Applied Biological Research 24(3): 280-287; (2022)
DOI: 10.5958/0974-4517.2022.00032.5

COMPUTATIONAL IDENTIFICATION OF CO-EXPRESSION NETWORKS, KEY GENES AND PATHWAYS IN ACUTE MYELOID LEUKEMIA

Nithya Rangasamy[1], P. Arulraj[2] and K.S. Santhy[3]*

[1,3]Department of Zoology, School of Biological Sciences, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore – 641 043, Tamil Nadu (India)
[2]Surgical Oncologist, GKNM Hospital, Coimbatore - 641 037, Tamil Nadu (India)
e-mail: [1]) 19phzof001@avinuty.ac.in; [2]) drarulp@gmail.com [3]) santhy_zoo@avinuty.ac.in

(Received 28 February, 2022; accepted 6 June, 2022)

ABSTRACT

Acute myeloid leukemia (AML) is characterized by proliferative, poorly differentiated cells of the hematopoietic system. The aim of this study was to identify biomarkers of AML by gene co-expression network analysis. Microarray dataset was retrieved from Gene Expression Omnibus (GEO) database. Weighted Co-expression Network Analysis (WGCNA) method was applied to evaluate key modules. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was performed using Database for Annotation, Visualization and Integrated Discovery (DAVID) web server. Then, hub genes were identified using Cytoscape software. Finally, expression of hub genes was validated using Gene Expression Profiling Interactive Analysis (GEPIA). A total of 1602 co-expressed genes with highest connectivity were identified in dark red and magenta modules. GO analysis suggested that co-expressed genes were mainly enriched in apoptotic process, signal transduction, cytosol, nucleus, protein binding and ATP binding activity. KEGG pathway analysis revealed that co-expressed genes were significantly enriched in MAPK signaling pathway and pathways in cancer. Protein-Protein Interaction network (PPI) demonstrated that the top 10 hub genes were identified. Survival analysis showed that genes downregulation of ITGB1, JUN and upregulation of ATM, MYC, NOTCH1, PTPN11 were associated with poor overall survival.

Keywords: Acute myeloid leukemia, biomarker, disease biology, gene co-expression, key genes

INTRODUCTION

Acute myeloid leukemia (AML) is a common type of hematologic neoplasm characterized by molecular characteristics; and its incidence rate is increasing every year (Estey, 2016). American Cancer Society in 2019 recorded 19,950 new cases of AML and 10,430 deaths due to AML in USA (Siegel *et al.,* 2019). Historically, genetic factors play an important role in cause of leukemia (Shih *et al.,* 2020). On the other hand, smoking, ionizing radiation, prior chemotherapy, Down's syndrome and exposure to chemicals such as benzene are major risk factors for leukemia (Puumala *et al.,* 2013). Indeed, molecular analysis of leukemic blasts from AML patients suggested that there was an obvious heterogeneity in the gene expression. However, the detailed pathophysiology of AML is still unclear, especially the complicated molecular mechanisms (Renneville *et al*., 2008). Therefore, the identification of more effective biomarkers and therapeutic targets, and better understanding of the mechanism of AML is the need of hour. In this context, co-expression analysis

uses gene expression levels to cluster genes into several modules based on similar biological functions and gene expression levels (Jang *et al.*, 2004). In last few decades, more innovative methods were proposed for co-expression network construction and analysis. WGCNA is an R package widely used to identify the key modules highly associated with clinical traits (Langfelder and Horvath, 2008). In addition, WGCNA is used to measure the relationship between modules and genes. Following identification of gene modules, gene enrichment analysis, PPI network analysis and survival analysis is usually performed to uncover if the gene set is associated for biological functions and overall survival. The purpose of this study is to investigate the co-expressed genes potentially involved in the AML prognosis.

Due to revolution in gene sequencing technology and emergence of open data-sharing platforms (GEO), an increasing number of ongoing studies are aimed at defining the gene expression profile associated with cancers. To this date, WGCNA method has been employed to construct co-expression modules for several studies, cutaneous systemic sclerosis-associated pulmonary arterial hypertension, prostate cancer, and gastric cancer (Fan *et al.*, 2018; Zhang., 2019; Zhang *et al.*, 2021). In this study, we considered a weighted gene co-expression network using AML expression profile data in GEO database, identified the key gene modules that affect the pathogenesis of AML, conducted enrichment analysis, and screened out six new biomarkers. This study was aimed to comprehensively analyze the co-expressed gene network so as to allow a better understanding of biological pathways and suggest six novel biomarkers for AML diagnosis and treatment for future personalized targeted therapy.

MATERIALS AND METHODS

Data collection and pre-processing

Gene expression data of 26 AML samples were collected from the microarray database GEO (*GSE9476, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE9476*).

Construction of gene co-expression network and hub-gene selection

Construction of gene co-expression network was performed by using R software with WGCNA package (Langfelder and Horvath, 2008). β was a soft threshold function that can emphasize a strong co-relation and make the network conform to the standard scale-free network. The function provides an appropriate β value ranging from 1 to 20. We utilized the pick soft threshold value to calculate the soft threshold by scale independence and mean connectivity analysis of the module. Then, through hierarchical clustering, co-expression modules were identified, and the tree was obtained. The minimum number of module genes was set as 30, and the modules were merged for the second time accordingly to the modules with a correlation more significant than 75%.

Functional enrichment analysis in the key modules

Functional enrichment analysis was performed on genes in the significant modules using the DAVID (*https://david.ncifcrf.gov/tools.jsp*) web server (Jiao *et al.*, 2012). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) biological pathway analysis was employed for annotating genes and gene products for identifying novel biological processes. The p-value < 0.05 was used as the cut-off criterion.

Identification of hub genes

The correlation weight value from each gene in the dark red and magenta modules was generated with the WGCNA package. The gene correlation was visualized using Cytoscape software by the cyto-Hubba plugin (Shannon *et al.*, 2003). The hub genes that were highly interconnected with nodes in a module were identified with the degree of connectivity method. The top 10 genes with a correlation weight value > 0.1 were defined as hub genes.

Validation of hub genes

GEPIA (*http://gepia.cancer-pku.cn/*) database (Tang *et al.*, 2017) was used to validate the expression difference of candidate genes among AML patients. Kaplan-Meier survival plots were applied to validate the hub genes prognosis.

RESULTS AND DISCUSSION

Data source and co-expression network

In present study, the dataset GSE9476 that included 26 AML samples was used. By bioinformatic method WGCNA, the genes were clustered into seven modules, of which two modules were the most closely related to the development and progression of AML. The scale-free network condition was satisfied when $\beta = 14$, the scale independence value achieved 0.14 and lower connectivity (Fig. 1). Hierarchical clustering dendrogram grouped the genes with similar expressions into same module (Fig. 2). A total of 1009 and 493 mutual gene probes were clustered into dark red and magenta gene modules. As previously mentioned, two methods were applied to find key modules. The results of first method showed that the Module Eigengene (ME) of dark red and magenta

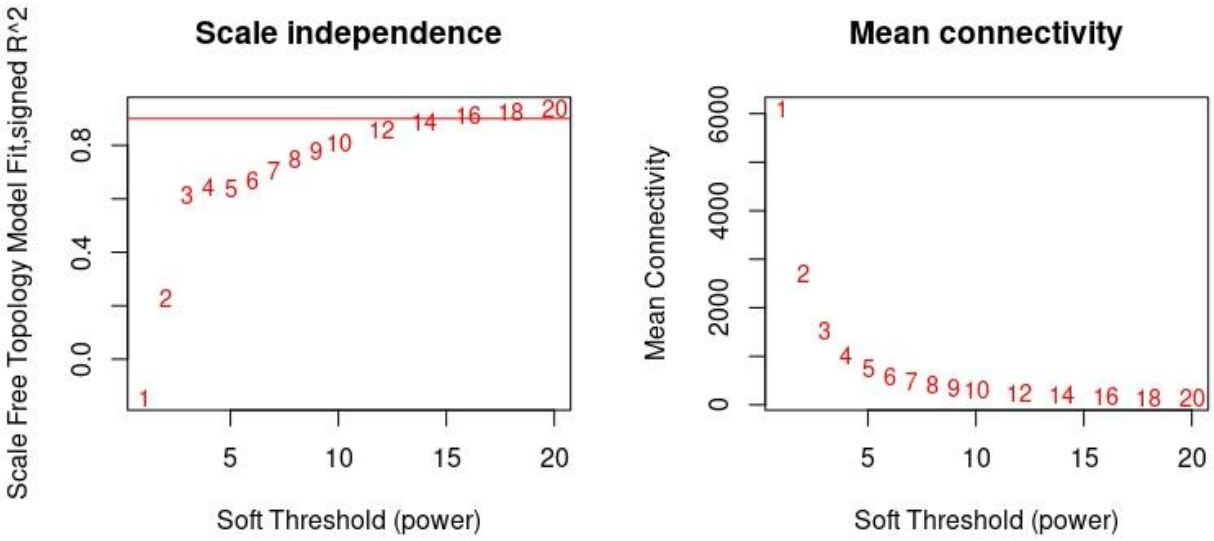

Fig. 1: Determination of network topology using various soft-threshold power

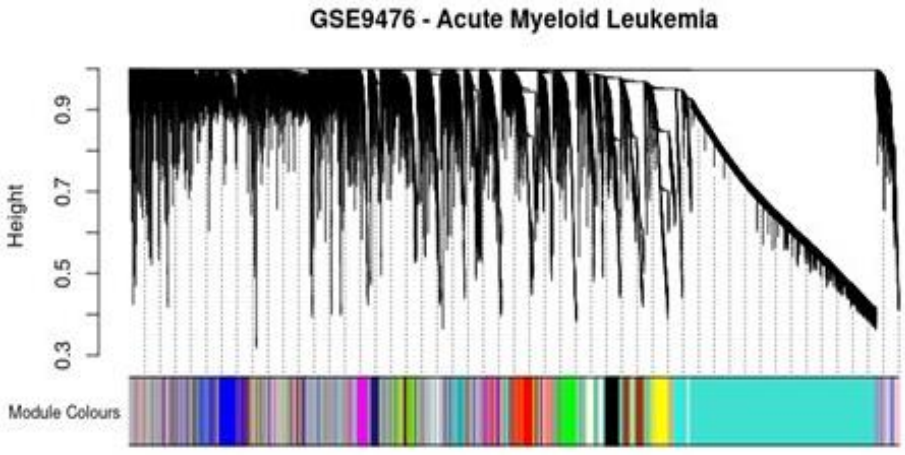

Fig. 2: Cluster dendrogram of gene co-expressed modules

modules, in its relation to AML (Fig. 3). The Gene Significance (GS) indicated that dark red and magenta modules were highly correlation (Fig. 4 A, B). These findings suggest that the dark red and magenta modules associated with metastasis in AML.

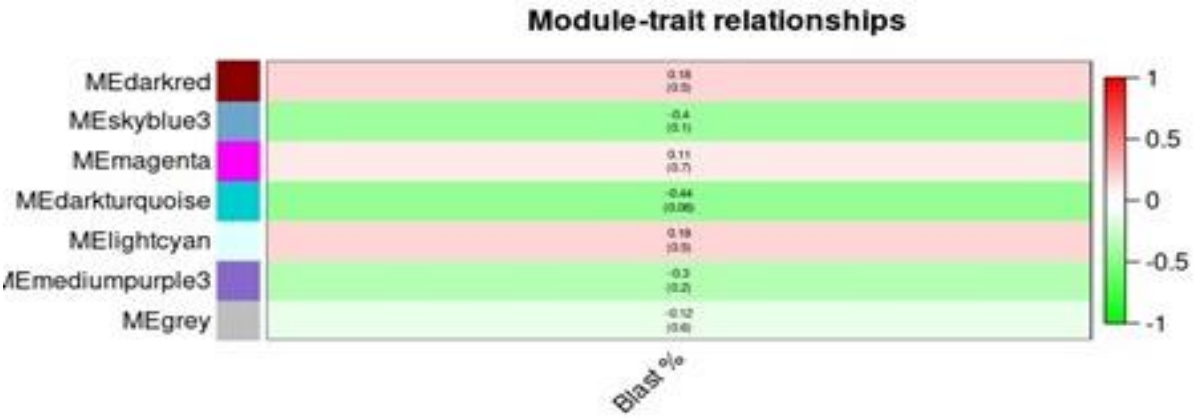

Fig. 3: Module-condition correlations with the significance level

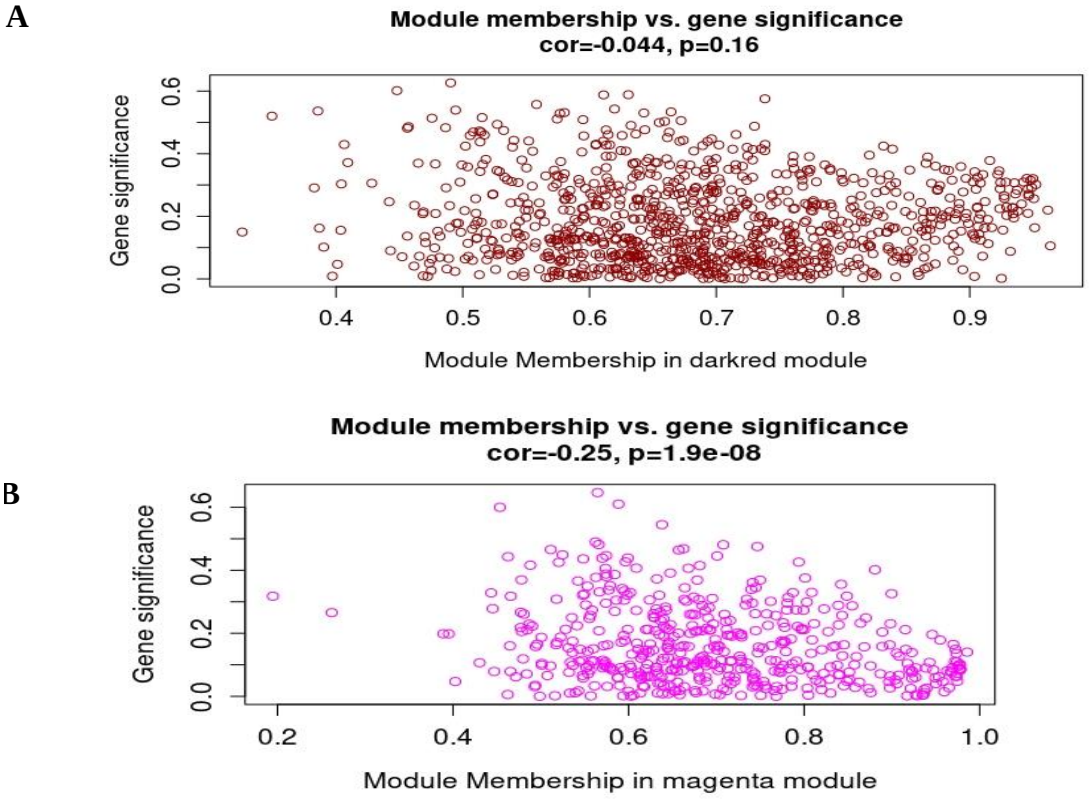

Fig. 4: Scattorplots of gene significance versus module membership for blast % associated samples

GO and KEGG pathway enrichment

Considering the value of weighted correlation 0.18, 1602 genes in the dark red and magenta modules were finally selected for functional enrichment analysis. GO analysis of all the genes in these two modules showed that the apoptotic process, positive and negative regulation of transcription, signal transduction, protein phosphorylation were the most abundantly represented in the biological process category. In the cellular component category, cytosol, nucleus, cytoplasm were the top three terms, while protein binding, metal ion binding and ATP binding activity were top molecular function terms (Table 1). KEGG enrichment analysis revealed that co-expressed genes were enriched in MAPK signaling pathway (hsa04010) and Pathways in cancer (hsa05200)

(Table 2). These findings, while primary, suggested that these biological processes and pathways plays important role in the AML progression. Additionally, the genes in the dark red and magenta modules were selected for hub gene screening.

Table 1: Functional enrichment analysis of co-expressed genes

Category	Terms	Count	Percent	P-value
GOTERM_BP	Apoptotic process	67	5.279748	4.48E-07
GOTERM_BP	Positive regulation of transcription, DNA-templated	73	5.752561	7.72E-07
GOTERM_BP	Protein phosphorylation	59	4.64933	1.33E-06
GOTERM_BP	Negative regulation of transcription, DNA-templated	54	4.255319	0.001353
GOTERM_BP	Signal transduction	94	7.407407	0.005991
GOTERM_BP	Regulation of transcription, DNA-templated	78	6.146572	0.007081
GOTERM_CC	Cytosol	468	36.87943	1.81E-22
GOTERM_CC	Nucleoplasm	332	26.16233	2.59E-14
GOTERM_CC	Nucleus	454	35.7762	2.47E-12
GOTERM_CC	Membrane	221	17.41529	4.68E-10
GOTERM_CC	Cytoplasm	402	31.67849	1.32E-08
GOTERM_CC	Extracellular exosome	156	12.29314	0.013336
GOTERM_MF	Protein binding	932	73.44366	5.23E-22
GOTERM_MF	ATP binding	147	11.58392	2.35E-07
GOTERM_MF	Metal ion binding	212	16.70607	8.29E-05
GOTERM_MF	Identical protein binding	144	11.34752	1.83E-04
GOTERM_MF	RNA binding	121	9.535067	0.001155
GOTERM_MF	DNA binding	107	8.431836	0.001414

BP: Biological Process, CC: Cellular Compound, MF: Molecular Function

Table 2: Pathway enrichment analysis of co-expressed genes

Category	Term	Count	%	P-value
KEGG_PATHWAY	hsa04010: MAPK signaling pathway	41	3.23089	3.92E-05
KEGG_PATHWAY	hsa05200: Pathways in cancer	56	4.412924	0.002232

KEGG: Kyoto Encyclopedia of Genes and Genomes

PPI network

A network of protein-protein interaction was constructed for all the genes by Cytoscape according to the STRING database in dark red and magenta modules. Ten genes namely AKT1, MYC, JUN, ATM, NOTCH1, PTPN11, ITGB1, MAPK14, TP53BP1, NFKBIA were identified as hub genes in PPI network (Fig. 5). Further, serine/ threonine kinase AKT promote the epithelial mesen-chymal transition, cellular motility and invasion in a wide variety of pathological conditions and it restricts invasive capacity of head and neck carcinoma cells (Brolih *et al.*, 2018). Meanwhile, previous studies have explored that the role of MYC in the pathogenesis of cancers and found that it has a positive and important role in transcription factor activation and molecular pathways (Gabay and Felsher, 2014). JUN activation is associated with the proliferation and angiogenesis in invasive breast cancer (Vleugel *et al.*, 2006). ATM has been associated with thymic lymphoma and several cancers (Cremona and Behrens, 2014).

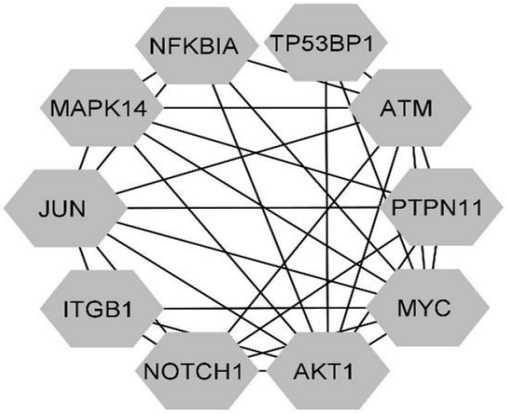

Fig. 5: Top 10 in network STRING network ranked by degree of connectivity method

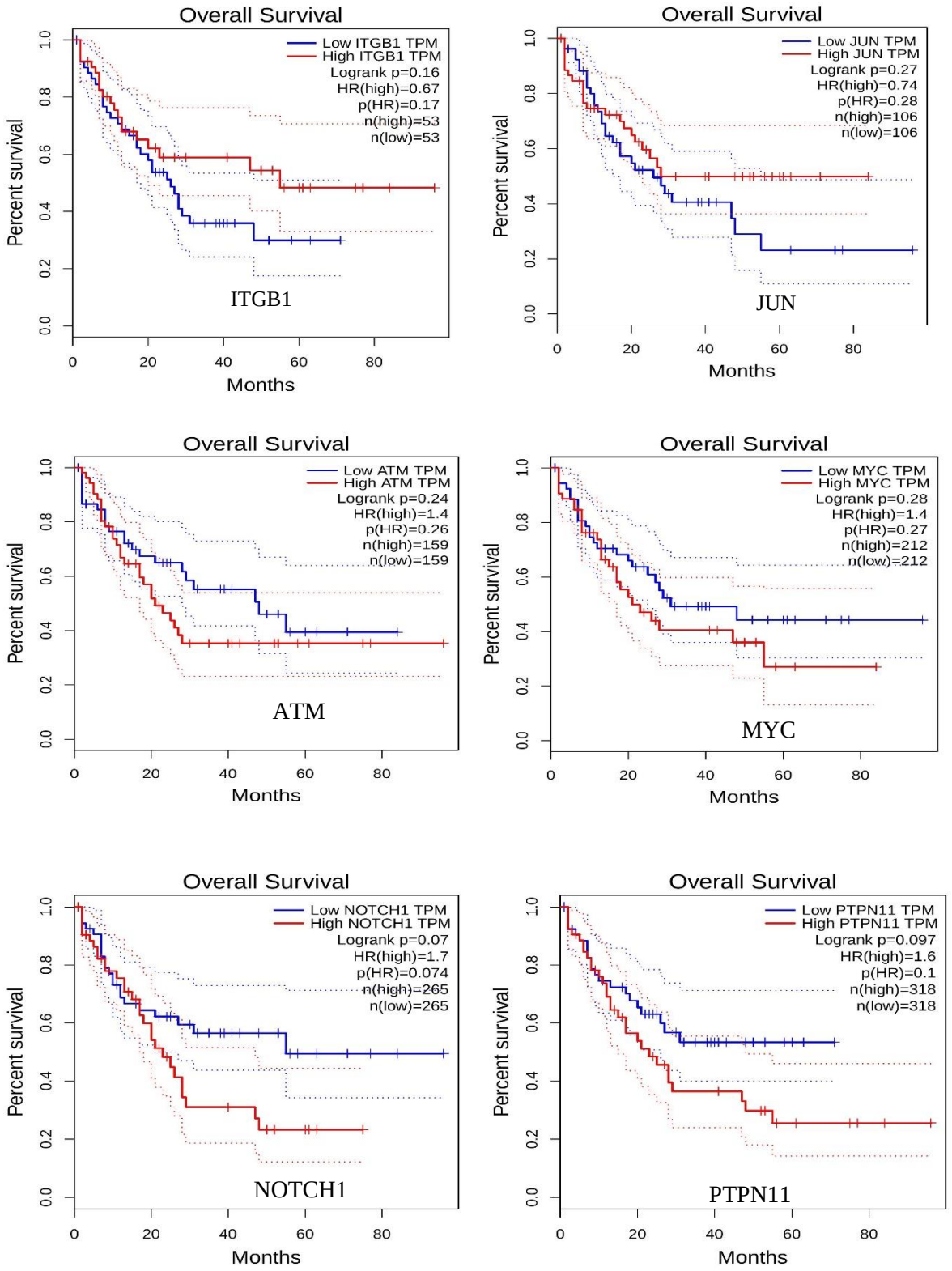

Fig. 6: Kaplan-Meier survival curve

Recent studies have shown that NOTCH1 is implicated in carcinogenesis in a variety of human malignancies by regulating many biological processes, including the cell growth, survival,

apoptosis, migration, and invasion (Guo *et al*., 2015). Several studies implicated that PTPN11 as a critical component of oncogenic signaling pathways and play a fairly broad role in cancer pathogenesis (Chan *et al.*, 2008). ITGB1 induces growth and invasion in a human colorectal cancer cell line through the hedgehog (Hh) signaling pathway in vitro and in vivo (Song *et al.*, 2015). MAPK14 and TP53BP1 were found in high levels in gastric cancer and breast cancer samples compared to normal samples (Mesquita *et al.*, 2020; De *et al.*, 2017). NFKBIA activity in gliomas is significantly higher than in normal brain tissues and phospho-IκBα protein levels have been shown to negatively correlate with tumor grade (Kinker *et al.*, 2016). Furthermore, to identify whether these ten hub genes were involved in the AML prognosis, we validated the relative mRNA expression of the ten genes in GEPIA database.

Hubgene validation

Finally, Kaplan Meier plots indicated that ATM, MYC, NOTCH1, PTPN11 genes were upregulated, of which ITGB1, JUN genes were downregulated and found to be correlated with poor overall survival in AML patients (Fig. 6). Therefore, these genes might be a promising diagnostic and prognostic biomarkers of AML proliferation and deserved for further studies.

Conclusion: In present study, we screened co-expressed genes and identified two modules and six potential genes associated with AML progression. The study may help in developing novel and promising targets for the diagnosis and treatment of AML in future. The limitations of relatively small sample size and the heterogeneity of AML the findings need to be interpreted with caution. The results need further validation by taking larger number of samples, and the mechanism needs to be investigated both *in vitro* and *in vivo*.

Acknowledgements: We thank the DST sponsored Centre for Machine Learning and Intelligence (CCMI), Avinashilingam Institute for Home Science and Higher Education for Women for providing the necessary computational facilities.

Conflict of interest: All the authors declare that they have no capitating interest.

Ethical statement: This study does not involve any active human participants as the openly accessible datasets have been utilized.

REFERENCES

Brolih, S., Parks, S.K., Vial, V., Durivault, J., Mostosi, L., Pouyssegur, J., Pages, G. and Picco, V. 2018. AKT1 restricts the invasive capacity of head and neck carcinoma cells harboring a constitutively active PI3 kinase activity. *BMC Cancer*, **18**(1): 249. [*https://doi.org/10.1186/s12885-018-4169-0*].

Chan, G., Kalaitzidis, D. and Neel, B.G. 2008. The tyrosine phosphatase Shp2 (PTPN11) in cancer. *Cancer Metastasis Reviews*, **27**(2): 179-192.

Cremona, C.A. and Behrens, A. 2014. ATM signalling and cancer. *Oncogene*, **33**(26): 3351-3360.

De Gregoriis, G., Ramos, J.A., Fernandes, P.V., Vignal, G.M., Brianese, R.C., Carraro, D.M., Monteiro, A.N., Struchiner, C.J., Suarez-Kurtz, G., Vianna-Jorge, R. and de Carvalho, M.A. 2017. DNA repair genes PAXIP1 and TP53BP1 expression is associated with breast cancer prognosis. *Cancer Biology and Therapy*, **18**(6): 439-449.

Estey, E. 2016. Acute myeloid leukemia: 2016 Update on risk-stratification and management. *American Journal of Hematology*, **91**(8): 824-846.

Fan, S., Liang, Z., Gao, Z., Pan, Z., Han, S., Liu, X., Zhao, C., Yang, W., Pan, Z. and Feng, W. 2018. Identification of the key genes and pathways in prostate cancer. *Oncology Letters*, **16**(5): 6663-6669.

Gabay, M., Li, Y. and Felsher, D.W. 2014. MYC activation is a hallmark of cancer initiation and maintenance. *Cold Spring Harbor Perspectives in Medicine*. **4**(6): a014241. [*https://doi.org/10.1101/cshperspect.a014241*].

Guo, L.X., Zhang, T., Xiong, Y. and Yang, Y.A. 2015. Roles of NOTCH1 as a therapeutic target and a biomarker for lung cancer: Controversies and perspectives. *Disease Markers*, **2015**: 520590. [*https://doi.org/10.1155/2015/520590*].

Jiang, D., Tang, C and Zhang, A. Cluster analysis for gene expression data: A survey. *IEEE Transactions on Knowledge and Data Engineering*, **16**(11): 1370-1386.

Jiao, X., Sherman, B.T., Stephens, R., Baseler, M.W., Lane, H.C. and Lempiciki, R.A. 2012. DAVID-WS: A stateful web service to facilitate gene/protein list analysis. *Bioinformatics*, **28**(13): 1805-1806.

Kinker, G.S., Thomas, A.M., Carvalho, V.J., Lima, F.P. and Fujita, A. 2016. Deletion and low expression of NFKBIA are associated with poor prognosis in lower-grade glioma patients. *Scientific Reports*, **6**: 24160. [*https://doi.org/10.1038/srep24160*].

Langfelder, P. and Horvath, S. 2008. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**: 559. [*https://doi.org/10.1186/1471-2105-9-559*].

Mesquita, F.P., Moreira-Nunes, C.A., da Silva, E.L., Lima, L.B., Daniel, J.P., Zuerker, W.J., Brayner, M., de Moraes, M. and Montenegro, R.C. 2020. MAPK14 (p38α) inhibition effects against metastatic gastric cancer cells: A potential biomarker and pharmacological target. *Toxicology In vitro*, **66**: 104839. [*https://doi.org/10.1016/j.tiv.2020.104839*].

Puumala, S.E., Ross, J.A., Aplenc, R. and Spector, L.G. 2013. Epidemiology of childhood acute myeloid leukemia. *Pediatric Blood and Cancer*, **60**(5): 728-733.

Renneville, A., Roumier, C., Biggio, V., Nibourel, O., Boissel, N., Fenaux, P. and Preudhomme, C. 2008. Cooperating gene mutations in acute myeloid leukemia: A review of the literature. *Leukemia*, **22**(5): 915-931.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13**: 2498-2504. [*https://doi.org/10.1101/gr.1239303*].

Shih, L., Delwel, A.H., Levine, R., Mason, R.L. and Melnick, A.M. 2020. Somatic mutations drive specific, but reversible, epigenetic heterogeneity states in AML. *Cancer Discovery*, **10**(12): 1934-1949.

Siegel, R.L., Miller, K.D. and Jemal, A. 2019. Cancer statistics, 2019. *CA: Cancer Journal for Clinicians*, **69**(1): 7-34.

Song, J., Zhang, J., Wang, J., Wang, J., Guo, X. and Dong, W. 2015. β1 integrin mediates colorectal cancer cell proliferation and migration through regulation of the Hedgehog pathway. *Tumour Biology, the Journal of the International Society for Oncodevelopmental Biology and Medicine*, **36**(3): 2013-2021.

Tang, Z., Li, C., Kang, B., Gao, G., Li, C. and Zhang, Z. 2017. GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Research*, **45**(W1): W98–W102. [*https://doi.org/10.1093/nar/gkx247*].

Vleugel, M.M., Greijer, A.E., Bos, R., van der Wall, E. and van Diest, P.J. 2006. c-Jun activation is associated with proliferation and angiogenesis in invasive breast cancer. *Human Pathology*, **37**(6): 668-674.

Zhang, L., Kang, W., Lu, X., Ma, S., Dong, L. and Zou, B. 2019. Weighted gene co-expression network analysis and connectivity map identifies lovastatin as a treatment option of gastric cancer by inhibiting HDAC2. *Gene*, **681**: 15-25.

Zhang, T., Huang, C., Luo, H., Li, J., Huang, H., Liu, X. and Zhan, S. 2021. Identification of key genes and immune profile in limited cutaneous systemic sclerosis-associated pulmonary arterial hypertension by bioinformatics analysis. *Life Sciences*, **271**: 119151. [*https://doi.org/10.1016/j.lfs.2021.119151*].