Identification of potential biomarkers of human bone marrow hematopoietic stem cells micro-array gene expression data

Address

Abstract: We characterized aging at the gene expression level using GSE32719 data set publicly available at gene expression omnibus (GEO) and ArrayExpress. Using Biobase, GEOquery, gplots packages out of which 453 genes chosen. The gene ontology (GO) molecular function, biological procces, cell component, KEGG and pathway enrichment analyses of DEGs were performed. 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 idetntified 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. Omics alterations of PTK2 in human cancers such Glioblastoma Multiforme and Glioma. Diseases associated with TCF7L2 include type 2 Diabetes Mellitus and Colorectal Cancer. Other hub genes discussed in this study, may be used as potential targets for AML and related diseases diagnosis and treatment.

Keywords: biomarker; differentially expressed genes; cancer; gene ontology pathway enrichment; diabetes mellitus

Reference to this paper should be made as follows: Author. (xxxx) 'Title', *Int. J. xxxxxxxxxx xxxxxxxxxx*,

Biographical notes:

*This paper is a revised and expanded version of a paper entitled **[title]** presented at [**name, location and date of conference**].*

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 and contributes to circulation. The specialized phenestra capillary, called sinusoid, penetrates the extracellular / extracellular matrix (ECM), and ECMs are sponge-like matrices produced by reticular fibroblasts (Pinho & Frenette, 2019).

These proteins and cells place the hematopoietic cells in separate compartments. Similarly, hematopoietic cells, bidothelial cells, stromal cells, ECM, cytokines, growth factors, and chemokines contain special microenvironment in the bone marrow (Dar et al., 2006; Morrison & Scadden, 2014). Recent advances include new markers for HSCs and niche stem cells, systematic analysis of expression patterns of niche factors, genetic tools for functional in vivo identification of niche cells, and improved imaging techniques. Stem cells have the ability to divide for a long time 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 & Esteller, 2007).

A detailed examination of the cellular and molecular properties of HSCs has made studies in clinical use, stem cell identification and use highly effective. Bone marrow microenvironment is an ideal place to support healthy and yet it might be also support malignant hematopoiesis that plays an important role in the development and progression of leukemia and other types of cancer (Fraga & Esteller, 2007; Kuranda et al., 2011; Perry & Li, 2007; Pinho & Frenette, 2019).

In acute myeloid leukemia (AML), blast cells, which deteriorate during normal maturation, begin to accumulate in the blood together with the bone marrow. The body remains vulnerable because white blood cells cannot form. Erythrocyte and platelet production are disrupted in the bone marrow due to abnormal proliferation of myeloblasts. As a result, anemia infection and platelet count decrease (Boyd et al., 2017). AML is not limited to a particular part of the body since its onset, but can spread to the blood, lymphatic tissue and all other organ systems from the bone marrow. As with many other leukemia diseases, it is defined as a malignant systemic disease. HSC and other hemotopoietic progenitor populations are evaluated and it was found that the elderly HSC increased frequently (Pang et al., 2011). Over the years, the detection of AML-specific antigen has been applied as one of the diagnostic biomarkers for monitoring AML.

Previous studies addressing age-related changes in human HSC, due to the indirect evaluation of root and progenitor populations, need to support experimental studies with numerical analysis and statistical methods in addition to previous HSC studies in mice . Several agents are estimated and used to prevent AML, including monoclonal antibodies, new formulation of old drugs, FLT3 and IDH1/2 inhibitors (Bohl et al., 2019; Luppi et al., 2018). Hundreds of studies have been performed to expose the central mechanism of AML occurrence and progression by microarray gene expression technology and target these processes for therapeutic strategy. It still remains a demand for more effective therapies or for ways that can ameliorate therapeutic responses to the medication of AML. In this study, we employed the microarray data sets of human bone marrow tissue publicly available transcriptome data sets and performed integrative analysis on DEGs by bioinformatics analysis. Our results will uncover the potential biomarkers and the promising therapeutic targets for AML although experimental studies are required to authenticate our findings.

2 Materials and Methods

2.1. Microarray data and preprocessing

Expression data from human bone marrow hematopoietic stem cells were downloaded from the gene expression omnibus (GEO) database with GSE32719 was used (Pang et al., 2011). Genomic information ranging from gene sequences to protein structure predictions were obtained. As described by Pang et al, these data sets contain a total of 50,000 gene expression of healthy human bone marrow hematopoietic stem cells in groups of 14 young (20–31 years), 5 middle age (42–61), 8 old (65–85) groups. The GSE32719 data set is analyzed by using the GEOquery package in Bioconductor following standard procedures in R studio (Durinck et al., 2005, 2009; Tarca et al., 2006; Warnes et al., 2009). The other packages we used in R studio are as the following; Biobase, biomaRT and gplots packages (Davis & Meltzer, 2007; Warnes et al., 2009). Multiple testing was corrected through the Benjamini–Hochberg procedure to calculate the adjusted *p*-value. Hypergeometric distribution test was used for DEGs in the functional and pathway enrichment analysis, and false discovery rate (FDR) was calculated to adjust the statistical tests locally (Benjamini & Hochberg, 1995; Dudoit et al., 2003; Hochberg & Tamhane, 1987) .

2.2. Experimental data and analysis codes

Analysis were conducted in the R statistical environment. Sample codes and analysis of GSE32719 data can be found at public github repository. We separate samples into three groups provided that young-old , young-middle, and middle-old aged. The data set was normalized by computing the means of the samples of each group in R programming language. The process on separated samples which is grouped by categories of young-old, young-middle, and middle-old aged was performed as computing fold-change (biological significance) difference between the means of the groups. A widely used statistical model is the t-distribution and its variants. A t-test compares the difference in the mean expression levels between the two groups, taking into account the variability of the data (difference in means between groups divided by the standard deviation). The genes are screened based on satisfaction of both fold and p-value filtering criteria. We highlight the significantly up-regulated and down-regulated differential expressed genes (DEGs) using both t-test and fold change criteria. Although methods to correct for multiple comparisons have been available for a long time (eg, Bonferroni correction), many of these methods are not well suited for the analysis of microarray data (Tarca et al., 2006). We perform statistical significance using t-test by taking p-value cutoff value 0.01 and log2(fold cut-off) > 1.2 for each group separately.

2.3. Hierarchical clustering analysis

The gene expression value was extracted from every sample of the data set and showed as base-2 logarithmic value by Affy package of R language. The heatmap of bidirectional hierarchical clustering was set up through heatmap.2 function of gplots package of R language for DEGs in the data set.

2.4. Functional and pathway enrichment analysis

Expression measurements annotations for up-regulated and down regulated DEGs for each groups probes mapped to gene names using Ensemble Biomart package in R. The common DEGs were classified according to their biological processes, molecular functions or cellular components by gene ontology (GO) of the database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang et al., 2007; Sherman & Lempicki, 2009). All characterized genes were carefully investigated and additional components like the Universal Protein resource, and physical properties Gene Ontology (GO) and annotation types were derived using ***G****ene* ***O****ntology En****ri****chment ana****l****ysis and Visual****iz****a****t****ion Tool* (GOrilla) and DAVID, the reference of database website from Kyoto Encyclopedia of Genes and Genomes (KEGG) (Eden et al., 2009; Supek et al., 2011). We then compared the results of DAVID with NetworkAnalyst enrichments performed with KEGG (Xia et al., 2015; Zhou et al., 2019).

2.5. Protein-Protein Interaction Network

NetworkAnalyst, available on the web, enables analysis of protein-protein interaction (PPI) networks for multiple gene lists using STRING Interactome (Szklarczyk et al., 2015). The DEGs from young-old, young-middle, and middle-old aged groups (See Table 1) were analyzed to construct PPI network, combined with previous reported GO classification and enrichment, for the purpose of comprehensively unraveling the molecular regulatory mechanisms in AML and related diseases.

Table 1 Summary of differentially expressed genes by groups gene symbols. Rows represent comparison of groups within samples whereas columns represent up-regulated and down-regulated DEGs.

groups compared	up-regulated DEGs gene symbols	down-regulated DEGs gene symbols
young-old	70	29
young-middle	232	59
middle-old	24	39

3 Results

3.1. Experimental data analysis

With gene expression result of the GSE32719 data set, we detect differentially expressed genes (**DEGs**) in total 453 genes from **young-old, young-middle**, and **middle-old** aged groups which was displayed in volcano plot (Figure 1). We find the up-regulated and down-regulated differentially expressed genes in each group. The gene expression values were extracted, and a hierarchical clustering heat map was plotted to present the DEGs of young-old **,** young-middle, and middle-old gene expressions groups. (Figure 2). DEGs were picked out with common *t* test, and statistically significant DEGs were labelled with $p < 0.01$ and log2-fold change (log2FC) >1.2 as the cut-off standard for each group. Numbers of up and down regulated gene expressions between young-old, young-middle, and middle-old groups shown in Table 1.

Here, we detected 326 differentially expressed genes of up regulation, whereas we find 127 down-regulated genes. However, we could not affirm if the data sets belonged to any specific gender, i.e., male or females.

Figure 1 Volcano plot showing all the gene expression change in young-old, young-middle and middle-old comparison. Black represents no change in expression (NO), blue represents downregulation (Down), and red represents upregulation (Up). logYO, logYM, and logMO on the x-axis labels represents : log2 fold changes respectively.

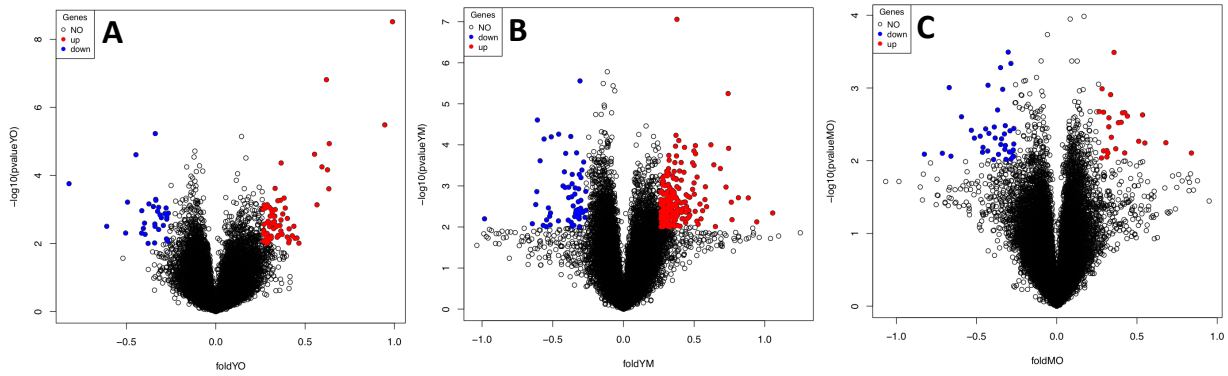
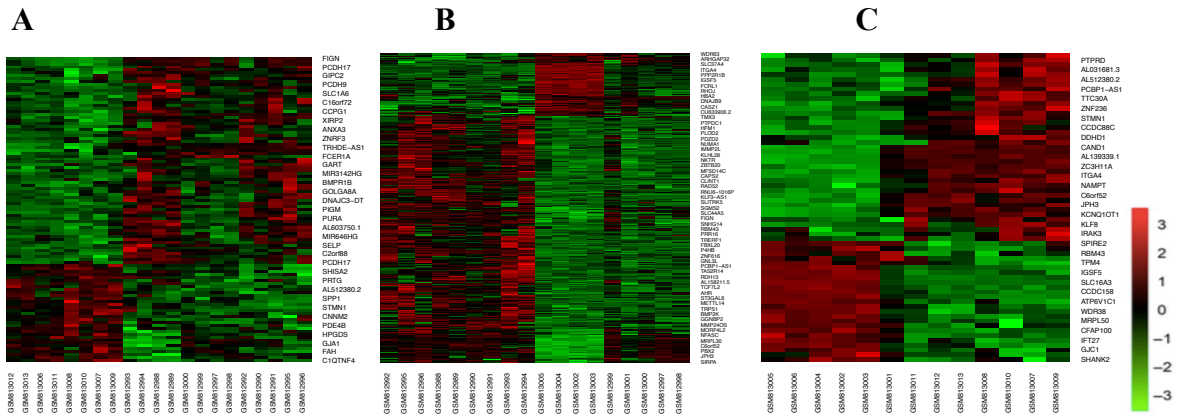


Figure 2 Heat map showing differentially expressed genes in (A) **young-old** and (B) **young-middle** (C) **middle-old** aged groups. Each column represents one sample, and each row represents one gene. The gene expression values of all samples are showed as base-2 logarithmic value. The gradual color ranging from green to red represents the changing process from down to upregulation differential expression.



3.2 Gene Ontology Enrichment Analysis

In Table 2, we can see that the DEGs were significantly enriched in biological processes including vasculogenesis, leukocyte migration, oligosaccharide metabolic process, and defense response to virus so on. For cell component, the DEGs were significantly enriched in DNA replication factor A complex, nucleoplasm, apical part of the cell and so on. In addition, the DEGs were significantly enriched in protein binding, GTPase activity and translation factor activity, RNA binding. KEGG signaling pathway analysis results showed that the DEGs were markedly enriched in hepatitis C pathway and hippo signaling pathway, and arrhythmogenic right ventricular cardiomyopathy (ARVC). Hippo signaling pathway has a critical role in stem cell and tissue specific progenitor cell and genes self-renewal and expansion (Table 2).

3.3. GO pathways results of biological processes of DEGs from the groups between young and middle aged groups

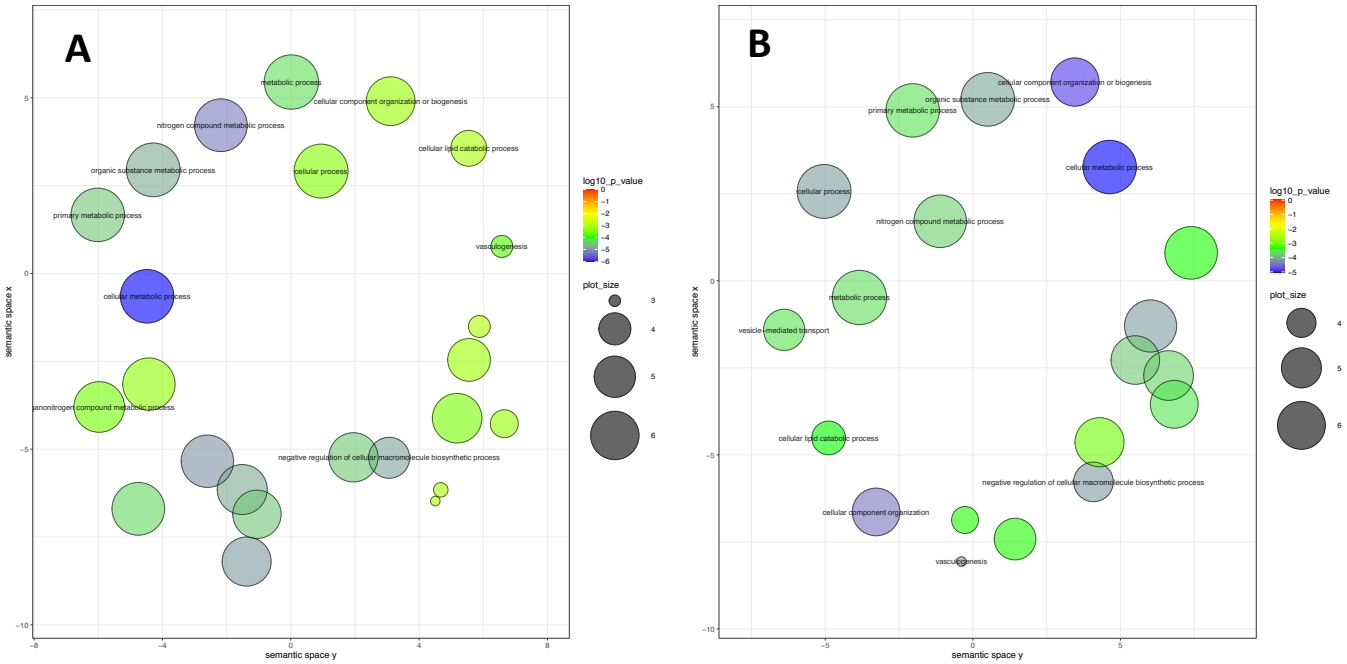
The most and most significant GO pathway results are obtained from the down-regulated DEGs between young and middle aged samples in Fig.3 (left). It demonstrates the most important pathways that play role in HSCs between young-middle and young-old aged samples. We used REVIGO that summarizes and visualizes long list of gene ontology terms using embedded R-script (Supek et al., 2011). Among 51 ontology terms, by visualizing in the scatter plot, 23 of the biological process terms seem to be the most significant ones are *cellular metabolic process, metabolic process,* and *negative regulation of cellular macromolecule biosynthetic process*. Similar to down-regulated DEGs, all the DEGs between young-old aged samples (Fig.3, right) can be summarized as the following; *cellular metabolic process*, *negative regulation of multicellular organismal process,* and *cellular component organization or biogenesis.*

Table 2 Gene ontology and pathway enrichment analysis of differentially expressed genes function in HSC data (top GO terms in each category)

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

Note. BP: biological process; CC: cell component; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; MF: molecular function (as ranked by the p-value)

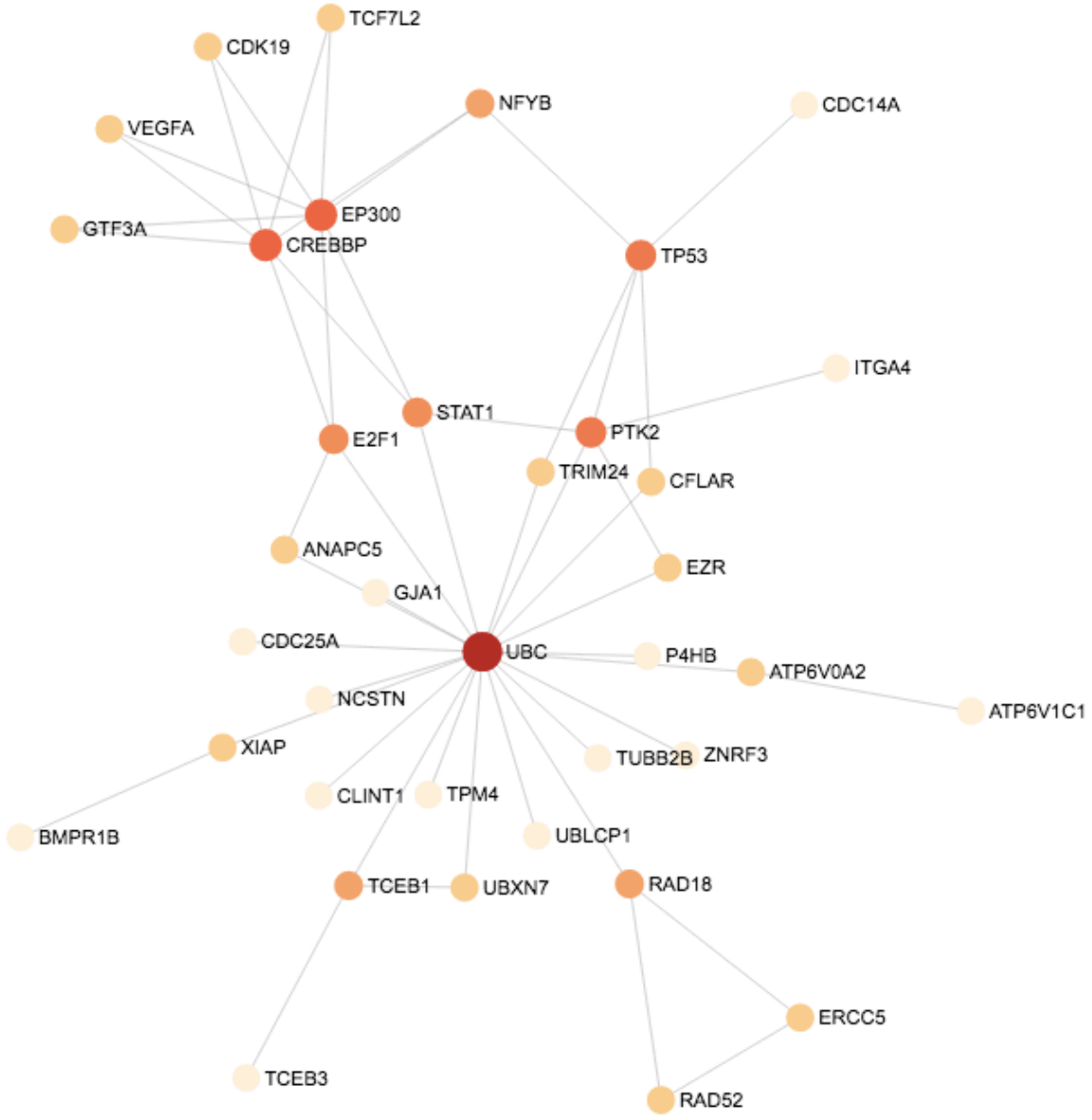
Figure 3 A ReViGo scatter plot of GO analyses in biological process from (A) down-regulated DEGs in the young-middle and (B) all the DEGs in the young-old aged groups. Related GO terms are clustered and presented with bubbles with similar color shades. Bubble colors represent p-values and bubble sizes indicate the relative frequency of the GO terms.



3.4. The PPI-network

In the visualized networks, the expressions and the degrees of connections of the nodes were symbolized by their colors from red to orange and areas, respectively. Figure 4 shows the PPI map between the set of input DEGs. Three best scoring genes by the p value are identified; these proteins define the functionality of PPI network. Best scoring genes belong to hepatitis C (hsa05160) and cell cycle pathway (hsa04110). Moreover, PPI network represents involvement of prostate cancer (hsa05215) and pathways in cancer (hsa05200) (See publicly available github repository).

Figure 4 Differentially expressed genes protein–protein interaction (PPI) network was constructed and visualized using NetworkAnalyst. UBC with the largest degree was suggested to be central to the PPI network associated with AML. Moreover, EP300, CREBBP, TP53, E2F1, and PTK2 with the secondary highest degree and STAT1 with the secondary Betweenness Centrality (BC) might be involved in the development AML and other substance diseases.



4 Discussions

While many studies have been carried out to uncover the causes and potential mechanisms of acute myeloid lymphoma (AML) and development during the past 10 years, the outcome is depressing. It is necessary to find particular biomarkers to assist early detection of AML which is central role in aiding patients the best likely result. Equivalently important, it is necessary to identify and validate new molecular drug targets in order to develop underlying therapeutic agents that may be successful in remedying AML and AML related cancers. To uncover potential biomarkers in the properties of an aged human hematopoietic system that may predispose to age-associated hematopoietic dysfunction, hematopoietic progenitor populations from healthy, hematologically normal young and elderly human bone marrow samples are evaluated.

Surprisingly, (Glorioso et al., 2011)most researches concentrate on an individual genetic result or the results are de- rived from an alone cohort research through microarray analysis which is not uniform with each other (Derosa et al., 2012; Mortensen et al., 2015; Satake et al., 2010; Wen, Geng, Li, Guo, & Zheng, 2014). Our study performed bi-group comparison profile from three category in which young, middle, and old aged samples within one data set and utilized bioinformatics methods to deeply analyze the data set and identified 453 significantly changed DEGs. The number of up-regulated genes was significantly higher than the down-regulated genes (326 versus 127).

Gene ontology information of these proteins is used to identify function and disease associated with proteins. 13 proteins UBC, EP300, CREBBP, TP53, PTK2, E2F1, STAT1, NFYB, RAD18, TCEB1, and ANAPC5, VEGFA proteins are predicted that are involved in various types of cancers like lung cancer, breast cancer, glioma, ovarian cancer, colorectal cancer, and leukemia. (Attar & Kurdistani, 2017; Tang et al., 2015; Rasheed et al., 1994; Olivier et al., 2006; Fan et al., 2019). Some proteins like GTF3A, STAT1, TFC7L2, and XIAP proteins are associated 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 can be used as markers for detection of early stage of type 1 and type 2 diabetes disease and can also act as potential 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).

Further work is required for wet lab verification of predicted genes that are expressed in HSC data set and express at the developmental stage AML and related diseases. More research is required in the field of cancer biology to identify AML and subset diseases at its budding stage. This paper also highlights the importance of microarray experiment in understanding AML and methodology to study various outcomes of gene expression data, like differentially expressed genes analysis, pathway and process identification, and protein-protein interaction network study.

References

Attar, N., & Kurdistani, S. K. (2017). Exploitation of EP300 and CREBBP lysine acetyltransferases by cancer. *Cold Spring Harbor Perspectives in Medicine*, *7*(3), a026534.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *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. *International Journal of Molecular Sciences*, *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. *Nature Cell Biology*, *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. *Experimental Hematology*, *34*(8), 967–975.

Davis, S., & Meltzer, P. (2007). GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics (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. *Statistical Science*, 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. *Nature Protocols*, *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 Bioinformatics*, *10*(1), 1–7.

Fraga, M., & Esteller, M. (2007). Fraga, M. F. & Esteller, M. Epigenetics and aging: the targets and the marks. Trends Genet. 23, 413-418. *Trends in Genetics : 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. *Neurobiology of Disease*, *41*(2), 279–290.

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. *Nature Genetics*, *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.

Hochberg, Y., & Tamhane, A. C. (1987). *Multiple comparison procedures*. John Wiley & Sons, Inc.

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 Research*, *35*(suppl_2), W169–W175.

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. *Antiviral Research*, *146*, 1–11.

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

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

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. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 20012–20017. https://doi.org/10.1073/pnas.1116110108

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. *Nature Reviews Molecular Cell Biology*, *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 Research*, *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.

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

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 Research*, *43*(D1), D447–D452.

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. *Scientific Reports*, *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. *American Journal of Obstetrics and Gynecology*, *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. *Molecular Pharmaceutics*, *7*(5), 1655–1666.

Xia, J., Gill, E. E., & Hancock, R. E. W. (2015). NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nature Protocols*, *10*(6), 823–844.

Zhou, G., Soufan, O., Ewald, J., Hancock, R. E. W., Basu, N., & Xia, J. (2019). NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Research*, *47*(W1), W234–W241.