

Comprehensive Gene Expression Analysis Reveals Factors that Influence the Progression, Occurrence and Development of AML-M4

Ying Qu

Second Affiliated Hospital of Qiqihar Medical University

Shuying Zhang

Second Affiliated Hospital of Qiqihar Medical University

Yanzhang Qu

Second Affiliated Hospital of Qiqihar Medical University

Heng Guo

Second Affiliated Hospital of Qiqihar Medical University

Suling Wang

Second Affiliated Hospital of Qiqihar Medical University

Xuemei Wang

Second Affiliated Hospital of Qiqihar Medical University

Tianjiao Huang

Second Affiliated Hospital of Qiqihar Medical University

Hong Zhou (✉ ZhouH_lab@163.com)

The Second Affiliated Hospital of Qiqihar Medical University <https://orcid.org/0000-0003-0156-8645>

Research article

Keywords: AML-M4, WGCNA, therapeutic targets, WT1, FLT3

Posted Date: December 3rd, 2019

DOI: <https://doi.org/10.21203/rs.2.12978/v2>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Acute myelomonocytic leukemia (M4), a special type of acute myeloid leukemia (AML), is a clonal malignant disease with poor prognosis and a low overall survival rate. Here, we aim to explore the molecular mechanism of AML-M4 development and progression using integrative bioinformatic analysis.

Methods

We used an integrative method to identify potential driven genes in AML-M4. We firstly identify DEGs using limma packages and annotated them by GO and KEGG. To avoid individual bias, GSEA was adopted to certify the results. Furthermore, we constructed the PPI network using WGCNA which was superimposed onto STRING database. We also assessed the correlation and mutation among hub genes to deeply explore the biological mechanism in AML-M4. Finally, we confirmed our results by experiments.

Results

The results show that FLT3, WT1 and TP53, which are involved in transcriptional misregulation were upregulated, while PPBP and CCR7, which regulate cytokine-cytokine receptor interaction, as well as CD24, which acts as a protein marker of AML, were downregulated. 12 hub genes were found through the TCGA an OncoPrint analysis, and the results also show that FLT3, CCR7 and MMP-9 can be potential targets for the detection and treatment of AML-M4.

Conclusion

The overall aim of this study was to identify critical pathways and genes associated with AML-M4, and to also provide potential therapeutic targets and new research ideas for the treatment and early detection of AML-M4.

Background

AML is one of the most common hematological malignancy, AML cells not only were proliferated and accumulated in bone marrow and other hematopoietic tissues, but also infiltrate other non-hematopoietic tissues and organs, as well as inhibit normal hematopoietic function due to uncontrolled proliferation. The clinical manifestations of AML are anemia, hemorrhage, fever caused by infection, enlargement of liver, spleen, lymph nodes and bone pain. [1]. A variety of chemotherapy regimens, biological agents, and stem cell transplantation are the main treatment options for AML [2–4]. However, chemotherapy drug toxicity may lead to acute and life-threatening complications. Compared with standard chemotherapy, allogeneic stem cell transplantation is a suitable method to reduce the risk of recurrence of AML, but also increase the risk of serious complications. Although continuously improved, the traditional method of treatment does not lead to a complete cure or an ideal duration of survival for AML-M4 in clinical practice [5]. Genomics, proteomics and bioinformatics analysis methods have been used to develop new

personalized treatment strategies, study of the functions of related biomolecules, and collection of information on emerging trends in genome matching of clinical data are effective methods to improve the prognosis of patients [6–8]. Although many studies have analyzed genome variation in AML, the association between genome variation and molecular mechanism of AML-M4 is still unclear. Therefore, a comprehensive study of AML-M4 was urgent.

In this study, the relationship between phenotype and functionally differentiated genes of AML-M4 was studied. The DEGs between 154 AML-M4 specimens and 69 normal non-leukemia specimens were studied through GO, KEGG and GSEA analysis of microarray data. The signaling pathways and physiological functional modules that are highly relevant to AML-M4 were analyzed, and the differentially expressed genes were verified and preliminarily analyzed. The objective of this study is to use a bioinformation analysis method to identify critical pathways and genes associated with AML-M4 and to provide potential targets and new research ideas for the treatment and early detection of AML-M4.

Methods

Sample collection

AML-M4 samples and non-leukemia samples were collected from 40 patients at The Second Affiliated Hospital of Qiqihar Medical University. The peripheral blood samples were obtained and then treated with liquid nitrogen followed stored at -80°C for further use. Three independent pathologists made the diagnosis of AML-M4 and assessed the samples using western blotting. Patients characteristics of the 40 samples were summarized in Supplementary table 2.

Microarray data source and pre-processing

The gene expression profiles of AML-M4 were obtained from three data sets, GSE6891, GSE10358 and GSE15061 of the NCBI GEO database, which are based on the Affymetrix HT HG-U133A and HG-U133A 2.0 Array. A total of 223 biochips from AML patients were analyzed, including 154 AML tumor samples and 69 non-leukemia samples. The raw data of the three datasets were downloaded from GEO, and the R package, Simpleaffy, was used for Affymetrix quality control and data analysis [9]. Annotations were made using gene symbols from each respective platform annotations. Then, using Perl scripts, data from all 223 samples were included into a united gene expression matrix. The mean value of gene expression was used in multiple probe sets with a single gene symbol. Normalization and batch correction were performed before the next analysis was conducted.

Functional analysis

The limma package was used to identify DEGs [10], then GO and KEGG analysis were performed using clusterProfiler [11]. An FDR value of <0.01 was considered to be statistically significant. GSEA was

utilized to deeply analyze the variation in biological functional and pathways between AML-M4 and non-leukemia samples [12]. The MSigDB gene sets file (c5.bp.v7.0.symbols.gmt) was selected for use as the reference gene set with permutations of 10^4 . The threshold was set at $P < 0.05$.

PPI Network analysis with WGCNA

The 1,084 DEGs between AML-M4 and control samples were retrieved to perform co-expression analysis which integrated WGCNA R package by default parameter [13] and PPI network from STRING database [14]. We used the dynamic tree cut package to lay out the co-expression analysis clusters with the minimum peak at 0.2 for every module [15]. The liability of every module was on the basis of eigengene, and each element in every module was assembled using Pearson correlation among DEGs and their interactors with a cutoff of r of at least 0.7 [16]. Furthermore, we also identified and characterized the system-level features of PPI network using topological overlapping matrix [17]. Ultimately, every unique module was annotated using clusterProfiler package [11] and visualized in Cytoscape [18].

Survival analysis

The clinical data of hub genes in AML prognosis were studied using the survival analysis. We downloaded the 187 TCGA-LAML expression profiles and clinical data from the TCGA database using TCGA biolinks [19]. The 187 AML specimen were divided into two group by expression value of every hub gene using the survminer package for the best separation. Furthermore, the association between every hub gene and overall survival (OS) was carried out using univariate Cox regression. To remove any hub gene that might not be independent factor in prognostic predictors, multivariate Cox regression was performed. A p adjust value of <0.05 was considered to be statistically significant and is indicated in the results.

Quantitative Real-time PCR (qRT-PCR)

In order to confirm the results of former analysis, we performed qPCR to verify them in mRNA level. The total RNA of 24 paired AML-M4 samples and non-leukemia samples were obtained using TRIzol reagent. The interest target genes were then assessed through qRT-PCR using a One-Step qPCR Kit (Invitrogen, USA). All RT reactions were run in a CFX Connect™ Real-Time System (BIO-RAD, USA), in accordance with the manufacturer's instructions. The results were reanalyzed using $2^{-\Delta\Delta CT}$ method using GAPDH as an internal control gene [20]. The primer sequences of the target genes are shown in Supplementary Table 1.

Western blotting analysis

The tissues were lysed, and total protein was quantified using the Pierce™ Detergent Compatible Bradford Assay Kit (Thermo Scientific). 20 µg of protein from each sample was used for SDS-PAGE. After transferring the sample onto a PVDF membrane, the blot was incubated with indicate antibodies. All antibodies were purchased from CST: CD11b (#45581, CST), c-KIT (#37805, CST), FLT3 (#3462, CST), CSF1R (#67455, CST), C/EBPB (#3082, CST) and GAPDH (#5174, CST).

Statistical analysis

All experiments involved in this article were replayed at least three times. The student's t test was used to compare the two groups analysis. Data are showed as mean SDs, except when indicated otherwise. A p value <0.05 was considered to be statistically significant.

Results

Identification of differential expression genes (DEGs) and functional variation

We used the limma package to screen out DEGs from 154 AML-M4 samples and 69 non-leukemia samples. The inclusion criteria of the DEGs was an absolute log₂ fold change value >1 and an FDR value <0.01 was used as the statistical filter conditions. 1,084 DEGs, containing 202 naturally upregulated genes and 882 downregulated genes, were obtained. The Heat map of DEGs are shown in Fig. 1A, while the volcano plot of the same is shown in Fig. 1B.

In order to farther analyze the trait of DEGs, we performed the functional variation analysis between the two groups using the cluster Profiler package. 107 GO terms were identified with an adjusted FDR of <0.01. We also used the GO SemSim package to remove rigmarole terms, only retain one unique term, which resulted in 49 unique GO terms [9]. The results of the GO analysis show that the most enriched GO targets were involved in neutrophil activation, neutrophil degranulation, neutrophil mediated immune response and leukocyte migration (Fig. 2A). The KEGG pathway enrichment analysis showed that transcriptional misregulation in cancer, hematopoietic cell lineage, cell cycle, and TH1, 2, 17 cell differentiation were the most significantly affected phases in AML-M4 (Fig. 2B). These results complemented the results of the GO enrichment analysis. To further check on the relationship between two phenotypes and functionally DEGs, we performed a GSEA analysis on gene expression matrix at whole transcription level. The transcripts of AML-M4 were found to be prominently associated with downregulated genes related to three pathways (Fig. 2C).

Integrative module analysis proved new regulatory mechanism

For the sake of comprehensive analysis of AML-M4, we simulated the kinetics of proteome changes as previously described [21]. We integrated WGCNA and PPI network analysis among DEGs to aggregate the relational proteins which were involved in analogous molecular functions or biological actions [22]. As a result, we identified 143 modules with the number of proteins in each ranging from 2 to 25 (Fig. 3B), and 122 of these modules were highly interconnected by their members (Fig. 3A). Each module was annotated using known functional terms or signaling pathways. We found that many modules, including module 3, 10, 15, 22 and 24 (Fig. 3C), were notably enriched in hematopoietic system related progression. In addition, module 27 was found to be involved in RNA splicing, module 38 was involved in autophagy, module 54 was involved in the regulation of transcription, while module 83 was involved in translational initiation (Fig. 3D). In summary, the progression of AML-M4 involves the balanced regulation and extensive reprogramming of mutually connected functional modules.

Comprehensive analysis of hub genes in AML

Based on the expression profile and clinical data of 187 AML samples from the TCGA database, the clinical outcomes of the hub genes were determined using the survival analysis. Univariate and multivariate Cox regression analyses were performed to explore the candidate gene which is a prognostic predictor applicable to AML patients. As a result, 6 of the 53 hub genes were found to be significantly associated with poor prognosis, as indicated by their positive or negative correlation with a higher risk by being upregulated or downregulated, respectively in AML (Fig. 4). To further deeply explore the biological mechanism in AML, we assessed the correlation and mutation among hub genes (Fig. 5, 6). These genes may be involved in the regulation of AML progression, and may also have the potential of being candidate biomarkers or drug targets for the disease.

Initial validation of transcriptome results through qRT-PCR

In order to confirm DEGs obtained, the total RNA of 24 paired AML-M4 samples were isolated for qRT-PCR validation. 40 target DEGs were selected, as shown in Fig. 7. The DEGs were successfully validated and showed a good correspondence with the results of the transcriptome analysis, indicating precise and reliable microarray results.

Protein level validation using western blotting

Finally, we confirmed DEGs at protein level. One of the most widely studied transcription factors in the hematopoiesis of AML-M4 is the leucine zipper ccaat-enhancer binding protein alpha (C/EBP α), which is mainly involved in bone marrow cell differentiation. In AML, patients often present with mutations or decreased expression of the C/EBP α gene [23]. CD11b is a member of the family of CAM integrins, which are mainly found in M1, M2, M4 and M5 subtypes of AML[24]. Its role is integral for the development of AML. In this paper, we emphasize the importance of downregulation of the expression of C/EBP α and

CD11b in the neutrophil related signaling pathway in AML, as shown in Fig. 8. FLT3, c-KIT and CSF1R are class III receptor tyrosine kinases that play a crucial role in hematopoiesis [25]. The pathogenesis of several malignant tumors are associated with the overexpression of CSF1R, c-KIT and FLT3 [26, 27]. In particular, the FLT3 and c-KIT genes have been intensively studied in childhood AML [28, 29]. In this study, FLT3, c-KIT and CSF1R protein expression levels were all found to be upregulated in AML, which is consistent with the results of the previous analysis, as shown in Fig. 8.

Discussion

AML is a clonal malignant disease with a poor prognosis and low overall rate of survival. It originates from hematopoietic bone marrow primordial cells. Immature leukocytes grow rapidly and it effect the production of normal blood cells. The median survival time of AML patients is only 5–10 months [1].

Acute myelomonocytic leukemia (M4) and acute monocytic leukemia (M5) are special types of AML and they are manifested in cytogenetics, immunology and clinical processes [30]. The overall survival rate (OS) as a result of traditional treatment (chemotherapy and stem cell transplantation) for AML-M4 is low, and chemotherapy is easily accompanied by complications, while stem cell transplantations are high in cost, with a risk of being rejected [31]. The molecular mechanism of AML-M4 development and progression is not fully understood, and it is particularly important to find new targets and strategies for individualized therapy.

In recent years, genomic bioinformatics analysis of differentially expressed genes in pathological samples has been used to study potential targets to provide an early theoretical basis for the treatment of AML-M4 [32–34]. In this study, DEGs between 154 AML-M4 specimens and 69 normal specimens were used to analyze GO, KEGG and GSEA. The results of the GO analysis show that neutrophil activation, neutrophil degranulation, neutrophil activation involved in immune response, neutrophil mediated immunity and leukocyte migration were the most abundant targets. The results of the GSEA and KEGG analyses show that transcriptional misregulation in cancer and the related signaling pathways of the hematopoietic cell lineage are important factors that influence the occurrence and development of AML-M4. The results of GO, KEGG and GSEA analyses are consistent in showing that neutrophil-related signaling pathways, immune responses and transcriptional misregulations are the key factors related to AML-M4. The results of PPI network analyses are consistent with the results of the GO, KEGG and GSEA analyses. The results indicated that the dynamic balance regulation of functional modules plays important role in the progression of AML-M4, such as leukocyte migration, T cell receptor signaling pathway, neutrophil mediated immunity, neutrophil activation, platelet degranulation, RNA splicing, autophagy, regulation of transcription and translational initiation.

In Figure 6. Only two genes, DNMT3A and FLT3, have a mutation rate greater than 10% in TCGA-LAML cohort, and DNMT3A was not differentially expressed. Both the GSEA analysis and TCGA analysis show that FLT3 is upregulated in AML-M4, which shows that it plays an important role in the development of AML-M4. FLT3 is considered to be a target of treatment for AML, and at present, the development of

clinical targets related with FLT3 is very active. FLT3 is characterized by the presence of five immunoglobulin-like motifs within their extracellular section. These motifs are exclusively expressed in hematopoietic cells [35]. FLT3 mutations occur as secondary events during AML clonal evolution [36]. FLT3-ITD mutation has a negative impact on the prognosis of AML, only a minority of patients with FLT3-ITD mutation in leukemic blasts are cured through chemotherapy.

The GSEA analysis found that CCR7 is downregulated in AML. CCR7 is one of the most important chemokine receptors for adaptive immune cell migration. Additionally, CCR7 also mediates the expression of signaling molecules, such as Th1 and Th2, that can affect the homeostasis, activation and polarization of T cells [37]. Interestingly, the KEGG analysis found that Th1, 2, 17 cell differentiation is one of the most significantly affected phased in AML-M4, It has been reported that the dysregulation of Th subset cytokines contribute to the pathogenesis of AML, while accumulating evidence also indicates that many diseases, including hematological malignancies, are involved the imbalance ratio of Th1/Th2. The Th1/Th2 balance plays an important role in normal immunity [38]. The study by Georgios Leandros Moschovakis elaborates that early T cell activation and Th1 differentiation may be promoted by activation of CCR7 signaling, while the Th2 polarization was increased after the CCR7 signaling was inhibited [39]. Xu reported that decreased CCR7 expression results in an increase of Th2 responses and increased expression of IgE and IL-13 in the lung [40]. It has been established that Th17 cells participate in certain autoimmune diseases and tumors [41–43]. Le Di Eu et al. have demonstrated that T cells had an impaired ability to form immune synapses that are critical for T cell activation [44]. Following their activation, the expression of CCR7 was upregulated by dendritic cells (DCs) [44]. It also reported that the apoptosis in mature DCs was inhibited PI3K/Akt pathway by CCR7 signaling. [45]. CCR7 signaling has also been reported to inhibit apoptosis in mature DCs through the PI3K/Akt pathway [46]. The balanced regulation between CCR7 and DC impact T cell activation. CCR7 may be used as a treatment target of AML-M4 through the regulation of DC and T cell activation. CCR7 may also be a therapeutic target for AML-M4 through its influence on T cell activation, polarization and the imbalance in the Th1/Th2 ratio and the role of CCR7 signaling in T cell homeostasis remains to be further verified.

We also found that most PPI modules are related to CCR7 and MMP-9, indicating that MMP-9 plays an important role in the occurrence and development of leukemia. MMP-9 belongs to matrix metalloproteinase (MMP) family and plays important role in cell growth, inflammation, angiogenesis and migration [47]. The Amir-Foroushani research team confirmed that MMP-9 is inadequately expressed and highly methylated in AML [48]. Many studies have shown that MMP-9 can be further applied for the detection, prognosis or treatment of AML. For example, Amir-Foroushani reported of a method to distinguish AML from MDS using MMP-9 in 2017 [48]. In our analysis, MMP-9 downregulation was found to be correlated with a higher risk of AML-M4. MMP-9 participates in neutrophil activation and leukocyte migration module, which is related with CXCL12, ITGAM, neutrophil mediated immunity and platelet threshing. It has been suggested that the expression of MMP-9 may be related to the progression and invasiveness of AML-M4. As an antimicrobial gene, Chemokine ligand 12 (CXCL12) is involved in many diverse cellular functions, including embryogenesis, immune surveillance, inflammation response, tissue homeostasis, and tumor growth and metastasis. Relationship of CXCL12 and MMP-9 has been

demonstrated to be important for normal hematopoietic cell migration in *in vitro* studies of normal human hematopoietic cells [49]. MMP-2 and MMP-9 in bone marrow cells can be induced by CXCL12 and then increase the cell migration [50]. MMP-9/MMP-2/CXCL12 plays as regulators in bone marrow cell migration and infiltration. CXCL12/ CXCL12 receptor together was reported as important regulators of AML cell migration in the bone marrow [51]. Several MMPs involved in the processing of CXCL12, MMP-2 and MMP-9 also can inhibit the chemoattractant activity in normal CD34+ HSCs and immature AML cell subsets [52]. Extracellular matrix protein 1 (ECM1) is one kind of secreted glycoprotein which may promote tumor development [53]. MMPs have the ability to degrade essentially all components of ECM. By breaking down ECM, MMPs may remove physical barriers, thus allowing cells to migrate and invade other tissues [54]. ECM1 has been shown to be directly associated with EGFR and is followed by EGF dependent EGFR and ERK activation. EGF was in turn shown to activate ERK signaling at transcription level, through the activation of transcription factor MMP-9 expression that proceeds downstream of these events [53]. ITGAM encodes CD11b and is involved in many inflammatory biological processes, the remarkable feature of ITGAM is its ability to combine various ligands. MMP-9 is associated with ITGAM and participates in neutrophil-mediated immunity. CXCL8 is activated by MMP-9 through proteolytic cleavage [55]. It was reported that relationship between these two mediators can regulate the development of human AML [56]. CXCL8, which can be potential therapeutic target, may regulate the angiogenesis of AML in bone marrow [57]. MMP-9 is also anti-angiogenetic; intertumoral delivery of MMP-9 decreases tumor growth, in particular angiogenesis in breast cancer mice [58]. But there are few reports of its involvement in leukemia. Therefore, it is necessary to understand the molecular mechanism of certain molecules, for which further research is needed.

Conclusion

Overall, we studied the factors that influence the occurrence and development of AML-M4 at the genome level using bioinformatics methods, qRT-PCR and western blotting assay was used to test verify those DEGs. The results show that pathways involved in immune response, neutrophil mediated immunity, leukocyte migration and transcriptional misregulation may be important factors that affect the occurrence and development of AML-M4. These results also show that FLT3, c-KIT and MMP-9 may act as potential targets for the detection and treatment of AML-M4 and further in-depth research is needed. In general, the results of this study may help to identify critical pathways and genes associated with AML-M4 and provide potential targets and new research ideas for the treatment and early detection of AML-M4.

Abbreviations

AML: Acute Myelocytic Leukemia; DEGs: Differential Expression Genes; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; WGCNA: Weighted Gene Co-expression Network Analysis; PPI: Protein-Protein Interaction; GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas.

Declarations

Availability of data and materials

The microarray data of AML-M4 samples were obtained from GEO database with the accession numbers GSE6891, GSE10358 and GSE15061. The RNA sequencing data, mutation data and clinical data of AML were downloaded from TCGA data portal (<https://portal.gdc.cancer.gov>). The mRNA expression data for correlation analysis of AML were obtained from Oncomine dataset (www.oncomine.org).

Acknowledgements

Not applicable.

Funding

This study was supported by Research Fund Project of Qiqihar Medical University (QY2015L-08). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

Authors' contributions

HZ designed the study and write the manuscript. YQ1 analyzed the data. SZ, YQ2 and HG collected the data. YQ1, SW, XW and TH performed the experiment. All authors read and approved the manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The study protocol was approved by the Ethics Board of the Second Affiliated Hospital of Qiqihar Medical University. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

References

1. Hansrivijit P, RP Gale, J Barrett, and SO Ciurea. Cellular therapy for acute myeloid Leukemia - Current status and future prospects. *Blood Rev.* 2019;
2. Liu Y, Z Cheng, Y Pang, L Cui, T Qian, L Quan, H Zhao, J Shi, X Ke, and L Fu. Role of microRNAs, circRNAs and long noncoding RNAs in acute myeloid leukemia. *J Hematol Oncol.* 2019;12(1):51.
3. Zhou JD, XX Li, TJ Zhang, ZJ Xu, ZH Zhang, Y Gu, XM Wen, W Zhang, RB Ji, ZQ Deng, J Lin, and J Qian. MicroRNA-335/ID4 dysregulation predicts clinical outcome and facilitates leukemogenesis by activating PI3K/Akt signaling pathway in acute myeloid leukemia. *Aging (Albany NY).* 2019;
4. Liu L, J Ma, L Qin, X Shi, H Si, and Y Wei. Interleukin-24 enhancing antitumor activity of chimeric oncolytic adenovirus for treating acute promyelocytic leukemia cell. *Medicine (Baltimore).* 2019;98(22):e15875.
5. Manola KN, F Panitsas, S Polychronopoulou, A Daraki, M Karakosta, C Stavropoulou, G Avgerinou, E Hatzipantelis, G Pantelias, C Sambani, and M Pagoni. Cytogenetic abnormalities and monosomal karyotypes in children and adolescents with acute myeloid leukemia: correlations with clinical characteristics and outcome. *Cancer Genet.* 2013;206(3):63–72.
6. Bret C, E Viziteu, A Kassambara, and J Moreaux. Identifying high-risk adult AML patients: epigenetic and genetic risk factors and their implications for therapy. *Expert Rev Hematol.* 2016;9(4):351–60.
7. Wang BH, YH Li, and L Yu. Genomics-based Approach and Prognostic Stratification Significance of Gene Mutations in Intermediate-risk Acute Myeloid Leukemia. *Chin Med J (Engl).* 2015;128(17):2395–403.
8. Cai SF and RL Levine. Genetic and epigenetic determinants of AML pathogenesis. *Semin Hematol.* 2019;56(2):84–89.
9. Yu G, F Li, Y Qin, X Bo, Y Wu, and S Wang. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics.* 2010;26(7):976–978.
10. Ritchie ME, B Phipson, D Wu, Y Hu, CW Law, W Shi, and GK Smyth. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
11. Yu G, LG Wang, Y Han, and QY He. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–7.
12. Subramanian A, P Tamayo, VK Mootha, S Mukherjee, BL Ebert, MA Gillette, A Paulovich, SL Pomeroy, TR Golub, ES Lander, and JP Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545–50.
13. Langfelder P and S Horvath. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9(559).

14. Szklarczyk D, A Franceschini, S Wyder, K Forslund, D Heller, J Huerta-Cepas, M Simonovic, A Roth, A Santos, KP Tsafou, M Kuhn, P Bork, LJ Jensen, and C von Mering. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43(Database issue):D447–52.
15. Langfelder P, B Zhang, and S Horvath. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics.* 2008;24(5):719–20.
16. Voineagu I, X Wang, P Johnston, JK Lowe, Y Tian, S Horvath, J Mill, RM Cantor, BJ Blencowe, and DH Geschwind. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature.* 2011;474(7351):380–4.
17. Ravasz E, AL Somera, DA Mongru, ZN Oltvai, and AL Barabasi. Hierarchical organization of modularity in metabolic networks. *Science.* 2002;297(5586):1551–5.
18. Shannon P, A Markiel, O Ozier, NS Baliga, JT Wang, D Ramage, N Amin, B Schwikowski, and T Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504.
19. Colaprico A, TC Silva, C Olsen, L Garofano, C Cava, D Garolini, TS Sabedot, TM Malta, SM Pagnotta, I Castiglioni, M Ceccarelli, G Bontempi, and H Noushmehr. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 2016;44(8):e71.
20. Livak KJ and TD Schmittgen. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods.* 2001;25(4):402–8.
21. Tan H, K Yang, Y Li, TI Shaw, Y Wang, DB Blanco, X Wang, JH Cho, H Wang, S Rankin, C Guy, J Peng, and H Chi. Integrative Proteomics and Phosphoproteomics Profiling Reveals Dynamic Signaling Networks and Bioenergetics Pathways Underlying T Cell Activation. *Immunity.* 2017;46(3):488–503.
22. Jansen R, D Greenbaum, and M Gerstein. Relating whole-genome expression data with protein-protein interactions. *Genome Res.* 2002;12(1):37–46.
23. Avellino R and R Delwel. Expression and regulation of C/EBPalpha in normal myelopoiesis and in malignant transformation. *Blood.* 2017;129(15):2083–2091.
24. Paietta E. Expression of cell-surface antigens in acute promyelocytic leukaemia. *Best Practice & Research Clinical Haematology.* 2003;16(3):17.
25. JT R. Class III receptor tyrosine kinases: role in leukaemogenesis. *Br J Haematol.* 2002;116(4):14.
26. Fassunke J, MC Blum, HU Schildhaus, M Zapatka, B Brors, H Kunstlinger, R Buttner, E Wardelmann, and S Merkelbach-Bruse. qPCR in gastrointestinal stromal tumors: Evaluation of reference genes and expression analysis of KIT and the alternative receptor tyrosine kinases FLT3, CSF1-R, PDGFRB, MET and AXL. *BMC Mol Biol.* 2010;11(100).

- 27.Kacinski BM. CSF-1 and its receptor in breast carcinomas and neoplasms of the female reproductive tract. *Mol Reprod Dev.* 1997;46(1):71-4.
- 28.Liang DC, LY Shih, IJ Hung, CP Yang, SH Chen, TH Jaing, HC Liu, and WH Chang. Clinical relevance of internal tandem duplication of the FLT3 gene in childhood acute myeloid leukemia. *Cancer.* 2002;94(12):3292-8.
- 29.Boissel N, H Leroy, B Brethon, N Philippe, S de Botton, A Auvrignon, E Raffoux, T Leblanc, X Thomas, O Hermine, B Quesnel, A Baruchel, G Leverger, H Dombret, C Preudhomme, A Acute Leukemia French, and G Leucemies Aigues Myeloblastiques de l'Enfant Cooperative. Incidence and prognostic impact of c-Kit, FLT3, and Ras gene mutations in core binding factor acute myeloid leukemia (CBF-AML). *Leukemia.* 2006;20(6):965-70.
- 30.Stam RW, I Hubeek, ML den Boer, JG Buijs-Gladdines, U Creutzig, GJ Kaspers, and R Pieters. MLL gene rearrangements have no direct impact on Ara-C sensitivity in infant acute lymphoblastic leukemia and childhood M4/M5 acute myeloid leukemia. *Leukemia.* 2006;20(1):179-82.
- 31.Vaughn JE, V Shankaran, and RB Walter. Trends in Clinical Benefits and Costs of Novel Therapeutics in AML: at What Price Does Progress Come? *Curr Hematol Malig Rep.* 2019;
- 32.Pollyea DA. New drugs for acute myeloid leukemia inspired by genomics and when to use them. *Hematology Am Soc Hematol Educ Program.* 2018;2018(1):45-50.
- 33.Bullinger L, K Dohner, and H Dohner. Genomics of Acute Myeloid Leukemia Diagnosis and Pathways. *J Clin Oncol.* 2017;35(9):934-946.
- 34.Papaemmanuil E, H Dohner, and PJ Campbell. Genomic Classification in Acute Myeloid Leukemia. *N Engl J Med.* 2016;375(9):900-1.
- 35.Blume-Jensen P and T Hunter. Oncogenic kinase signalling. *Nature.* 2001;411(6835):355-65.
- 36.Shlush LI, S Zandi, A Mitchell, WC Chen, JM Brandwein, V Gupta, JA Kennedy, AD Schimmer, AC Schuh, KW Yee, JL McLeod, M Doedens, JJ Medeiros, R Marke, HJ Kim, K Lee, JD McPherson, TJ Hudson, HP-LGP Consortium, AM Brown, F Yousif, QM Trinh, LD Stein, MD Minden, JC Wang, and JE Dick. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature.* 2014;506(7488):328-33.
- 37.Sun YX, HL Kong, CF Liu, S Yu, T Tian, DX Ma, and CY Ji. The imbalanced profile and clinical significance of T helper associated cytokines in bone marrow microenvironment of the patients with acute myeloid leukemia. *Hum Immunol.* 2014;75(2):113-8.
- 38.Tian T, S Yu, L Liu, F Xue, C Yuan, M Wang, C Ji, and D Ma. The Profile of T Helper Subsets in Bone Marrow Microenvironment Is Distinct for Different Stages of Acute Myeloid Leukemia Patients and Chemotherapy Partly Ameliorates These Variations. *PLoS One.* 2015;10(7):e0131761.

39. Moschovakis GL and R Forster. Multifaceted activities of CCR7 regulate T-cell homeostasis in health and disease. *Eur J Immunol.* 2012;42(8):1949–55.
40. Xu B, K Aoyama, M Kusumoto, A Matsuzawa, EC Butcher, SA Michie, T Matsuyama, and T Takeuchi. Lack of lymphoid chemokines CCL19 and CCL21 enhances allergic airway inflammation in mice. *Int Immunol.* 2007;19(6):775–84.
41. Bettelli E, M Oukka, and VK Kuchroo. T(H)–17 cells in the circle of immunity and autoimmunity. *Nat Immunol.* 2007;8(4):345–50.
42. Tesmer LA, SK Lundy, S Sarkar, and DA Fox. Th17 cells in human disease. *Immunol Rev.* 2008;223(87–113).
43. Cua DJ, J Sherlock, Y Chen, CA Murphy, B Joyce, B Seymour, L Lucian, W To, S Kwan, T Churakova, S Zurawski, M Wiekowski, SA Lira, D Gorman, RA Kastelein, and JD Sedgwick. Interleukin–23 rather than interleukin–12 is the critical cytokine for autoimmune inflammation of the brain. *Nature.* 2003;421(6924):744–8.
44. Le Dieu R, DC Taussig, AG Ramsay, R Mitter, F Miraki-Moud, R Fatah, AM Lee, TA Lister, and JG Gribben. Peripheral blood T cells in acute myeloid leukemia (AML) patients at diagnosis have abnormal phenotype and genotype and form defective immune synapses with AML blasts. *Blood.* 2009;114(18):3909–16.
45. AD L. The role of chemokines in linking innate and adaptive immunity. *CURR OPIN IMMUNOL.* 2002;14(1):7.
46. Noelia Sánchez-Sánchez LR-B, Gonzalo de la Rosa, Amaya Puig-Kröger, Julio García-Bordas, Daniel Martín, Natividad Longo, Antonio Cuadrado, Carlos Cabañas, Angel L. Corbí, Paloma Sánchez-Mateos and José Luis Rodríguez-Fernández. Chemokine receptor CCR7 induce intracellular signaling that inhibits apoptosis of mature dendritic cells. *blood.* 2004;104(3):7.
47. Bauvois B. New facets of matrix metalloproteinases MMP–2 and MMP–9 as cell surface transducers: outside-in signaling and relationship to tumor progression. *Biochim Biophys Acta.* 2012;1825(1):29–36.
48. Foroushani A, R Agrahari, R Docking, L Chang, G Duns, M Hudoba, A Karsan, and H Zare. Large-scale gene network analysis reveals the significance of extracellular matrix pathway and homeobox genes in acute myeloid leukemia: an introduction to the Pigengene package and its applications. *BMC Med Genomics.* 2017;10(1):16.
49. Lane WJ, S Dias, K Hattori, B Heissig, M Choy, SY Rabbany, J Wood, MA Moore, and S Rafii. Stromal-derived factor 1-induced megakaryocyte migration and platelet production is dependent on matrix metalloproteinases. *Blood.* 2000;96(13):4152–9.

50. Janowska-Wieczorek A, LA Marquez, A Dobrowsky, MZ Ratajczak, and ML Cabuhat. Differential MMP and TIMP production by human marrow and peripheral blood CD34(+) cells in response to chemokines. *Exp Hematol.* 2000;28(11):1274–85.
51. Kittang AO, K Hatfield, K Sand, H Reikvam, and O Bruserud. The chemokine network in acute myelogenous leukemia: molecular mechanisms involved in leukemogenesis and therapeutic implications. *Curr Top Microbiol Immunol.* 2010;341(149–72).
52. McQuibban GA, GS Butler, JH Gong, L Bendall, C Power, I Clark-Lewis, and CM Overall. Matrix metalloproteinase activity inactivates the CXC chemokine stromal cell-derived factor-1. *J Biol Chem.* 2001;276(47):43503–8.
53. Lee KM, K Nam, S Oh, J Lim, YP Kim, JW Lee, JH Yu, SH Ahn, SB Kim, DY Noh, T Lee, and I Shin. Extracellular matrix protein 1 regulates cell proliferation and trastuzumab resistance through activation of epidermal growth factor signaling. *Breast Cancer Res.* 2014;16(6):479.
54. Hatfield KJ, H Reikvam, and O Bruserud. The crosstalk between the matrix metalloprotease system and the chemokine network in acute myeloid leukemia. *Curr Med Chem.* 2010;17(36):4448–61.
55. Van den Steen PE, P Proost, A Wuyts, J Van Damme, and G Opdenakker. Neutrophil gelatinase B potentiates interleukin-8 tenfold by aminoterminal processing, whereas it degrades CTAP-III, PF-4, and GRO-alpha and leaves RANTES and MCP-2 intact. *Blood.* 2000;96(8):2673–81.
56. Reikvam H, KJ Hatfield, AM Oyan, KH Kalland, AO Kittang, and O Bruserud. Primary human acute myelogenous leukemia cells release matrix metalloproteases and their inhibitors: release profile and pharmacological modulation. *Eur J Haematol.* 2010;84(3):239–51.
57. Hatfield KJ, AM Olsnes, BT Gjertsen, and O Bruserud. Antiangiogenic therapy in acute myelogenous leukemia: targeting of vascular endothelial growth factor and interleukin 8 as possible antileukemic strategies. *Curr Cancer Drug Targets.* 2005;5(4):229–48.
58. Leifler KS, S Svensson, A Abrahamsson, C Bendrik, J Robertson, J Gauldie, AK Olsson, and C Dabrosin. Inflammation induced by MMP-9 enhances tumor regression of experimental breast cancer. *J Immunol.* 2013;190(8):4420–30.

Figures

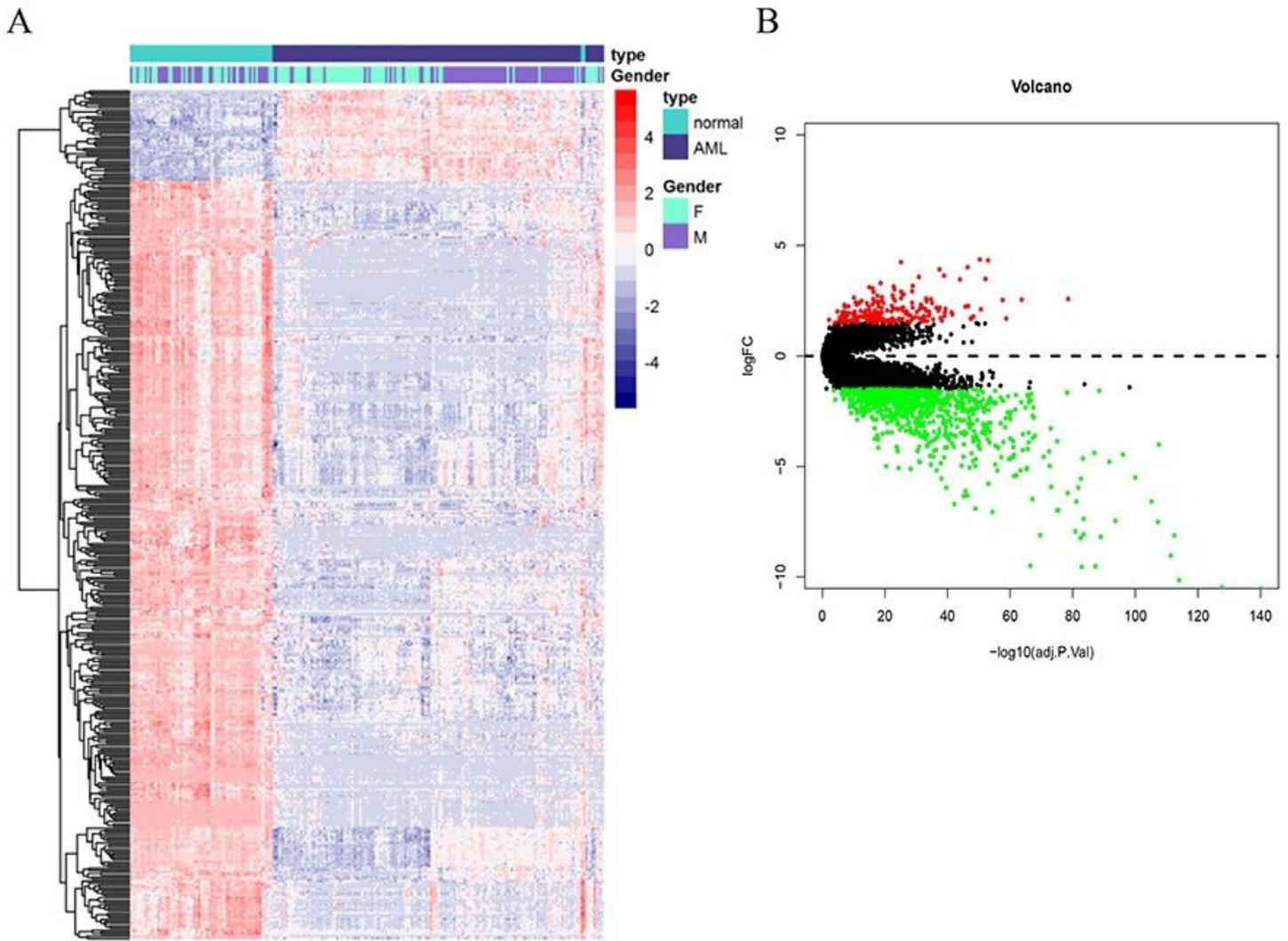


Figure 1

Identification of differential expression genes (DEGs). A. Heatmap of the DEGs between 154 AML-M4 tumor samples and 69 non-leukemia samples, based on read counts at transcription level. The inclusion criteria for the DEGs was an absolute \log_2FC value of >1 and an adjusted p value of <0.01 , as the statistical filter conditions. 1,084 DEGs, including 202 significantly upregulated genes and 882 significantly downregulated genes, were found; B. The volcano plot of the 1,084 DEGs between 154 AML-M4 tumor samples and 69 non-leukemia samples.

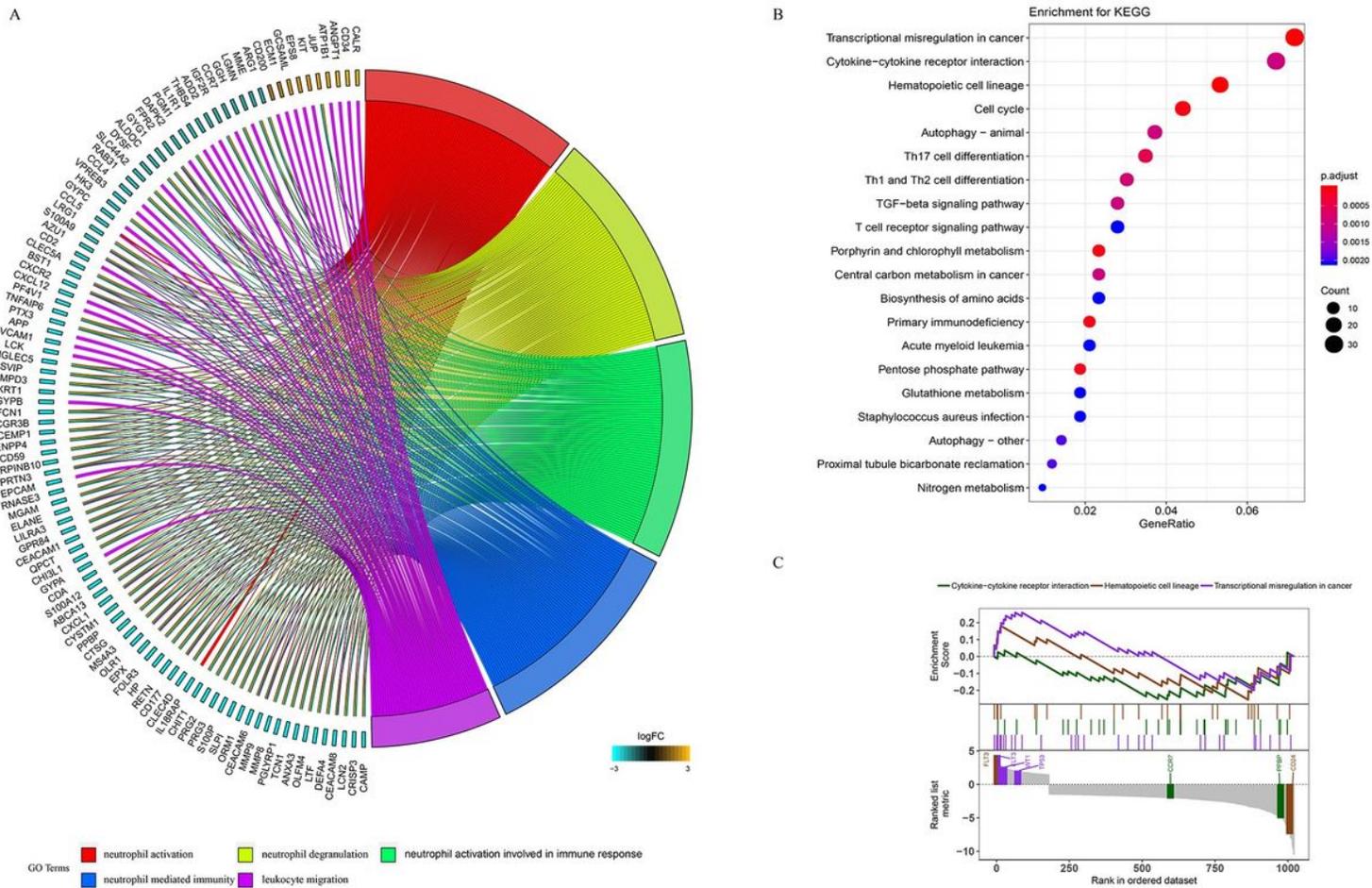


Figure 2

Functional analysis. A. The most enriched GO targets were involved in neutrophil activation, neutrophil degranulation, neutrophil activation involved in immune response, neutrophil mediated immunity and leukocyte migration; B. The KEGG pathway enrichment analysis showed that transcriptional misregulation in cancer, hematopoietic cell lineage, cell cycle, and TH1, 2, 17 cell differentiation were the most significantly affected phases in AML-M4; C. The GSEA results of AML-M4 patients and non-leukemia tissues performed on all genes at transcription level.

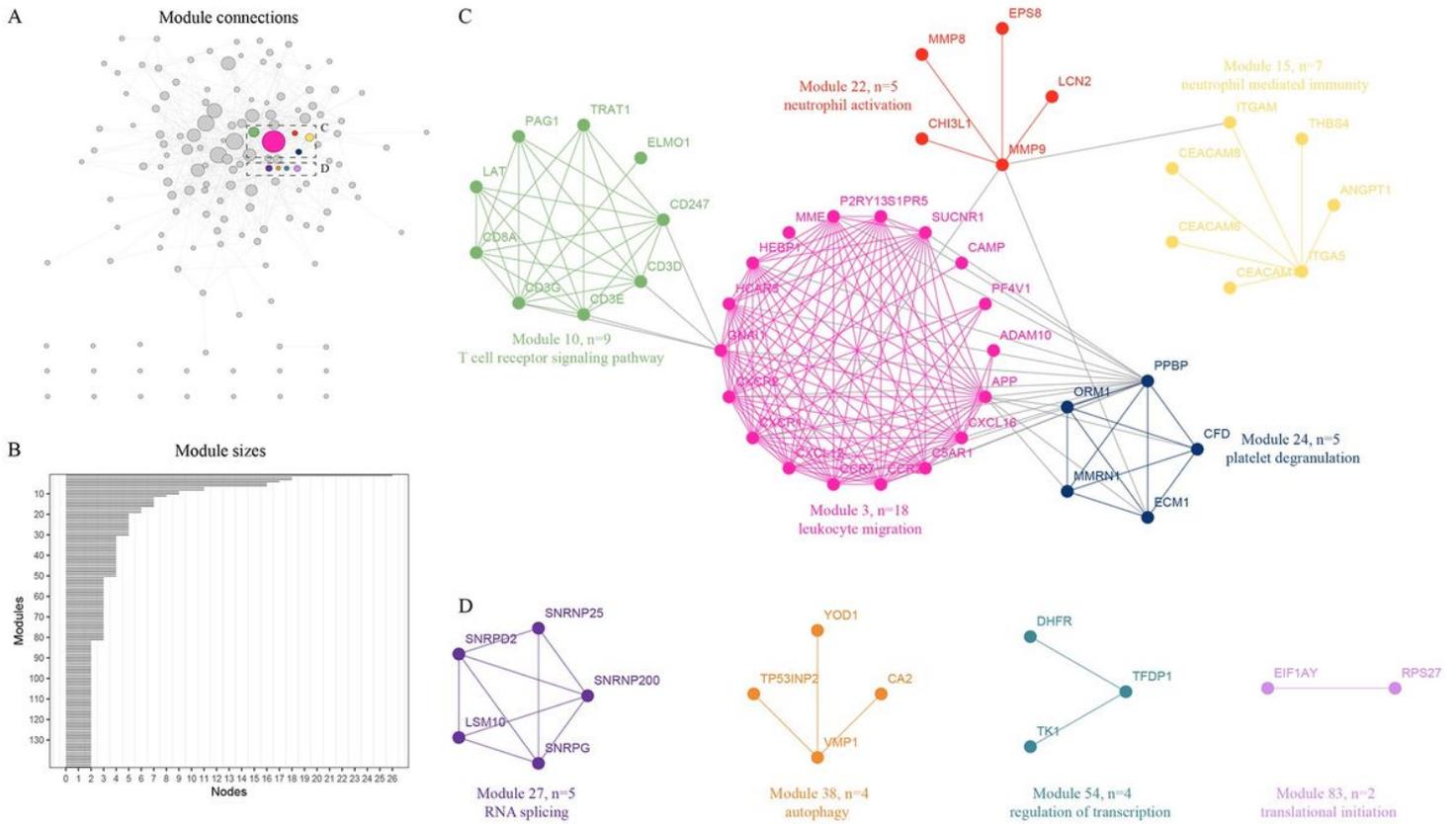


Figure 3

PPI network analysis. A. 122 out of the 143 modules had a high degree of interconnection among their members; B. 143 modules were identified; C. Module 3, 10, 15, 22 and 24 were notably enriched in hematopoietic system related progression; D. Module 27 was found to be involved in RNA splicing, module 38 was found to be involved in autophagy, module 54 was found to be involved in the regulation of transcription, while module 83 was found to be involved in translational initiation.

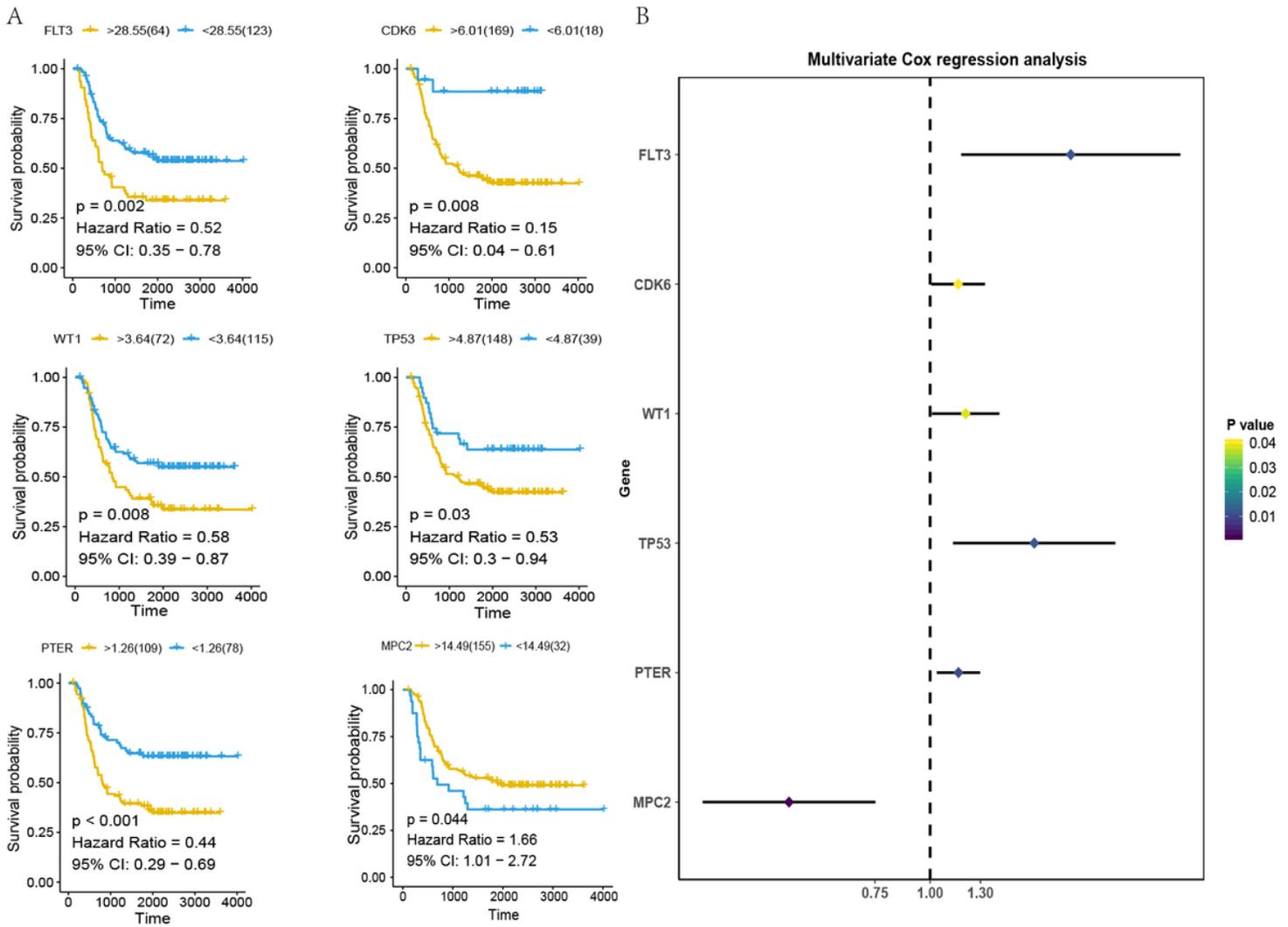


Figure 4

Survival analysis. A. TCGA database was used to select 612 hub genes which were correlated with a higher risk of AML, based on the expression profiles and clinical data of 187 AML samples, the genes were shown by their upregulation or downregulation, respectively, which were positively or negatively correlated with a higher risk of AML, as shown by their upregulation or downregulation, respectively, in AML. B. Hazard ratios of 6 hub genes with survival performed by multivariate Cox regression analysis.

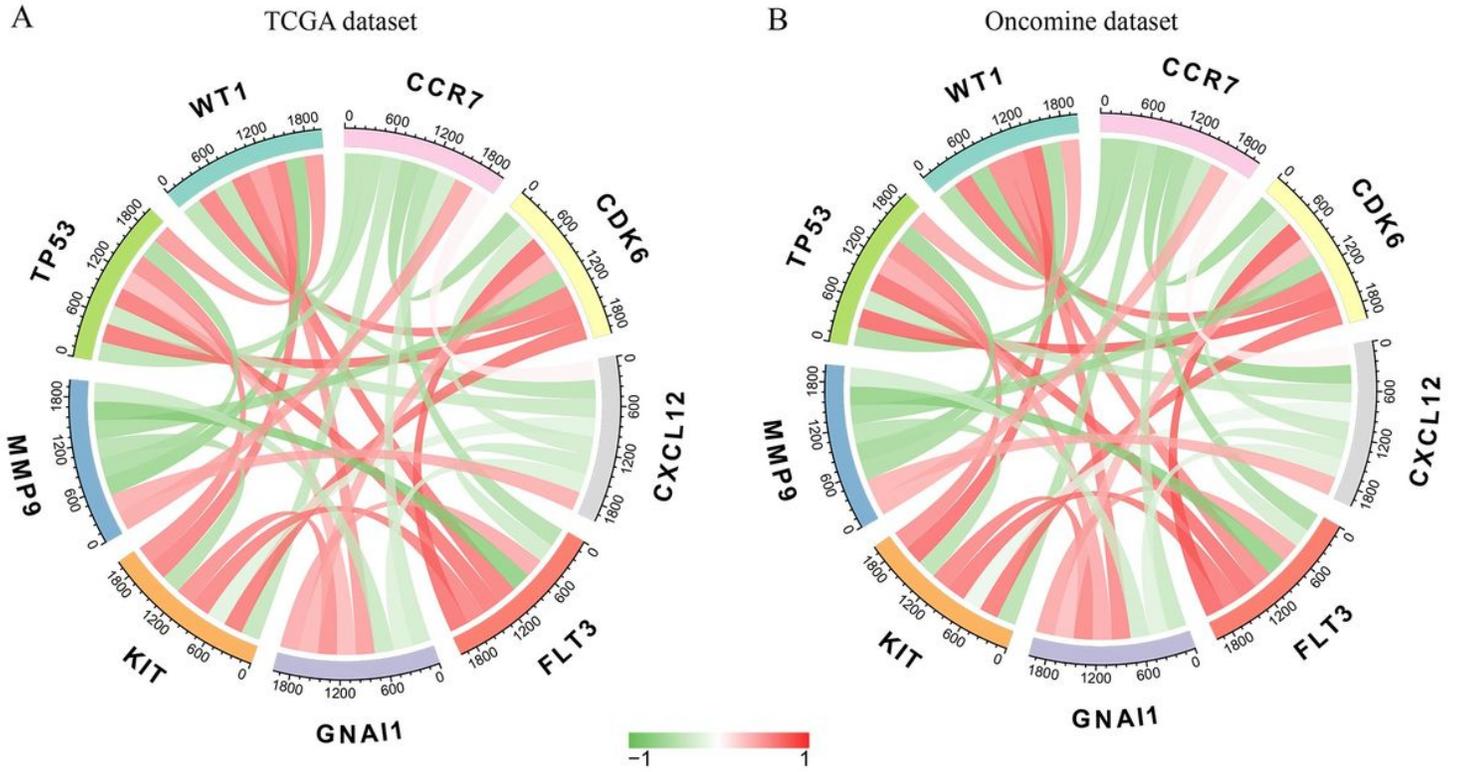


Figure 5

Correlation among hub genes. A. Correlation in TCGA dataset. B. Correlation in Oncomine dataset.

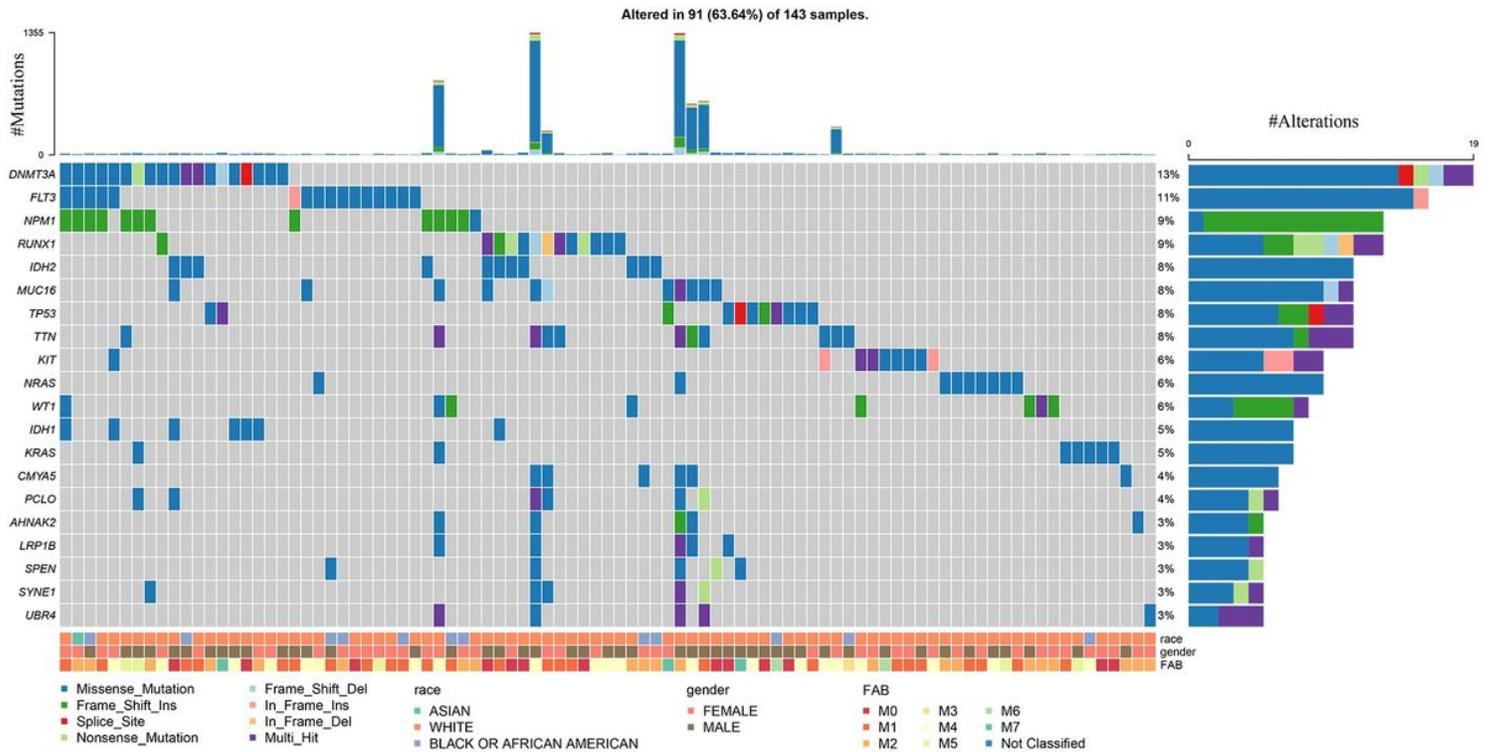


Figure 6

The Genomic landscape of AML and mutational signatures. The top panel shows individual sample mutation rates, the bottom panel details race, gender and FAB for 143 AML patients.

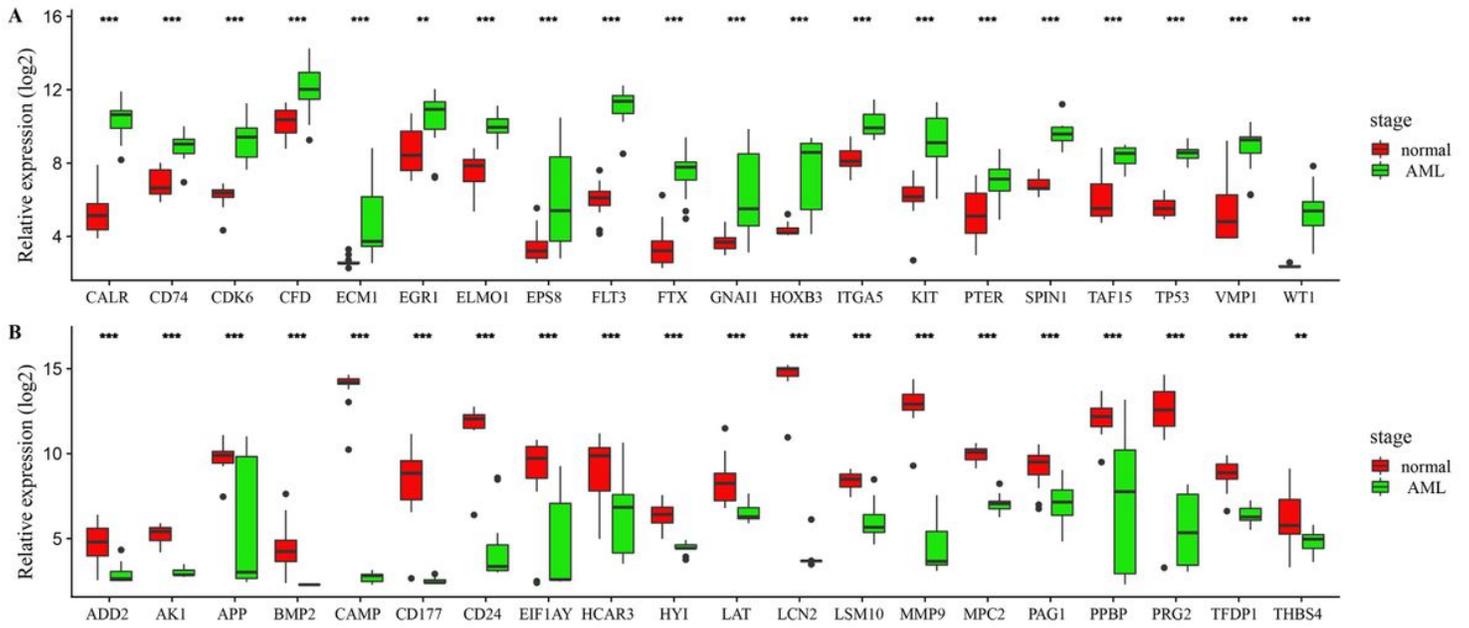


Figure 7

Validation by qPCR. The Validation of DEGs in 40 AML-M4 and normal samples were validated by qRT-PCR assay. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

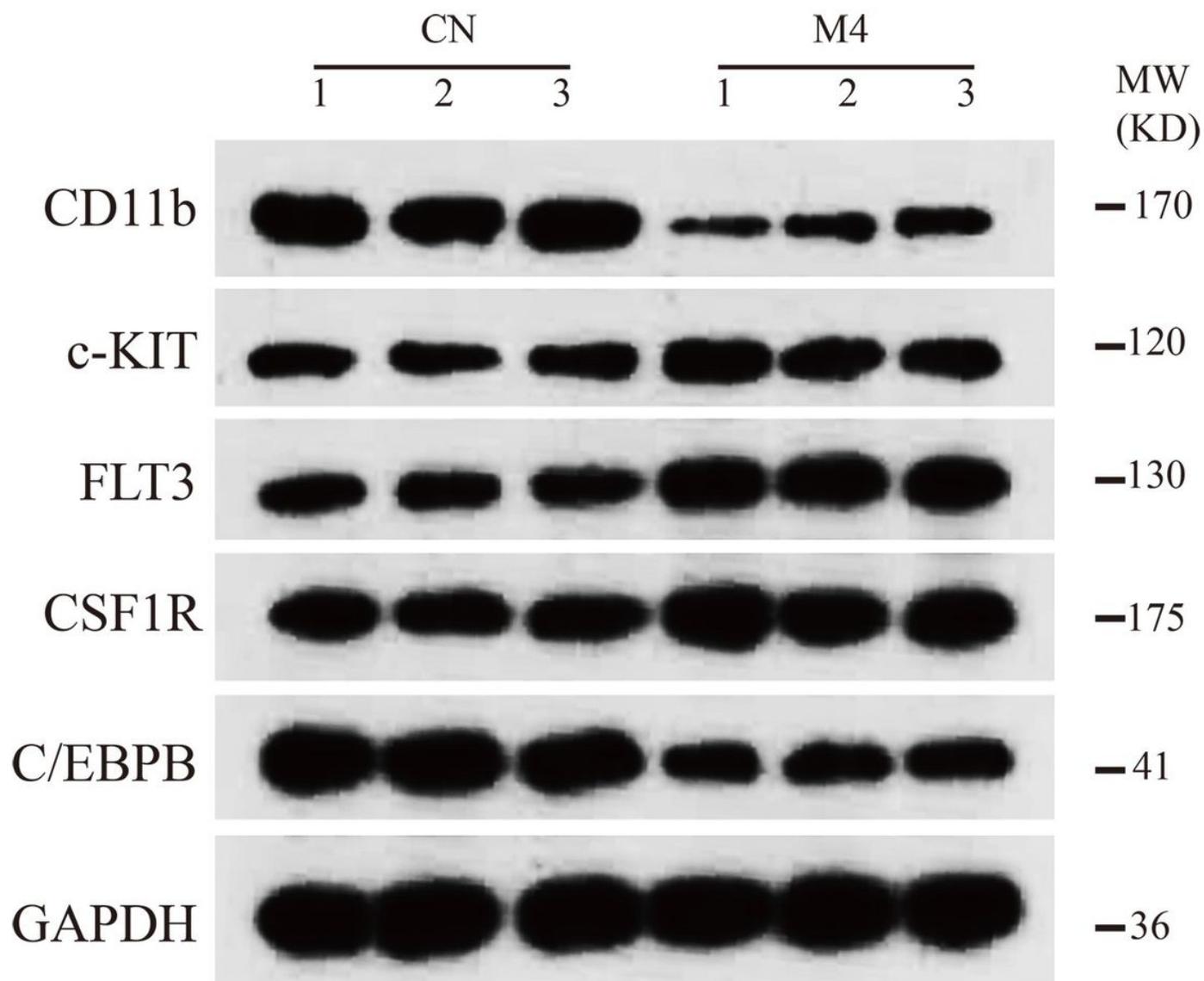


Figure 8

Detection in protein level. Western blotting detection of indicated protein. Lysates from three pairs of AML-M4 and normal samples were subjected to western blotting with antibody to CD11b, c-KIT, FLT3, CSF1R, C/EBPB, and GAPDH. GAPDH is a reference gene.