33 (2022) 201068

Contents lists available at ScienceDirect

Human Gene

journal homepage: www.journals.elsevier.com/human-gene

# Bioinformatics analysis of molecular pathways and key candidate biomarkers associated with human bone marrow hematopoietic stem cells (HSCs) micro-array gene expression data

Emine Güven [a], Sevinç Akçay [b,*]

[a] Department of Biomedical Engineering, Düzce University, Düzce, Turkey
[b] Department of Molecular Biology and Genetics, Kırşehir Ahi Evran University, Kırşehir, Turkey

ARTICLE INFO

*Keywords:*
HSC
Aging
Differentially expressed genes
Gene ontology pathway enrichment
UBC
PTK2
TCF7L2

ABSTRACT

An identification of the molecular properties of glioma stem cells (GSCs) has led studies in clinical use, stem cell identification, and highly-effective usage. The GSE32719 contains a total expression of 54,676 genes of healthy human bone marrow HSCs in 14 young (20–31 years), 5 middle (42–61 years), and a lot of old (65–85 years) age groups. The researchers of this study described age-related changes in the human HSC population, using the gene expression profile of significance analysis to discover differentially expressed genes (DEGs) between each age group. The DEGs were subjected to significantly enriched biological processing that decoded the increase and functional decline in the HSC population. The GSE dataset analysis was conducted by the GEOquery package in Bioconductor. Using the Biobase and gplots packages, 453 DEGs were screened. DEGs analyses were conducted by gene ontology (GO) pathway enrichment and Kyoto Encylopedia of Genes and Genomes (KEGG) enrichment analysis. The Hippo signaling pathway was observed to be significant using the GO pathway enrichment analysis, which was previously reported as an effective pathway in cancers, including AML. A protein-protein interaction (PPI) network was constructed; then based on that, a subnetwork module analysis for the Hippo signaling pathway was made. Additionally, the GO pathway enrichment analysis revealed 'cellular process', 'cellular metabolic process', 'metabolic process', 'biogenesis', and 'vasculogenesis biological processes', which are involved in a wide of biological activities such as metabolic regulation, cell growth, and proliferation. Our findings offer silico evidence for candidate genes, such as the UBC, PTK2, and TCF7L2, that may be promising biomarkers for the translation approach associated with HSC population age-related diseases.

## 1. Introduction

Bone marrow is a special site where blood cells are covered with structural stromal cells. It is a spongy adipose tissue found in the bones such as femurs, rib cage, ribs, pelvis, and human skull. Bone marrow is fed with specialized blood vessels to contribute to the circulation. The specialized phenestra capillary, called sinusoid, penetrates the extra-cellular/extracellular matrix (ECM), a sponge-like matrix produced by reticular fibroblasts (Pinho and Frenette, 2019).

The aforementioned proteins and cells place the hematopoietic cells in separate compartments. Similarly, hematopoietic cells, cytokines, stromal cells, ECM, bidothelial cells, growth factors, and chemokines contain special microenvironment in the bone marrow (Dar et al., 2006; Morrison and Scadden, 2014). Recent improvements cover new markers for Human Bone Marrow Hematopoietic Stem Cells (HSCs) and niche stem cells, systematic analysis of expression data sets of niche factors, implementation of genetics for functional in vivo identification of niche cells, and improved imaging techniques. Stem cells have the ability to divide for long hours in the living body, to be able to regenerate and transform into other tissue cells by differentiating according to the needs of the body (Fraga and Esteller, 2007).

A detailed examination of the cellular and molecular properties of HSCs has conducted studies in clinical use, stem cell identification, highly effective use. Bone marrow microenvironment is an ideal area to support healthy and yet it might also support malignant hematopoiesis, which performs a fundamental position in the growth and development of leukemia and different cancer types (Fraga and Esteller, 2007; Kuranda et al., 2011; Perry and Li, 2007; Pinho and Frenette, 2019).

* Corresponding author.
*E-mail address:* sevinc.akcay@ahievran.edu.tr (S. Akçay).

https://doi.org/10.1016/j.humgen.2022.201068
Received 15 April 2022; Received in revised form 28 April 2022; Accepted 1 June 2022
Available online 10 June 2022

Blast cells decline during normal maturation, and they begin to accumulate in the blood, with the bone marrow, in acute myeloid leukemia (AML). The body remains vulnerable because leukocytes cannot form. Red blood cell and thrombocyte production is impaired in the bone marrow due to abnormal proliferation of myeloblasts. Thus, anemia infection and platelet count reduces (Boyd et al., 2017). Age-related HSC diseases, such as anemia and AML, are not limited to a particular part of the body since its onset, but they can spread to the blood, lymphatic tissue, and all other organ systems from the bone marrow. As with many other leukemia, they are defined as malignant systemic diseases. When HSC and other hematopoietic progenitor cells are evaluated, it was discovered that the elderly HSC frequently increased (Pang et al., 2011).

Microarray technology has several applications including identification of therapeutic drugs and biomarkers in cancer research. Recent research suggests performing a comprehensive gene expression study of HSCs using publically accessible gene expression data to uncover age-related illnesses gene expression changes linked to AML. (Roushangar and Mias, 2019). Furthermore, relative research on differentially expressed genes (DEGs) is limited, and a reliable biomarker profile would be required to design distinct medications for different age groups. (Appelbaum et al., 2006; Kuranda et al., 2011). The expression alteration of proteins in the development and growth of AML and related diseases requires broad investigation.

Raw data (GSE32719) of healthy samples are retrieved through GEO, that is a public data repository for the storing of microarray and sequence-based data and their retrieval (Pang et al., 2011). The DEGs of HSC data between different age groups were determined by comparing gene expression. The DEGs were screened using Gorilla and DAVID Gene Ontology (GO) (Eden et al., 2009; Huang et al., 2007). By studying their hub nodes globally and between different age groups developing PPI networks, the objective of this project is to investigate the molecular mechanisms associated with age-related diseases in HSC population.

Previous research investigating age-related changes in human HSC should be validated by in vivo and in vitro studies studies with in silico analysis. Various agents are assessed and employed to prevent AML, including monoclonal antibodies, new formulation of old drugs, FLT3, and IDH1/2 inhibitors (Bohl et al., 2019; Luppi et al., 2018). Several microarray research has been carried out to discover the pathobiology of HSCs age-related formation and possible therapeutic targets. More efficient treatments are needed to improve therapeutic responses to HSC-related diseases.

Our investigation was specified as an “in silico” strategy to produce novel understanding related to the missing genetic drivers and key agents of the HSCs as well as to locate potential key molecular pathways, which may be targeted for therapeutic strategies and used for HSC-related diseases. In this project, the researchers used microarray data sets of human bone marrow tissue and conducted connective analysis on the DEGs with bioinformatics tools. This study also focused on hippo signaling pathway transcription factors that take a key task in regulating organ growth, regeneration, homeostasis, and gene expression control (Zheng and Pan, 2019).

## 2. Materials and methods

### 2.1. Data set and analysis codes

The publicly available gene expression data sets from human bone HSCs were accessed from the GEO database with GSE32719 (Pang et al., 2011). The microarray data was produced by the Affymetrix Human Genome U133 Plus 2.0 Array (GPL570). This data set contains a total of 54,676 gene expressions of HSCs in 14 young (20–31 years), five middle-aged (42–61), 8 old (65–85) groups. The GSE32719 data set is investigated by the GEOquery package in Bioconductor (Tarca et al., 2006; Warnes et al., 2009; Durinck et al., 2005, 2009). The GEOquery package is a helpful tool to analyze GEO datasets quickly and easily, because it creates a bridge between GEO datasets and Bioconductor. The following are the other R studio packages that the researchers used: gplots, ggplot2, Biobase, and biomaRT packages (Warnes et al., 2009; Durinck et al., 2005; Gomez-Rubio, 2016; Davis and Meltzer, 2007). Gplot and ggplot2 packages are useful packages for creatin graphic and visualizing the statistical data. BiomaRT package makes it possible to retrieve vast volumes of data in a consistent manner without knowing the underlying database structures. Biobase package is a helpful package that provides basic functions in Bioconductor.

Quality control of genes were done based on the following criteria: genes with the poor-quality and genes with low number of reads were removed prior to DEG analysis. This study screened probes according to the criterion that the sum of expression values across samples of each probe should be greater than zero. The researchers employed the Benjamini-Hochberg Procedure to decrease the false discovery rate (FDR) to calculate the corrected $p$ value. Moreover, the researchers employed hypergeometric model to the down-regulated and up-regulated DEGs in GO and KEGG enrichment analysis. (Benjamini and Hochberg, 1995; Dudoit et al., 2003).

Analysis of the study was carried out in the R programme. The researchers separated the samples into three groups: young-old, young-middle, and middle-old. The process on separated samples was carried out as calculating fold-change difference among the means of each group. Fold change and p value are used as criteria to filter out the genes. To identify up-regulated and down-regulated DEGs in each group, we used a $t$-test with a p value $<0.01$ and $\text{Log2} > 1.2$ to emphasize statistical significance.

### 2.2. Differentially expressed genes and cluster analysis of DEGs

Gene expression levels for each sample were accessed by the GEOquery program in Bioconductor and transformed to a base-2 logarithmic scale using R. Heatmaps of DEGs were constructed by gplots package of R by the heatmap.2 function. To compare the expression pattern of DEGs in each bi-group i.e. young-old, young-middle, and middle-old aged groups, clustering analysis of DEGs was carried out.

### 2.3. GO and KEGG pathway analysis

The Ensemble Biomart program in R was used to map expression measurements annotations for up-regulated and down-regulated DEGs for each group probe to gene names. We characterized DEGs in terms of their biological processes (BP), molecular functions (MF), and cellular components (CC) of Gene Ontology (GO) terms by the DAVID database (The Database for Annotation, Visualization and Integrated Discovery) (Huang et al., 2007; Sherman and Lempicki, 2009). Universal Protein resource, and annotation types were analyzed by GORILLA (Gene Ontology Enrichment Analysis and Visualization Tool), DAVID, and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Supek et al., 2011; Eden et al., 2009).

### 2.4. Protein-protein interaction network

NetworkAnalyst is a web-based tool that analyzes PPI networks for visualization of single gene lists created by STRING (Szklarczyk et al., 2015). Since the STRING database purposes to incorporate all known and forecasted relationships between proteins, along with both somatic connections as well as functional associations, this study further have explored the functional enrichment analyses via STRING. DEGs from young-old, young-middle, and middle-aged groups were examined to build a PPI network with previously reported GO categorization and enrichment in order to thoroughly interpret the regulatory mechanisms in the HSC population. (Table 1).

**Table 1**
DEGs by groups and the number of genes. Rows represent a comparison of groups within samples, whereas columns portray down-regulated and up-regulated DEGs.

| Groups | Down-regulated DEGs | Up-regulated DEGs |
|---|---|---|
| Young-old | 29 | 70 |
| Young-middle | 59 | 232 |
| Middle-old | 39 | 24 |

## 3. Results

### 3.1. Experimental data analysis

The researchers found out DEGs in 453 genes from young-old, young-middle, and middle-old age groups in the GSE32719 data set, which is demonstrated in Figs. 1 and 2. The researchers found the down-regulated and up-regulated DEGs in each group. The expression values were extracted and a heat map was constructed to show the young-old, young-middle, and middle-old differentially expressed genes by groups. DEGs were labeled with $p < 0.01$ and $|\text{Log2 (FC)}| > 1.2$, which is a screening cut-off for each group. Numbers of up- and down-regulated gene expressions between young-old, young-middle, and middle-old groups are shown in Table 1. Here, 326 differentially expressed gene regulation was detected, whereas 127 down-regulated genes were found. However, information on the specific gender of the samples could not be asserted.

### 3.2. Gene ontology and enrichment analysis

The mostly enriched KEGG terms are “hepatitis C pathway”, “arrhythmogenic right ventricular cardiomyopathy (ARVC)”, and “hippo signaling pathway” (Fig. 2A). In Fig. 2B, the significant enrichment of DEGs using BP involving vasculogenesis, leukocyte migration, oligosaccharide metabolic process, and defense response to virus is shown. The important enrichments of DEGs in the CC term are “DNA replication factor A complex”, “nucleoplasm”, “apical part of the cell for cell component” (Fig. 2C). Finally, the important enrichments GO terms in MF are “protein binding”, “GTPase activity”, “translation factor activity”, and “RNA binding” (Fig. 2D). Hippo signaling pathway has a vital function in stem cells and tissue, particularly in progenitor cells and genes self-regeneration and growth (Table 2).

### 3.3. GO Pathways' results of the biological processes of DEGs from different age groups

The most significant GO pathway results are obtained from the down-regulated DEGs between young and middle-aged samples in Fig. 3A. It demonstrates the most important pathways that play a role in HSCs of young-middle and young-old age samples. The long list of GO terms is outlined and pictured using the REVIGO tool embedded R-script. Among 51 ontology terms, by visualizing in the scatter plot, 23 of the BP terms were found; the most significant ones are “cellular metabolic process”,” metabolic process”, and “negative regulation of cellular macromolecule biosynthetic process”. BP terms of down-regulated DEGs between young-old age samples (Fig. 3B) can be summarized as “cellular metabolic process”, “cellular component organization or biogenesis”, and “negative regulation of multicellular organismal process”. These pathways are involved in several cellular functions such as metabolic regulation, cell growth, and proliferation (Zhu and Thompson, 2019).

### 3.4. PPI network and KEGG pathway enrichment

The PPI map between the sets of input DEGs demonstrated in Fig. 4. In the displayed networks, red to orange fields represented the expression of the nodes and their degree of connectedness, accordingly. The genes were determined according to *p*-values with the best three scores. Best-scoring genes comprised hepatitis C (hsa05160) and the cell cycle pathway (hsa04110).

Furthermore, the KEGG enrichment analysis of the PPI network reveals the involvement of prostate cancer (hsa05215) and pathways in cancer (hsa05200) (Supplementary Table 1). The most significantly enriched pathways associated with HSC data were ‘hepatitis C', ‘ARVC', ‘sphingolipid signaling pathway’, ‘one carbon pool by folate’, ‘cell cycle’, ‘N-Glycan biosynthesis', ‘small cell lung cancer’, ‘collecting duct acid secretion’, and ‘cell adhesion molecules' (Table 4). Moreover, the ‘Hippo signaling pathway’ was included in the enriched KEGG pathways since the former plays a significant role in cell and tissue growth, which is also linked to malignant cells.

Fig. 5 demonstrates the PPI network of DEGs in each age group comparison. Hub genes of the young-old age group in Fig. 5A can be listed as EZR, GJA1, IFIT1, ZNRF3, and KDM6A. EZR gene is active in biological processes such as cell surface structure adhesion, migration, and organization, and it has been involved in several human cancers. IFIT1 involves biological processes such as *RNA binding*. PIP-network of the young-middle age group (Fig. 5B) is a sample of all DEGs (see Fig. 4 and Table 3). STAT1 and E2F1 proteins are observed in Fig. 5B and C as common genes.

### 3.5. Analysis of the hippo signaling pathway

There is a strong association of the Hippo signaling pathway with the cancer, so we detected the DEGs particularly associated with this pathway. We discovered 6 DEGs related to the Hippo signaling pathway: TCF7L2, PPP2R1B, PPP2R2B, PPP2R2D, BMPR1B, and TEAD2 (Fig. 6 and Tables 2 and 4). The Hippo signaling pathway was performed in the primary DEGs, as shown in Fig. 6A. The Hippo signaling pathway is previously reported as an effective pathway in cancer including AML (Kornblau et al., 2013). We implemented hippo signaling pathway DEGs to construct PPI Subnetwork 1 and Subnetwork 2 (Fig. 6A and B to further investigate the roles of related genes in the subnetworks.

Related genes with DEGs from the dataset enriched with the hippo signaling pathway were identified as novel hub genes. In subnetwork 1 of the hippo signaling pathway (Fig. 6A), PPP2R1B is the most significant gene in terms of BC. Furthermore, the protein phosphatase 2 regulatory subunits gene family has a key role in the hippo signaling pathway's subnetwork. As shown in subnetwork 2, TEAD2 is a gene that is related to the hippo signaling pathway, as previously reported by Zheng and Pan, 2019. In Fig. 6B, associated proteins with TEAD2 in the hippo signaling pathway of the HSC gene expression data set are revealed using NetworkAnalyst. Drug therapies such as verteporfin have side effects on cells in terms of toxicity, and stopping YAP/TAZ in consistently using little molecules could have fall outs in other cells and tissues.

The researchers suggest an alternative approach on the basis of the structure of the NCOA1-TEAD compound, which has been previously formulated based on the organizations of YAP-TEAD and VGLL4-TEAD compounds. Our results results validate the key role of the Hippo signaling pathway engaged in HSCs and treatment of age-related diseases, while also pointing to new possible therapeutic targets.

## 4. Discussion

Even though many researches have been conducted over the past decade to reveal the causes and potential mechanisms of AML and related diseases, the results still remain bleak. While most cases arise among adults, AML is a widespread kind of leukemia detected in children and old-aged people. AML accounts for 32% of all adult and 20% of all children leukemia cases (Shallis et al., 2019).

Discovering specific biomarkers to advance early diagnosis of AML, which is a main task in assisting patients with the best possible result, is vital. Equally essential, it is a requirement to determine and confirm

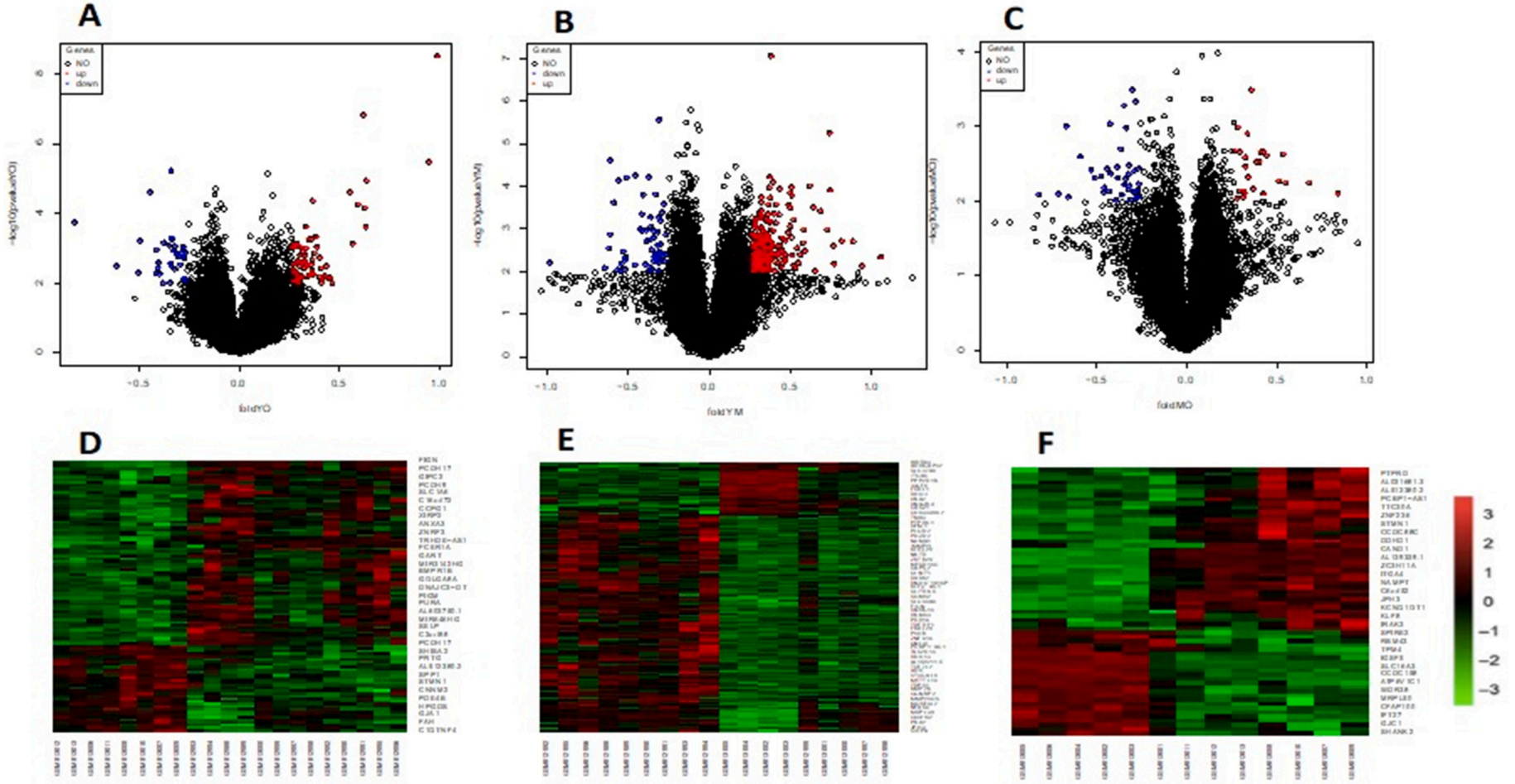


**Fig. 1.** Plots displaying the gene expression discrepancy in (A) young-old, (B) young-middle, and (C) middle-old comparison. (Top row) Black color shows no change (NO), blue color demonstrates down-regulated, and red color demonstrates up-regulated DEGs. logYO, logYM, and logMO on the x-axis labels represent: Log2 fold changes, respectively. (Bottom row) Heat map demonstrating DEGs in (D) young-old and (E) young-middle (F) middle-old age groups. Each column presents samples from age groups, while each row represent genes. The color change from red to green demonstrates the range of up- to down-regulated DEGs.

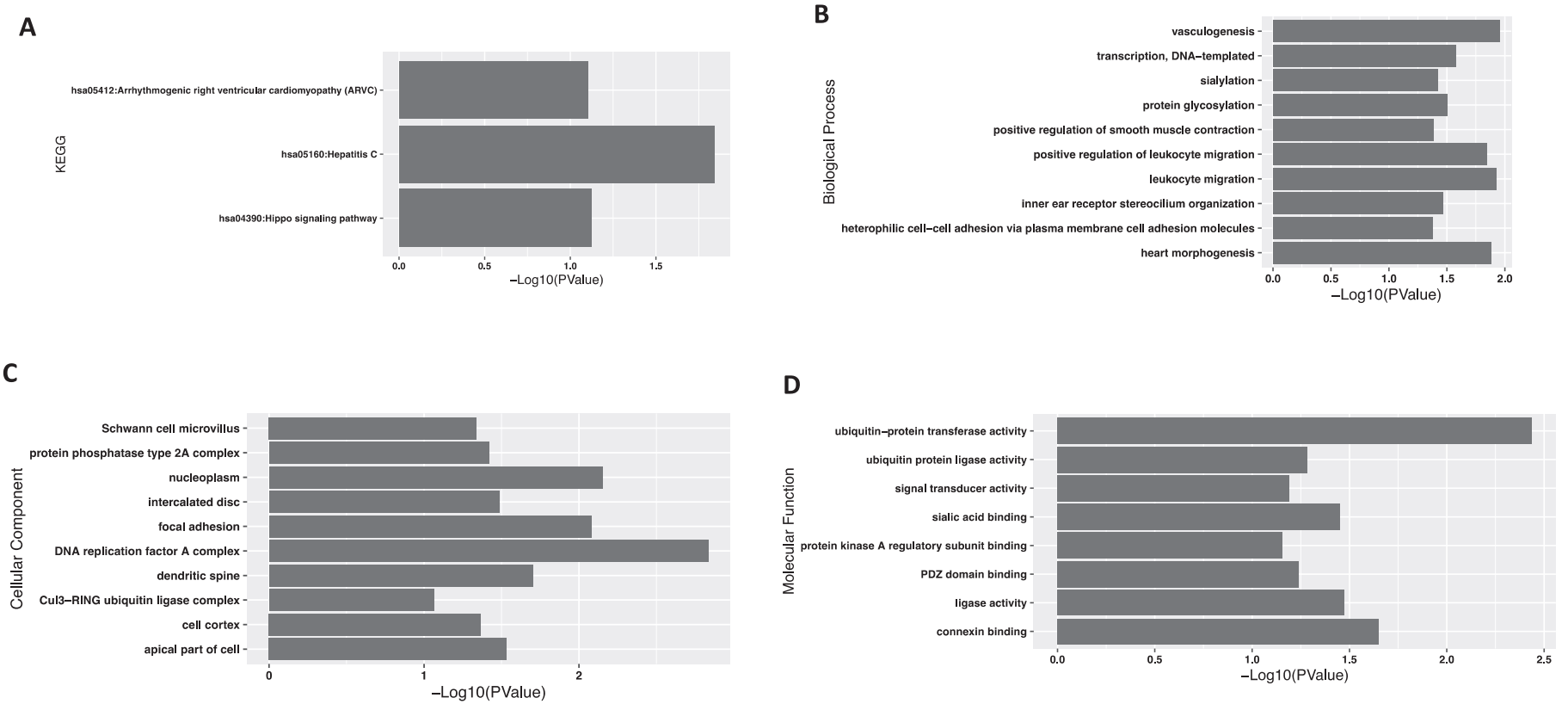


**Fig. 2.** The important enrichments of DEGs (A) KEGG pathways (B) Gene ontology (GO) enrichment of biological processes (BP), (C) GO enrichment cellular component (CC), and (D) GO enrichment molecular function (MF).

**Table 2**
Top significant pathways GO enrichment (BP and CC) and KEGG pathway analysis of all DEGs.

| Category | Term | Count | *p*-value | Genes |
|---|---|---|---|---|
| BP | GO:0001570 ~ vasculogenesis | 5 | 0.01 | GJC1, ACKR3, PTK2, TEAD2, VEGFA |
| BP | GO:0050900 ~ leukocyte migration | 7 | 0.011 | SELP, CXADR, ITGA4, PDE4B, SIRPA, SLC16A3, NKX2–3 |
| BP | GO:0051607 ~ defense response to virus | 7 | 0.04 | DNAJC3, RNASEL, CXADR, OAS1, STAT1, NCBP3, IFIT1 |
| BP | GO:0007507 ~ heart development | 7 | 0.06 | GJA1, CXADR, LOX, TRPS1, XIRP2, BCOR, FBN1 |
| BP | GO:0009311 ~ oligosaccharide metabolic process | 3 | 0.056 | ST6GAL2, ST8SIA4, ST3GAL6 |
| BP | GO:0007159 ~ leukocyte cell-cell adhesion | 3 | 0.056 | SELP, ITGA4, EZR |
| BP | GO:0009615 ~ response to virus | 4 | 0.09 | GJA1, STMN1, TRIM13, GNG11 |
| BP | GO:0030198 ~ extracellular matrix organization | 7 | 0.084 | ITGA4, LOX, ABI3BP, ITGA2, SPP1, PTK2, FBN1 |
| CC | GO:0005662 ~ DNA replication factor A complex | 4 | 1.4E-3 | PURB, PURA, RPA4, ERCC5 |
| CC | GO:0005654 ~ nucleoplasm | 60 | 0.7E-2 | NUMA1, C2ORF88, ZBTB20, EFCAB13, AHR, CLINT1, PTPDC1, CDC14A, DCAF7, CDC14B, SPRED1, METTL14, TRPS1, NAMPT, UBXN7, TEAD2, C8ORF44, KDM6A, FNBP4, CXADR, RMI1, GTF3A, HCFC2, CDC25A, EMSY, SGO2, ZC3H11A, MORF4L2, BMP2K, AAGAB, ANAPC5, MCTP2, CASZ1, XIAP, NMD3, TGOLN2, CAND1, NAA25, E2F1, SRSF11, SCAI, TCF7L2, RBM39, NFYB, STAT1, PCIF1, ANKRD23, BTBD8, SMARCA2, SCAF4, FOSL2, SELP, RAD52, NFIA, RPA4, ZNF217, ERCC5, NANOG, RAD18, FERMT2 |
| CC | GO:0005925 ~ focal adhesion | 14 | 0.8E-2 | TPM4, ITGA4, ITGA2, LPP, TRIOBP, PTK2, GJA1,NFASC, NCSTN, CSRP2, ATP6V0A2, P4HB, EZR, FERMT2 |
| CC | GO:0043197 ~ dendritic spine | 6 | 1.9E-2 | FARP1, ARHGAP32, PDE4B, CRIPT, STRN, SHANK2 |
| CC | GO:0045177 ~ apical part of cell | 5 | 2.9E-2 | SRR, NUMA1, FAT4, EZR, ATP6V1C1 |
| KEGG | hsa05160:Hepatitis C | 7 | 1.4E-2 | RNASEL, OAS1, PPP2R1B, PPP2R2B, STAT1, PPP2R2D, IFIT1 |
| KEGG | hsa04390:Hippo signaling pathway | 6 | 7.5E-2 | TCF7L2, PPP2R1B, PPP2R2B, PPP2R2D, BMPR1B, TEAD2 |
| KEGG | hsa05412: Arrhythmogenic right ventricular cardiomyopathy (ARVC) | 4 | 7.8E-2 | TCF7L2, GJA1, ITGA4, ITGA2 |

GO: gene ontology; BP: biological process; CC: cell component; KEGG: Kyoto Encyclopedia of Genes and Genomes.

novel therapeutic targets that may be prosperous in curing AML and AML-related diseases. Healthy and hematologically normal young, middle-aged and old human bone marrow specimens of hematopoietic progenitor populations are evaluated to uncover potential biomarkers that may affect age-related hematopoietic dysfunction in the elderly human hematopoietic system.

Astonishingly, most studies focus on multiple genetic results that are derived from multiple sample studies through microarray analysis that are not uniform with one another (Figueroa et al., 2009; Haferlach et al., 2007; Shivarov and Bullinger, 2014; Takahashi et al., 2018). This study performed bi-group comparison profile of three categories (young, middle, and old-aged healthy samples) within one data set and employed bioinformatics techniques to strongly analyze the dataset. 453 DEGs were identified. The number of down-regulated DEGs was importantly less than the up-regulated DEGs (127 versus 326).

By using computational methods, we have identified genes that are expressed in the early stage of AML and related cancers and also involved in type-2 diabetes and viral infectious diseases. Subsequently, the DEGs were subjected to biological process and functional analysis which deciphers the increase in hematopoietic stem cell population and functional decline. More importantly, the results reveal UBC, PTK2, and TCF7L2 genes are identified as hub genes. UBC plays a critical role in maintaining ubiquitin (Ub) homeostasis that also a delay in cell-cycle progression and increased susceptibility to cellular stress. Abnormalities in Ub system might result in carcinogenesis and metastasis. Upregulated expression of UBC gene was observed in many cancer types that result in enhanced cellular stress (Tang et al., 2015). UBC with the largest Betweenness Centrality (BC) was suggested to be central to the PPI network associated with HSCs gene expression data. Moreover, STAT1, PTK2, E2F1, RHOJ, RAD52, and CDK19 with the secondary highest degree and VEGFA with the secondary BC might be engaged in the progression AML and other substance diseases.

Omics alterations of PTK2 in human cancers such Glioblastoma Multiforme and Glioma. Among its related pathways are association between physico-chemical features and toxicity associated pathways and transcriptional misregulation in cancer. Gene Ontology (GO) annotations related to this gene include transferase activity, transferring phosphorus-containing groups *and* protein tyrosine kinase activity. Diseases associated with TCF7L2 include type 2 Diabetes Mellitus and Colorectal Cancer. Moreover, other hub genes discussed in this study, may be used as potential targets for AML and related diseases diagnosis and treatment.

Gene ontology analysis, which reveals information about these genes, is used to discover the functions and diseases related to proteins. The top 12 proteins UBC, EP300, CREBBP, TP53, PTK2, E2F1, STAT1, NFYB, RAD18, TCEB1, NAPC5, and VEGFA are predicted to be identified in many cancer types such as lung cancer, leukemia, breast cancer, glioma, ovarian cancer, and colorectal cancer. (Attar and Kurdistani, 2017; Rasheed et al., 1994; Tang et al., 2015). Some genes like GTF3A, STAT1, TFC7L2, and XIAP proteins are related with mental retardation, type 2 diabetes, and mycobacterial and viral infections (Durbin et al., 1996; Glorioso et al., 2011; Grant et al., 2006; Wu et al., 2010). These genes have the potential to be used as biomarkers for diagnosis of early stage of type 1 and type 2 diabetes and can also operate as promising drug targets for the drug to strengthen immunity to viral diseases (Durbin et al., 1996; Groves et al., 2006; Kamiyama et al., 2017; Redondo et al., 2018).

The top hub genes that we identified in our study are STAT1, E2F1, RHOJ, RAD52, CDK19, UBXN7, GNL3L, CDC25A, EZR, XIAP, SMARCA2, UBC and STRN. Hub genes of the young-old age group are EZR, GJA1, IFIT1, ZNRF3, and KDM6A. EZR gene is active in biological processes such as cell surface structure adhesion, migration, and organization, and it has been involved in several human cancers. IFIT1 involves biological processes such as *RNA binding*. PIP-network of the

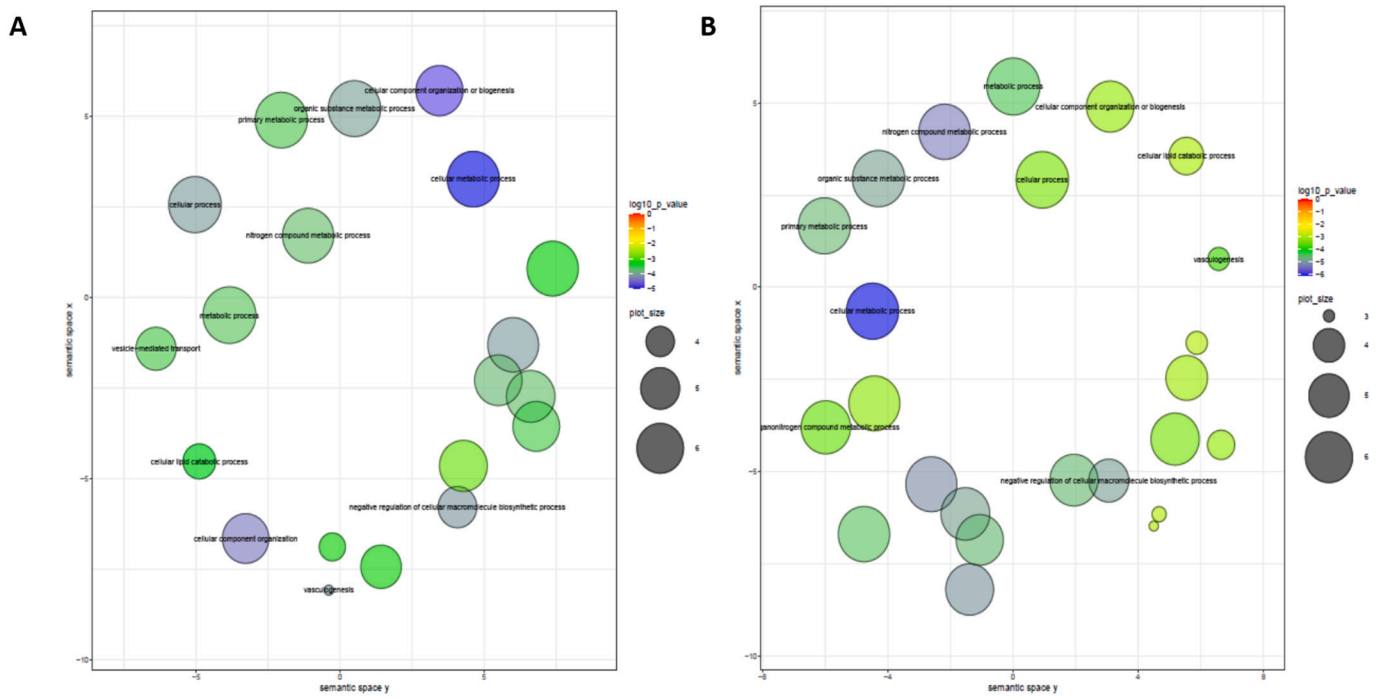


**Fig. 3.** Scatter plots done with REVIGO that shows GO analyses in the biological process (A) down-regulated DEGs in the young-middle (B) down-regulated DEGs in the young-old age groups. Relevant GO terms are aggregated and demonstrate in bubbles of identical hues. Bubble colors show *p values*, while the bubble sizes demonstrate the relative frequency of GO terms.

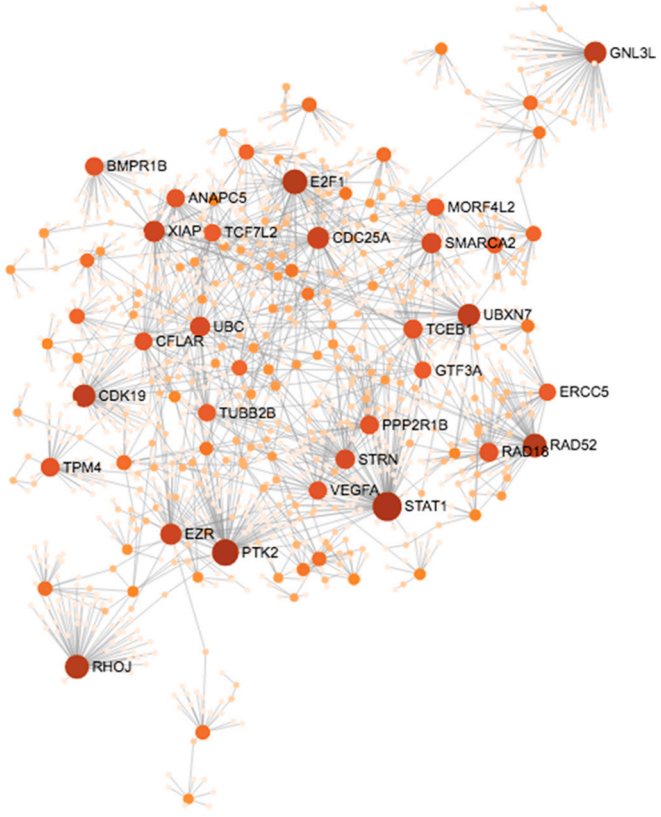


**Fig. 4.** A PPI network of DEGs was created by NetworkAnalyst..

young-middle age group is a sample of all DEGs. PIP network of middle-old age group hub genes are STAT1, E2F1, XIAP, TPM4 and PPP2R1B. STAT1 acts as a transcription factor and increases inflammation and acts as a tumor suppressor (Pensa et al., 2000-2013). It has been shown that E2F1 that acts as a transcription factor has been involved in cellular proliferation and apoptosis in colon cancer (Fang et al., 2020). XIAP has been shown to be involved in tumor growth and XIAP expression has been reported in several cancers (Huang et al., 2019).

Overall, many of the identified genes in our study were shown previously related to age-related diseases like cancers, mental diseases and type 2 diabetes mellitus.

GO pathway analysis discovered ‘cellular process’, ‘cellular metabolic process’, ‘metabolic process’, ‘biogenesis’, and ‘vasculogenesis biological processes’, which are involved in a wide of biological activities such as metabolic regulation, cell growth, and proliferation (Zhu and Thompson, 2019). The significant enrichment of DEGs in BP category are “vasculogenesis”, “leukocyte migration”, “oligosaccharide metabolic process”, and “defense response to virus”. The important enrichments of DEGs in the CC term are “DNA replication factor A complex”, “nucleoplasm”, “apical part of the cell for cell component”. Finally, the important enrichments GO terms in MF category are “protein binding”, “GTPase activity”, “translation factor activity”, and “RNA binding”.

The mostly enriched KEGG terms are “hepatitis C pathway”, “arrhythmogenic right ventricular cardiomyopathy (ARVC)”, and “hippo signaling pathway”. Our research found out that Hippo signaling pathway involving differentially expressed genes in several kinds of cancer and immune cells (Zheng and Pan, 2019). Hippo signaling pathway has a vital function in stem cells and tissue, particularly in progenitor cells and genes self-regeneration and growth. The major role of the Hippo signaling network in cell life and tissue growth has been increasingly investigated in the last 20 years. Simultaneously, knowledge on the entanglement of this network and its partners in other pathways has enhanced enormously. The Hippo signaling pathway is now regarded as an important integrator of tag from the biophysical and biochemical surroundings of cells (Misra and Irvine, 2018). The Hippo signaling pathway is controlled by negative regulation of YAP transcription factors, which are protein partners of TEAD2. Drug therapies can disturb the connection between YAP proteins and TEAD and inhibit

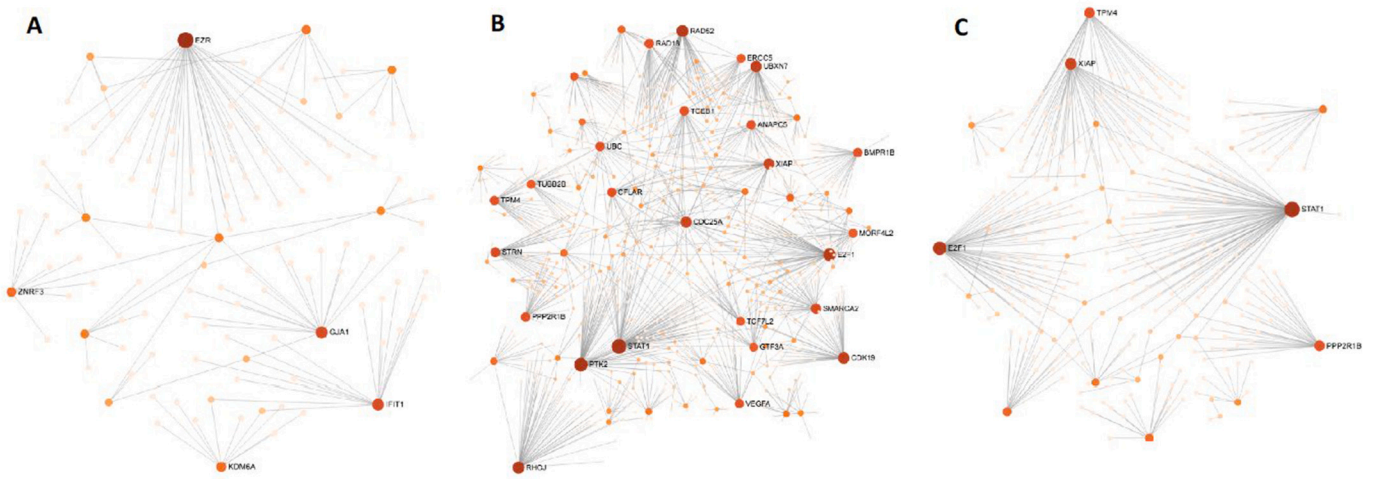


**Fig. 5.** PPI networks of (A) young-old group DEGs (B) young-middle group DEGs and finally (C) middle-old age group DEGs were constructed and pictured using NetworkAnalyst. Light orange and red colors represent |Log2 (FC)| expression value for the DEGs in each bi-group. PIP-network of young-old comparison (B) reflects the characteristics of all DEGs.

**Table 3**
Top 15 hub genes of the PPI network.

| Gene ID | Gene name | Node degree | Betweenness centrality (BC) |
|---|---|---|---|
| 6772 | STAT1 | 84 | 175,186.36 |
| 5747 | PTK2 | 68 | 148,475.01 |
| 1869 | E2F1 | 56 | 91,229.24 |
| 57,381 | RHOJ | 51 | 57,433.11 |
| 5893 | RAD52 | 50 | 40,048.94 |
| 23,097 | CDK19 | 45 | 45,505.58 |
| 26,043 | UBXN7 | 43 | 54,800.8 |
| 54,552 | GNL3L | 43 | 50,432.01 |
| 993 | CDC25A | 41 | 59,535.81 |
| 7430 | EZR | 38 | 97,757.97 |
| 331 | XIAP | 38 | 63,169.02 |
| 6595 | SMARCA2 | 34 | 24,992.22 |
| 7316 | UBC | 33 | 340,166.09 |
| 6801 | STRN | 31 | 19,748.33 |
| 56,852 | RAD18 | 29 | 52,481.81 |

expression of YAP target genes (Liu-Chittenden et al., 2012). However, certain drug therapies such as verteporfin have high cytotoxicity, and stopping YAP/TAZ in consistently using little molecules could have fall outs in other cells and tissues. A different technique has been formulated

**Table 4**
Top 10 KEGG pathways of DEGs.

| Pathway | Gene count | p value |
|---|---|---|
| Hepatitis C | 155 | 0.000445 |
| ARVC | 72 | 0.0214 |
| Hippo signaling pathway | 154 | 0.0259 |
| Sphingolipid signaling pathway | 119 | 0.0308 |
| One carbon pool by folate | 20 | 0.0344 |
| Cell cycle | 124 | 0.0359 |
| N-Glycan biosynthesis | 50 | 0.037 |
| Small cell lung cancer | 93 | 0.048 |
| Collecting duct acid secretion | 27 | 0.0595 |
| Cell adhesion molecules (CAMs) | 146 | 0.0641 |

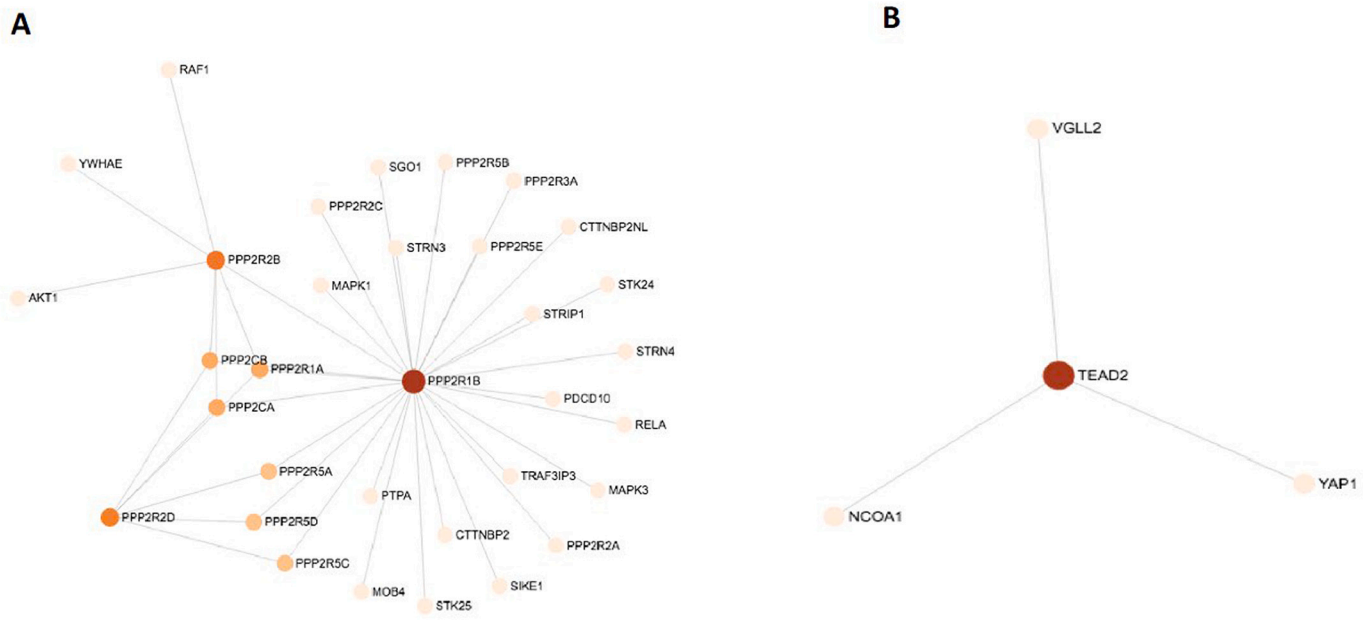


**Fig. 6.** (A) The hippo signaling pathway's subnetwork 1 of hub genes. (B) The hippo signaling pathway's subnetwork 2 of hub genes.

based on the organizations of YAP-TEAD and VGLL4-TEAD complexes (Jiao et al., 2014). The researchers of this study proposed that through silico analysis, there could be an alternative approach on the basis of the structure of the NCOA1-TEAD complex. We also discovered that TCF7L2, PPP2R1B, PPP2R2B, PPP2R2D, BMPR1B and TEAD2 genes are associated with hippo signaling pathway. TCF7L2 and PPP2R1B genes are also one of the most significant hub genes that we found in our study. TEAD2 is a gene that is related to the hippo signaling pathway, as previously reported by Zheng and Pan, 2019 which supports our study.

In vivo and in vitro experiments are needed for predicted genes that are expressed in HSC data set and at the developmental stage of HSC age-related diseases, particularly AML. In the future, validation would be performed using AML samples or at least AML cells. Studies looking for early diagnosis of AML is an urgent need to treat the disease effectively. Our findings offer silico evidence for candidate genes, such as the UBC, PTK2, and TCF7L2, that may be promising biomarkers for the translation approach associated with HSC population age-related diseases.

## Author contributions

CONCEPTION: Emine Güven INTERPRETATION OR ANALYSIS OF DATA: Emine Güven PREPARATION OF THE MANUSCRIPT: Emine Güven and Sevinç Akçay REVISION FOR IMPORTANT INTELLECTUAL CONTENT: Sevinç Akçay and Emine Güven SUPERVISION: Emine Güven and Sevinç Akçay.

## Declaration of Competing Interest

The authors report no conflicts of interest.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.humgen.2022.201068.

## References


Appelbaum, F.R., Gundacker, H., Head, D.R., Slovak, M.L., Willman, C.L., Godwin, J.E., Anderson, J.E., Petersdorf, S.H., 2006. Age and acute myeloid leukemia. Blood 107 (9), 3481–3485.

Attar, N., Kurdistani, S.K., 2017. Exploitation of EP300 and CREBBP lysine acetyltransferases by cancer. Cold Spring Harb. Perspect. Med. 7 (3), a026534.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 57 (1), 289–300.

Bohl, S.R., Bullinger, L., Rücker, F.G., 2019. New targeted agents in acute myeloid leukemia: new Hope on the rise. Int. J. Mol. Sci. 20 (8), 1983. https://doi.org/10.3390/ijms20081983.

Boyd, A.L., Reid, J.C., Salci, K.R., Aslostovar, L., Benoit, Y.D., Shapovalova, Z., Nakanishi, M., Porras, D.P., Almakadi, M., Campbell, C.J.V., 2017. Acute myeloid leukaemia disrupts endogenous myelo-erythropoiesis by compromising the adipocyte bone marrow niche. Nat. Cell Biol. 19 (11), 1336–1347.

Dar, A., Kollet, O., Lapidot, T., 2006. Mutual, reciprocal SDF-1/CXCR4 interactions between hematopoietic and bone marrow stromal cells regulate human stem cell migration and development in NOD/SCID chimeric mice. Exp. Hematol. 34 (8), 967–975.

Davis, S., Meltzer, P., 2007. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. Bioinform. (Oxford, England) 23, 1846–1847. https://doi.org/10.1093/bioinformatics/btm254.

Dudoit, S., Shaffer, J.P., Boldrick, J.C., 2003. Multiple hypothesis testing in microarray experiments. Stat. Sci. 71–103.

Durbin, J.E., Hackenmiller, R., Simon, M.C., Levy, D.E., 1996. Targeted disruption of the mouse Stat1 gene results in compromised innate immunity to viral disease. Cell 84 (3), 443–450.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., Huber, W., 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics 21 (16), 3439–3440.

Durinck, S., Spellman, P.T., Birney, E., Huber, W., 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4 (8), 1184.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., Yakhini, Z., 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinform. 10 (1), 1–7.

Fang, Z., Lin, M., Li, C., Liu, H., Gong, C., 2020. A comprehensive review of the roles of E2F1 in colon cancer. Am. J. Cancer Res. 10 (3), 757–768.

Figueroa, M.E., Wouters, B.J., Skrabanek, L., Glass, J., Li, Y., Erpelinck-Verschueren, C. A., Langerak, A.W., Löwenberg, B., Fazzari, M., Greally, J.M., 2009. Genome-wide epigenetic analysis delineates a biologically distinct immature acute leukemia with myeloid/T-lymphoid features. Blood J. Am. Soc. Hematol. 113 (12), 2795–2804.

Fraga, M., Esteller, M., 2007. Fraga, M. F. & Esteller, M. epigenetics and aging: the targets and the marks. Trends genet. 23, 413-418. Trends Genet. TIG 23, 413–418. https://doi.org/10.1016/j.tig.2007.05.008.

Glorioso, C., Oh, S., Douillard, G.G., Sibille, E., 2011. Brain molecular aging, promotion of neurological disease and modulation by Sirtuin5 longevity gene polymorphism. Neurobiol. Dis. 41 (2), 279–290.

Gomez-Rubio, V., 2016. ggplot2: Elegant Graphics for Data Analysis, 2nd ed. Springer, New York, NY.

Grant, S.F.A., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Manolescu, A., Sainz, J., Helgason, A., Stefansson, H., Emilsson, V., Helgadottir, A., 2006. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat. Genet. 38 (3), 320–323.

Groves, C.J., Zeggini, E., Minton, J., Frayling, T.M., Weedon, M.N., Rayner, N.W., Hitman, G.A., Walker, M., Wiltshire, S., Hattersley, A.T., 2006. Association analysis of 6,736 UK subjects provides replication and confirms TCF7L2 as a type 2 diabetes susceptibility gene with a substantial effect on individual risk. Diabetes 55 (9), 2640–2644.

Haferlach, T., Kohlmann, A., Bacher, U., Schnittger, S., Haferlach, C., Kern, W., 2007. Gene expression profiling for the diagnosis of acute leukaemia. Br. J. Cancer 96 (4), 535–540.

Huang, D.W., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M.W., Lane, H.C., 2007. DAVID bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res. 35 (suppl_2), W169–W175.

Huang, X., Wang, Xn, Yuan, Xd., et al., 2019. XIAP facilitates breast and colon carcinoma growth via promotion of p62 depletion through ubiquitination-dependent proteasomal degradation. Oncogene 38, 1448–1460. https://doi.org/10.1038/s41388-018-0513-8.

Jiao, S., Wang, H., Shi, Z., Dong, A., Zhang, W., Song, X., He, F., Wang, Y., Zhang, Z., Wang, W., 2014. A peptide mimicking VGLL4 function acts as a YAP antagonist therapy against gastric cancer. Cancer Cell 25 (2), 166–180.

Kamiyama, N., Soma, R., Hidano, S., Watanabe, K., Umekita, H., Fukuda, C., Noguchi, K., Gendo, Y., Ozaki, T., Sonoda, A., 2017. Ribavirin inhibits Zika virus (ZIKV) replication in vitro and suppresses viremia in ZIKV-infected STAT1-deficient mice. Antivir. Res. 146, 1–11.

Kornblau, S.M., Hu, C.W., Qiu, Y.H., Yoo, S.Y., Zhang, N., Kadia, T.M., Ferrajoli, A., Coombes, K.R., Qutub, A.A., 2013. Hippo pathway (HP) activity in acute myelogenous leukemia (AML): different prognostic implications of TAZ versus YAP inactivation by phosphorylation. Blood 122 (21), 1337. https://doi.org/10.1182/blood.V122.21.1337.1337.

Kuranda, K., Vargaftig, J., de la Rochere, P., Dosquet, C., Charron, D., Bardin, F., Tonnelle, C., Bonnet, D., Goodhardt, M., 2011. Age-related changes in human hematopoietic stem/progenitor cells. Aging Cell 10 (3), 542–546. https://doi.org/10.1111/j.1474-9726.2011.00675.x.

Liu-Chittenden, Y., Huang, B., Shim, J.S., Chen, Q., Lee, S.-J., Anders, R.A., Liu, J.O., Pan, D., 2012. Genetic and pharmacological disruption of the TEAD–YAP complex suppresses the oncogenic activity of YAP. Genes Dev. 26 (12), 1300–1305.

Luppi, M., Fabbiano, F., Visani, G., Martinelli, G., Venditti, A., 2018. Novel agents for acute myeloid leukemia. Cancers 10 (11), 429. https://doi.org/10.3390/cancers10110429.

Misra, J.R., Irvine, K.D., 2018. The hippo signaling network and its biological functions. Annu. Rev. Genet. 52, 65–87. https://doi.org/10.1146/annurev-genet-120417-031621.

Morrison, S.J., Scadden, D.T., 2014. The bone marrow niche for haematopoietic stem cells. Nature 505 (7483), 327–334.

Pang, W.W., Price, E.A., Sahoo, D., Beerman, I., Maloney, W.J., Rossi, D.J., Schrier, S.L., Weissman, I.L., 2011. Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. Proc. Natl. Acad. Sci. U. S. A. 108 (50), 20012–20017. https://doi.org/10.1073/pnas.1116110108.

Pensa, S., Regis, G., Boselli, D., et al., 2000-2013. STAT1 and STAT3 in Tumorigenesis: Two Sides of the Same Coin? In: Madame Curie Bioscience Database [Internet]. Landes Bioscience, Austin (TX). Available from: https://www.ncbi.nlm.nih.gov/books/NBK6568/.

Perry, J.M., Li, L., 2007. Disrupting the stem cell niche: good seeds in bad soil. Cell 129 (6), 1045–1047.

Pinho, S., Frenette, P.S., 2019. Haematopoietic stem cell activity and interactions with the niche. Nat. Rev. Mol. Cell Biol. 20 (5), 303–320.

Rasheed, B.K.A., McLendon, R.E., Herndon, J.E., Friedman, H.S., Friedman, A.H., Bigner, D.D., Bigner, S.H., 1994. Alterations of the TP53 gene in human gliomas. Cancer Res. 54 (5), 1324–1330.

Redondo, M.J., Geyer, S., Steck, A.K., Sosenko, J., Anderson, M., Antinozzi, P., Michels, A., Wentworth, J., Xu, P., Pugliese, A., 2018. TCF7L2 genetic variants contribute to phenotypic heterogeneity of type 1 diabetes. Diabetes Care 41 (2), 311–317.

Roushangar, R., Mias, G.I., 2019. Multi-study reanalysis of 2,213 acute myeloid leukemia patients reveals age-and sex-dependent gene expression signatures. Sci. Rep. 9 (1), 1–17.

Shallis, R.M., Wang, R., Davidoff, A., Ma, X., Zeidan, A.M., 2019. Epidemiology of acute myeloid leukemia: recent progress and enduring challenges. Blood Rev. 36, 70–87.

Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4 (1), 44.

Shivarov, V., Bullinger, L., 2014. Expression profiling of leukemia patients: key lessons and future directions. Exp. Hematol. 42 (8), 651–660.
Supek, F., Bošnjak, M., Škunca, N., Šmuc, T., 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One 6 (7), e21800.
Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., 2015. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43 (D1), D447–D452.
Takahashi, K., Wang, F., Morita, K., Yan, Y., Hu, P., Zhao, P., Abou Zhar, A., Wu, C.J., Gumbs, C., Little, L., 2018. Integrative genomic analysis of adult mixed phenotype acute leukemia delineates lineage associated molecular subtypes. Nat. Commun. 9 (1), 1–12.
Tang, Y., Geng, Y., Luo, J., Shen, W., Zhu, W., Meng, C., Li, M., Zhou, X., Zhang, S., Cao, J., 2015. Downregulation of ubiquitin inhibits the proliferation and radioresistance of non-small cell lung cancer cells in vitro and in vivo. Sci. Rep. 5 (1), 9476. https://doi.org/10.1038/srep09476.
Tarca, A.L., Romero, R., Draghici, S., 2006. Analysis of microarray experiments of gene expression profiling. Am. J. Obstet. Gynecol. 195 (2), 373–388. https://doi.org/10.1016/j.ajog.2006.07.001.
Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., 2009. gplots: various R programming tools for plotting data. R Package Version 2 (4), 1.
Wu, H., Panakanti, R., Li, F., Mahato, R.I., 2010. XIAP gene expression protects β-cells and human islets from apoptotic cell death. Mol. Pharm. 7 (5), 1655–1666.
Zheng, Y., Pan, D., 2019. The hippo signaling pathway in development and disease. Dev. Cell 50 (3), 264–282. https://doi.org/10.1016/j.devcel.2019.06.003.
Zhu, J., Thompson, C.B., 2019. Metabolic regulation of cell growth and proliferation. Nat. Rev. Mol. Cell Biol. 20 (7), 436–450. https://doi.org/10.1038/s41580-019-0123-5.