



## Identification of differently expressed genes in leukemia using multiple microarray datasets

Z.Y. Zhang, R.Q. Xu, T.J. Guo, M. Zhang, D.X. Li and X.Y. Lu

Department of Hematology, Heze Hospital of Traditional Chinese Medicine, Heze, China

Corresponding author: Z.Y. Zhang  
E-mail: ziyanzhang12345@126.com

Genet. Mol. Res. 13 (4): 10482-10489 (2014)  
Received January 21, 2014  
Accepted May 15, 2014  
Published December 12, 2014  
DOI <http://dx.doi.org/10.4238/2014.December.12.9>

**ABSTRACT.** The purpose of this study was to identify differentially expressed genes and analyze biological processes related to leukemia. A meta-analysis was performed using the Rank Product package of Gene Expression Omnibus datasets for leukemia. Next, Gene Ontology-enrichment analysis and pathway analysis were performed using the Gene Ontology website and Kyoto Encyclopedia of Genes and Genomes. A protein-protein interaction network was constructed using the Cytoscape software. Using the Rank Product package for leukemia, we identified a total of 1294 differentially expressed genes, 357 of which were not involved in individual differentially expressed genes. Gene Ontology-enrichment analyses showed that these 357 genes were enriched in biological processes such as mRNA metabolism, RNA splicing, and mRNA processing. Pathway-enrichment analysis showed that the genes were involved in the intestinal immune network for IgA production, endocytosis, and the mitogen-activated protein kinase signaling pathway. The protein-protein interaction network indicated that *HRAS*, *CD44*, *STAT1*, *SMAD2*, and *COPS5* were important in many interactions. Our study revealed genes that were consistently differentially expressed

in leukemia, as well as the biological pathways and protein-protein interaction network associated with these genes.

**Key words:** Bioinformatics; Differentially expressed genes; Leukemia

## INTRODUCTION

Leukemia is a type of malignant tumor that seriously threatens human health and includes acute lymphoblastic leukemia (ALL) and chronic lymphocytic leukemia (CLL). ALL is a clonal disease arising from somatic mutations in lymphoid progenitor cells that alter the regulation of cellular proliferation, differentiation, and apoptosis (Daver and Udden, 2011). ALL is the most common type of cancer during childhood (Grupp et al., 2013), and constitutes 15% of adult leukemia cases (Chao et al., 2011). CLL is the most common leukemia in adults; its pathogenesis remains largely unknown (Fabbri et al., 2011).

Analysis of gene regulation mechanisms can increase the understanding of leukemia. Identification of a robust genetic marker signature would be beneficial for diagnosis and targeted treatment of leukemia in clinical practice. The use of gene expression analysis with high-throughput methods such as microarray has increased in recent years. Microarray technology has been used as an advanced strategy for identifying diagnostic gene signatures of human diseases on the genome-wide scale. The genome-wide discovery of such a signature would provide important insight into the underlying biological mechanisms of leukemia.

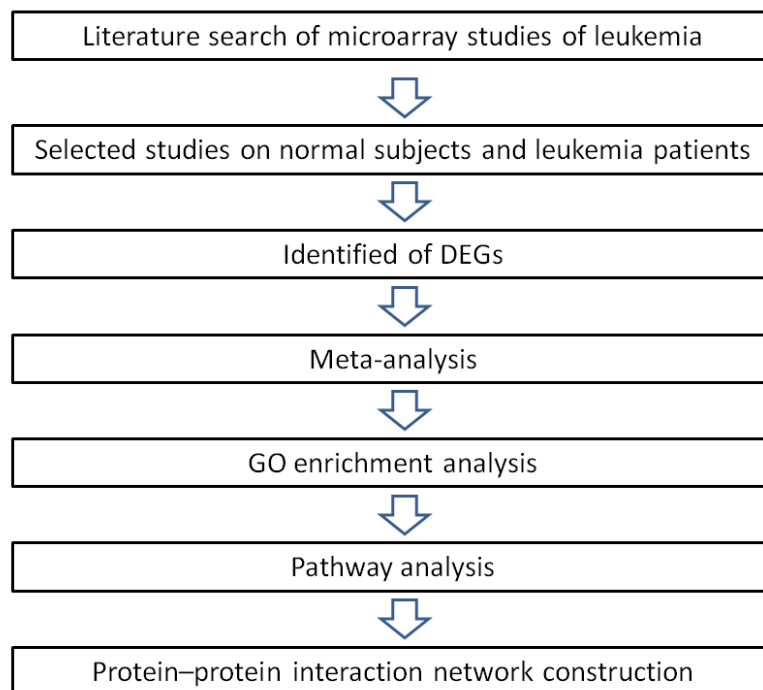
A significant amount of microarray data has been produced and deposited in publically available data repositories, including the Gene Expression Omnibus and ArrayExpress Archive. However, the results are inconsistent between studies because of factors such as small samples size. The Rank Product (RP) package, which was initially used to detect differently expressed genes (DEGs) in a single experiment, was improved to identify genes that were differently expressed under multiple-experimental conditions (Breitling et al., 2004; Hong et al., 2006).

To identify a more robust gene biomarker signature for leukemia, as well as to better understand the complex pathology associated with leukemia and identify molecular networks involved in the disease, we used a system biology approach to identify and integrate changes at the mRNA level between biopsy samples in patients with leukemia and in normal subjects. We integrated the microarray data for leukemia using RP method to build a more precise target network for the selected biomarkers of leukemia and identified DEGs through functional enrichment analysis, pathway-enrichment analysis, and protein-protein interaction (PPI). Our results may provide information that will increase the understanding of leukemia.

## MATERIAL AND METHODS

### Identification of gene expression datasets

In the present study, we examined DEGs between normal subjects and leukemia patients. Two microarray datasets were extracted from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus database (accession Nos. GSE32962, GSE22529). The experimental protocol of this study is shown in Figure 1. We excluded studies that included samples with other serious diseases such as nephropathy, diabetes, and hepatitis. We also excluded animal studies and studies in which microarray data were unclear.



**Figure 1.** Experimental protocol of this study.

### **Integrated analysis of DEGs**

The identification of DEGs and the meta-analysis were performed using the RP package (Hong et al., 2006). Using the RP method, a list of up- or down-regulated genes was selected based on the estimated percentage of false-positive predictions, which was also known as the false-discovery rate. Genes with a false-positive prediction  $\leq 0.05$  were considered to be differentially expressed between cases and controls.

### **Functional- and pathway-enrichment analysis**

To further investigate the functions of DEGs, we performed GO-enrichment analysis using the Gene Ontology database (<http://www.geneontology.org/>). To further investigate the signaling pathways of the DEGs, we performed pathway analysis using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database ([www.genome.jp/kegg/](http://www.genome.jp/kegg/)). The 2 analyses were performed using DAVID (Huang et al., 2009).

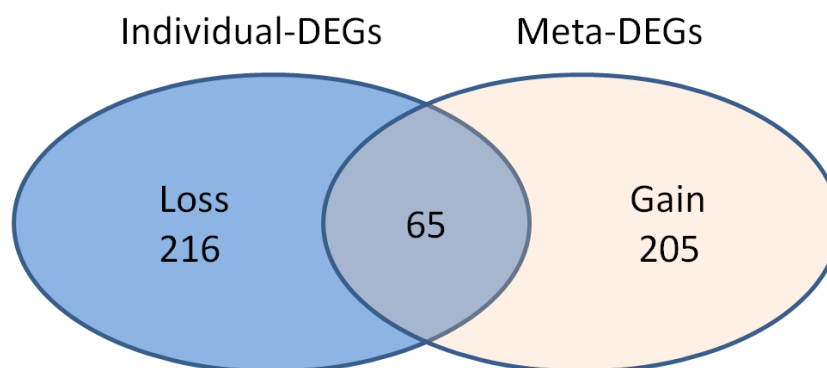
### **PPI network construction**

PPI data were downloaded from STRING (<http://string.embl.de/>). Next, the DEGs were imported into the interaction network and the interactions were screened with both end nodes containing DEGs. Networks were constructed using the Cytoscape software.

## RESULTS

### Integrated analysis of DEGs

We identified 1741 and 1265 DEGs from the datasets of GSE9476 and GSE22529, respectively. The following meta-analysis identified 1294 DEGs, including 634 up-regulated DEGs and 660 down-regulated DEGs. There were 357 gained genes and 2069 lost genes in this meta-analysis ([Table S1](#) and Figure 2). Lost genes were genes identified as DEGs in any individual analysis, but not in the meta-analysis.



**Figure 2.** Venn diagram showing overlap between DEGs identified from the meta analysis (Meta-DEGs) and those from each individual data analysis (individual-DEGs).

### Functional- and pathway-enrichment analysis

To further investigate the functions of the newly identified 357 genes, we performed GO analysis and pathway analysis. The most significant GO term was GO: 0016071 (mRNA metabolic process). The most significant term in pathway analysis was the intestinal immune network for IgA production. The top 10 GO terms and pathway terms are shown in Tables 1 and 2.

**Table 1.** Top 10 GO terms of Meta-DEGs.

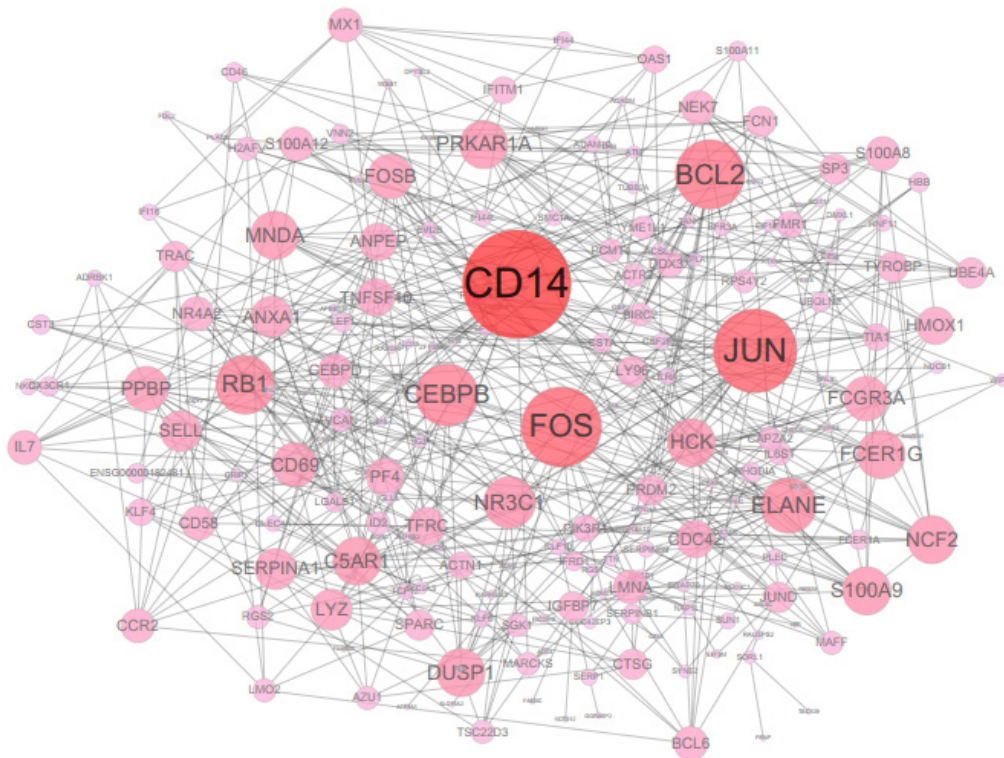
ID	Term	P value	Count
GO:0002376	Immune system process	1.53E-11	91
GO:0006955	Immune response	1.99E-11	68
GO:0006952	Defense response	3.59E-10	63
GO:0009611	Response to wounding	9.97E-08	49
GO:0006954	Inflammatory response	1.36E-07	29
GO:0002252	Immune effector process	1.41E-07	35
GO:0009617	Response to bacterium	1.80E-07	24
GO:0002682	Regulation of immune system process	3.10E-07	51
GO:0051707	Response to other organism	4.37E-07	34
GO:0044707	Single-multicellular organism process	5.80E-07	124

**Table 2.** Top 10 pathways analysis based on KEGG.

Term	P value	Count
Systemic lupus erythematosus	0.000214	8
Asthma	0.001531	4
Cytokine-cytokine receptor interaction	0.007311	9
Amoebiasis	0.011013	6
Osteoclast differentiation	0.011385	9
Leishmaniasis	0.016331	6
Apoptosis	0.018472	7
Renin-angiotensin system	0.023184	2
Chemokine signaling pathway	0.027262	9
Toxoplasmosis	0.028494	7

### Interaction network of DEGs

PPI data were downloaded from STRING (<http://string.embl.de/>). Next, the DEGs were imported into the interaction network and the interactions were screened with both end nodes containing DEGs. Using the Cytoscape software, the interaction network was identified (Figure 3). HRAS showed the highest degree in the network, followed by CD44, STAT1, SMAD2, and COPSS5. Genes with a degree greater than 10 in the PPI network are shown in Table 3.



**Figure 3.** PPI networks of Meta-DEGs. The nodes stand for the protein (gene), edges stand for the interaction of proteins. The size of the nodes represent the degree of node, the bigger nodes with higher degree.

**Table 3.** Genes that degree greater than 10 in PPI network.

Genes	Degree
<i>CD14</i>	30
<i>JUN</i>	23
<i>FOS</i>	22
<i>BCL2</i>	19
<i>CEBPB</i>	17
<i>RB1</i>	16
<i>ELANE</i>	15
<i>NR3C1</i>	14
<i>DUSP1, S100A9, MND4, FCER1G,</i> <i>C5AR1, NCF2, HCK, PRKARIA</i>	13
<i>FOSB, CD69, FCGR3A, PPBP</i>	12
<i>LYZ, ANXA1, SELL, SERPINA1</i>	11

## DISCUSSION

Leukemia is a complex disease with an unclear pathogenesis. The goal of this study was to identify important genes in leukemia to contribute to the understanding of its pathogenesis. We first combined the DEGs of microarray data using meta-analysis, and then analyzed these DEGs by functiona-enrichment analysis, pathway-enrichment analysis, and PPI.

A total of 357 DEGs were identified in this meta-analysis. These genes mainly participated in mRNA metabolic processes, RNA splicing, and mRNA processing (Table 1). Pathway-enrichment analysis revealed the involvement of DEGs in the intestinal immune network for IgA production, endocytosis, and the mitogen-activated protein kinase signaling pathway (Table 2).

HRAS showed the highest degree (26) in the PPI network (Figure 3). The HRAS gene encodes a protein with a molecular weight of 21 kDa, which is primarily involved in regulating cell growth, division, and apoptosis (Sun et al., 2012). HRAS belongs to the Ras gene superfamily (Tomei et al., 2012). Germline mutations in HRAS are thought to perturb human development and increase susceptibility to tumors (Aoki et al., 2005). HRAS overexpression was also found to be associated with an anaplastic phenotype in mammary adenocarcinomas (Jung et al., 2013).

CD44 participated in 23 interactions in the PPI network. CD44 is a cell membrane glycoprotein that mediates the cellular response to the cellular microenvironment and regulates growth, survival, differentiation, and motility (Loh et al., 2014). CD44 is a negative cell surface marker of pluripotent stem cell identification during human fibroblast reprogramming (Quintanilla Jr. et al., 2014) and regulates endothelial cell proliferation and apoptosis by modulating CD31 and VE-cadherin expression (Tsuneki and Madri, 2014). A previous study suggested that CD44 contributed to lenalidomide resistance in multiple myeloma (Bjorklund et al., 2014). Overexpression of the CD44 membrane receptor results in tumor initiation, growth, and cancer stem cell-specific behavior (Shah et al., 2013).

STAT1 belongs to the signal transducer and activator of transcription family (Samsonov et al., 2013). *STAT1* proteins are key mediators of type I and type III interferon signaling and are essential components of the cellular antiviral response and adaptive immunity (Au-Yeung et al., 2013). STAT1 regulates neural stem-cell function through transcriptional regulation of SOX9 (Imitola, 2013). STAT1 transcription factor activity enhances breast tumor growth and immune suppression (Hix et al., 2013). However, STAT1 was also reported to negatively regulate hepatocellular carcinoma cell proliferation (Chen et al., 2013).

SMAD2 and COPS5 also participated in various interactions in the PPI network. SMAD2 is a member of the transforming growth factor (TGF)- $\beta$  and activin-signaling pathway (Heyer et al., 1999). It is essential for maintaining the human and mouse primed pluripotent stem cell state (Sakaki-Yumoto et al., 2013). This protein is also a direct mediator of TGF- $\beta$  signaling (Brown et al., 2007). COPS5, which is also known as JAB1, is a component of the COP9 signalosome (Gemmill et al., 2002) and is thought to be a specificity factor for E2F1-induced apoptosis (Hallstrom and Nevins, 2006). Knock-down cells showed reduced mitotic arrest following exposure to nocodazole, resulting in cellular resistance to the drug (Kwak et al., 2005).

Several other genes also showed high degrees of association with leukemia, including SMARCA5 and SRSF3. SMARCA5 is a member of the SWI/SNF protein family, which is involved in the cell cycle (Leite et al., 2013). It is required for DNA double-strand break repair and DNA replication (Yamaguchi et al., 2013). SRSF3 is a member of a highly conserved family of splicing factors, which plays critical roles in key biological processes, including tumor progression (Tang et al., 2012). SRSF3 is a serine/arginine-rich protein that has been implicated in the misregulation of alternative splicing in cancer (Zhang and Manley, 2013).

In summary, the detailed mechanism of leukemia is not clear. Several genes associated with leukemia were identified in our study, and the function and signaling pathways in which these genes participated were examined systematically. Many of the genes identified have not been previously associated with leukemia. Those genes may play a role in leukemia, and additional studies should be conducted to further examine their roles.

### [Supplementary material](#)

## REFERENCES

- Aoki Y, Niihori T, Kawame H, Kurosawa K, et al. (2005). Germline mutations in HRAS proto-oncogene cause Costello syndrome. *Nat. Genet.* 37: 1038-1040.
- Au-Yeung N, Mandhana R and Horvath CM (2013). Transcriptional regulation by STAT1 and STAT2 in the interferon JAK-STAT pathway. *JAKSTAT* 2: e23931.
- Bjorklund CC, Baladandayuthapani V, Lin HY, Jones RJ, et al. (2014). Evidence of a role for CD44 and cell adhesion in mediating resistance to lenalidomide in multiple myeloma: therapeutic implications. *Leukemia* 28: 373-383.
- Breitling R, Armengaud P, Amtmann A and Herzyk P (2004). Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.* 573: 83-92.
- Brown KA, Pietenpol JA and Moses HL (2007). A tale of two proteins: differential roles and regulation of Smad2 and Smad3 in TGF-beta signaling. *J. Cell Biochem.* 101: 9-33.
- Chao MP, Alizadeh AA, Tang C, Jan M, et al. (2011). Therapeutic antibody targeting of CD47 eliminates human acute lymphoblastic leukemia. *Cancer Res.* 71: 1374-1384.
- Chen G, Wang H, Xie S, Ma J, et al. (2013). STAT1 negatively regulates hepatocellular carcinoma cell proliferation. *Oncol. Rep.* 29: 2303-2310.
- Daver NG and Udden MM (2011). Acute lymphoblastic leukemia. In: *Tumor Board Review: Guideline and Case Reviews in Oncology* Demos Medical Publishing, New York, 298.
- Fabbri G, Rasi S, Rossi D, Trifonov V, et al. (2011). Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *J. Exp. Med.* 208: 1389-1401.
- Gemmill RM, Bemis LT, Lee JP, Sozen MA, et al. (2002). The TRC8 hereditary kidney cancer gene suppresses growth and functions with VHL in a common pathway. *Oncogene* 21: 3507-3516.
- Grupp SA, Kalos M, Barrett D, Aplenc R, et al. (2013). Chimeric antigen receptor-modified T cells for acute lymphoid leukemia. *N. Engl. J. Med.* 368: 1509-1518.
- Hallstrom TC and Nevins JR (2006). Jab1 is a specificity factor for E2F1-induced apoptosis. *Genes Dev.* 20: 613-623.
- Heyer J, Escalante-Alcalde D, Lia M, Boettinger E, et al. (1999). Postgastrulation Smad2-deficient embryos show defects in embryo turning and anterior morphogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 96: 12595-12600.

- Hix LM, Karavitis J, Khan MW, Shi YH, et al. (2013). Tumor STAT1 transcription factor activity enhances breast tumor growth and immune suppression mediated by myeloid-derived suppressor cells. *J. Biol. Chem.* 288: 11676-11688.
- Hong F, Breitling R, McEntee CW, Wittner BS, et al. (2006). RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* 22: 2825-2827.
- Huang da W, Sherman BT and Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44-57.
- Imitola J (2013). Stat1 Regulates Neural Stem-cell Function by Transcriptional Regulation of SOX9. The 27th Annual Meeting of the CMSC and the 5th Cooperative Meeting of the CMSC-ACRIMS.
- Jung S, Ro SW, Jung G, Ju HL, et al. (2013). Sleeping Beauty transposon system harboring HRAS, c-Myc and shp53 induces sarcomatoid carcinomas in mouse skin. *Oncol. Rep.* 29: 1293-1298.
- Kwak HJ, Kim SH, Yoo HG, Park SH, et al. (2005). Jun activation domain-binding protein 1 is required for mitotic checkpoint activation via its involvement in hyperphosphorylation of 53BP1. *J. Cancer Res. Clin. Oncol.* 131: 789-796.
- Leite KR, Morais DR, Reis ST, Viana N, et al. (2013). MicroRNA 100: a context dependent miRNA in prostate cancer. *Clinics* 68: 797-802.
- Loh TJ, Moon H, Cho S, Jung DW, et al. (2014). SC35 promotes splicing of the C5-V6-C6 isoform of CD44 pre-mRNA. *Oncol. Rep.* 31: 273-279.
- Quintanilla RH Jr, Asprer JS, Vaz C, Tanavde V, et al. (2014). CD44 is a negative cell surface marker for pluripotent stem cell identification during human fibroblast reprogramming. *PLoS One* 9: e85419.
- Sakaki-Yumoto M, Liu J, Ramalho-Santos M, Yoshida N, et al. (2013). Smad2 is essential for maintenance of the human and mouse primed pluripotent stem cell state. *J. Biol. Chem.* 288: 18546-18560.
- Samsonov A, Zenser N, Zhang F, Zhang H, et al. (2013). Tagging of genomic STAT3 and STAT1 with fluorescent proteins and insertion of a luciferase reporter in the cyclin D1 gene provides a modified A549 cell line to screen for selective STAT3 inhibitors. *PLoS One* 8: e68391.
- Shah V, Taratula O, Garbuzenko OB, Taratula OR, et al. (2013). Targeted nanomedicine for suppression of CD44 and simultaneous cell death induction in ovarian cancer: an optimal delivery of siRNA and anticancer drug. *Clin. Cancer Res.* 19: 6193-6204.
- Sun XF, Li L, Li XJ and Shen W (2012). Methylation pattern of oncogene HRAS gene promoter region and its clinical relevance to urocytic tumorigenesis. *Mol. Biol. Rep.* 39: 8431-8437.
- Tang Y, Horikawa I, Ajiro M, Robles AI, et al. (2013). Downregulation of splicing factor SRSF3 induces p53beta, an alternatively spliced isoform of p53 that promotes cellular senescence. *Oncogene* 32: 2792-2798.
- Tomei S, Adams S, Uccellini L, Bedognetti D, et al. (2012). Association between HRAS rs12628 and rs112587690 polymorphisms with the risk of melanoma in the North American population. *Med. Oncol.* 29: 3456-3461.
- Tsuneki M and Madri JA (2014). CD44 regulation of endothelial cell proliferation and apoptosis via modulation of CD31 and VE-cadherin expression. *J. Biol. Chem.* 289: 5357-5370.
- Yamaguchi H, Matsumoto S, Ishibashi M, Hasegawa K, et al. (2013).  $\beta$ -glucuronidase is a suitable internal control gene for mRNA quantitation in pathophysiological and non-pathological livers. *Exp. Mol. Pathol.* 95: 131-135.
- Zhang J and Manley JL (2013). Misregulation of pre-mRNA alternative splicing in cancer. *Cancer Discov.* 3: 1228-1237.