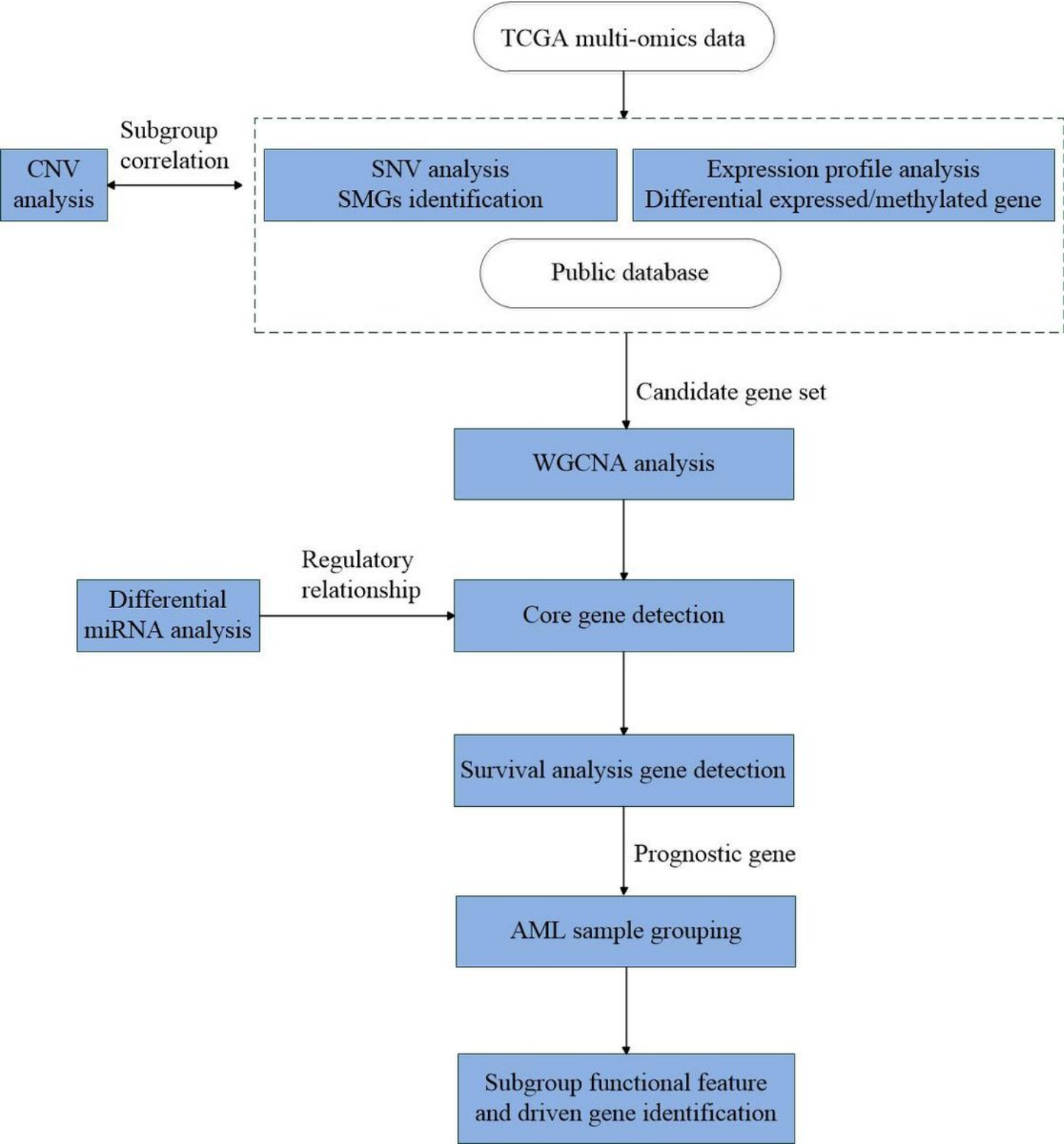
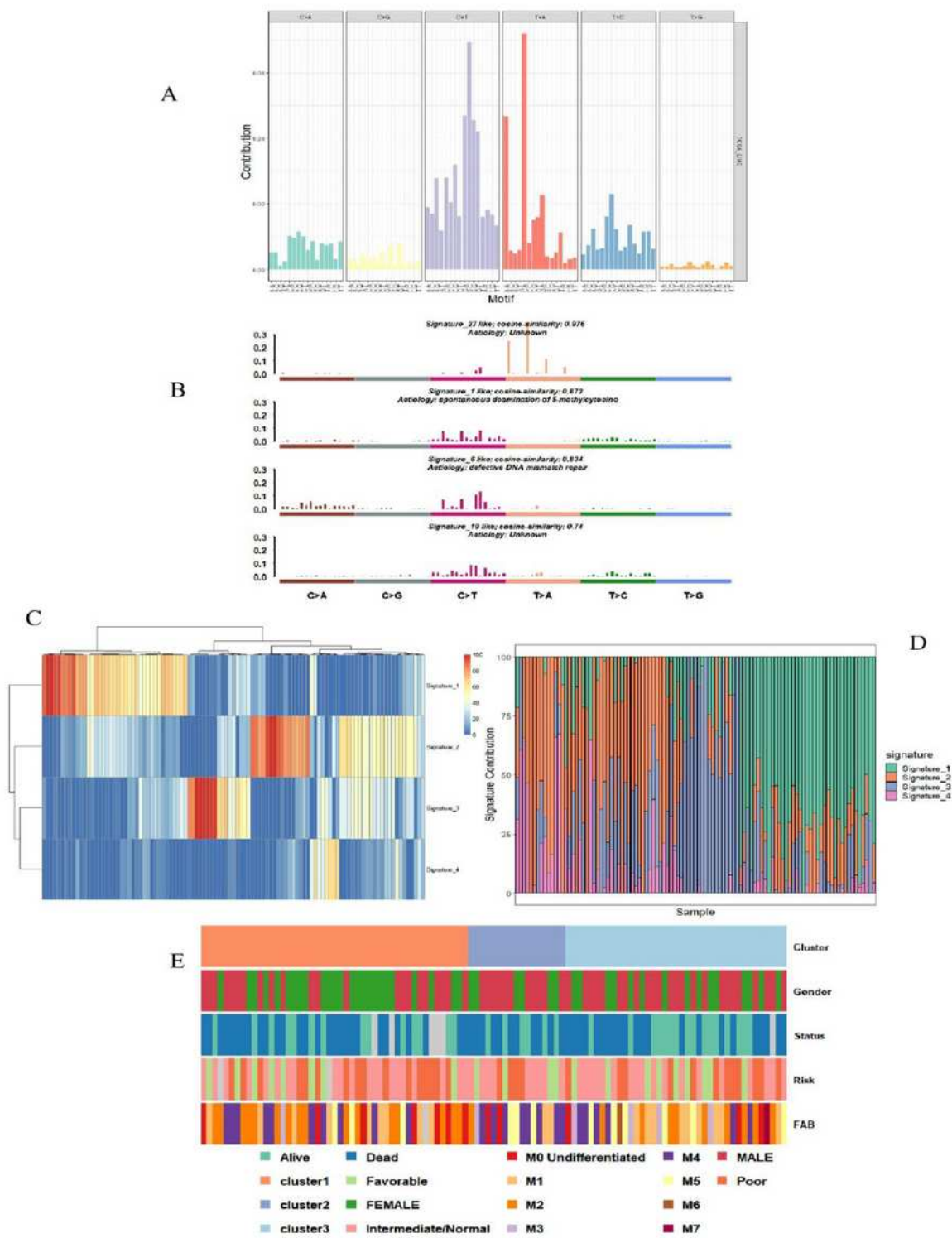


Figure 1

Workflow



**Figure 2**

(A) Mutant mapping. The abscissa indicates the type of 96 mutations, and the ordinate indicates the proportion of the corresponding mutation. (B) AML mutation feature distribution. The abscissa indicates the type of 96 mutations, and the ordinate indicates the proportion of the corresponding mutation. (C) Mutation feature proportional cluster analysis. (D) Distribution of four mutation characteristics in the

AML population. The abscissa represents the sample, and the ordinate represents the contribution ratio of each signature in the sample. (E) Distribution of the clinical features in the three subgroups.

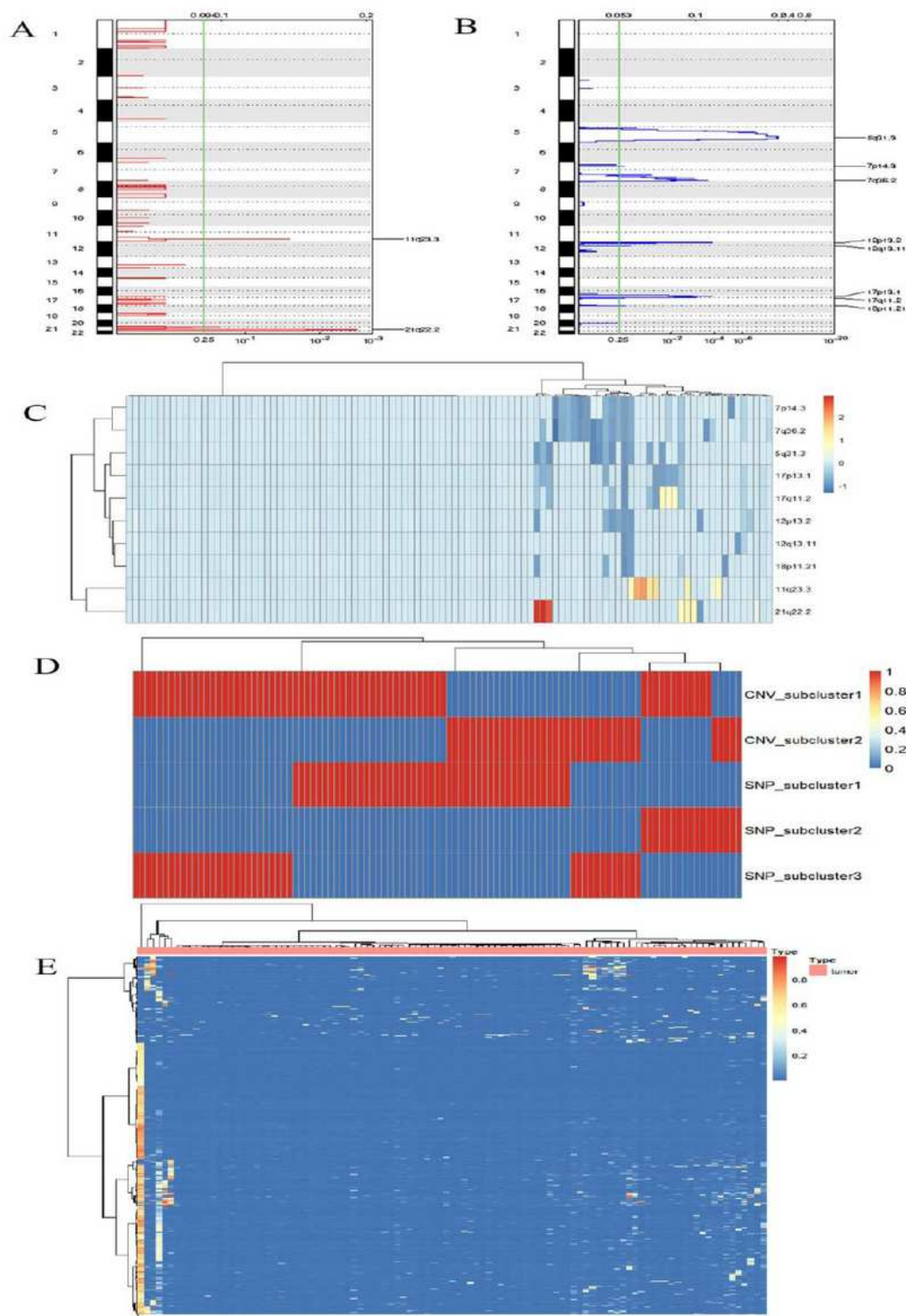
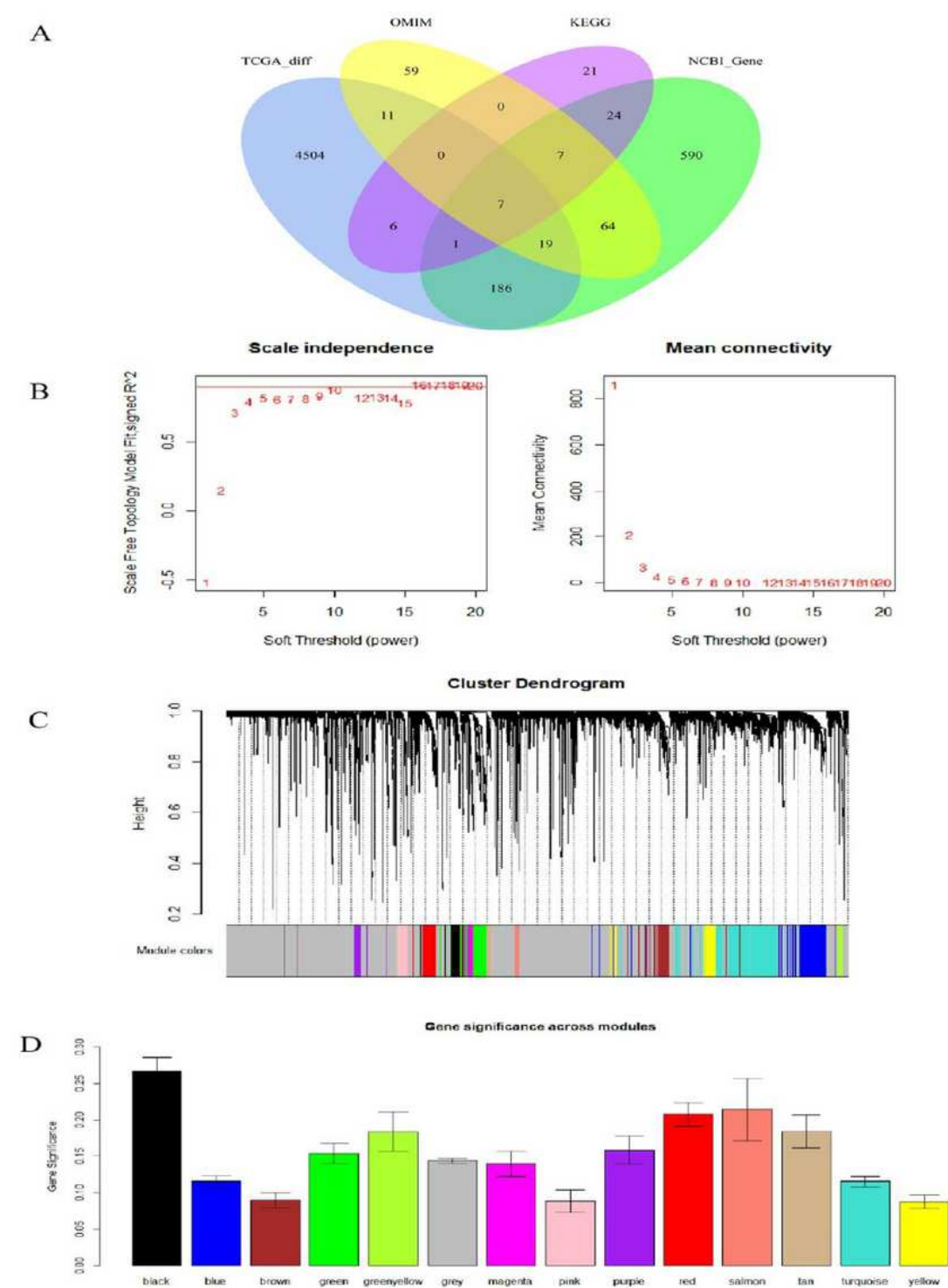


Figure 3

(A) CNV amplification distribution. (B) CNV missing distribution. (C) CNV unsupervised hierarchical clustering. (D) Comparison of different platform classification results. The red color indicates that the sample belongs to the subgroup corresponding to the horizontal axis, and the blue color indicates that

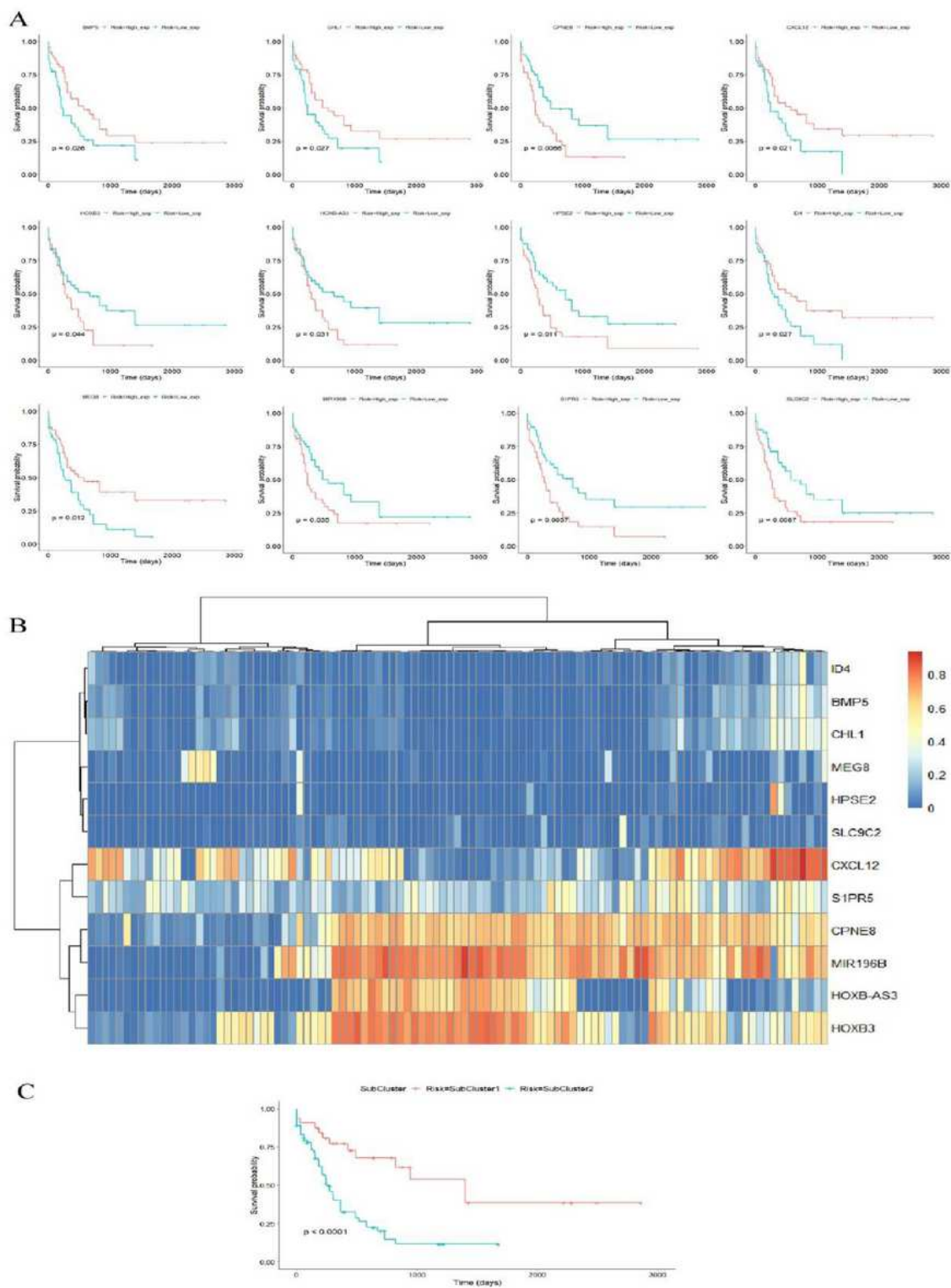
the sample is not in the corresponding subgroup of the horizontal axis. (E) High heterogeneity methylation site hierarchical clustering.



**Figure 4**

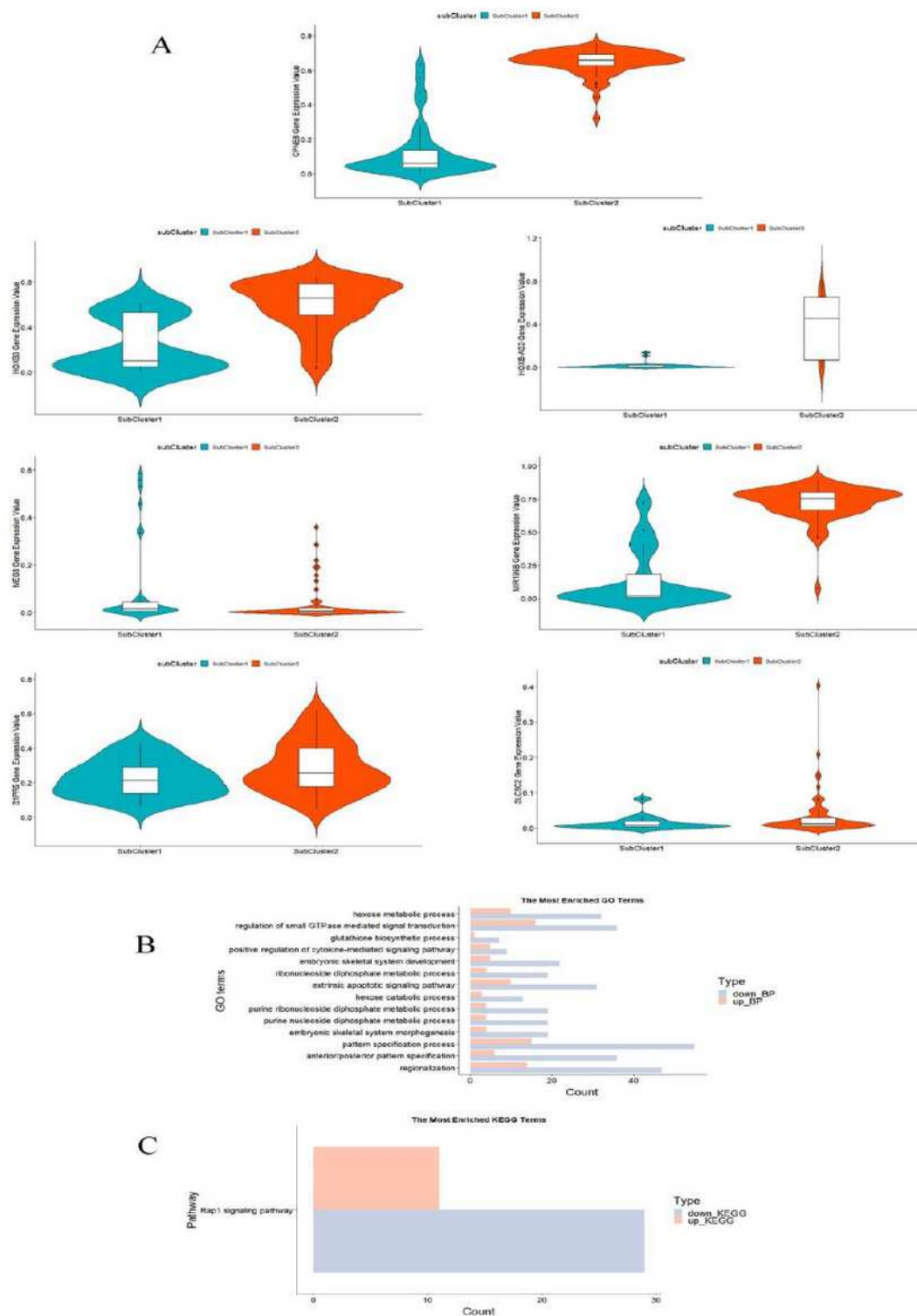
(A) AML-related genes from different platforms (B) Network soft threshold  $\beta$  determination. (C) Co-expression module. (D) GS mean distribution of different gene modules.





**Figure 5**

(A) Survival curve of 12 module core gene expression. (B) Prognosis-related gene expression in AML samples. (C) Differences in the prognosis of the two subgroups.



**Figure 6**

(A) Seven genes—HOXB-AS3—HOXB3—SLC9C2—CPNE8—MEG8—S1PR5—MIR196B—differentially expressed in two subgroups. (B) Differential gene GO enrichment analysis. (C) Differential gene KEGG enrichment analysis.

## Supplementary Files



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

- [SupplementaryTable21datExpr.txt](#)
- [SupplementaryTable26GOenrichBP.csv](#)
- [SupplementaryTable1clinicalinformation.docx](#)
- [SupplementaryTable22cluster.txt](#)
- [SupplementaryTable23driverGeneSubcluster.txt](#)
- [SupplementaryTable24driverGenedrug.txt](#)
- [SupplementaryTable25nrDEGmRNA.txt](#)
- [SupplementaryTable2sigMutSigCVgene.txt](#)
- [SupplementaryTable3Signaturedistribution.txt](#)
- [SupplementaryTable4alllesions.conf95.txt](#)
- [SupplementaryTable5CNVpeakcluster.txt](#)
- [SupplementaryTable6SNPCNVcluster.txt](#)
- [SupplementaryTable7methcv.txt](#)
- [SupplementaryTable8diffgene.txt](#)
- [SupplementaryTable9mutationtype.txt](#)
- [SupplementaryTable10mutationgene.txt](#)
- [SupplementaryTable11pathwaygene.txt](#)
- [SupplementaryTable12totalgenelist.txt](#)
- [SupplementaryTable13modulegene.txt](#)
- [SupplementaryTable14ModuleSignificance.txt](#)
- [SupplementaryTable15CytoscapeInputedgesblack.txt](#)
- [SupplementaryTable16blackdegree.txt](#)
- [SupplementaryTable17cvmiRNA.txt](#)
- [SupplementaryTable18miRNApivotsig.txt](#)
- [SupplementaryTable19hubgenecvmiRNA.txt](#)
- [SupplementaryTable20Survgene.txt](#)