

# Genome-scale integrated analysis to identify prospective molecular mechanisms and therapeutic targets in isocitrate dehydrogenase 2 R140Q-mutated acute myeloid leukemia

RUI HUANG<sup>1</sup>, XIWEN LIAO<sup>2</sup>, JING LI<sup>1</sup>, JIEMIN WEI<sup>1</sup>, XIAYUN SU<sup>1</sup>, XIAOXUAN LAI<sup>1</sup>,  
BEICAI LIU<sup>1</sup>, FANGXIAO ZHU<sup>1,3</sup>, YUMEI HUANG<sup>1,4</sup> and QIAOCHUAN LI<sup>1</sup>

Departments of <sup>1</sup>Hematology and <sup>2</sup>Hepatobiliary Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region 530021; <sup>3</sup>Department of Rheumatology and Immunology, Affiliated Hospital of Guilin Medical College, Guilin, Guangxi Zhuang Autonomous Region 541001; <sup>4</sup>Department of Oncology, The Second Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region 530007, P.R. China

Received December 8, 2018; Accepted March 7, 2019

DOI: 10.3892/or.2019.7075

**Abstract.** The aim of the present study was to identify potential molecular mechanisms and therapeutic targets in regards to isocitrate dehydrogenase 2 (*IDH2*) *R140Q*-mutated acute myeloid leukemia (AML). An RNA sequencing dataset of *IDH2* wild-type and *R140Q*-mutated adult *de novo* AML bone marrow samples was obtained from The Cancer Genome Atlas (TCGA) database. The edgeR package was used to screen for the differentially expressed genes (DEGs), and the potential molecular mechanisms and therapeutic targets were identified using Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8, Biological Networks Gene Ontology tool, Connectivity Map (CMap), Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) and GeneMANIA. A total of 230 DEGs were identified between the bone marrow tissues of *IDH2 R140Q*-mutated and wild-type AML patients, of which 31 were significantly associated with overall survival (OS). Functional assessment of DEGs showed significant enrichment in multiple biological processes, including angiogenesis and cell differentiation. STRING and GeneMANIA were used to identify the hub genes of these DEGs. CMap analysis identified 13 potential small-molecule drugs against *IDH2 R140Q*-mutated adult *de novo* AML. Genome-wide co-expression network analysis identified several *IDH2 R140Q* co-expressed genes, of which 56 were significantly associated with AML OS. The difference

in *IDH2* mRNA expression levels and OS between the *IDH2 R140Q*-mutated and wild-type AML were not statistically significant in our cohort. In conclusion, we identified several co-expressing genes and potential molecular mechanisms that are instrumental in *IDH2 R140Q*-mutated adult *de novo* AML, along with 13 candidate targeted therapeutic drugs.

## Introduction

Acute myeloid leukemia (AML) is a cancer of the myeloid lineage hematopoietic cells, and is clonally heterogeneous since all myeloid precursors and progenitors can potentially undergo malignant transformation (1,2). The uncontrolled proliferation and accumulation of immature myeloid cells or blasts impair normal hematopoiesis, thereby increasing the risk of severe infections and hemorrhage (1,2). The genomic landscape of AML is highly diverse, and some mutations can be predictive of prognosis, as well as potential therapeutic targets. Most leukemias are characterized by specific oncoproteins encoded by fusion genes resulting from chromosomal translocations. Therefore, these oncoproteins not only serve as diagnostic markers of specific types of leukemia but also as potential therapeutic targets. In addition, some chromosomal abnormalities are associated with characteristic histopathological and clinical features, which form the cornerstone of the accurate diagnosis and treatment of leukemia (1). However, the heterogeneity of AML has stymied its targeted treatment, with current mortality rates well over 50% (1).

Isocitrate dehydrogenase 2 (*IDH2*) mutations are frequently observed in AML (3) and are not only potential targets for personalized therapies (4,5), but also indicators of AML prognosis (6,7). *R140Q* is the most frequent *IDH2* mutation in AML, and is correlated to advanced age, normal karyotype, and the French-American-British (FAB) classification M2 at diagnosis (8). Wiseman *et al* observed that multi-lineage hematopoiesis from *IDH2 R140Q* clones was frequently reconstituted after chemotherapy in AML patients (9). However, the key genes and pathways related to *IDH2* mutations in

**Correspondence to:** Professor Qiaochuan Li, Department of Hematology, The First Affiliated Hospital of Guangxi Medical University, 6 Shuang Yong Road, Nanning, Guangxi Zhuang Autonomous Region 530021, P.R. China  
E-mail: liqiaochuan@sohu.com

**Key words:** acute myeloid leukemia, The Cancer Genome Atlas, RNA sequencing, isocitrate dehydrogenase 2, mutation

AML are not completely clear. In a previous study, we used RNA sequencing datasets from The Cancer Genome Atlas (TCGA) database and multiple bioinformatic analyses to identify potential molecular mechanisms in tumor protein p53 (*TP53*)-mutated AML, as well as potential prognostic biomarkers (10,11). The aim of the present study was to identify the potential molecular mechanisms and genes associated with *IDH2 R140Q* mutation in adult *de novo* AML using a similar approach, in addition to potential targeted therapeutic drugs using the Connectivity Map (CMap).

## Materials and methods

**RNA-seq data of adult *de novo* AML patients.** The RNA-seq dataset of the bone marrow tissues of an adult *de novo* AML patient cohort collected at diagnosis, as well as the corresponding survival information, were obtained from TCGA (<https://gdc-portal.nci.nih.gov/>; accessed August 10, 2018) database (12). The corresponding information on *IDH2 R140Q* mutation status was obtained from the cBioPortal for Cancer Genomics website (<http://www.cbioportal.org/index.do>; accessed August 10, 2018) (13,14). Since all the data in this study were retrieved from TCGA, the present study did not require the approval of the ethics committee. The RNA-seq data are available on a public domain, and were acquired and analyzed according to the published guidelines of TCGA (<https://cancergenome.nih.gov/publications/publicationguidelines>).

**Identification of differentially expressed genes (DEGs) and their prognostic value in AML.** DEGs between the *IDH2 R140Q*-mutated and wild-type adult *de novo* AML were identified by edgeR using the following criteria:  $|\log_2$  fold change (FC)|  $\geq 1$ , and both P-value and false discovery rate (FDR)  $< 0.05$  (15,16). The heat map and volcano plot of the DEGs were generated using the 'gplots' package in R platform, and their prognostic value in AML was determined using the 'survival' package in the R platform (17,18).

**Functional assessment.** Functional assessment of the DEGs, in terms of Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment, was performed by Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8 (<https://david.ncifcrf.gov/tools.jsp>; accessed August 10, 2018) (19,20) and those with P-values  $< 0.05$  were considered statistically significant. The directed acyclic graph of GO terms was drawn using the Biological Networks Gene Ontology (BiNGO) tool in Cytoscape\_v3.6.1, a plugin used to assess the overrepresentation of GO categories in biological networks (21).

**Construction of protein-protein and gene-gene interaction networks.** The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (<http://string.embl.de/>; accessed August 10, 2018) was used to construct the protein-protein interaction (PPI) networks (22-24), and GeneMANIA (<http://genemania.org/>; accessed August 10, 2018) was used for gene-gene interaction (GGI) networks (25,26).

**Connectivity Map analysis.** Connectivity Map (CMap, <https://portals.broadinstitute.org/cmap/>; accessed August 10, 2018) is

an online tool and data source for analyzing the mechanism of action and localization of drugs based on transcriptome data (27,28). A positive score indicates an inducement effect of a small-molecule drug on the query signatures, and a negative score reflects a repression effect. CMap was used to screen for putative small-molecule drugs against *IDH2 R140Q*-mutated AML, and those with a connective score  $< -0.2$  were identified as the potential therapeutic drugs. The chemical structures of these drugs were obtained from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>; accessed August 10, 2018) (29,30) and the Search Tool for Interacting Chemicals (STITCH: <http://stitch.embl.de/>; accessed August 10, 2018) was used to construct protein-chemical interaction networks based on the existing literature and databases (31-33).

**Genome-wide co-expression network analysis.** To assess the biological relevance of the *IDH2 R140Q* mutation in AML, the *IDH2* co-expressing genes were screened using the Pearson correlation test. The genome-wide co-expression analysis was performed using the 'cor' function in the R platform, and the genes with a Pearson correlation coefficient  $> 0.75$  and  $P < 0.05$  were identified as the *IDH2 R140Q* co-expressing genes. In addition, the prognostic values of these genes was evaluated by the 'survival' package in the R platform (17,18).

**Statistical analysis.** The *IDH2* mRNA expression levels in the *R140Q*-mutated and wild-type AML bone marrow tissues was compared by the Student's t-test. The clinical outcomes of the *R140Q*-mutated and wild-type AML patients was compared by the Kaplan-Meier method and log-rank test. Hazard ratio (HR) and 95% confidence interval (CI) were calculated using the Cox proportional hazards regression model. FDR in edgeR was calculated by multiple testing with the Benjamini-Hochberg procedure (34-36). P-value  $< 0.05$  was regarded as statistically significant. All statistical analyses were conducted using the SPSS version 20.0 (IBM Corp., Armonk, NY, USA) and R 3.5.0 (<https://www.r-project.org/>; accessed June 21, 2018).

## Results

**Patient data.** RNA-seq datasets from 151 AML patients were downloaded from TCGA (12), and data of 3 patients with a second *IDH2* mutation was excluded. Thirteen of the remaining 148 patients had the *IDH2 R140Q* mutation and 135 patients were *IDH2* wild-type. Their corresponding survival information, including overall survival duration and survival status, was also download from TCGA. The survival information for 11 patients (10 *IDH2* wild-type and 1 *IDH2 R140Q*-mutated) was missing. The bioinformatic analysis flowchart for the present study is shown in Fig. 1.

**DEG screening.** A total of 230 DEGs were identified between the *IDH2* wild-type and *R140Q*-mutated AML samples (Table SI), of which 133 were downregulated and 97 were upregulated in the bone marrow of the *IDH2 R140Q*-mutated AML patients. The heat map and volcano plot of the DEGs are shown in Fig. S1 and Fig. 2, respectively.

**Association of the DEGs with survival.** To investigate the prognostic value of the *IDH2 R140Q*-specific DEGs, we

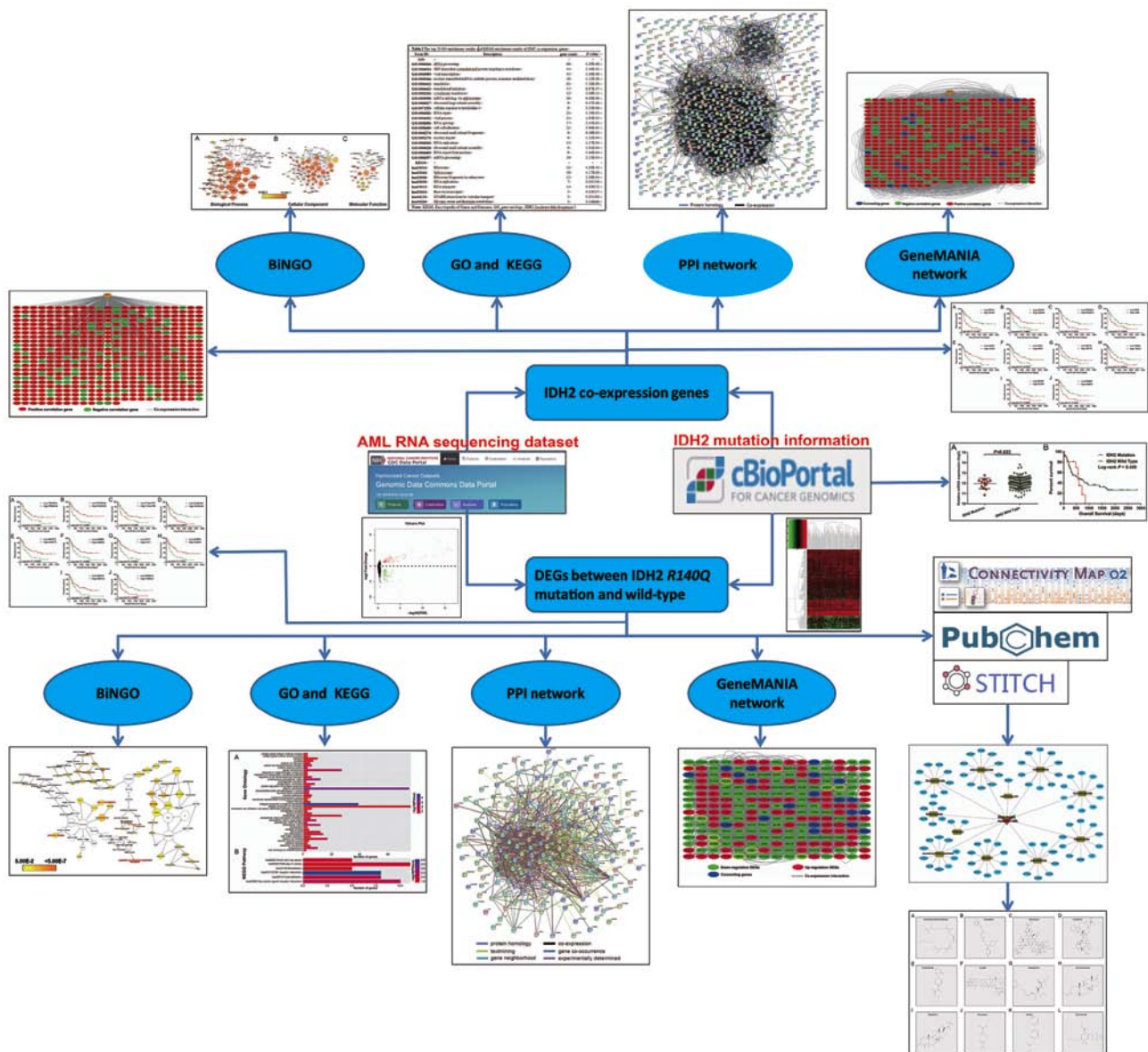


Figure 1. Flowchart of the bioinformatic analysis in the present study. AML, acute myeloid leukemia; *IDH2*, isocitrate dehydrogenase 2; BiNGO, Biological Networks Gene Ontology tool; DEGs, differentially expressed genes; GO, Gene Ontology, KEGG, Kyoto Encyclopedia of Genes and Genomes; STITCH, Search Tool for Interacting Chemicals; PPI, protein-protein interaction.

stratified the patients into a high- and low-expression groups according to the median values of each of these DEGs, and compared their survival status. Due to the incomplete survival information, we could not perform a multivariate Cox proportional risk regression model, and therefore assessed their prognostic values through a univariate analysis. Thirty-one DEGs were significantly associated with the OS (Table SII), and the top 10 significant prognostic DEGs were tetraspanin 10 (*TSPAN10*), pleckstrin homology domain containing A5 (*PLEKHA5*), coiled-coil domain containing 198 (*CCDC198*, also known as *C14orf105*), pleckstrin homology domain containing A6 (*PLEKHA6*), A-kinase anchoring protein 12 (*AKAP12*), gamma-aminobutyric acid type A receptor delta subunit (*GABRD*), fucosyltransferase 1 (*FUT1*), potassium voltage-gated channel interacting protein 3 (*KCNIP3*), immunoglobulin superfamily DCC subclass member 4 (*IGDCC4*) and podocalyxin like 2 (*PODXL2*) (Fig. 3A-J).

**Functional assessment of DEGs.** The DEGs were functionally assessed using the GO terms and KEGG pathways. The significantly enriched GO terms were angiogenesis, cell differentiation, cell-matrix adhesion, homophilic cell adhesion via plasma membrane adhesion molecules, cell junction, signal transduction, phosphoinositide phospholipase C activity and integrin binding (Fig. 4A). KEGG analysis showed that the DEGs were significantly enriched in ECM-receptor interaction, focal adhesion and pathways in cancer (Fig. 4B). The BiNGO analysis also demonstrated that these DEGs were significantly enriched in regulation of calcium ion-dependent exocytosis, cell differentiation, signal transmission, signaling process and phosphoinositide phospholipase C activity (Fig. S2).

**GGI and PPI interaction network construction of DEGs.** The GGI network indicated complex interactions between the DEGs (Fig. 5), and the top 10 degree genes that were

Table I. Potential targeted therapeutic drugs for *IDH2* R140Q-mutated AML based on Connectivity Map analysis.

CMap name	Mean connective score	n	Enrichment	P-value	Specificity	Percent non-null
Arachidonyltrifluoromethane	-0.616	2	-0.92	0.01296	0.0261	100
Vanoxerine	-0.586	4	-0.718	0.01281	0.025	75
Vancomycin	-0.48	4	-0.71	0.01428	0.0552	75
Vinblastine	-0.436	3	-0.87	0.00433	0.0534	66
Thioperamide	-0.386	5	-0.697	0.00555	0.0373	60
Nystatin	-0.381	3	-0.797	0.01711	0.0216	66
Meteneprost	-0.364	4	-0.648	0.0366	0.1797	50
Etiocholanolone	-0.351	6	-0.678	0.00302	0.0844	50
Finasteride	-0.348	6	-0.664	0.00407	0.0854	50
PF-00539745-00	-0.345	3	-0.855	0.00603	0.0593	66
Benzocaine	-0.34	4	-0.829	0.00163	0.0134	50
Felbinac	-0.337	4	-0.68	0.02316	0.2199	50
Brinzolamide	-0.282	4	-0.651	0.03503	0.0541	50

*IDH2*, isocitrate dehydrogenase 2; AML, acute myeloid leukemia; CMap, Connectivity Map.

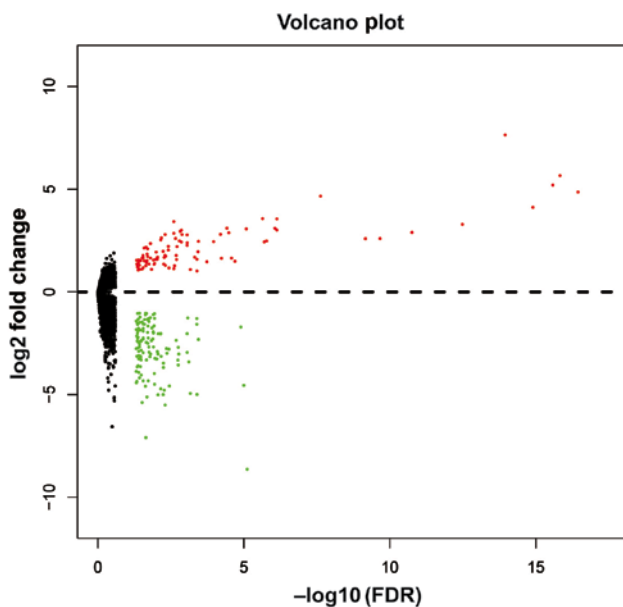


Figure 2. Volcano plot of DEGs between *IDH2* R140Q-mutated and wild-type AML bone marrow tissues. Red dots indicate upregulated DEGs; green dots indicate downregulated DEGs; black dots indicate non-differentially expressed genes. DEGs, differentially expressed genes; *IDH2*, isocitrate dehydrogenase 2; AML, acute myeloid leukemia; FDR, false discovery rate.

identified as the hub genes were collagen type IV  $\alpha 2$  chain (*COL4A2*), lumican (*LUM*), fibulin 2 (*FBLN2*), thrombospondin 2 (*THBS2*), lysyl oxidase like 1 (*LOXLI*), myosin light chain 9 (*MYL9*), heparan sulfate proteoglycan 2 (*HSPG2*), gamma-glutamyltransferase 5 (*GGT5*), integrin subunit alpha 1 (*ITGA1*), protein tyrosine phosphatase, receptor type M (*PTPRM*), and leucine rich repeat containing 17 (*LRRCL17*); the highest degree was that of *COL4A2* (degree=127). The PPI network supported the GGI results, and substantiated complex interactions among these DEGs at the protein level as well (Fig. S3). The top 10 degree genes identified as hub genes

in the PPI networks were insulin receptor related receptor (*INSRR*), *LUM*, AXL receptor tyrosine kinase (*AXL*), cyclin D1 (*CCND1*), CDC42 binding protein kinase beta (*CDC42BPB*), cyclin dependent kinase 15 (*CDK15*), nitric oxide synthase 2 (*NOS2*), t-complex-associated-testis-expressed 1 (*TCTE1*), SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1 (*SMARCA1*) and espin (*ESPN*); the highest degree was that of *INSRR* (degree=39). The *LUM* gene was common to both GGI and PPI networks, indicating its critical role in *IDH2* R140Q-mutated AML.

**CMap analysis.** The CMap analysis was conducted to screen for small-molecule drugs targeting the *IDH2* R140Q mutation. Thirteen drugs were identified: arachidonyltrifluoromethane, vanoxerine, vancomycin, vinblastine, thioperamide, nystatin, meteneprost, etiocholanolone, finasteride, PF-00539745-00, benzocaine, felbinac and brinzolamide (Table I and Fig. 6). (PF-00539745-00 does not appear in the figure as its chemical structure is not available in PubChem). The protein-chemical interaction networks of 11 drugs were obtained from STITCH (Fig. 7), and cytochrome P450 family 2 subfamily C member 8 (*CYP2C8*), *NOS2* and adrenoceptor  $\alpha$  2A (*ADRA2A*) were common to the DEGs and protein-chemical interaction networks. In addition, KEGG analysis showed that *ADRA2A* was involved in the neuroactive ligand-receptor interaction pathway, and *NOS2* in pathways in cancer and amoebiasis pathway. Based on these findings, we hypothesized that nystatin is a potential therapeutic agent against *IDH2* R140Q-mutated AML that acts through the nystatin-*NOS2* pathway to regulate cancer/amoebiasis pathways. Similarly, thioperamide possibly functions through the thioperamide-*ADRA2A*-neuroactive ligand-receptor interaction pathway axis, and vanoxerine through the vanoxerine-*CYP2C8* axis.

**Genome-wide co-expression network analysis.** Genome-wide co-expression network analysis showed that 542 genes were co-expressed with *IDH2* in the *IDH2* R140Q-mutated AML



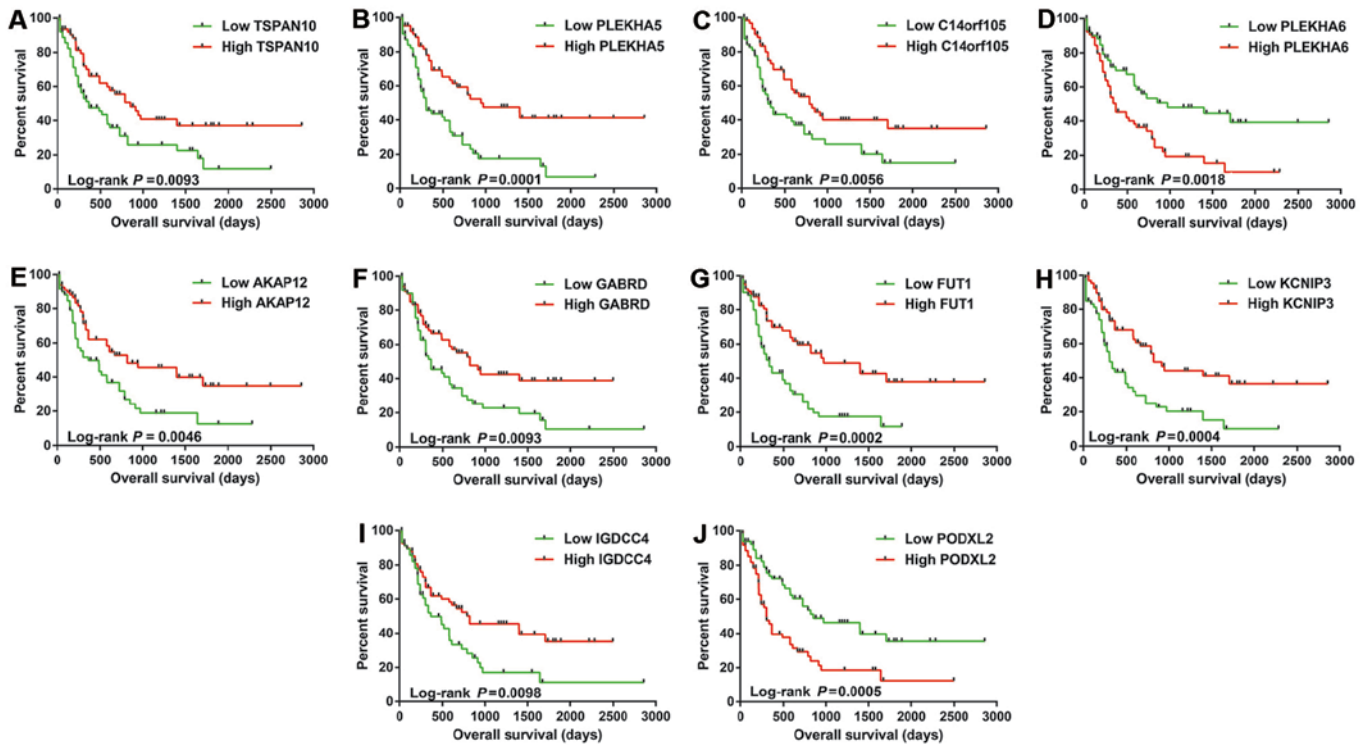


Figure 3. The Kaplan-Meier curves of the top 10 significantly prognostic DEGs. (A) Overall survival stratified by *TSPAN10*, (B) *PLEKHA5*, (C) *C14orf105*, (D) *PLEKHA6*, (E) *AKAP12*, (F) *GABRD*, (G) *FUT1*, (H) *KCNIP3*, (I) *IGDCC4* and (J) *PODXL2*. DEGs, differentially expressed genes; *TSPAN10*, tetraspanin 10; *PLEKHA5*, pleckstrin homology domain containing A5; *C14orf105*, coiled-coil domain containing 198 (also known as *C14orf105*); *PLEKHA6*, pleckstrin homology domain containing A6; *AKAP12*, A-kinase anchoring protein 12; *GABRD*, gamma-aminobutyric acid type A receptor delta subunit; *FUT1*, fucosyltransferase 1; *KCNIP3*, potassium voltage-gated channel interacting protein 3; *IGDCC4*, immunoglobulin superfamily DCC subclass member 4; *PODXL2*, podocalyxin like 2.

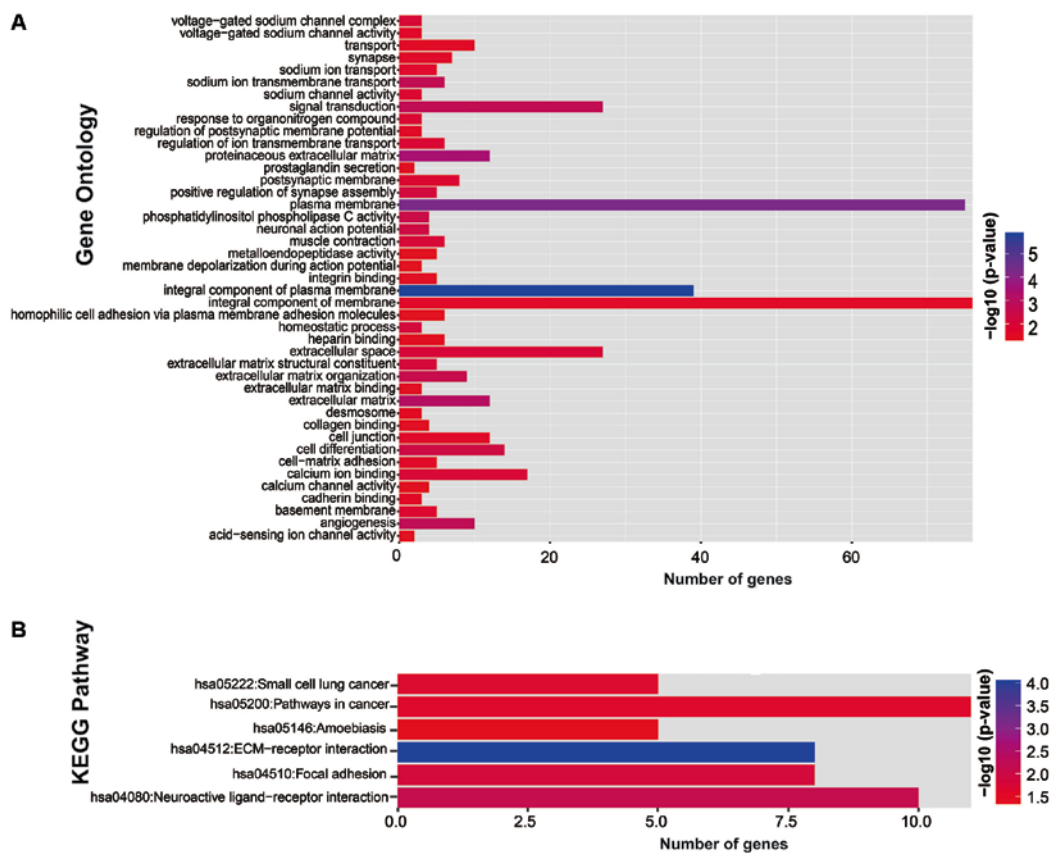


Figure 4. Functional assessment of DEGs. (A) GO term enrichment. (B) KEGG enrichment. DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

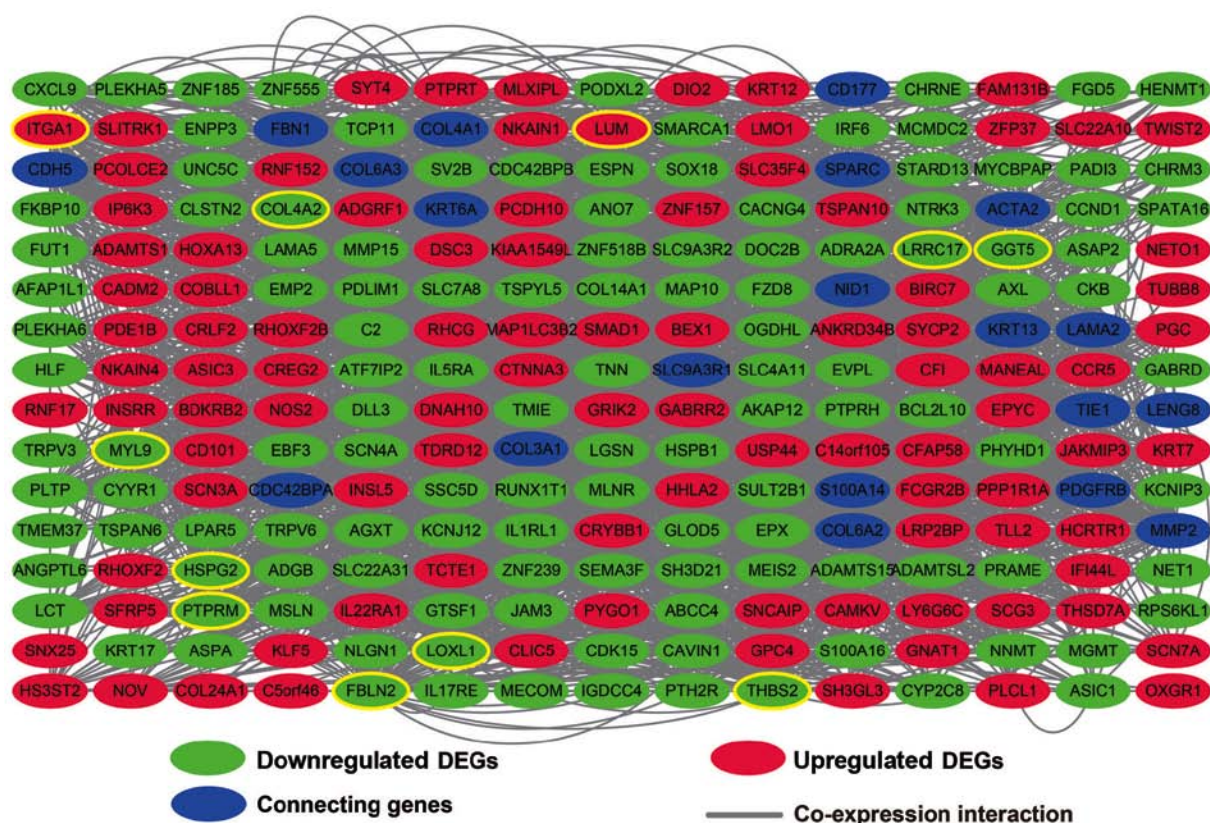


Figure 5. Gene-gene interaction network of DEGs identified by GeneMANIA. The nodes with yellow rings are the hub genes in the network. DEGs, differentially expressed genes.

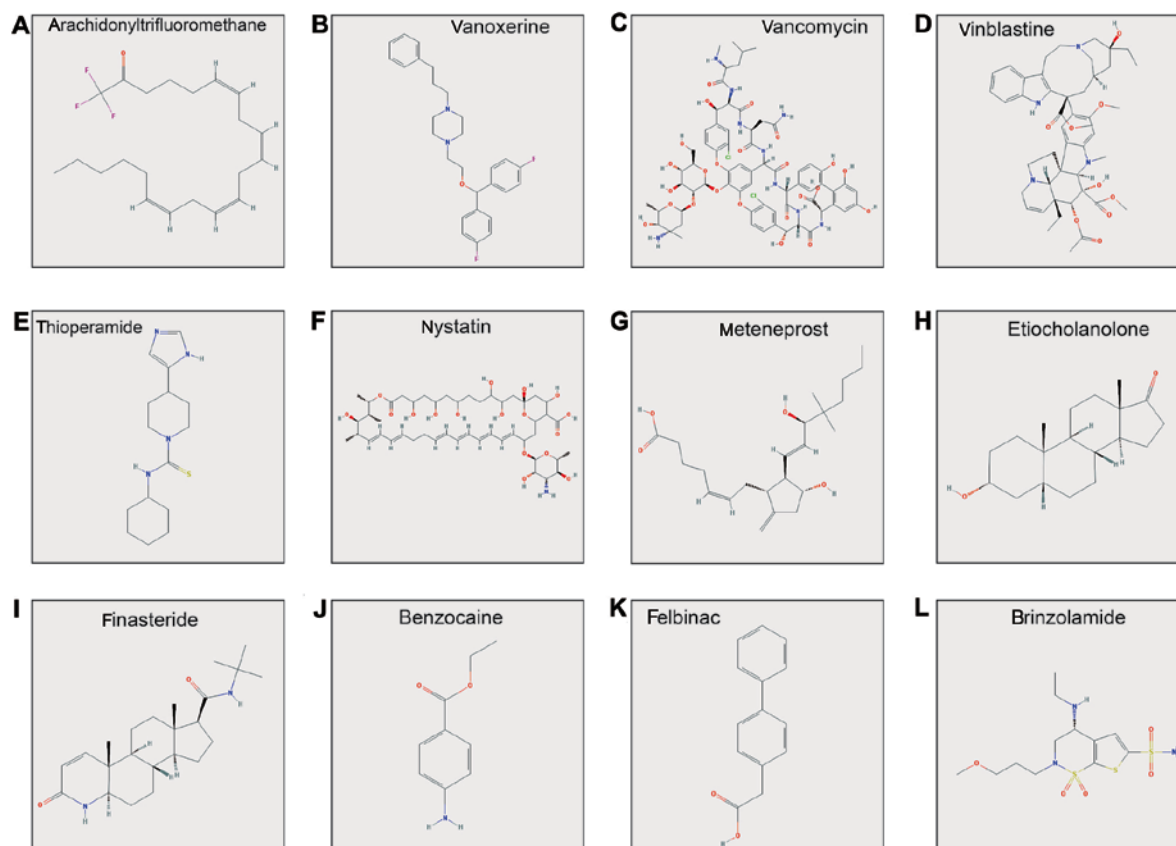


Figure 6. The chemical structures of the small-molecule drugs. The order of these small-molecule drugs were: (A) arachidonyltrifluoromethane, (B) vanoxerine, (C) vancomycin, (D) vinblastine, (E) thioperamide (F) nystatin, (G) meteneprost, (H) etiocholanolone, (I) finasteride, (J) benzocaine, (K) felbinac and (L) brinzolamide. (PF-00539745-00 does not appear in the figure as its chemical structure is not available in PubChem).



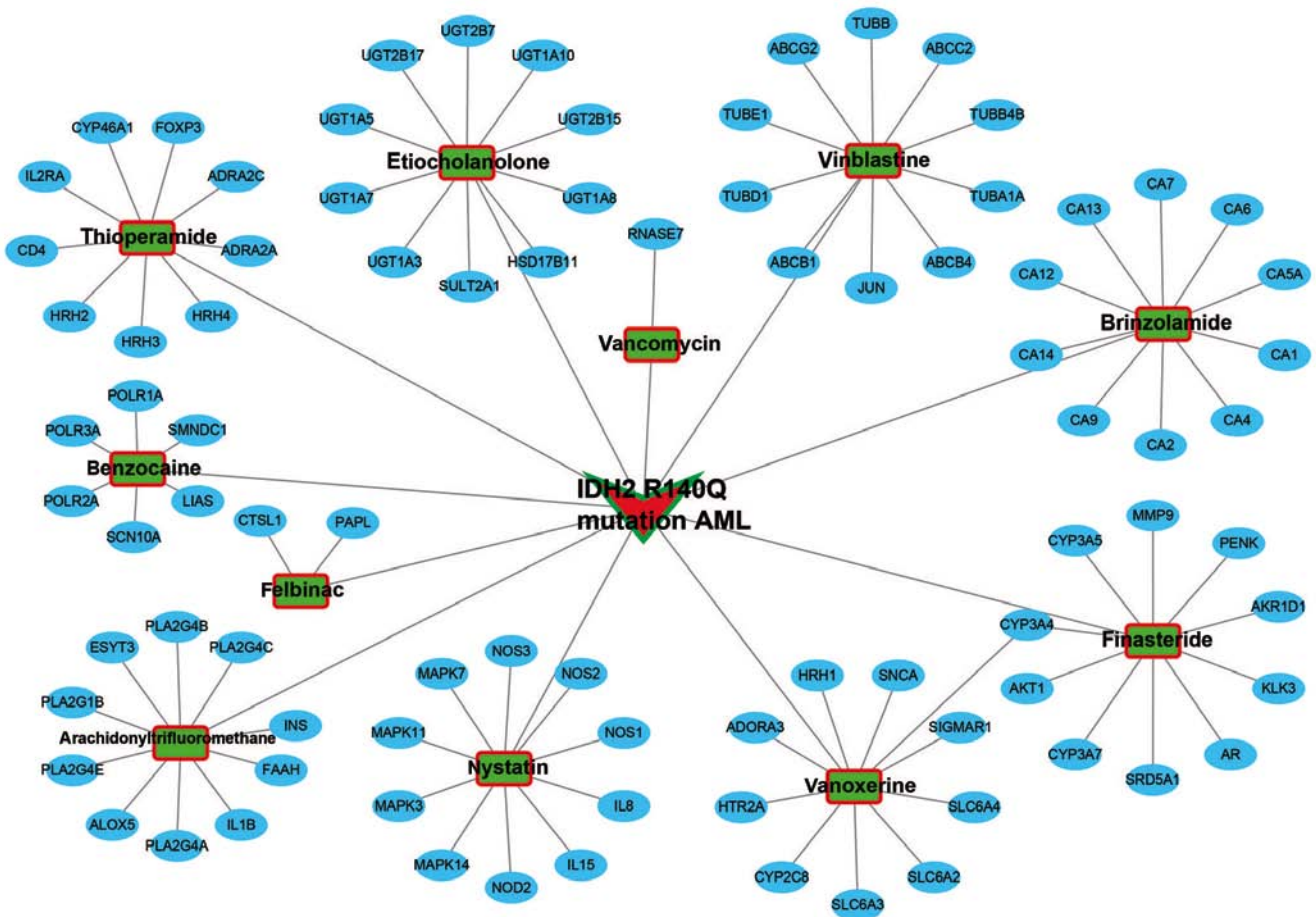


Figure 7. Interaction networks of *IDH2 R140Q*-mutated AML-related small-molecule drugs and target genes. *IDH2*, isocitrate dehydrogenase 2; AML, acute myeloid leukemia.x

bone marrow tissues, of which 455 were positively correlated and 87 were negatively correlated with *IDH2 R140Q* (Fig. 8 and Table SIII). GO enrichment analysis showed that *IDH2* and the co-expressed genes were significantly enriched in DNA replication, DNA repair, cell-cell adhesion, cell division and cell cycle-related biological processes (Table II and Table SIV), while KEGG analysis showed significant enrichment in DNA replication, SNARE interactions in vesicular transport and base excision repair pathway (Table II and Table SV). Furthermore, these results were verified by BiNGO, which indicated significant enrichment in the cell cycle process and DNA replication (Fig. S4). GGI (Fig. 9) and PPI (Fig. S5) interaction networks were respectively constructed using geneMANIA and STRING, and further verified the relationship between *IDH2* and the co-expressing genes. The prognostic value of these co-expressing genes in AML was determined using the 'survival' package in the R platform, and 56 genes were significantly associated with the OS (Table SVI). The top ten significantly prognostic genes were splicing factor 3a subunit 1 (*SF3A1*), centromere protein V (*CENPV*), interferon stimulated exonuclease gene 20 like 2 (*ISG20L2*), prohibitin (*PHB*), glycine cleavage system protein H (*GCSH*), heat shock transcription factor 2 (*HSF2*), zinc finger protein 19 (*ZNF19*), EARP complex and GARP complex interacting protein 1 (*EIPRI*, also known as *TSSCI*), enoyl-CoA hydratase, short chain 1 (*ECHS1*), and nucleic acid

binding protein 2 (*NABP2*). The Kaplan-Meier curves of the patients stratified into the low-and high-expressing groups are shown in Fig. 10A-J.

**Survival analysis of *IDH2*.** To better understand the role of the *IDH2 R140Q* mutation in AML, we compared the expression levels of the wild-type and mutated *IDH2*, and observed no significant differences (Fig. 11A). Furthermore, the OS of the AML patients harboring the wild-type or *IDH2 R140Q* mutant was also not significantly different (Fig. 11B).

## Discussion

Although studies have reported fewer mutations in acute myeloid leukemia (AML) compared to other cancers, those specifically related to AML pathogenesis and prognoses have not been fully elucidated. An integrative analysis based on multi-genomic data is necessary to determine the relationship between specific gene mutations and cancer progression. Isocitrate dehydrogenase (IDH) is a family of enzymes that catalyze the oxidative decarboxylation of isocitrate to  $\alpha$ -ketoglutarate and NADPH in the tricarboxylic acid (TCA) cycle. The human genome has five IDH genes that encode three isozymes, IDH1, IDH2 and IDH3 of which *IDH2* is mitochondrial and frequently mutated in different cancers, including AML, brain tumors and gliomas (4). The most common mutations in IDH are located



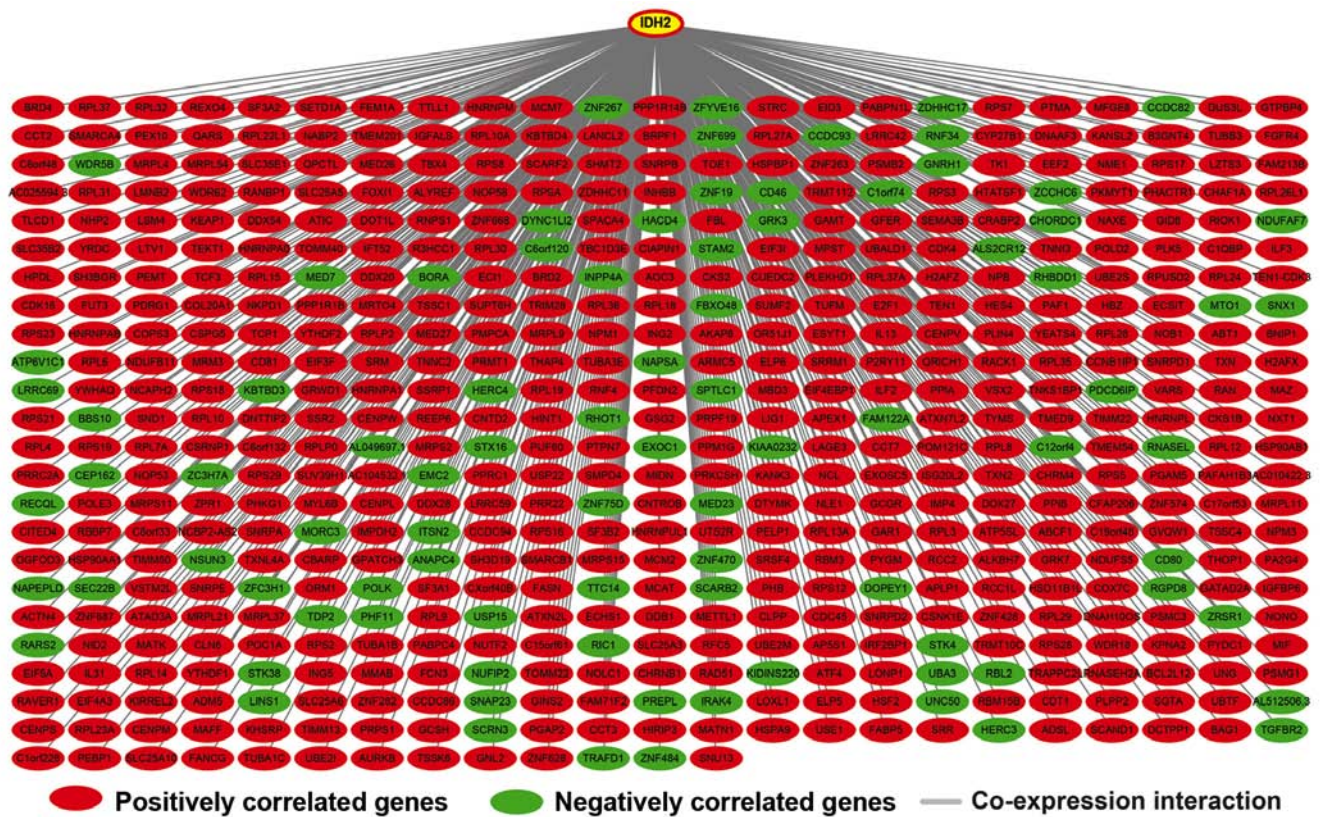


Figure 8. Co-expression network of *IDH2* in the *IDH2* R140Q-mutated AML bone marrow tissue. *IDH2*, isocitrate dehydrogenase 2; AML, acute myeloid leukemia.

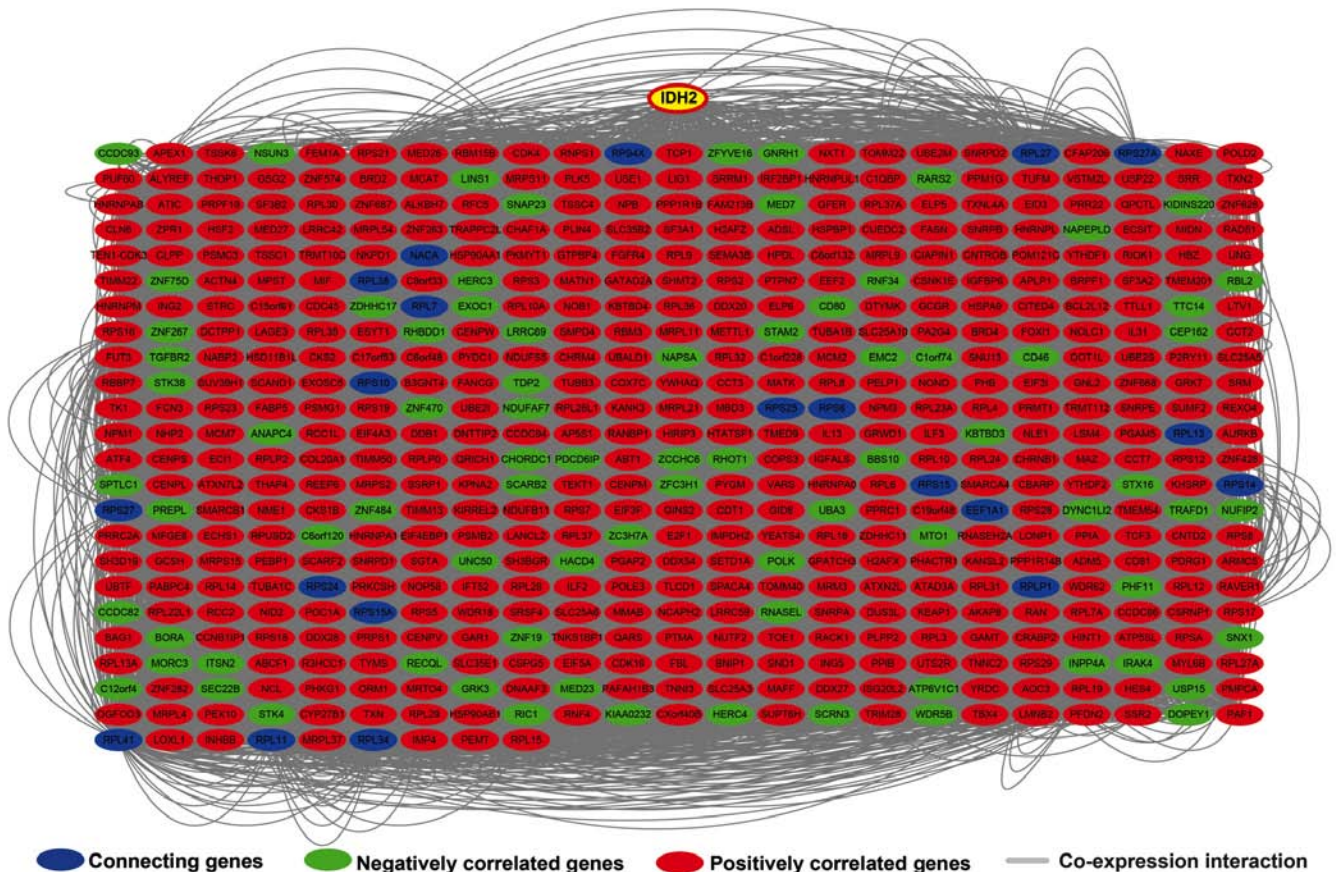


Figure 9. Gene-gene interaction network of *IDH2* and its co-expression genes in *IDH2* R140Q-mutated AML identified by GeneMANIA. *IDH2*, isocitrate dehydrogenase 2; AML, acute myeloid leukemia.



Table II. The top 20 GO enriched and KEGG enriched *IDH2* co-expression genes.

Term ID	Description	Gene count	P-value
GO			
GO:0006364	rRNA processing	66	1.29E-48
GO:0006614	SRP-dependent cotranslational protein targeting to membrane	44	2.46E-41
GO:0019083	Viral transcription	45	1.06E-38
GO:0000184	Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	46	1.12E-38
GO:0006412	Translation	61	5.18E-38
GO:0006413	Translational initiation	47	8.97E-37
GO:0002181	Cytoplasmic translation	12	3.09E-11
GO:0000398	mRNA splicing, via spliceosome	26	6.08E-09
GO:0000027	Ribosomal large subunit assembly	9	6.47E-08
GO:0071353	Cellular response to interleukin-4	8	3.53E-06
GO:0006281	DNA repair	21	1.59E-05
GO:0016032	Viral process	24	1.94E-05
GO:0008380	RNA splicing	17	2.45E-05
GO:0098609	Cell-cell adhesion	22	3.99E-05
GO:0042274	Ribosomal small subunit biogenesis	6	6.48E-05
GO:0051170	Nuclear import	6	1.21E-04
GO:0006260	DNA replication	15	1.57E-04
GO:0000028	Ribosomal small subunit assembly	6	1.61E-04
GO:0006405	RNA export from nucleus	9	1.64E-04
GO:0006397	mRNA processing	16	2.13E-04
KEGG			
hsa03010	Ribosome	52	4.30E-40
hsa03040	Spliceosome	18	4.17E-06
hsa03008	Ribosome biogenesis in eukaryotes	12	2.38E-04
hsa03030	DNA replication	7	0.001519
hsa03013	RNA transport	14	0.008172
hsa03410	Base excision repair	5	0.028527
hsa04130	SNARE interactions in vesicular transport	5	0.031456
hsa00260	Glycine, serine and threonine metabolism	5	0.048668

KEGG, Encyclopedia of Genes and Genomes; GO, Gene Ontology; *IDH2*, isocitrate dehydrogenase 2.

at R132 (*IDH1*), R140 and R172 (*IDH2*) (3,8,37). In the present study, the RNA-seq dataset from The Cancer Genome Atlas (TCGA) was utilized to identify the key genes and pathways associated with the *IDH2 R140Q* mutation in adult *de novo* AML for the first time using multiple bioinformatic methods. In addition, potential small-molecule drugs targeting the mutated *IDH2* were identified.

Functional assessment of the *IDH2 R140Q*-specific differentially expressed genes (DEGs) showed significant association with angiogenesis, cell differentiation, cell-matrix adhesion, homophilic cell adhesion via plasma membrane adhesion molecules and cell junction. Previous studies have shown a prognostic role of marrow angiogenesis-associated factors in AML patients, and a promising therapeutic role of angiogenesis inhibition (38-40). *IDH* mutations induce tumorigenesis by epigenetic alterations, as well as by disrupting the tricarboxylic acid (TCA) cycle and activating hypoxia-related signaling pathways, which increase anabolic

processes and angiogenesis (41,42). Functional assessment of *IDH2* co-expressing genes showed enrichment in DNA replication, DNA repair, cell-cell adhesion, cell division, and cell cycle-related processes. Therefore, the biological and clinical manifestations of *IDH2 R140Q* mutation are distinct from wild-type AML at the cellular level. Further studies are needed to validate the functions of these *IDH2* co-expressing genes.

Among the hub differentially expressed genes (DEGs) identified by protein-protein interaction (PPI) and gene-gene interaction (GGI), several have been correlated to AML, especially in patients with numerous mutations. Previous studies have identified *MYH9* as a direct target of *RUNX1*, an important transcription factor that regulates the differentiation of hematopoietic stem cells into mature blood cells, and is frequently mutated in AML (43,44). Zhang *et al* reported downregulation of *CCND1* in an AML cell line treated with amifostine, and may therefore be its direct target (45). *AXL* is

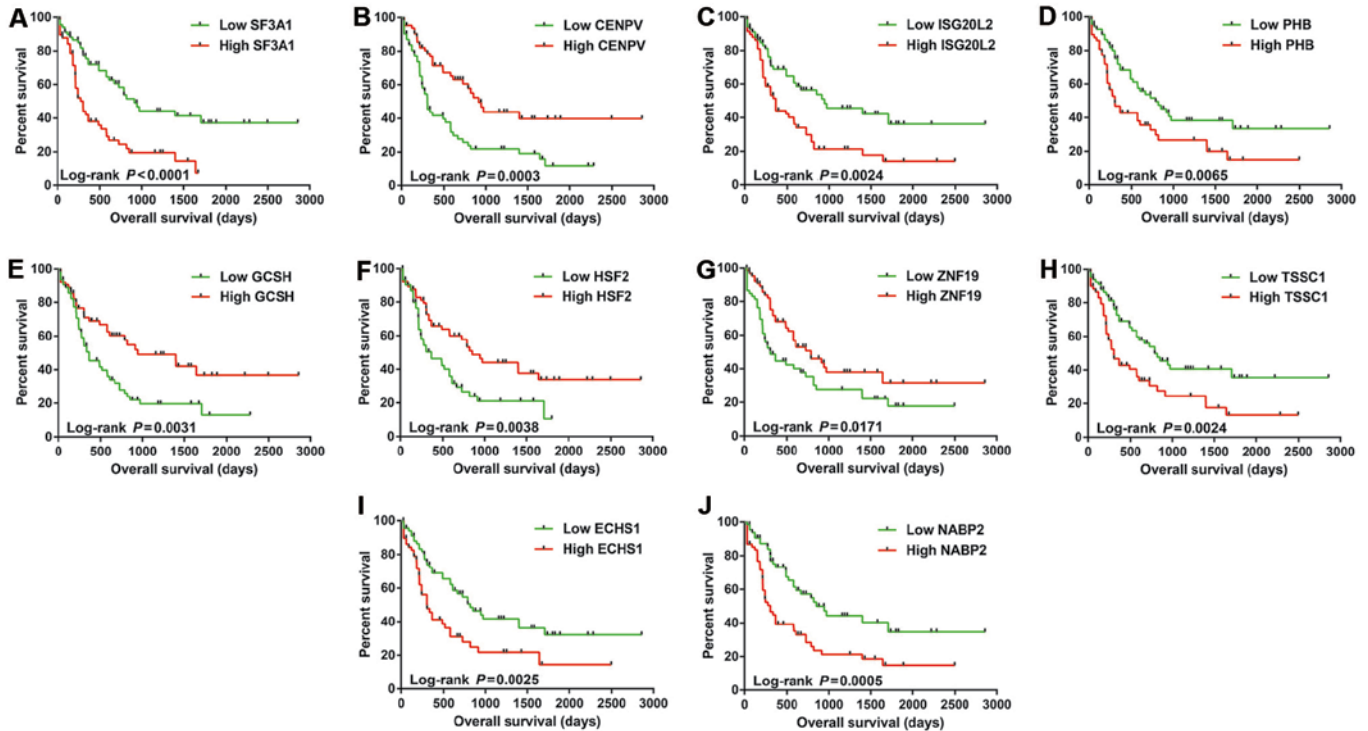


Figure 10. The Kaplan-Meier curves of the top 10 significantly prognostic *IDH2* co-expression genes in *IDH2* R140Q-mutated AML. Overall survival stratified by (A) *SF3A1*, (B) *CENPV*, (C) *ISG20L2*, (D) *PHB*, (E) *GCSH*, (F) *HSF2*, (G) *ZNF19*, (H) *TSSC1*, (I) *ECHS1* and (J) *NABP2*. *IDH2*, isocitrate dehydrogenase 2; *SF3A1*, splicing factor 3a subunit 1; *CENPV*, centromere protein V; *ISG20L2*, interferon stimulated exonuclease gene 20 like 2; *PHB*, prohibitin; *GCSH*, glycine cleavage system protein H; *HSF2*, heat shock transcription factor 2; *ZNF19*, zinc finger protein 19; *TSSC1*, EARP complex and GARP complex interacting protein 1; *ECHS1*, enoyl-CoA hydratase, short chain 1; *NABP2*, nucleic acid binding protein 2.

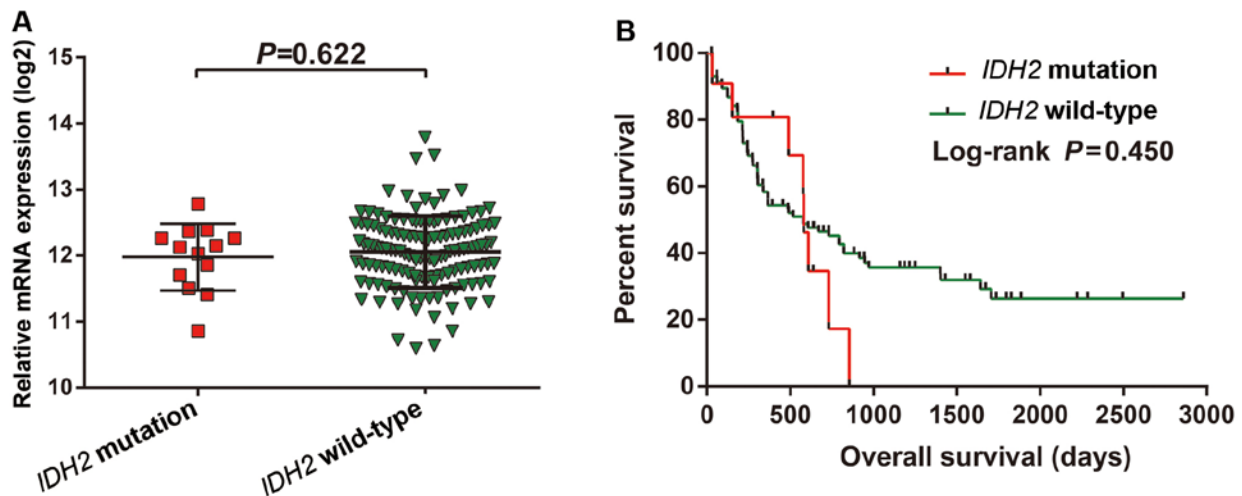


Figure 11. The comparison of mRNA expression levels and survival of *IDH2* R140Q-mutated and wild-type AML patients. (A) The *IDH2* mRNA expression levels in the bone marrow tissues; (B) Kaplan-Meier survival curves of AML patients stratified by *IDH2* R140Q mutation. AML, acute myeloid leukemia; *IDH2*, isocitrate dehydrogenase 2.

essential for the constitutive Fms-like tyrosine kinase-3 (*FLT3*) phosphorylation in *FLT3*-internal tandem duplication AML, which is responsible for the aberrant blast cell proliferation and the clinical outcome of AML, and its inhibition decreases *FLT3* phosphorylation (46). Therefore, *AXL* is a potential target for treating *FLT3*-mutated AML, and its specific inhibitor BGB324 has been tested against leukemia cell proliferation and therapy resistance (47,48). *AKAP12* is a tumor-suppressor gene that is inactivated in childhood myeloid malignancies via

epigenetic silencing through promoter DNA methylation (49). *SF3A1* is a core spliceosomal gene that plays a fundamental role in the processing of nascent RNA transcripts. Although a somatic mutation of *SF3A1* has been reported in AML, its rarity has precluded analysis of its function and molecular mechanism (50,51). In the present study, the prognostic DEGs and *IDH2* co-expressing genes were identified based on a single cohort of TCGA. Therefore, our results have to be verified in additional larger cohorts with complete clinical parameters.



Although we identified 13 candidate drugs targeting the *IDH2 R140Q* mutation, we could not find any evidence of their application against AML that were based on previous studies. Vinblastine is one of the most widely used plant-derived chemotherapeutic agents used to treat cancers, including AML. The combination of cytosine arabinoside, VP 16-213, vincristine and vinblastine (A-triple-V) is used to treat AML relapse, and results in complete remission in most cases (52,53). Finasteride is a 5- $\alpha$ -reductase inhibitor used for the treatment of alopecia and prostate cancer. Chau *et al* demonstrated that while finasteride may not have a dose-dependent effect on prostate cancer, it may decrease the risk (54). One of the potential mechanisms of finasteride action in prostate cancer is the inhibition of cancer cell invasion and metastasis by downregulation of matrix metalloproteinase (MMP)2 and MMP9 (55). Wiebe *et al* reported that progesterone-induced stimulation of mammary tumorigenesis can also be inhibited by finasteride (56). In addition, it also decreased melanogenesis in both melanocytes and melanoma cells by inhibiting adenylate cyclase and the melanocortin 1 receptor (57). The potential therapeutic effects of these drugs on AML need to be studied further.

The present study has some limitations that need to be clarified. First, our results are from a single cohort of TCGA and generated by bioinformatic analysis, and thus need to be verified in additional cohorts, as well as in experimental studies, such as *in vitro* validation. Second, due to the relatively small sample size of our cohort, and only 13 patients with the *IDH2 R140Q* mutation, the results generated in this study still need to be further verified using a larger sample size in future research. Third, we did not observe any significant differences in the *IDH2* mRNA expression levels and OS duration between the *IDH2 R140Q* mutant and wild-type AML patients, which may be due to the small sample size. Despite these limitations, we identified multiple genes associated with the *IDH2 R140Q* mutation, and its potential molecular mechanisms in adult *de novo* AML. In addition, we also identified 13 drugs specific for *IDH2 R140Q*-mutated AML, which once verified can be used for AML patients harboring this mutation.

In conclusion, we identified several prognostic biomarkers and potential molecular mechanisms, that may play an essential role in *IDH2 R140Q*-mutated adult *de novo* AML and improve prognosis, as well as 13 candidate drugs for *IDH2 R140Q*-mutated adult *de novo* AML. However, our findings need to be verified by further experimental and clinical studies for future clinical application.

## Acknowledgements

The authors thank the contributors of The Cancer Genome Atlas (<https://portal.gdc.cancer.gov/>) and the cBioPortal for Cancer Genomics (<http://www.cbioportal.org/index.do>) for their contribution to share the AML dataset on open access. In addition, we also would like to acknowledge the helpful comments on this paper received from our reviewers.

## Funding

The present study was supported in part by the National Natural Science Foundation of China (no. 81160075), the Natural Science Foundation of Guangxi (no. 0728124) and the

Self-raised Scientific Research Fund of the Ministry of Health of Guangxi Province (Z2014035).

## Availability of data and materials

The datasets used during the present study are available from the corresponding author upon reasonable request. All raw *IDH2* mutation and RNA-seq dataset of AML, which were included in the present study, can be downloaded from TCGA (<https://portal.gdc.cancer.gov/>) and cBioPortal for Cancer Genomics (<http://www.cbioportal.org/index.do>).

## Authors' contributions

RH, XiL and QL acquired the data and created a draft of the manuscript; RH, XiL, JL, JW, XS, XiaL, BL, FZ, YH and QL conducted and further performed the study, processed and analyzed the data; RH and QL revised and approved the final version of the manuscript. All authors read and approved the manuscript and agree to be accountable for all aspects of the research in ensuring that the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## Ethics approval and consent to participate

This study does not contain any studies with human participants or animals performed by any of the authors. Since all AML dataset included in this manuscript were obtained from The Cancer Genome Atlas and cBioPortal for Cancer Genomics, therefore, additional approval by an Ethics Committee is not necessary. In addition, the procedures of this manuscript were in accordance with the Helsinki declaration of 1964 and its later amendments.

## Patient consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Authors' information

Professor Qiaochuan Li: ORCID: <https://orcid.org/0000-0001-6255-8155>.

## References

- Short NJ, Rytting ME and Cortes JE: Acute myeloid leukaemia. *Lancet* 392: 593-606, 2018.
- Estey E and Döhner H: Acute myeloid leukaemia. *Lancet* 368: 1894-1907, 2006.
- Paschka P, Schlenk RF, Gaidzik VI, Habdank M, Krönke J, Bullinger L, Späth D, Kayser S, Zucknick M, Götze K, *et al*: *IDH1* and *IDH2* mutations are frequent genetic alterations in acute myeloid leukemia and confer adverse prognosis in cytogenetically normal acute myeloid leukemia with *NPM1* mutation without *FLT3* internal tandem duplication. *J Clin Oncol* 28: 3636-3643, 2010.
- Dang L, Yen K and Attar EC: *IDH* mutations in cancer and progress toward development of targeted therapeutics. *Ann Oncol* 27: 599-608, 2016.

5. Shih AH, Meydan C, Shank K, Garrett-Bakelman FE, Ward PS, Intlekofer AM, Nazir A, Stein EM, Knapp K, Glass J, *et al*: Combination targeted therapy to disrupt aberrant oncogenic signaling and reverse epigenetic dysfunction in *IDH2*- and *TET2*-mutant acute myeloid leukemia. *Cancer Discov* 7: 494-505, 2017.
6. Green CL, Evans CM, Zhao L, Hills RK, Burnett AK, Linch DC and Gale RE: The prognostic significance of *IDH2* mutations in AML depends on the location of the mutation. *Blood* 118: 409-412, 2011.
7. DiNardo CD, Ravandi F, Agresta S, Konopleva M, Takahashi K, Kadia T, Routbort M, Patel KP, Mark Brandt, Pierce S, *et al*: Characteristics, clinical outcome, and prognostic significance of *IDH* mutations in AML. *Am J Hematol* 90: 732-736, 2015.
8. Chao HY, Jia ZX, Chen T, Lu XZ, Cen L, Xiao R, Jiang NK, Ying JH, Zhou M and Zhang R: *IDH2* mutations are frequent in Chinese patients with acute myeloid leukemia and associated with *NPM1* mutations and *FAB-M2* subtype. *Int J Lab Hematol* 34: 502-509, 2012.
9. Wiseman DH, Williams EL, Wilks DP, Sun Leong H, Somerville TD, Dennis MW, Struys EA, Bakkali A, Salomons GS and Somervaille TC: Frequent reconstitution of *IDH2(R140Q)* mutant clonal multilineage hematopoiesis following chemotherapy for acute myeloid leukemia. *Leukemia* 30: 1946-1950, 2016.
10. Huang R, Liao X and Li Q: Identification of key pathways and genes in TP53 mutation acute myeloid leukemia: Evidence from bioinformatics analysis. *Oncotargets Ther* 11: 163-173, 2017.
11. Huang R, Liao X and Li Q: Identification and validation of potential prognostic gene biomarkers for predicting survival in patients with acute myeloid leukemia. *Oncotargets Ther* 10: 5243-5254, 2017.
12. Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson A, Hoadley K, Triche TJ Jr, Laird PW, Baty JD, *et al*: Cancer Genome Atlas Research Network: Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 368: 2059-2074, 2013.
13. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, *et al*: The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2: 401-404, 2012.
14. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, *et al*: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6: pii, 2013.
15. Robinson MD, McCarthy DJ and Smyth GK: edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139-140, 2010.
16. McCarthy DJ, Chen Y and Smyth GK: Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40: 4288-4297, 2012.
17. Liao X, Yang C, Huang R, Han C, Yu T, Huang K, Liu X, Yu L, Zhu G, Su H, *et al*: Identification of potential prognostic long non-coding rna biomarkers for predicting survival in patients with hepatocellular carcinoma. *Cell Physiol Biochem* 48: 1854-1869, 2018.
18. Wei HT, Guo EN, Liao XW, Chen LS, Wang JL, Ni M and Liang C: Genome-scale analysis to identify potential prognostic microRNA biomarkers for predicting overall survival in patients with colon adenocarcinoma. *Oncol Rep* 40: 1947-1958, 2018.
19. Huang W, Sherman BT and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44-57, 2009.
20. Jiao X, Sherman BT, Huang W, Stephens R, Baseler MW, Lane HC and Lempicki RA: DAVID-WS: A stateful web service to facilitate gene/protein list analysis. *Bioinformatics* 28: 1805-1806, 2012.
21. Maere S, Heymans K and Kuiper M: BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448-3449, 2005.
22. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, *et al*: STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43: D447-D452, 2015.
23. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, *et al*: The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 45 (D1): D362-D368, 2017.
24. von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen MA and Bork P: STRING: Known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res* 33: D433-D437, 2005.
25. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, *et al*: The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 38 (Suppl 2): W214-W220, 2010.
26. Mostafavi S, Ray D, Warde-Farley D, Grouios C and Morris Q: GeneMANIA: A real-time multiple association network integration algorithm for predicting gene function. *Genome Biol* 9 (Suppl 1): S4, 2008.
27. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, *et al*: The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313: 1929-1935, 2006.
28. Lamb J: The Connectivity Map: A new tool for biomedical research. *Nat Rev Cancer* 7: 54-60, 2007.
29. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, *et al*: PubChem substance and compound databases. *Nucleic Acids Res* 44: D1202-D1213, 2016.
30. Wang Y, Bryant SH, Cheng T, Wang J, Gindulyte A, Shoemaker BA, Thiessen PA, He S and Zhang J: PubChem BioAssay: 2017 update. *Nucleic Acids Res* 45: D955-D963, 2017.
31. Kuhn M, von Mering C, Campillos M, Jensen LJ and Bork P: STITCH: Interaction networks of chemicals and proteins. *Nucleic Acids Res* 36: D684-D688, 2008.
32. Kuhn M, Szklarczyk D, Franceschini A, Campillos M, von Mering C, Jensen LJ, Beyer A and Bork P: STITCH 2: An interaction network database for small molecules and proteins. *Nucleic Acids Res* 38 (Suppl 1): D552-D556, 2010.
33. Szklarczyk D, Santos A, von Mering C, Jensen LJ, Bork P and Kuhn M: STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res* 44: D380-D384, 2016.
34. Benjamini Y and Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57: 289-300, 1995.
35. Reiner A, Yekutieli D and Benjamini Y: Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 19: 368-375, 2003.
36. Benjamini Y, Drai D, Elmer G, Kafkafi N and Golani I: Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 125: 279-284, 2001.
37. Ahmad F, Mohota R, Sanap S, Mandava S and Das BR: Molecular evaluation of DNMT3A and IDH1/2 gene mutation: Frequency, distribution pattern and associations with additional molecular markers in normal karyotype Indian acute myeloid leukemia patients. *Asian Pac J Cancer Prev* 15: 1247-1253, 2014.
38. Lee CY, Tien HF, Hu CY, Chou WC and Lin LI: Marrow angiogenesis-associated factors as prognostic biomarkers in patients with acute myelogenous leukaemia. *Br J Cancer* 97: 877-882, 2007.
39. Trujillo A, McGee C and Cogle CR: Angiogenesis in acute myeloid leukemia and opportunities for novel therapies. *J Oncol* 2012: 128608, 2012.
40. Hussong JW, Rodgers GM and Shami PJ: Evidence of increased angiogenesis in patients with acute myeloid leukemia. *Blood* 95: 309-313, 2000.
41. Krell D, Mulholland P, Frampton AE, Krell J, Stebbing J and Bardella C: IDH mutations in tumorigenesis and their potential role as novel therapeutic targets. *Future Oncol* 9: 1923-1935, 2013.
42. Schaap FG, French PJ and Bovée JV: Mutations in the isocitrate dehydrogenase genes *IDH1* and *IDH2* in tumors. *Adv Anat Pathol* 20: 32-38, 2013.
43. Bluteau D, Glembofsky AC, Raimbault A, Balayn N, Gilles L, Rameau P, Nurden P, Alessi MC, Debili N, Vainchenker W, *et al*: Dysmegakaryopoiesis of FPD/AML pedigrees with constitutional *RUNX1* mutations is linked to myosin II deregulated expression. *Blood* 120: 2708-2718, 2012.
44. Okuda T, Nishimura M, Nakao M and Fujita Y: *RUNX1/AML1*: A central player in hematopoiesis. *Int J Hematol* 74: 252-257, 2001.
45. Zhang F, Yang B, Zhang K, Hou ML, Lu XC and Li YX: CCND1-BCL2 Gene Network: A direct target of Amifostine in human acute megakaryocytic leukemia cells. *Chem Biol Drug Des* 89: 681-693, 2017.



46. Park IK, Mishra A, Chandler J, Whitman SP, Marcucci G and Caligiuri MA: Inhibition of the receptor tyrosine kinase Axl impedes activation of the FLT3 internal tandem duplication in human acute myeloid leukemia: Implications for Axl as a potential therapeutic target. *Blood* 121: 2064-2073, 2013.
47. Janning M, Ben-Batalla I and Loges S: Axl inhibition: A potential road to a novel acute myeloid leukemia therapy? *Expert Rev Hematol* 8: 135-138, 2015.
48. Ben-Batalla I, Schultze A, Wroblewski M, Erdmann R, Heuser M, Waizenegger JS, Riecken K, Binder M, Schewe D, Sawall S, *et al*: Axl, a prognostic and therapeutic target in acute myeloid leukemia mediates paracrine crosstalk of leukemia cells with bone marrow stroma. *Blood* 122: 2443-2452, 2013.
49. Flotho C, Paulun A, Batz C and Niemeyer CM: AKAP12, a gene with tumour suppressor properties, is a target of promoter DNA methylation in childhood myeloid malignancies. *Br J Haematol* 138: 644-650, 2007.
50. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, *et al*: Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 478: 64-69, 2011.
51. Larsson CA, Cote G and Quintás-Cardama A: The changing mutational landscape of acute myeloid leukemia and myelodysplastic syndrome. *Mol Cancer Res* 11: 815-827, 2013.
52. Sauter C, Fehr J, Frick P, Gmuier J, Honegger H and Martz G: Acute myelogenous leukemia: Successful treatment of relapse with cytosine arabinoside, VP 16-213, vincristine and vinblastine (A-triple-V). *Eur J Cancer Clin Oncol* 18: 733-737, 1982.
53. Müller MR, Sauter C, Erni J and Martz G: Influence of a new relapse treatment for acute myeloid leukemia (AML) on in vitro granulopoiesis. *Anticancer Res* 3: 127-131, 1983.
54. Chau CH, Price DK, Till C, Goodman PJ, Chen X, Leach RJ, Johnson-Pais TL, Hsing AW, Hoque A, Tangen CM, *et al*: Finasteride concentrations and prostate cancer risk: Results from the Prostate Cancer Prevention Trial. *PLoS One* 10: e0126672, 2015.
55. Moroz A, Delella FK, Almeida R, Lacorte LM, Fávaro WJ, Deffune E and Felisbino SL: Finasteride inhibits human prostate cancer cell invasion through MMP2 and MMP9 downregulation. *PLoS One* 8: e84757, 2013.
56. Wiebe JP, Rivas MA, Mercogliano MF, Elizalde PV and Schillaci R: Progesterone-induced stimulation of mammary tumorigenesis is due to the progesterone metabolite, 5 $\alpha$ -dihydroprogesterone (5 $\alpha$ P) and can be suppressed by the 5 $\alpha$ -reductase inhibitor, finasteride. *J Steroid Biochem Mol Biol* 149: 27-34, 2015.
57. Seo JO, Yumnam S, Jeong KW and Kim SY: Finasteride inhibits melanogenesis through regulation of the adenylate cyclase in melanocytes and melanoma cells. *Arch Pharm Res* 41: 324-332, 2018.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.