

# The Comparison of Genes and Molecular Pathway between Newly Diagnosed and Relapsed Acute Lymphocytic Leukemia (ALL) Patients: A Bioinformatics Analysis

Hadi Rezaeeyan PhD<sup>1</sup>, Negin Yousefi Chermehini MD<sup>2</sup>, Shana Ahadi MD<sup>3</sup>, Ehsan Kamali Yazdi MD<sup>4</sup>, Majid Ghanavat MD<sup>5\*</sup>

1. Asadabad school of medical sciences, Asadabad, Iran

2. Emergency Medicine Department, Faculty of Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

3. School of medicines, Jundishapur University of Medical Sciences, Ahvaz, Iran

4. Alborz University of Medical Sciences, Alborz, Iran

5. Assistant Professor of Pediatric Hematology and Oncology Department of Pediatrics, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

\*Corresponding author: Dr. Majid Ghanavat, Assistant Professor of Pediatric Hematology and Oncology Department of Pediatrics, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran. Email: ghanavatmajid@yahoo.com. ORCID ID: 0000-0002-9011-8453.

Received: July 05, 2025;  
Accepted: November 30, 2025

## Abstract

**Background:** Acute lymphoblastic leukemia (ALL) is among the most common cancers in children, which leads to the death of many patients. ALL patients include new cases and the relapsed ones. The common genes between these two conditions can largely help in managing patients and preventing future relapses. Therefore, this study investigated the genes and molecular pathways between the two conditions using bioinformatics.

**Materials and Methods:** In this bioinformatic study, GSE102301 database was used. After determining differentially expressed genes (DEGs), gene ontology (GO) and the related Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were analyzed (using GeneCodis4 software). Interactions between proteins were evaluated using STRING database and hub genes were identified based on the highest score. Finally, ROC curve analysis was used to determine the diagnostic value of hub genes.

**Results:** Based on the interaction between DEGs, 10 genes were selected as hubs, which included WDR75, MYBBP1A, DDX10, URB1, HEATR1, NOP14, PUM3, VARS1, EARS2 and SYK  $|\log_{2}FC| > 1$  and  $p < 0.05$ . The results showed that all the hub genes had diagnostic value (AUC=1).

**Conclusion:** Hub genes identified through bioinformatics analyses may serve as potential biomarkers; however, future studies need to examine these genes in clinical settings among the patients.

**Keywords:** Acute Lymphocytic Leukemia, Bioinformatics, Diagnosis, Relapse



## Introduction

Acute lymphoblastic leukemia (ALL) is one of the most common cancers in children, which has high mortality rate (1, 2). The average age of ALL diagnosis in patients is 15 years old. ALL is characterized by uncontrolled proliferation of immature lymphoid cells in bone marrow and peripheral blood (3). In this malignancy, the function and development of blood lymphocytes are disturbed. According to the statistical evidence, 6000 new cases are diagnosed annually and 1500 people die from the disease per year (1, 4).

ALL includes two groups of patients: 1- Patients who have recently been diagnosed with ALL and are so-called newly diagnosed; and 2- Patients who have been previously diagnosed and have undergone treatment; however, the disease has relapsed. Treatment strategies are different for each group of patients (5). For this purpose, treatment protocols are designed based on the patient's history, clinical symptoms, severity of the disease and response to the treatment. In addition, treatment strategies are determined in many cases based on the pathogenesis of the disease. Genes and molecular pathways in the category of hereditary factors are identified as factors of the pathogenesis of ALL. Recent diagnostic approaches have shown that changes in gene expression differ between newly diagnosed and relapsed patients. Therefore, their determination and differentiation can be helpful in better management of these patients (6-8).

Bioinformatics approaches are among the methods used nowadays to identify genes and molecular pathways in various diseases. Differentiating genes in newly diagnosed and relapsed cases can help in applying strategies for patient management. In ALL, this distinction has been made through bioinformatics approaches to a limited extent, so we investigated the issue in this study.

## Material and Methods

### *Characteristic of data*

In this study, the database (GSE102301) extracted from Gene Expression Omnibus (GEO) was used. The platform of GSE102301 was

GPL16791 Illumina HiSeq 2500 (Homo sapiens). The samples in this database included 3 newly diagnosed ALL samples and 4 relapsed ALL samples. In our study, differentially expressed genes (DEGs) were identified using the Limma package, with selection criteria set at  $|\log_2FC| > 1$  and  $p < 0.05$  (9). Limma employs an empirical Bayes approach to moderate variance estimates, which enhanced statistical reliability.

### *GO and KEGG pathways*

GeneCodis4 software was used for gene ontology (GO) analysis related to DEGs. For this purpose, molecular functions (MF), cellular components (CC) and biological process (BP) factors related to DEGs were identified. Also, GeneCodis4 software was used to analyze the Kyoto Encyclopedia of Genes and Genomes (KEGG) and determine the related molecular pathways.  $p < 0.05$  was determined as a significant level for the analyses.

### *Designing the protein-protein interaction*

Protein-protein interaction (PPI) was used to evaluate the interaction and network between proteins and DEGs. STRING database was used to design and specify the interaction between proteins. For PPI analysis, minimum required interaction score of  $> 0.4$  was used. Cytoscape software was used to observe these interactions.

### *ROC curve analysis*

The diagnostic value of hub genes was evaluated using receiver operating characteristic (ROC) and area under the curve (AUC) analyses. The criteria used for evaluation were ROC and  $AUC > 0.90$ , at  $p\text{-value} < 0.05$ .

## Results

### *Identifying DEGs*

In this study, 72 DEGs were identified in relapsed ALL compared to the newly diagnosed subjects. Out of 72 DEGs, 35 genes were up-regulated and 37 genes were down-regulated. The list of DEGs and additional information is shown in Figure 1.

### *GO and KEGG enrichment analyses*

The results showed that, in the BP category, the genes were mainly involved in processes such as DNA repair, positive regulation of transcription by RNA polymerase I and de novo biosynthetic processes of XMP, AMP and IMP.

In the MF category, they were primarily involved in ATP binding, aminoacyl-tRNA ligase activity and G-quadruplex DNA binding. In the CC category, they were mainly enriched in the nucleoplasm, nucleolus and 90S preribosome. In the KEGG pathway analysis, they were predominantly involved in aminoacyl-tRNA biosynthesis, cellular senescence and Fc gamma R-mediated phagocytosis (Figure 2 and 3).

#### *Designing the PPI network and identifying hub genes*

PPI of DEGs was evaluated based on STRING. The results showed that the total nodes and edges in PPI were 56 and 88, respectively. Based on the interaction between DEGs, 10 genes were selected as hub genes, including WDR75, MYBBP1A, DDX10, URB1, HEATR1, NOP14, PUM3, VARS1, EARS2 and SYK (Figure 4).

#### *Results of ROC curve analysis*

Based on the analyses, the results showed that all 10 hub genes had AUC=1. Therefore, all the hub genes had a diagnostic value (Figure 5).

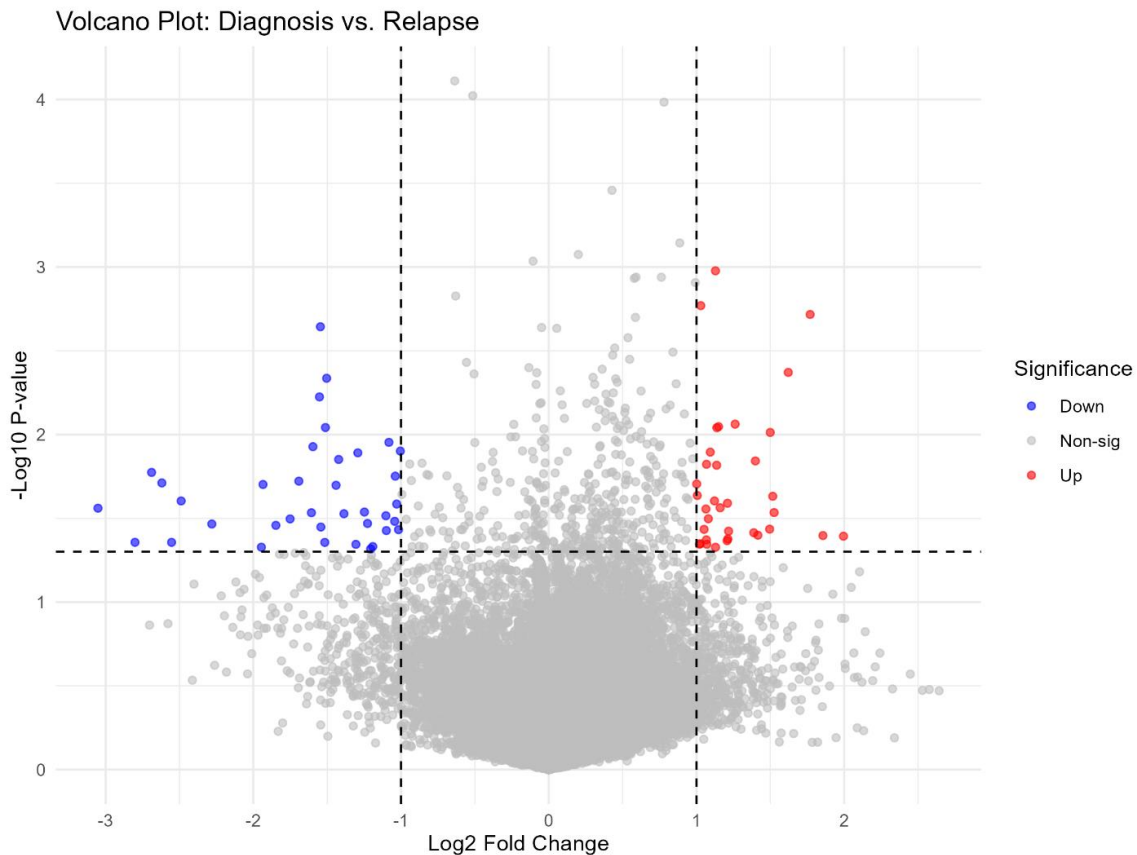
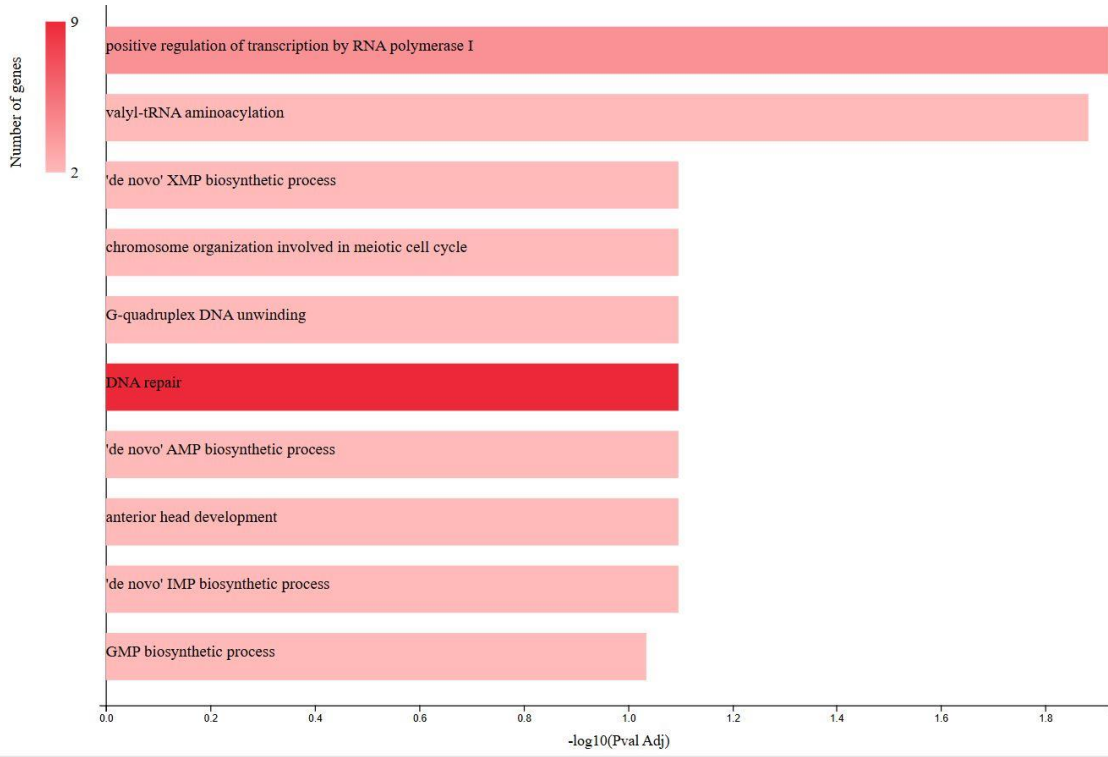
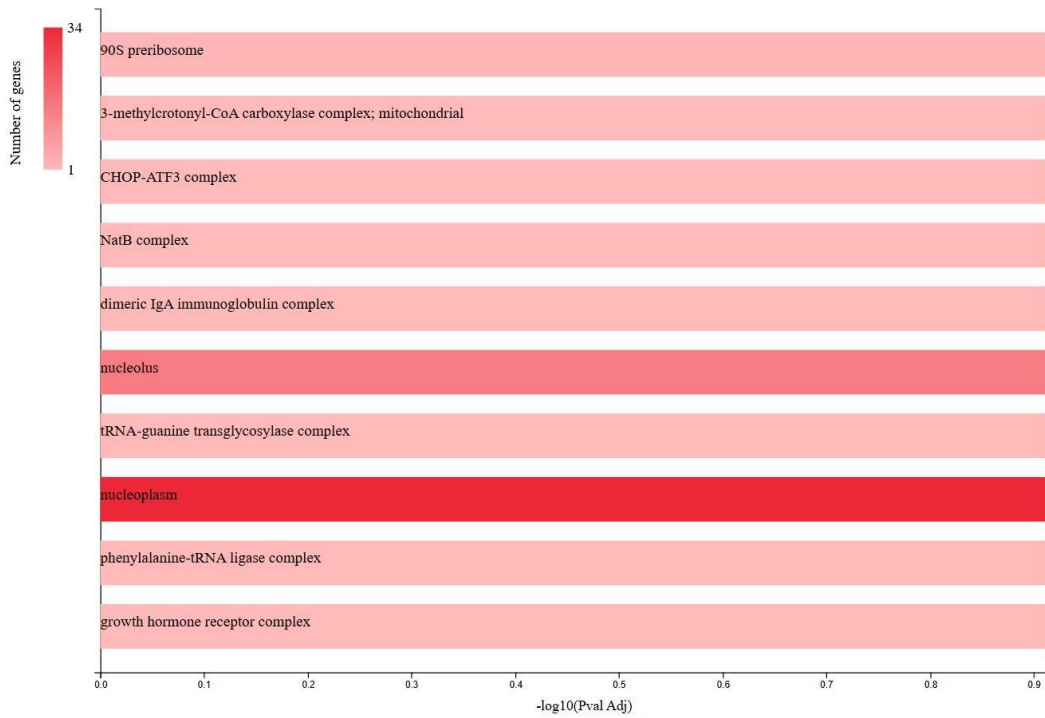


Figure 1. Volcano plot of differentially expressed genes. This figure shows DEGs which are up- or down-regulated as well as the non-significance between newly diagnosed and relapsed patients.



(A)



(B)

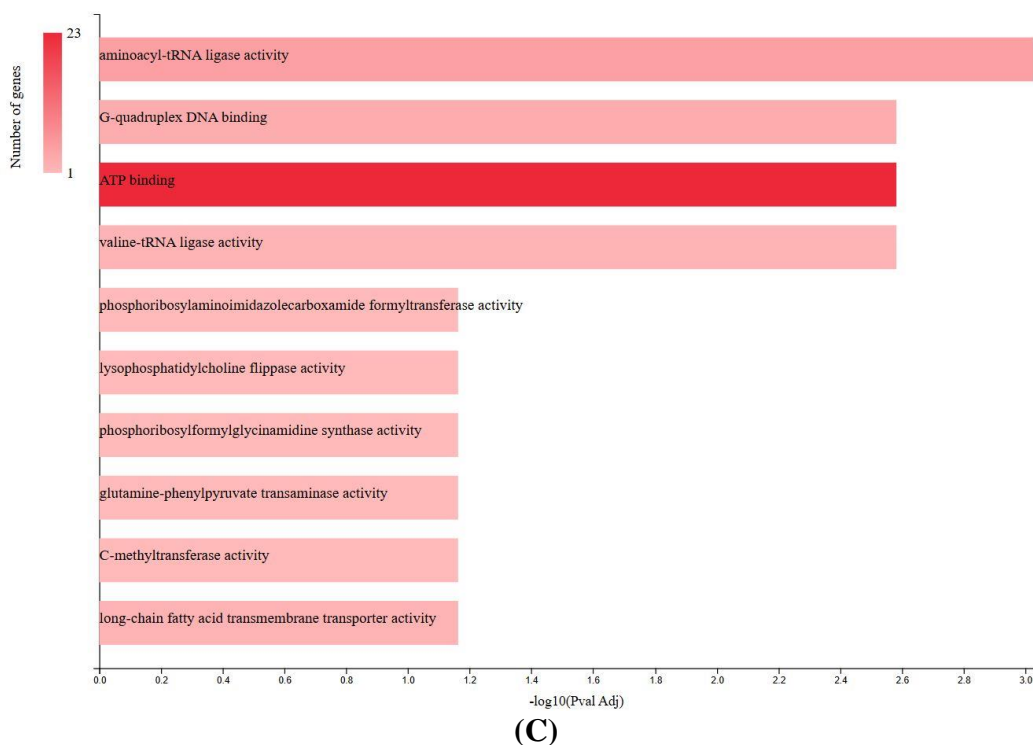


Figure 2. Gene ontology-related DEGs: A: Biological process, B: Cellular component, C: Molecular function. The colors are determined based on the number of genes involved in each pathway, with increasing the number of genes elevating the color intensity, while decreasing color intensity indicating a lower number of genes in each pathway.

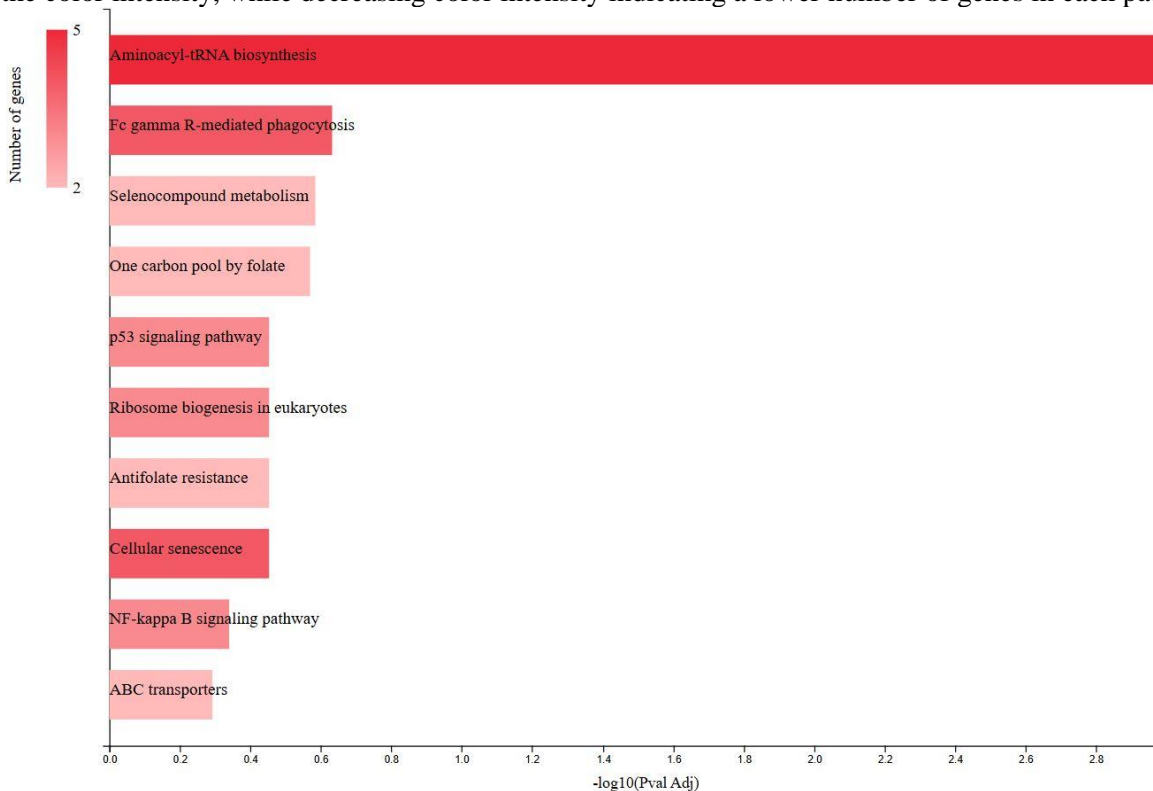


Figure 3. KEGG analysis of DEGs. The KEGG mostly involved in phagocytosis, biosynthesis and cellular senescence.

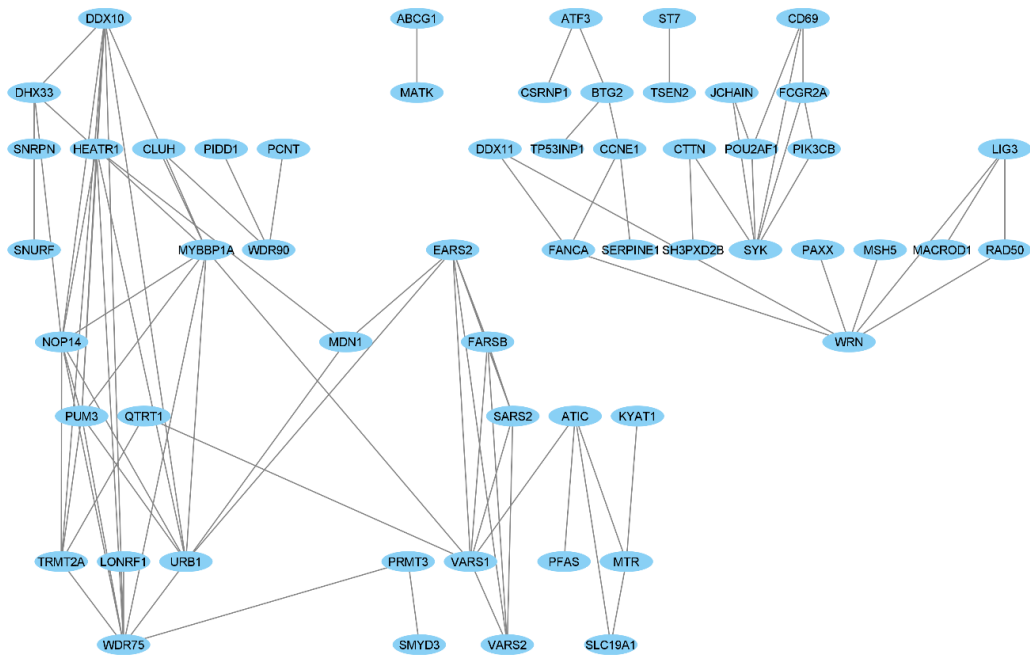


Figure 4. Protein-protein interaction (PPI) between DEGs. PPI was evaluated by STRING and Cytoscape. Nodes represent proteins encoded by DEGs and edges represent functional interactions.

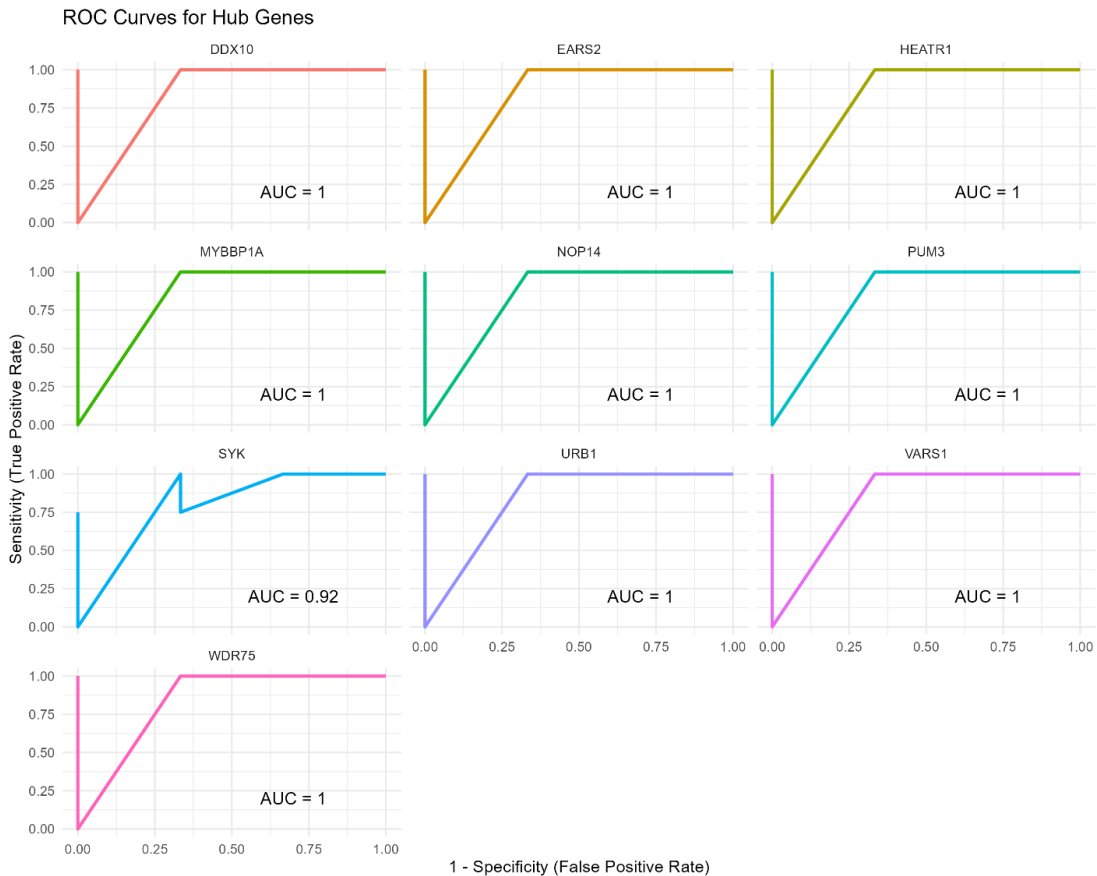


Figure 5. ROC curve analysis for evaluating survival of patients. The results showed that all hub genes had AUC=1.

## Discussion

In this study, the main genes, the expression of which was altered among newly diagnosed and relapsed ALL patients, included WDR75, MYBBP1A, DDX10, URB1, HEATR1, NOP14, PUM3, VARS1, EARS2 and SYK. In other words, these genes were shown to have a major role as hub genes in ALL.

The results demonstrated that WDR75 had the highest score among the hub genes. Previous studies have shown that the WDR gene family plays an important role in the pathogenesis of malignancies. Moreover, WDR can act as an oncogene in solid tumors, including colorectal and bladder, and cause cell proliferation (10, 11). Based on the role of WDR75, this gene is likely to play a role in the occurrence and recurrence of ALL because, in addition to affecting malignant cells, it prevents their apoptosis. Therefore, its targeting could be a therapeutic approach for managing patients.

MYBBP1A was another gene that was shown to be altered in expression among newly diagnosed and relapsed ALL patients. Previous evidence has shown that MYBBP1A is involved in asparagine metabolism. The breakdown of asparagine to aspartic acid and ammonia may be required for the cell cycle. Previous evidence has also demonstrated that mutations in MYBBP1A can confer resistance to asparaginase, a treatment strategy for ALL patients (12, 13). In addition, the increased expression of MYBBP1A in ALL patients can significantly increase the risk of thrombosis in patients (14). According to the literature, that uncontrolled asparagine levels can increase the risk of relapse in patients. Therefore, the MYBBP1A gene could play a significant role in ALL relapse (15).

DDX10 was another gene identified as a hub gene in this study. Evidence suggests that DDX10 is involved in T-cell differentiation. The NUP98-DDX10 fusion can promote the proliferation of human CD34+ cells and prevent their differentiation (16). In other words, this fusion causes leukemogenesis (17). Therefore, DDX10 could play a more prominent role in disease recurrence because if it is not targeted, recurrence and progression will occur.

URB1 also plays a role similar to DDX10. Decreased expression of URB1 can lead to decreased numbers of HSCs and, ultimately, reduced differentiation of T-cells. Further evaluations have shown that URB1 is the downstream of mTOR. Therefore, impaired URB1 function can be associated with decreased proliferation of HSCs and, ultimately, reduced differentiation of T-cells (18, 19). On the other hand, NOP14 can play a role in the biogenesis of ribosomal components, including URB1. Evidence suggests that NOP14 increases the proliferation of ALL cells and their resistance to rapamycin through the mTOR/AKT/ERK pathway (20). PUM3 was another hub gene that scored highly in this study. The PUM family is a group of genes that play a role in the pathogenesis of cancers, especially ALL. Bai et al. showed that PUM3 expression increased in ALL patients compared to controls. They also demonstrated that increased PUM3 expression in patients could be associated with poor prognosis. In other words, PUM3 was an oncogene associated with malignant cell proliferation (21). Other works have shown the PUM family controls the cell cycle and growth of malignant cells by regulating the expression of FOXp1/p21/p27 (22).

Given that PUM3 can be a regulator of URB1 and cause proliferation and self-renewal of cancer stem cells, its activity can cause relapse despite disease improvement.

Thandapani et al. represented that VARS1 gene expression increased in T-ALL samples. VARS1 causes the production of valine tRNAs which itself cause the assembly of complex I through the consumption of the amino acid valine. The assembly of complex I leads to mitochondrial synthesis and increased oxidative phosphorylation and cellular respiration. These factors lead to the proliferation of T-ALL cells. Therefore, targeting VARS1 or reducing the consumption of valine-containing foods could disrupt the assembly of complex I and prevent cell proliferation. Therefore, this gene seems to play a greater role in newly diagnosed patients (23).

This study has certain limitations; it is entirely based on publicly available RNA-seq data and relies solely on bioinformatics analyses. Therefore,

experimental validation could further strengthen the findings. Since the dataset corresponds to the relapsed form of the disease, the number of available samples is limited. Conducting similar analyses with a larger sample size in future studies could enhance the reliability and robustness of the results.

## Conclusion

The results of bioinformatics analysis showed that ten genes including WDR75, MYBBP1A, DDX10, URB1, HEATR1, NOP14, PUM3, VARS1, EARS2 and SYK could play a significant role as hub genes in patients with new diagnosis or relapse of ALL. However, future works need to evaluate the role of these genes in patients in clinical settings.

## Availability of Data

Data availability is the corresponding author's responsibility.

## Ethical Considerations

IR.ASAUMS.REC.1404.004.

## Acknowledgements

The authors thank all the sites and participants involved in this work. Hereby, we would like to thank the Research Administration of Asadabad School of Medical Sciences. Non AI used for write the manuscript.

## Authors' Contributions

H.R and M.Gh design the study and revise the manuscript. N.Y.Ch, Sh.A, E.K.Y extract the data, analysis the data and write the manuscript.

## Funding

This research was supported by the Asadabad School of Medical Sciences (Research project number 195).

## Conflict of Interest

The authors declare that they have no conflict of interest.

## References

- Ekpa QL, Akahara PC, Anderson AM, Adekoya OO, Ajayi OO, Alabi PO, et al. A Review of Acute Lymphocytic Leukemia (ALL) in the Pediatric Population: Evaluating Current Trends and Changes in Guidelines in the Past Decade. *Cureus* 2023; 15(12):e49930- e49930.
- Sohrevardi S, Ghodsi boushehri Y, Sheikhpour E, Nourian A, Dehghan Mongabadi H, Hashemi A. The effect of echinacea in the prevention of upper respiratory infections in children with acute lymphoblastic leukemia. *Iran J Ped Hematol Oncol* 2023; 13(1):1-6.
- Inaba H, Mullighan CG. Pediatric acute lymphoblastic leukemia. *Haematologica* 2020; 105(11):2524-2539.
- Brown P, Inaba H, Annesley C, Beck J, Colace S, Dallas M, et al. Pediatric acute lymphoblastic leukemia, version 2.2020, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* 2020; 18(1):81-112.
- Buchmann S, Schrappe M, Baruchel A, Biondi A, Borowitz M, Campbell M, et al. Remission, treatment failure, and relapse in pediatric ALL: an international consensus of the Ponte-di-Legno Consortium. *Blood* 2022; 139(12):1785-1793.
- Dobson SM, García-Prat L, Vanner RJ, Wintersinger J, Waanders E, Gu Z, et al. Relapse-fated latent diagnosis subclones in acute B lineage leukemia are drug tolerant and possess distinct metabolic programs. *Cancer Discov* 2020; 10(4):568-587.
- Koschade SE, Stratmann JA, Finkelmeier F, Wagner S, Chromik J, Steffen B, et al. Relapse surveillance of acute myeloid leukemia patients in first remission after consolidation chemotherapy: diagnostic value of regular bone marrow aspirations. *Ann Hematol* 2022; 101(8):1703-1710.
- Vicente C, Cools J. The origin of relapse in pediatric T-cell acute lymphoblastic leukemia. *haematologica* 2015;100(11):1373-1375.
- Chen C, Ma Y, Gao Y, Ge H, Zhang X. Prognostic significance of neutrophil extracellular trap-related genes in childhood acute lymphoblastic leukemia: insights from

multi-omics and in vitro experiment. *Hematology* 2025; 30(1):2452701-2452701.

10. Yuan Y, Qi G, Shen H, Guo A, Cao F, Zhu Y, et al. Clinical significance and biological function of WD repeat domain 54 as an oncogene in colorectal cancer. *Int J Cancer* 2019; 144(7):1584-1595.

11. Wei X, Wang B, Wu Z, Yang X, Guo Y, Yang Y, et al. WD repeat protein 54-mediator of ErbB2-driven cell motility 1 axis promotes bladder cancer tumorigenesis and metastasis and impairs chemosensitivity. *Cancer Lett* 2023; 556:216058-216058.

12. Abaji R, Roux V, Yssaad IR, Kalegari P, Gagné V, Gioia R, et al. Characterization of the impact of the MYBBP1A gene and rs3809849 on asparaginase sensitivity and cellular functions. *Pharmacogenomics* 2022; 23(7):415-430.

13. Abaji R, Gagné V, Xu CJ, Spinella J-F, Ceppi F, Laverdière C, et al. Whole-exome sequencing identified genetic risk factors for asparaginase-related complications in childhood ALL patients. *Oncotarget* 2017; 8(27):43752-43752.

14. Mootosamy C, Kondyli M, Serfaty SA, Tremblay D-É, Gagné V, Ribère M, et al. IL16 and factor V gene variations are associated with asparaginase-related thrombosis in childhood acute lymphoblastic leukemia patients. *Pharmacogenomics* 2023; 24(4):199-206.

15. Kawedia JD, Liu C, Pei D, Cheng C, Fernandez CA, Howard SC, et al. Dexamethasone exposure and asparaginase antibodies affect relapse risk in acute lymphoblastic leukemia. *Blood* 2012; 119(7):1658-1664.

16. Crescenzi B, Nofrini V, Barba G, Matteucci C, Di Giacomo D, Gorello P, et al. NUP98/11p15 translocations affect CD34+ cells in myeloid and T lymphoid leukemias. *Leuk Res* 2015; 39(7):769-772.

17. Yassin ER, Abdul-Nabi AM, Takeda A, Yaseen NR. Effects of the NUP98-DDX10 oncogene on primary human CD34+ cells: role of a conserved helicase motif. *Leukemia* 2010; 24(5):1001-1011.

18. Jiao L, Liu Y, Yu X-Y, Pan X, Zhang Y, Tu J, et al. Ribosome biogenesis in disease: new players and therapeutic targets. *Signal Transduct Target Ther* 2023; 8(1):15-15.

19. Cai P, Mao X, Zhao J, Luo L. Ribosome biogenesis protein Urb2 regulates hematopoietic stem cells development via P53 pathway in zebrafish. *Biochem Biophys Res Commun* 2018; 497(2):776-782.

20. Yan X, Kuang B-H, Ma S, Wang R, Lin J, Zeng Y-X, et al. NOP14-mediated ribosome biogenesis is required for mTORC2 activation and predicts rapamycin sensitivity. *J Biol Chem* 2024; 300(3):105681-105681.

21. Bai T, Liu N. RNA-binding protein PUM2 promotes T-cell acute lymphoblastic leukemia via competitively binding to RBM5 3' UTR with miR-28-5p. *Eur J Haematol* 2023; 110(5):498-509.

22. Díaz-Muñoz MD, Turner M. Uncovering the role of RNA-binding proteins in gene expression in the immune system. *Front Immunol* 2018; 9:1094-1094.

23. Thandapani P, Kloetgen A, Witkowski MT, Glytsou C, Lee AK, Wang E, et al. Valine tRNA levels and availability regulate complex I assembly in leukaemia. *Nature* 2022; 601(7893):428-433.