# Integrated bioinformatic analysis of microarray data reveals shared gene signature between MDS and AML

ZHEN ZHANG[1*], LIN ZHAO[1*], XIJIN WEI[2], QIANG GUO[1], XIAOXIAO ZHU[1], RAN WEI[1], XUNQIANG YIN[1,3], YUNHONG ZHANG[1,3], BIN WANG[2] and XIA LI[1]

[1]Laboratory for Molecular Immunology, Institute of Basic Medicine, Shandong Academy of Medical Sciences, Jinan, Shandong 250062; [2]Department of Peripheral Vascular Surgery, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, Shandong 250011; [3]School of Medicine and Life Sciences, University of Jinan-Shandong Academy of Medical Sciences, Jinan, Shandong 250062, P.R. China

Received March 2, 2018; Accepted June 20, 2018

DOI: 10.3892/ol.2018.9237

**Abstract.** Myeloid disorders, especially myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML), cause significant mobility and high mortality worldwide. Despite numerous attempts, the common molecular events underlying the development of MDS and AML remain to be established. In the present study, 18 microarray datasets were selected, and a meta-analysis was conducted to identify shared gene signatures and biological processes between MDS and AML. Using NetworkAnalyst, 191 upregulated and 139 downregulated genes were identified in MDS and AML, among which, *PTH2R*, *TEC*, and *GPX1* were the most upregulated genes, while *MME*, *RAG1*, and *CD79B* were mostly downregulated. Comprehensive functional enrichment analyses revealed oncogenic signaling related pathway, fibroblast growth factor receptor (FGFR) and immune response related events, 'interleukine-6/interferon signaling pathway, and B cell receptor signaling pathway', were the most upregulated and downregulated biological processes, respectively. Network based meta-analysis ascertained that *HSP90AA1* and *CUL1* were the most important hub genes. Interestingly, our study has largely clarified the link between MDS and AML in terms of potential pathways, and genetic markers, which shed light on the molecular mechanisms underlying the development and transition of MDS and AML, and facilitate the understanding of novel diagnostic, therapeutic and prognostic biomarkers.

*Correspondence to:* Dr Bin Wang, Department of Peripheral Vascular Surgery, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, 16369 Jingshi Road, Jinan, Shandong 250011, P.R. China
E-mail: 2801505631@qq.com

Dr Xia Li, Laboratory for Molecular Immunology, Institute of Basic Medicine, Shandong Academy of Medical Sciences, 18877 Jingshi Road, Jinan, Shandong 250062, P.R. China
E-mail: 786735868@qq.com

*Contributed equally

*Key words:* myelodysplastic syndrome, acute myeloid leukemia, microarray, gene expression profile, meta analysis

## Introduction

Hematopoietic stem cells (HSCs) reside in bone marrow (BM), and the BM microenvironment provides essential extrinsic signals to maintain the repopulation and differentiation of HSC (1). In addition, lineage specific transcriptional factors account fundamentally for the intrinsic control of HSC (2). When normal differentiation is hampered, the accumulation of immature HSCs and malignant neoplastic proliferation will occur. Myeloid disorders, especially myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML), are the most frequently reported malignant cases, which cause high mortality in adults (3). MDS is defined as clonal HSC malignancies characterized by ineffective hematopoiesis and dysplasia, with clinical manifestations of peripheral cytopenias, hypercellular BM, and variable degrees of increased blasts (4). AML normally originates from a small number of hematopoietic leukemic stem cells (LSCs) in BM, and their self-renewal and differentiation will generate leukemic progenitors, which will produce a considerable amounts of immature clonogenic leukemic blasts and interfere the normal hematopoiesis (5). Heterogeneous subsets of MDS and AML patients have been classified following World Health Organization (WHO) or French-American-British criteria (6), mainly based on subjective clinical findings (number of cytopenias and percentage of marrow blasts) and biological properties (specific cytogenetic and molecular lesions) (4). In addition, approximately 30% of the patients with MDS would progress into AML (7). Thus, a number of approaches have been developed to compare MDS and AML in the cytogenetic and molecular aspects, and some hallmark genes or phenotypes being established (8). However, the molecular mechanisms and biological events underlying MDS and AML development and their transition remain to be addressed.

Gene expression profile analysis is a powerful research strategy, which integrates data in genetics, molecular transcription, and functional genomics to reveal dysregulated genes between patients and healthy donors. Microarray provides

increasing body of gene-wide transcriptional data regarding MDS and AML. However, results vary between studies due to diversity in cohort selection, specimen source, and experimental designs. Therefore, meta-analysis is advantageous to enhance statistical power to detect the dysregulated genes and biological pathways by combining different publically available datasets.

Microarray data integration-based meta-analyses rely on efficient *in silico* tools. With the advances of ever-growing theories and bioinformatics tools, we can now employ *in silico* tools to efficiently combine multiple microarray datasets regardless of different populations, experimental designs, and diseases (9). NetworkAnalyst is a powerful web-based tool, which supports robust and reliable gene expression analysis through approaches including preliminary data processing, sample annotation, batch effect adjustment, dataset integration, and results visualization (10). To maximally overcome the impact caused by the differences in study design and platform usage among different datasets, 'Combining Effect Size (ES)' analysis and Random Effect Modeling were applied to achieve more consistent and accurate results by taking into consideration of both direction and magnitude of gene expression changes. In the present study, we have selected 8 and 10 eligible microarray datasets for MDS and AML, respectively, from publicly available dataset repositories. To our knowledge, this is the first time that common transcriptional signature of MDS and AML are illuminated in patients vs. healthy individuals based on meta-analysis. Our results will shed light on the mechanistic foundations for MDS progression into AML, and propose novel targets for the prevention and development of both MDS and AML (11).

## Materials and methods

*Search strategy*. Microarray-based gene expression profile studies were identified in the Pubmed database (http://www.pubmed.com), Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/gds/) and ArrayExpress dataset of the European Molecular Biology Laboratory-European Bioinformatics Institute (http://www.ebi.ac.uk/arrayexpress/) for MDS and AML. The following key words were used for MDS and AML respectively: myelodysplastic syndrome or dysmyelopoietic syndrome or hematopoetic myelodysplasia or MDS, and microarray or gene expression profile or gene expression profiling and acute myeloid leukemia or acute myeloblastic leukemia or acute myelocytic leukemia or acute nonlymphoblastic leukemia or acute nonlymphocytic leukemia) or acute myelogenous leukemia or AML or ANLL and microarray or gene expression profile or gene expression profiling.

*Inclusion and exclusion criteria*. Eligible studies and datasets should follow these inclusive criteria: i) patient and healthy control studies of human; ii) analysis of gene expression profiling; iii) comparable experimental conditions and untreated; and iv) available complete raw and processed microarray data. Studies were excluded if they were: i) letters, abstracts, meta-analysis, review articles and case repor; ii) cell lines were used in experimental design; iii) RT-PCR only for profiling studies; and iv) studies without healthy control. All the datasets and references, which are conformed to the criteria mentioned above, were carefully screened. The latest search was performed on May 30, 2017.

*Data extraction and processing*. Full text and supplementary materials of selected articles were screened and the key items were extracts as listed: GEO series accession number, type of disease, number of patients and healthy donors, specimen sources and platform of microarray (Table I). The series matrix files were downloaded from GEO datasets for all studies, with the exception of GSE983, whose CEL files were obtained and processed further by R platform to generate preliminary series matrix file. Common Entrez IDs were used to substitute all the gene probes in accordance with the corresponding microarray platforms. Before integrative meta-analysis, individual dataset was normalized by log2 transformation and R-mediated mean, and quantile normalization. Expression data of patients with MDS or AML and healthy donors were defined as class 2, and 1, respectively, according to the guidelines of NetworkAnalyst (10).

*Batch effect adjustment*. NetworkAnalyst is capable of massive datasets integration on the premise of batch effect adjust option. The processed and normalized datasets were uploaded and subjected to the well-established ComBat procedures to remove study-specific batch effects, which uses the Emperical Bayes method to adjust the extreme expression ratios, alleviate gene variances across all other genes, and possibly removing their inference without compromising the biological covariates (12). The sample clustering patterns with and without batch effect adjustment were visualized and compared by principal component analysis (PCA) to assess the efficiency of batch effect removal.

*Meta-analysis*. We conducted the meta-analysis using NetworkAnalyst, a web interface for integrative statistical and visualizing tool. In the option of 'multiple gene expression data' for the web interface, All datasets were uploaded to the 'multiple gene expression data' input area and analyzed in a streamlined manner, including data processing for Entrez ID, annotation check, view by both Boxplot and PCA plot, confirmation of normalization, individual DEGs analysis and data summary. For the DEGs discrimination, the cut-off of P-value was adjusted to 0.05, using the false discovery rate (FDR) based on Benjamini-Hochberg procedure and moderated t-test based on the Limma algorithm. Furthermore, all datasets were subjected to integrity check to ensure that the merged data could be carried out by ES combination, which allowed the generation of more biologically consistent meta-based DEGs by incorporation of both the magnitude and direction of gene fold change. Between the two popular methods: Fixed (FEM) and random effects models, REM was used in the current study. REM model assumes that each study contains a random ES that incorporates unknown cross-study heterogeneities, as demonstrated by Cochran's Q tests. 'Define custom signature' tool from NetworkAnalyst was used to produce the heatmap visualization for both top 25 up- and downregulated genes.

*PPI Network analysis*. PPI Network-based analysis was performed using NetworkAnalyst (10). In PPI networks,

Table I. Summary of individual studies included in the meta-analysis.

| Author, year | GEO accession no. | Disease | Sample source | Platform | (Refs.) |
|---|---|---|---|---|---|
| Del Rey *et al*, 2013 | GSE41130 | MDS | Bone marrow mononuclear cells | Affymetrix Human Genome U133 Plus 2.0 Array | (8) |
| Pellagatti *et al*, 2010 | GSE19429 | MDS | Bone marrow $CD34^+$ cells | Affymetrix human genome U133 plus 2.0 array | (13) |
| Sternberg *et al*, 2005 | GSE2779 | MDS | Bone marrow CD34+ cells | Affymetrix human genome U133A array | (14) |
| Graubert *et al*, 2011 | GSE30195 | MDS | Bone marrow CD34+ cells | Affymetrix human genome U133 plus 2.0 array | (15) |
| Pellagatti *et al*, 2006 | GSE4619 | MDS | Bone marrow CD34+ cells | Affymetrix human genome U133 plus 2.0 array | (16) |
| Wang *et al*, 2013 | GSE51757 | MDS | Bone marrow | Agilent-028004 surePrint G3 human GE 8x60K microarray | Unpublished |
| Gerstung *et al*, 2015 | GSE58831 | MDS | Bone marrow CD34+ cells | Affymetrix human genome U133 plus 2.0 srray | (17) |
| Xu *et al*, 2016 | GSE81173 | MDS | Bone marrow CD34+ cells | Affymetrix human gene expression array | Unpublished |
| Kikushige *et al*, 2010 | GSE24395 | AML | Bone marrow $CD34+CD38^-$ cells | Sentrix human-6 v2 expression beadchip | (18) |
| de Jonge *et al*, 2011 | GSE30029 | AML | Bone marrow CD34+ cells | Illumina human HT-12 V3.0 expression beadchip | (19) |
| Bacher *et al*, 2012 | GSE33223 | AML | Bone marrow CD34+ cells | Affymetrix human genome U133 plus 2.0 array | (20) |
| Stirewalt *et al*, 2012 | GSE37307 | AML | Bone marrow CD34+ and peripheral blood cells | Affymetrix human genome U133A array | Unpublished |
| Schneider *et al*, 2015 | GSE68172 | AML | Bone marrow | Affymetrix human genome U133 plus 2.0 array | (21) |
| Virtaneva *et al*, 2001 | GSE70284 | AML | Bone marrow | Affymetrix human full length HuGeneFL array | (22) |
| Zheng *et al*, 2016 | GSE79605 | AML | Bone marrow mononuclear cells | Agilent-014850 whole Human genome microarray | Unpublished |
| von der Heide *et al*, 2016 | GSE84881 | AML | Bone marrow mesenchymal stromal cells | Affymetrix human genome U133 plus 2.0 array | (23) |
| Stirewalt *et al*, 2008 | GSE9476 | AML | Bone marrow CD34+ and peripheral blood cells | Affymetrix human genome U133A array | (24) |
| Stegmaier *et al*, 2004 | GSE983 | AML | Primary patient AML cells | Affymetrix human full length HuGeneFL array | (25) |

MDS, myelodysplastic syndrome; AML, acute myeloid leukemia; NA, not available.

nodes stand for proteins, while edges represent known interactions between the linked proteins. Topology and subnetwork analyses were performed in this study, in purpose to demonstrate its overall structural properties and highlight part of the network that had shown significant changes. In short, the subnetwork analysis was executed in three steps: i) obtaining the list of DEGs by meta-analysis; ii) proceeding the list to the IMEx interactom-based PPI network analysis; and iii) choosing zero-order and minimum network to avoid 'hairball effect' and overall presentation of the structure. In addition, there are two kinds of complementary measurements in NetworkAnalyst to reveal the most important nodes, also called hub genes: Degree and betweenness centrality. Degree centrality is the number of connections that a node has to other nodes, whereas betweenness centrality corresponds to the number of shortest paths passing through the node. From the parent network, the most significant modules of hub genes were extracted for both up- and downregulated

DEGs using the 'module explorer tool', based on the random walks-dependent Walktrap algorithm.

*Functional gene set enrichment analysis of shared DEGs*. To demonstrate the implication of shared DEGs in MDS and AML, we conducted an enriched pathway group analysis by ClueGO of Cytoscape, a software with significantly expanded annotated gene sets of Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome pathway databases. The processes were submitted to the followed settings and parameters: enrichment (right-sided) hyper-geometric distribution tests, with a P-value ≤0.05; Benjamini and Hochberg adjustment for the terms and the groups with Kappa-statistics score threshold set to 0.4; leading term groups were selected based on the highest significance. The gene ontology (GO) terms enrichment analysis was conducted by topGO of Bioconductor version 2.14, R 3.1.1, and summarized by REVIGO. The enriched GO terms or pathways were considered to be significant with adjusted P-value <0.05.

*Statistical analysis*. The web-server NetworkAnalyst analyzed the data and performed meta-analysis by REM-based combined ES statistics. DEGs were selected by FDR guided Benjamini-Hochberg procedure with adjusted P-value <0.05. Hypergeometric test (right-sided) and Benjamini & Hochberg FDR correction were used to identify significantly enriched biological pathways in DEGs by ClueGO of Cytoscape. Significantly enriched biological processes were monitored using top GO and REVIGO analysis and P<0.05 was considered to indicate a statistically significant difference.

## Results

*Selection of eligible microarray datasets*. 2012 and 476 literatures were retrieved from Pubmed for AML and MDS, respectively, according to our search strategy. And after a series of inclusion and exclusion screening, a total of 8 and 10 microarray datasets were selected for MDS (GSE19429, GSE2779, GSE30195, GSE41130, GSE4619, GSE51757, GSE58831 and GSE81173) (8,13-17) and AML (GSE24395, GSE30029, GSE33223, GSE37307, GSE68172, GSE70284, GSE79605, GSE84881, GSE9476 and GSE983) (18-25) respectively, as indicated in Materials and methods (Fig. 1A). All the 18 datasets were carefully screened to meet the inclusion and exclusion criteria for meta-analysis. Collectively, a total of 106/281 and 75/465 (healthy control/patient) samples were included for MDS and AML, respectively. There are 8 different microarray platforms applied in the selected datasets, which include Affymetrix Human Genome U133 Plus 2.0 Array, Affymetrix Human Genome U133A Array, Affymetrix Human Full Length HuGeneFL Array, Affymetrix Human Gene Expression Array, Illumina HumanHT-12 v.3.0 expression beadchip, Sentrix Human-6 v2 Expression BeadChip, Agilent-014850 Whole Human Genome Microarray, and Agilent-028004 SurePrint G3 Human GE 8x60 K Microarray. In addition, various tissues, such as total BM cells, FACS-sorted $CD34^+$ hematopoietic progenitors, and peripheral blood cells, were used for the microarray analysis. In the process of data integration, patient samples of MDS and AML were not distinguished, for the purpose to uncover the common gene signature shared between MDS and AML. Table I presents the detailed information for each dataset, including GEO accession number, type of disease, sample composition, sample source, and the corresponding references.

*Identification of common differentially expressed genes (DEGs) between MDS and AML*. The workflow for meta-analysis used in the present study was illustrated in Fig. 1B. To identify the shared transcriptional signatures between MDS and AML, the total of 18 datasets were simultaneously analyzed by NetworkAnalyst. When the 18 microarray datasets were analyzed individually, a total of 7579 DEGs were revealed. For meta-analysis, we uploaded the 18 datasets in succession, and each dataset was processed by Entrez ID matching, sample (control/patient) annotation and individual DEGs identification. To remove batch effects among different datasets, we performed 'ComBat'-based batch effect adjustment (12), and the sample clustering patterns with or without batch effect adjustment were visualized by PCA plots (Fig. 2A and B). All the 18 gene expressional microarray data were then integrated and merged. Meta-analysis was conducted following the Cochran's Q test, REM and ES statistical methods, which facilitated to reveal the DEGs between healthy donors and patients across different microarray datasets by permitting variable true ES and integrating unknown cross-study heterogeneities. Finally, we found a total of 330 DEGs, including 191 up- and 139 downregulated genes across the 18 datasets with significance threshold of adjusted P-value <0.05. By comparing the list of individually identified DEGs with meta-analysis based DEGs, 211 DEGs were revealed by both analyses. Of note, 119 DEGs were uniquely discovered by meta-analysis, which were referred as gained genes, and 7378 DEGs were only found in individual analysis, which were defined as lost genes (Fig. 2C). Expression profile of the top 25 up- and downregulated genes among the 330 identified DEGs was visualized by heatmap (Fig. 2D). Due to the large number of healthy donor and patient samples, the heatmap was divided into 3 parts by GEO datasets. The gene expressional profiles of the top 25 up- and downregulated genes could be easily visualized to be consistent across different datasets, except those with small sample numbers, including GSE33223, GSE79605, GSE41130 and GSE84881. The *Parathyroid hormone 2 receptor* (*PTH2R*), *Tec protein tyrosine kinase* (*TEC*), and *Glutathione peroxidase 1* (*GPX1*) were among the most significantly upregulated genes, and *PTH2R* had the highest combined ES of 1.0994, which was upregulated consistently with ES ranging from 0.47068 to 2.8508 across all the 18 datasets (data not shown), while *Matrix metalloproteinase 12* (*MME*), *Recombination activating 1* (*RAG1*), and *CD79b molecule* (*CD79B)* were among the most significantly downregulated genes, and *MME* had the relatively highest combined ES of -1.29, which was consistent in 15 out of 18 datasets with ER ranging from -3.7844 to -0.33961 (data not shown). Additionally, top 10 up- and downregulated DEGs were listed in Table II.

*Hub genes identification by network based meta-analysis*. Many studies have shown that 'hub' proteins are more likely to be encoded by pleiotropic genes or genes that are related to certain diseases (26). Hub nodes are potentially key molecules in signaling, as they are highly interconnected with dysregulated

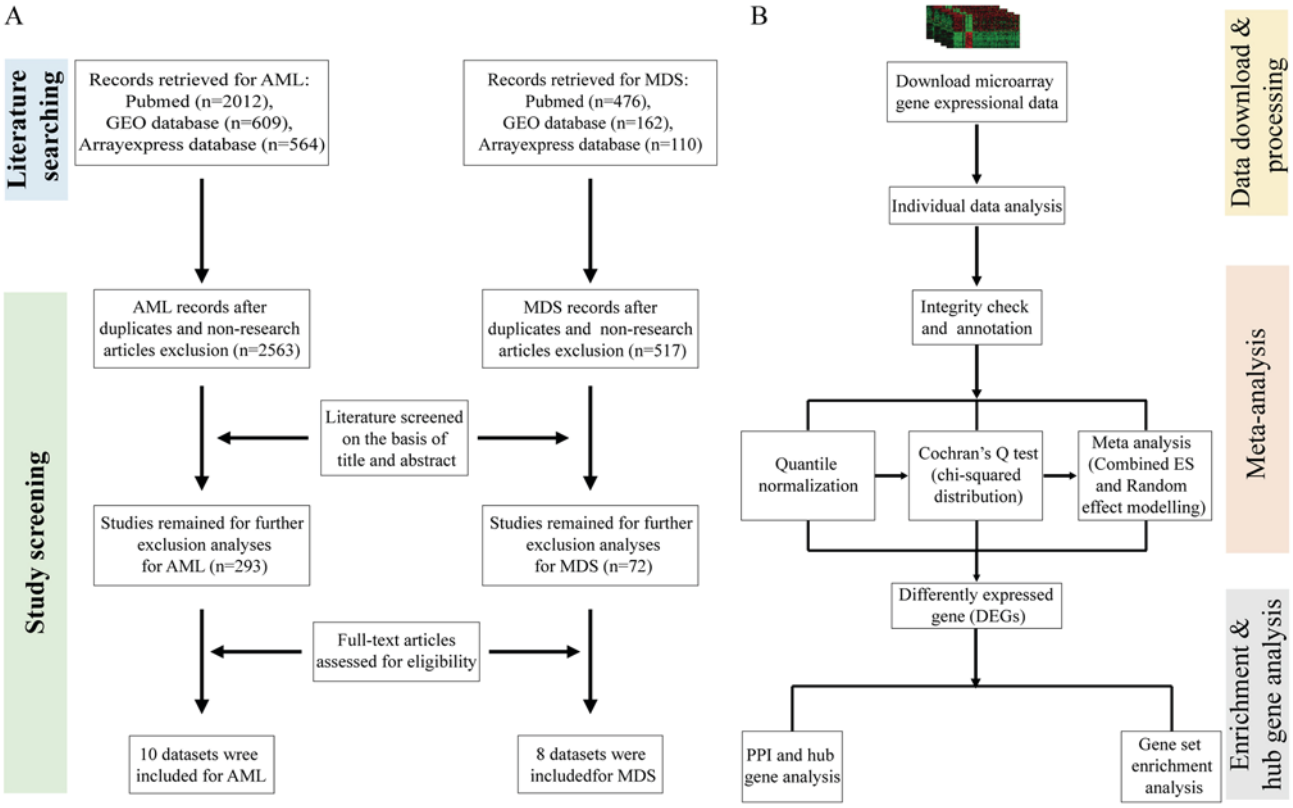


Figure 1. Flowcharts for microarray datasets selection and meta-analysis. (A) Selection process of microarray datasets for meta-analysis of shared gene expressional signature between MDS and AML. (B) Process of meta-analysis based data exploration. MDS, myelodysplastic syndrome; AML, acute myeloid leukemia.

genes and they often receive and integrate multiple signals and pass them onto downstream nodes (27). In order to identify the key hub genes among the common DEGs shared by MDS and AML, we intentionally performed a network based meta-analysis. NetworkAnalyst provided us with a protein-protein interaction (PPI) network analysis tool by integrating the IMEx (International Molecular Exchange Consortium) interactome with the original seed of 330 DEGs. The NetworkAnalyst-based PPI network analysis generated first-order, minimum-order and zero-order networks with (4507 nodes, 9421 edges), (873 nodes, 3510 edges), and (123 nodes, 157 edges), respectively. The minimum-order network shows the overall structure (Fig. 3A), while the zero-order network helps to visualize the detailed interactions between different seeds (Fig. 3B). *HSP90AA1* (combined ES: 0.36197, adjusted P-value: 0.0026418) and *CUL1* (combined ES: -0.54806, adjusted P-value: 0.002118) were found to be the top ranked hub genes by degree and betweenness centrality analysis among the up- and downregulated DEGs. We also listed the top 10 ranked hub genes with detailed information in Table III. Additionally, we used the 'module explorer' tool to highlight two most significant modules as two sub-networks composed of HSP90AA1 (19 nodes and 23 edges) and CUL1 (20 nodes and 25 edges), respectively (Fig. 3C and D).

*Gene set enrichment analysis for identification of overrepresented biological pathways*. In order to thoroughly understand the pathways involved in MDS and AML development, the enriched biological pathways for up- and downregulated DEGs were functionally grouped with the threshold of P-value <0.05 by the ClueGO plugin of Cytoscape v.3.4.0, and only KEGG and Reactome pathway databases were selected. For upregulated DEGs, we found 88 enriched biological pathways, which can be divided into 16 groups, mainly involved in 'Signaling by fibroblast growth factor receptor (FGFR) in disease', 'EPH-ephrin mediated repulsion of cells', 'Histidine, lysine, phenylalanine, tyrosine, proline and tryptophan catabolism', 'Positive epigenetic regulation of rRNA expression', and 'Neurophilin interactions with VEGF and VEGFR' (Fig. 4A). For downregulated DEGs, we found 37 enriched biological pathways, which can be divided into 17 groups, mainly involved in 'Interleukine-6 signaling', 'Integration of provirus', 'Graft-versus-host disease', 'B cell receptor signaling pathway', and 'Effects of PIP2 hydrolysis' (Fig. 4B).

*Gene set enrichment analysis for identification of overrepresented GO terms*. Gene set enrichment analyses were conducted for the analysis of overrepresented gene ontology (GO) terms, which were enriched by topGO package of R for the 191 up- and 139 downregulated DEGs with the threshold of P-value <0.05. REVIGO is a web server that summarizes long, unintelligible lists of GO terms by finding a representative subset of the terms using a simple clustering algorithm that relies on semantic similarity measures (28). Using REVIGO, the top 200 enriched GO terms for upregulated DEGs were shown in 20 subsets, including 'positive regulation of tyrosine phosphorylation of Stat4 protein' (comprised by 58 terms), 'adrenal cortex formation' (comprised by 33 terms), 'establishment or maintenance of transmembrane electrochemical gradient' (comprised by 20 terms), 'heart contraction' (comprised by 10 terms), etc, which were visualized by treemap generated by R (Fig. 5A). Similarly, 13 subsets were shown for downregulated

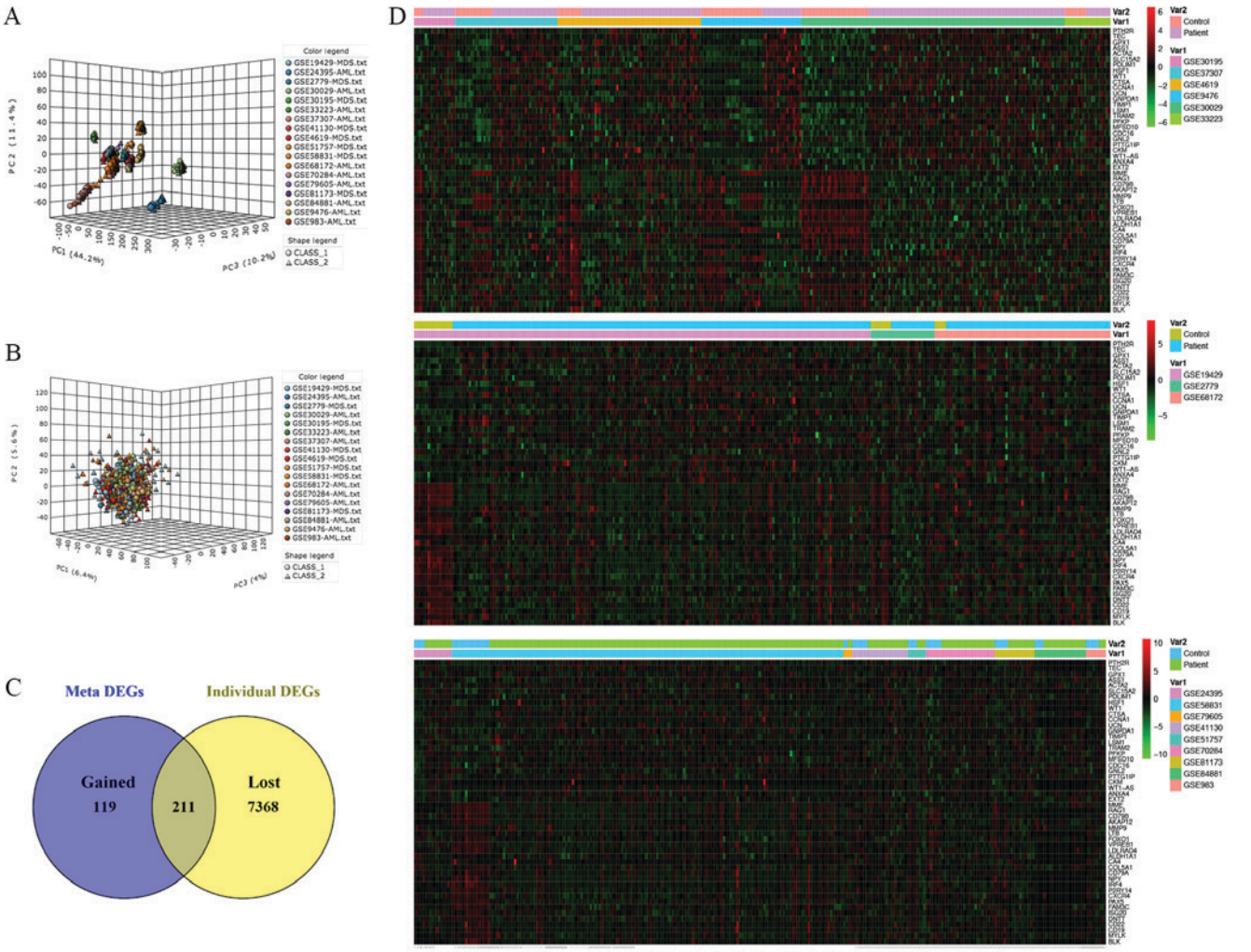


Figure 2. Meta-analysis based DEGs and gene expression profiles. (A) PCA-3D plot for sample clustering of microarray datasets without batch effect adjustment. (B) PCA-3D plot for sample clustering of microarray datasets with batch effect adjustment. (C) Venn diagram of DEGs by meta-analysis (meta DEGs) and individual microarray dataset analysis (individual DEGs). (D) Heat-map visualization of expressional profiles for top 25 up- and downregulated DEGs identified by meta-analysis. Genes were ranked by combined ES value. DEGs, DEGs, differentially expressed genes; Var1: variate 1, represents different datasets by colors; Var2: variate 2, represents control and patient samples by colors.

DEGs, including 'B cell receptor signaling pathway' (comprised by 75 terms), 'embryo implantation' (comprised by 24 terms), 'fructose metabolism' (comprised by 23 terms), 'response to lipopolysaccharide' (comprised by 18 terms), etc, which were also visualized by treemap (Fig. 5B).

## Discussion

Myeloid disorders, including MDS and AML, represent a group of hematopoietic malignancies with monoclonal expansion of immature myeloid lineages. Both MDS and AML are characterized by abnormal accumulation of defective or immature blasts in the BM, and MDS patients with 10-19% blasts are considered as high-risk of progressing to AML (>20% blasts) (29). The transformed immature cells in both diseases are biologically, genetically, and molecularly similar, thus identification of common molecular markers may best indicate the appropriate risk factors, as well as preventive, diagnostic and therapeutic decisions for these patients. Despite significant amounts of studies have used microarray-based technology to identify molecular markers in MDS and AML, inconsistent results have been reported due to diversity in patient selection, tissue source and study designs. Therefore, in the present study, we attempted to identify the common gene signature underlying MDS and AML by a comprehensive meta-analysis of 18 publically available microarray datasets. We found that there were 330 DEGs of P-value <0.05 in total shared by both MDS and AML, with 191 up- and 139 down-regulated. Importantly, 119 out of 330 DEGs were identified uniquely by meta-analysis, not individual studies. By uncovering shared gene expressional profiles, this study highlight potential diagnostic and prognostic biomarkers in MDS and AML, and may aid in understanding the molecular mechanisms of their development and progression.

Among the top ten upregulated DEGs, *PTH2R* encodes a receptor for parathyroid hormone (PTH), which belongs to the G-protein coupled receptor 2 families, and has been previously suggested as novel marker for AML (30). TEC, a non-receptor type protein-tyrosine kinase, was revealed to be highly expressed in MDS patients (31). *GPX1*, which encodes a glutathione peroxidase and helps to reduce organic hydroperoxides and hydrogen peroxide ($H_2O_2$) by glutathione, has been reported to be dramatically upregulated in

Table II. Top 20 DEGs shared by MDS and AML.

A, Top 10 upregulated genes

| Entrez ID | Gene symbol | Combined ES | Adjusted P-value |
|---|---|---|---|
| 5746 | *PTH2R* | 1.0994 | P<0.001 |
| 7006 | *TEC* | 0.8975 | $1.3x10^{-06}$ |
| 2876 | *GPX1* | 0.8564 | $3.2x10^{-03}$ |
| 445 | *ASS1* | 0.8543 | $1.2x10^{-06}$ |
| 59 | A*CTA2* | 0.7831 | $7.8x10^{-04}$ |
| 6565 | *SLC15A2* | 0.7372 | $2.2x10^{-03}$ |
| 9124 | *PDLIM1* | 0.7348 | $1.4x10^{-04}$ |
| 3297 | *HSF1* | 0.7343 | $8.0x10^{-04}$ |
| 7490 | *WT1* | 0.7320 | $3.3x10^{-03}$ |
| 5476 | *CTSA* | 0.7271 | $1.3x10^{-04}$ |

B, Top 10 downregulated genes

| Entrez ID | Gene symbol | Combined ES | Adjusted P-value |
|---|---|---|---|
| 4311 | *MME* | -1.2867 | $2.9x10^{-06}$ |
| 5896 | *RAG1* | -1.1663 | $2.2x10^{-03}$ |
| 974 | *CD79B* | -1.1597 | $2.0x10^{-05}$ |
| 9590 | *AKAP12* | -1.1201 | $1.9x10^{-03}$ |
| 4318 | *MMP9* | -1.1189 | $2.9x10^{-02}$ |
| 4050 | *LTB* | -1.0662 | $1.9x10^{-06}$ |
| 2308 | *FOXO1* | -1.0603 | $6.1x10^{-12}$ |
| 7441 | *VPREB1* | -1.0197 | $1.7x10^{-02}$ |
| 753 | *C18orf1* | -1.0169 | $1.5x10^{-03}$ |
| 216 | *ALDH1A1* | -0.9986 | $3.7x10^{-04}$ |

DEGs, differentially expressed genes; ES, effect size.

AML and relate with MDS by SNPs assay and DNA methylation (32,33). *ASS1* (Argininosuccinate synthetase 1) gene product is responsible for the process of arginine biosynthesis; however, Miraki-Moud *et al* (34), reported that most AML lacked ASS1 expression, which may be due to small sample size and the different detection methods. Other genes, such as HSF1 (Heat-shock transcription factor 1), WT1 (Wilms tumor 1), ACTA2 (Actin, alpha 2), SLC15A2 (Solute carrier family 15 member 2), PDLIM1 (C terminal LIM domain protein 1), and CTSA (Cathepsin A) were revealed by our meta-analysis, and even some of them have been reported previously with increased expression in AML and MDS (35,36), their potential as diagnostic or prognostic markers in MDS and AML needs further exploration. Among the top ten downregulated DEGs, although *MME* encoded protein MMP12 was believed to be related with acute lymphocytic leukemia (ALL) diagnosis (37), its role in MDS and AML remains to be demonstrated. *AKAP12* encodes a kinase, which is a part of the holoenzyme of PKA and serves as a scaffold protein for signaling transduction. The low expression of AKAP12 had been reported in both MDS and AML (38), which was consistent with our results. *MMP9* encodes a matrix metalloproteinase protein, and has been shown to express at reduced levels in AML. *LTB* (Lymphotoxin beta) gene expression has been reported to decrease in malignant myeloid cells, and is potentially involved in AML (39). *FOXO1* (Forkhead box O1) is a tumor suppressor gene and its low expression level was shown to correlate with AML with FLT3 internal tandem duplication mutation (Flt3-ITD) (40). *ALDH1A1* (Aldehyde dehydrogenase 1 family member A1) was also found to be minimally expressed or undetectable in about 25% AML patients (41). In addition, our meta analysis also revealed a number of novel genes, such as *RAG1* (Recombination activating 1); *CD79B*; *VPREB1* (V-set pre-B cell surrogate light chain 1) and *C18orf1* (Low density lipoprotein receptor class A domain containing 4), which have not been previously reported in MDS and AML. These candidate genes require further studies to evaluate their biological functions and biomarker potentials in both MDS and AML.

The network biology analysis is an efficient way for systematically investigating the molecular complexity of a particular disease, facilitating the discovery of biomarker and drug targets (42). Integrating the list of DEGs by meta-analysis with IMEx interactome-based PPI network, we identified *HSP90AA1* and *CUL1* as the most important hub genes among up- and downregulated DEGs, respectively, based on network centrality scoring across the 18 datasets. *HSP90AA1*, also known as *LAP2/ HSPC1*, encodes a product that serves as protein chaperone with major functions in protein folding and stabilization as a homodimer, and by regulating a number of cancer related proteins such as AKT, CDK4, HIF-1, VEGFR, ERBB2 and MMPs, *HSP90AA1* largely dictates tumor proliferation, survival, invasion, metastasis and angiogenesis (43). Consistent with our meta-based DEGs analysis, increased expression of *HSP90AA*1 has been well stated in both MDS and AML (44,45). The protein level of *HSP90AA*1 was higher in patients with higher grade MDS, which is associated with short survival and increased risk of progression into AML (44). Similarly, AML patient with higher HSP90AA1 level showed lower remission rates, and was correlated with poorer AML prognosis (46). The network-based hub analyses also revealed that HSP90AA1 served as the central upregulated protein, which may coordinate a cohort of signaling pathways mediated by LTK, MAP3K3, MAP2 K, etc, and promote MDS and AML formation. Our results validated and supported the importance of targeting of HSP90AA1 for MDS and AML therapy in the clinic (47,48). The downregulatedhub gene, *CUL1*, encodes an essential component of the SCF (SKP1-CUL1-F-box protein) E3 ubiquitin ligase complex, in which it serves as the scaffold protein to organize SKP-1-F-box protein and RBX1 subunit (49). The complex mediates the ubiquitination involved in cell cycle progression, signal transduction and transcription (50). Thus, loss of CUL1 expression may impair signaling cascades involved cell cycle, signaling pathways and protein expression, which may lead to the development of MDS and AML.

In order to decipher the enriched biological pathways participating in MDS and AML, we submitted the up- and downregulated DEGs separately to Clue GO plugin of Cytoscape to group functionally related pathways. Interestingly, we found that 'Signaling by FGFR in disease'

Table III. Top 10 shared hub genes identified by network based meta-analysis.

| Gene symbol | Regulation | Degree | Betweenness | Combined ES |
|---|---|---|---|---|
| HSP90AA1 | Up | 722 | 2416884.66 | 0.3620 |
| CUL1 | Down | 603 | 1592425.89 | -0.5481 |
| CUL5 | Up | 366 | 700674.50 | 0.3516 |
| IL7R | Down | 199 | 349830.53 | -0.7247 |
| MAP3K3 | Down | 176 | 304739.58 | -0.4870 |
| XRCC5 | Down | 169 | 372212.88 | -0.4065 |
| CDKN2A | Up | 148 | 316047.11 | 0.5783 |
| RPL11 | Up | 134 | 157024.97 | 0.4898 |
| SP3 | Down | 132 | 376542.88 | -0.5418 |
| TLE1 | Down | 108 | 302176.33 | -0.6311 |

ES, effect size.

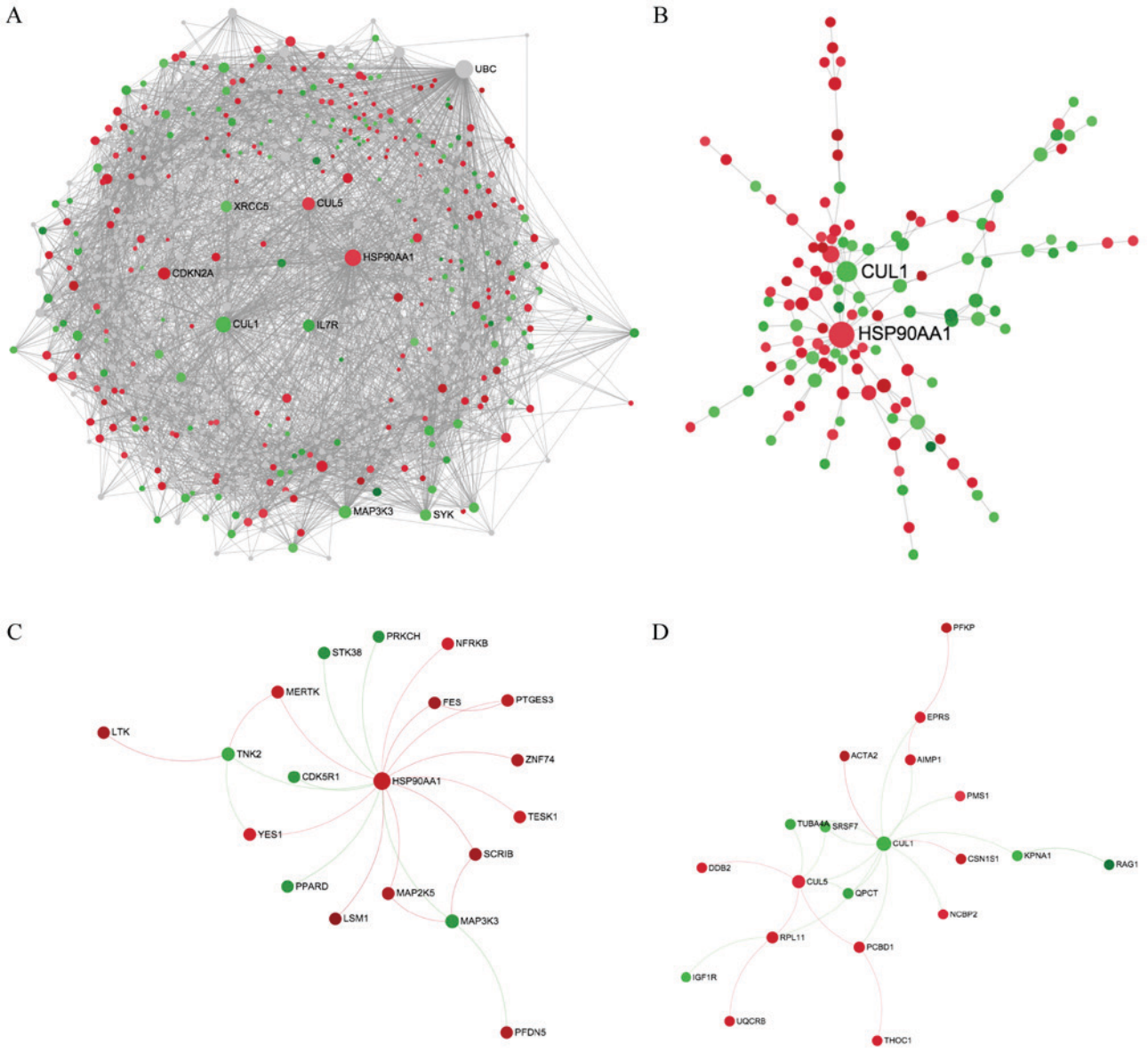


Figure 3. PPI network based hub gene analysis. (A) Minimum order of PPI network structure of DEGs identified by meta-analysis with Fruchterman-Rengold layout. Red nodes represent upregulated and green nodes represent downregulated DEGs. (B) Zero order PPI network of shared DEGs by meta-analysis. (C and D) PPI subnetworks representative of up- and downregulated DEGs. WalkTrap algorithm based 'module explorer' of NetworkAnalyst extracted the module. PPI, protein-protein interaction; DEGs, DEGs, differentially expressed genes.

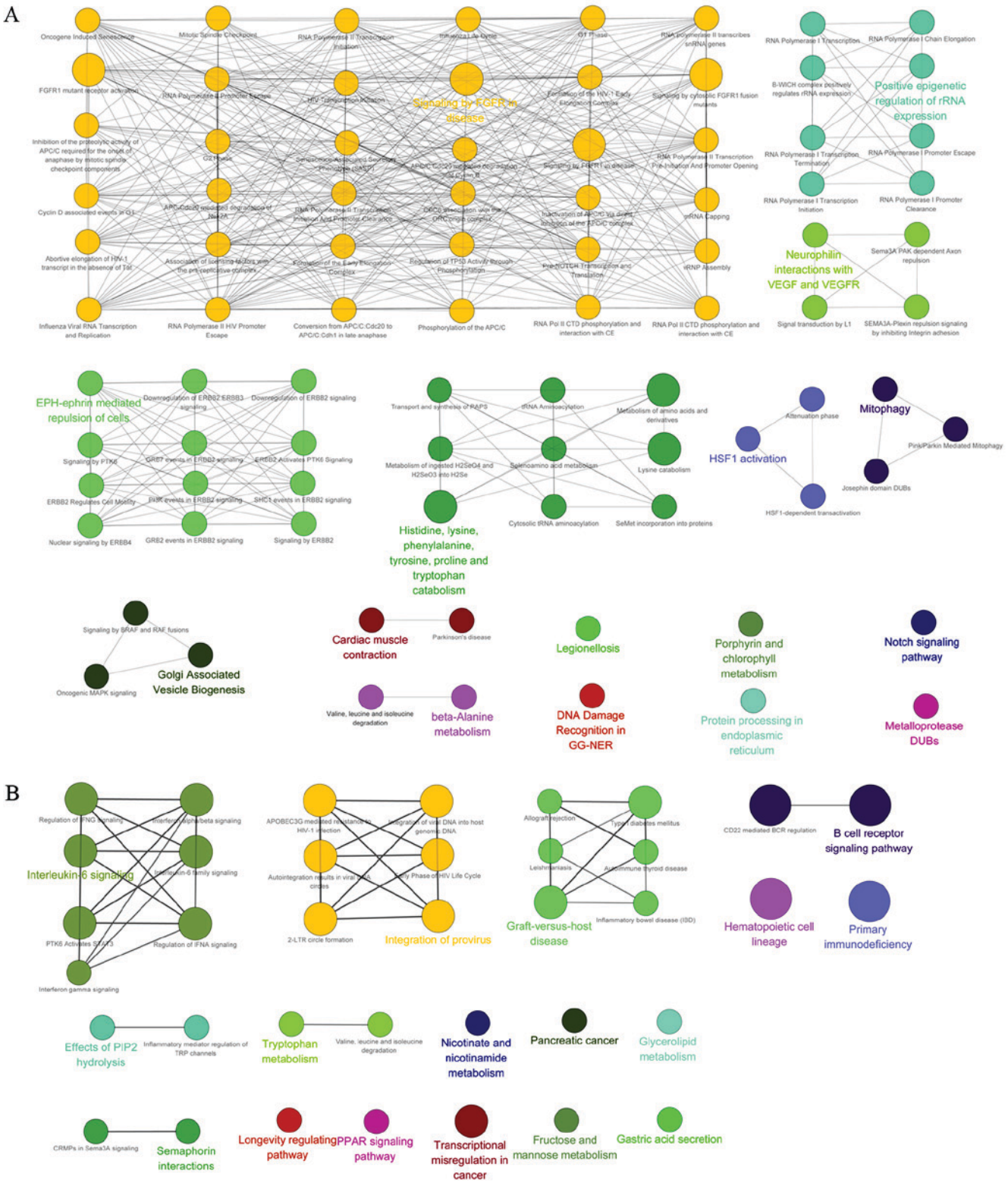


Figure 4. Over representation of enriched pathways for DEGs. (A) Enriched pathway groups were generated with Cytoscape plug-in (ClueGO) by integrating the upregulated genes with KEGG and Reactome pathways. (B) Enriched pathway groups were generated by integrating the downregulated genes with KEGG and Reactome pathways. The node size indicates greater significance of enrichment, and the colors represent different groups. The pathways with adjusted P-value <0.05 are shown in the network. DEGs, DEGs, differentially expressed genes; KEGG, Kyoto Encyclopedia of Genes and Genomes.

related pathways group was top enriched. The FGFR family is composed of four kinase receptor members: FGFR1-4, which are universally expressed and the FGFR signaling activation had been demonstrated to promote survival, migration of AML cells, and their resistance to chemotherapy. In the DEGs of our meta-analysis, FGF7 and FGFR3 gene expression were both upregulated. FGF7 is a ligand of FGFR2b with high affinity, which have been recently reported to serve as a potentially niche factor for hematopoietic stem and progenitor cells (HSPCs) support and leukemic growth by activating

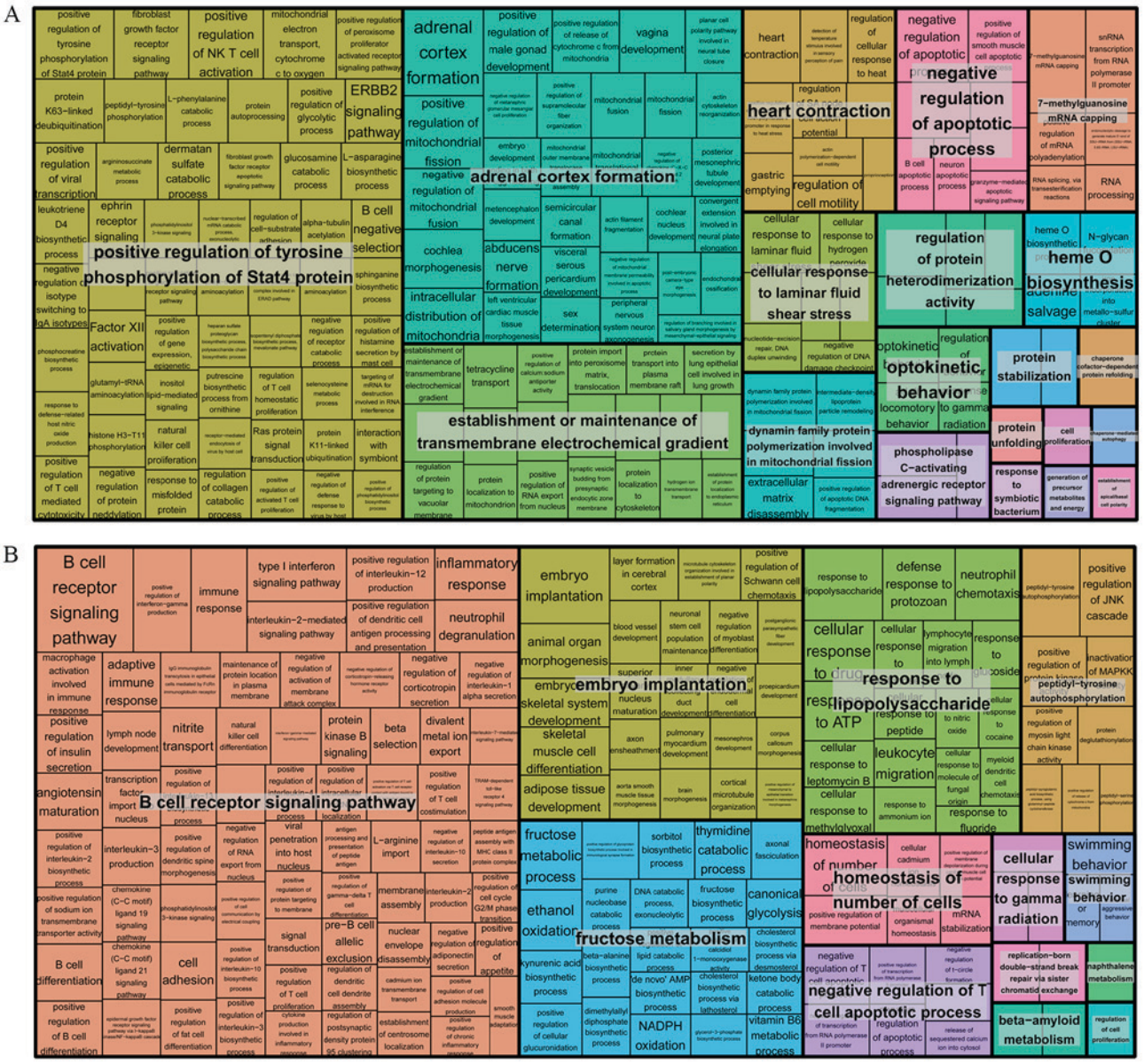


Figure 5. Over representation of enriched biological processes for DEGs. (A) REVIGO gene ontology treemap for upregulated DEGs by meta-analysis. (B) REVIGO gene ontology treemap for downregulated DEGs by meta-analysis. DEGs, DEGs, differentially expressed genes.

FGFR2b signaling pathway (51). Besides, the truncated recombinant human FGF7, palifermin, which compete for binding of FGFR2b, has been FDA-approved for the treatment of patients with oral mucositis (52). It is meaningful to explore the potential application of palifermin in MDS and AML intervention or even therapy according to our meta analysis. FGFR3 belongs to a family of receptor tyrosine kinases (RTKs) responding to FGF, and will stimulate the downstream signaling modules, including the phosphatidylinositol 3-kinase (PI3K)/AKT and phospholipase C-γ (PLC-γ) pathways (53,54). And in murine model of leukemia, it has been revealed that FGFR-3 activity is important for hematopoietic transformation (55). For the downregulated genes, 'Interleukine-6 signaling' related pathways were the top enriched, which were well represented by immune related genes: *IL12RB2*, *IRF4* and *IRF8* down-regulation. *IL12RB2* encodes a type I transmembrane protein identified as a subunit of the interleukin 12, which interacts with IL12RB1 and form the high-affinity binding site for IL12 to reconstitute IL12 dependent signaling. Although *IL12RB2* down regulation have not yet been reported in MDS or AML, its silencing in B cell malignant tumors illustrated that neoplastic B cells would escape the IL-12-mediated apoptosis and growth inhibition in the absence of IL12RB2 (56). From this point of view, the down regulation of *IL12RB2* in myeloid cells might help them to evade the anti-tumor activity of IL-12. Besides, *IRF4* and *IRF8* gene products belong to IRF (interferon regulatory factor) family of transcription factors. It has been reported that the *Irf4*$^{-/-}$ *Irf8*$^{-/-}$ double knockout mice can develop aggressive myeloid disorders rather like CML (57). Furthermore, IRF4 induction by the long non-coding RNA linc-223 would inhibit cell proliferation and stimulate AML cell differentiation (58). The promoter of IRF8 was found to be hypermethylated in MDS and AML patients (59), which might associate with its transcriptional repression revealed by our meta analysis. In addition, *IRF8* expression level was potential prognostic

biomarkers for adult patients with AML (60). Similar results were also observed by REVIGO analysis, that 'FGFR signaling pathway' was highly enriched, whereas a series of immune related pathways were downregulated including 'B cell receptor signaling pathway', 'positive regulation of interferon-gamma production', 'adaptive immune response', etc, which further highlighted the significance of immune responses down-regulation in MDS and AML development.

Our meta-analysis is based on gene expressional microarray, which has become a principle technology for transcriptome analysis to support drug screening and health evaluation. With the development of genetic detecting technology, next-generation sequencing technologies present new ways of genetic mutation analysis, especially for whole genome sequencing. Currently, there was no comparable research focusing on common gene expressional profiles of MDS and AML by transcriptome or genomic sequencing; therefore, we compared the DEGs of MDS or AML uncovered by our meta-analysis with that of whole genome sequencing individually. We found that 4 (*ZFHX2, PTPRD, STAG2* and *ALAS2*) out of 105 somatic mutated genes for MDS, and 2 (*FLT3* and *WT1*) out of 23 significantly mutated genes for AML were covered by DEGs of our meta-analysis respectively (data not shown) (61,62). The finding indicated that the potential diagnostic or prognostic biomarkers obtained by our meta-analysis are more likely to undergo transcriptional regulations instead of genetic mutations.

In summary, our meta-analysis revealed 330 DEGs, some of which have been proved critical for MDS or AML progress, and some others deserve further exploration for their potential as biomarkers for both MDS and AML. Functional enrichment analysis demonstrated that tumor related processes or pathways were upregulated, such as 'Signaling by FGFR in disease'; however, the immune response related pathway, such as 'Interleukine-6 signaling' was downregulated in both MDS and AML, which may predict to the common events during MDS or AML development.

**Acknowledgements**

Not applicable.

**Funding**

The present study was supported by the Natural Science Foundation of China (grant nos. 81601442, 81373670, 81673981 and 81704116), the Primary Research and Development Plan of Shandong Province (grant nos. 2016GSF202016, 2017GSF19118 and 2017GSF218013), the Project of Transformation in High-tech Achievements (grant no. 2013ZHZX2A0405), the Natural Science Foundation of Shandong Province (grant nos. ZR2017PH008 and ZR2018PH042), the Science and Technology Development Grant of the State Administration of traditional Chinese medicine of Shandong Province (grant nos. 2013-2016 and 2017-174), the Family Planning Committee of Shandong Province [grant no. (2014) 14], the Project for Shandong Medical and Health Science and Technology Plan (grant nos. 2015WS0191 and 2016WS0524), the Project of Science and Technology of Shandong Academy of Medical Sciences (grant nos. 2016-34, 2016-35 and 2017-15) and the Innovation Project of Shandong Academy of Medical Sciences.

**Availability of data and materials**

The datasets analyzed during the current study are available in the Gene Expression Omnibus repository (https://www.ncbi.nlm.nih.gov/gds/) (GSE19429, GSE2779, GSE30195, GSE4619, GSE51757, GSE58831, GSE81173, GSE24395, GSE30029, GSE33223, GSE37307, GSE68172, GSE70284, GSE79605, GSE84881, GSE9476 and GSE983).

**Authors' contributions**

ZZ, XW, LZ and QG individually searched the eligible datasets suitable for meta-analysis. ZZ, XZ and RW analyzed the data; ZZ, XY and YZ produced the graphs, BW and XL conceived and designed the project; ZZ, BW and XL wrote the manuscript. All authors have read and approved the final manuscript.

**Ethics approval and consent to participate**

Not applicable.

**Patient consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**References**

1. Morrison SJ and Scadden DT: The bone marrow niche for haematopoietic stem cells. Nature 505: 327-334, 2014.
2. Tenen DG, Hromas R, Licht JD and Zhang DE: Transcription factors, normal myeloid development, and leukemia. Blood 90: 489-519, 1997.
3. Ozbalak M, Cetiner M, Bekoz H, Atesoglu EB, Ar C, Salihoglu A, Tuzuner N and Ferhanoglu B: Azacitidine has limited activity in 'real life' patients with MDS and AML: A single centre experience. Hematol Oncol 30: 76-81, 2012.
4. Greenberg PL, Young NS and Gattermann N: Myelodysplastic syndromes. Hematology Am Soc Hematol Educ Program: 136-161, 2002.
5. Hope KJ, Jin L and Dick JE: Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. Nat Immunol 5: 738-743, 2004.
6. Vardiman JW, Harris NL and Brunning RD: The World Health Organization (WHO) classification of the myeloid neoplasms. Blood 100: 2292-2302, 2002.
7. Adès L, Itzykson R and Fenaux P: Myelodysplastic syndromes. Lancet 383: 2239-2252, 2014.
8. Del Rey M, O'hagan K, Dellett M, Aibar S, Colyer HA, Alonso ME, Díez-Campelo M, Armstrong RN, Sharpe DJ, Gutiérrez NC, *et al*: Genome-wide profiling of methylation identifies novel targets with aberrant hypermethylation and reduced expression in low-risk myelodysplastic syndromes. Leukemia 27: 610-618, 2013.
9. Gomez-Cabrero D, Abugessaisa I, Maier D, Teschendorff A, Merkenschlager M, Gisel A, Ballestar E, Bongcam-Rudloff E, Conesa A and Tegnér J: Data integration in the era of omics: Current and future challenges. BMC Syst Biol 8 (Suppl 2): I1, 2014.
10. Xia J, Gill EE and Hancock RE: NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. Nat Protoc 10: 823-844, 2015.

11. Haidich AB: Meta-analysis in medical research. Hippokratia 14 (Suppl 1): S29-S37, 2010.
12. Johnson WE, Li C and Rabinovic A: Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8: 118-127, 2007.
13. Pellagatti A, Cazzola M, Giagounidis A, Perry J, Malcovati L, Della Porta MG, Jädersten M, Killick S, Verma A, Norbury CJ, *et al*: Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. Leukemia 24: 756-764, 2010.
14. Sternberg A, Killick S, Littlewood T, Hatton C, Peniket A, Seidl T, Soneji S, Leach J, Bowen D, Chapman C, *et al*: Evidence for reduced B-cell progenitors in early (low-risk) myelodysplastic syndrome. Blood 106: 2982-2991, 2005.
15. Graubert TA, Shen D, Ding L, Okeyo-Owuor T, Lunn CL, Shao J, Krysiak K, Harris CC, Koboldt DC, Larson DE, *et al*: Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat Genet 44: 53-57, 2011.
16. Pellagatti A, Cazzola M, Giagounidis AA, Malcovati L, Porta MG, Killick S, Campbell LJ, Wang L, Langford CF, Fidler C, *et al*: Gene expression profiles of CD34+ cells in myelodysplastic syndromes: involvement of interferon-stimulated genes and correlation to FAB subtype and karyotype. Blood 108: 337-345, 2006.
17. Gerstung M, Pellagatti A, Malcovati L, Giagounidis A, Porta MG, Jadersten M, Dolatshad H, Verma A, Cross NC, Vyas P, *et al*: Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. Nat Commun 6: 5901, 2015.
18. Kikushige Y, Shima T, Takayanagi S, Urata S, Miyamoto T, Iwasaki H, Takenaka K, Teshima T, Tanaka T, Inagaki Y and Akashi K: TIM-3 is a promising target to selectively kill acute myeloid leukemia stem cells. Cell Stem Cell 7: 708-717, 2010.
19. de Jonge HJ, Woolthuis CM, Vos AZ, Mulder A, van den Berg E, Kluin PM, van der Weide K, de Bont ES, Huls G, Vellenga E and Schuringa JJ: Gene expression profiling in the leukemic stem cell-enriched CD34+ fraction identifies target genes that predict prognosis in normal karyotype AML. Leukemia 25: 1825-1833, 2011.
20. Bacher U, Schnittger S, Macijewski K, Grossmann V, Kohlmann A, Alpermann T, Kowarsch A, Nadarajah N, Kern W, Haferlach C and Haferlach T: Multilineage dysplasia does not influence prognosis in CEBPA-mutated AML, supporting the WHO proposal to classify these patients as a unique entity. Blood 119: 4719-4722, 2012.
21. Schneider V, Zhang L, Rojewski M, Fekete N, Schrezenmeier H, Erle A, Bullinger L, Hofmann S, Götz M, Döhner K, *et al*: Leukemic progenitor cells are susceptible to targeting by stimulated cytotoxic T cells against immunogenic leukemia-associated antigens. Int J Cancer 137: 2083-2092, 2015.
22. Virtaneva K, Wright FA, Tanner SM, Yuan B, Lemon WJ, Caligiuri MA, Bloomfield CD, De La Chapelle A and Krahe R: Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics Proc Natl Acad Sci USA 98: 1124-1129, 2001.
23. Von Der Heide EK, Neumann M, Vosberg S, James AR, Schroeder MP, Ortiz-Tanchez J, Isaakidis K, Schlee C, Luther M, Jöhrens K, *et al*: Molecular alterations in bone marrow mesenchymal stromal cells derived from acute myeloid leukemia patients. Leukemia 31: 1069-1078, 2017.
24. Stirewalt DL, Meshinchi S, Kopecky KJ, Fan W, Pogosova-Agadjanyan EL, Engel JH, Cronk MR, Dorcy KS, McQuary AR, Hockenbery D, *et al*: Identification of genes with abnormal expression changes in acute myeloid leukemia. Genes Chromosomes Cancer 47: 8-20, 2008.
25. Stegmaier K, Ross KN, Colavito SA, O'malley S, Stockwell BR and Golub TR: Gene expression-based high-throughput screening(GE-HTS) and application to leukemia differentiation. Nat Genet 36: 257-263, 2004.
26. Vidal M, Cusick ME and Barabási AL: Interactome networks and human disease. Cell 144: 986-998, 2011.
27. Schramm SJ, Li SS, Jayaswal V, Fung DC, Campain AE, Pang CN, Scolyer RA, Yang YH, Mann GJ and Wilkins MR: Disturbed protein-protein interaction networks in metastatic melanoma are associated with worse prognosis and increased functional mutation burden. Pigment Cell Melanoma Res 26: 708-722, 2013.
28. Zhang L, Zhang X and Fan S: Meta-analysis of salt-related gene expression profiles identifies common signatures of salt stress responses in Arabidopsis. Plant Syst Evol 303: 757-774, 2017.
29. Greenberg P, Cox C, Lebeau MM, Fenaux P, Morel P, Sanz G, Sanz M, Vallespi T, Hamblin T, Oscier D, *et al*: International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood 89: 2079-2088, 1997.
30. Bonardi F, Fusetti F, Deelen P, Van Gosliga D, Vellenga E and Schuringa JJ: A proteomics and transcriptomics approach to identify leukemic stem cell (LSC) markers. Mol Cell Proteomics 12: 626-637, 2013.
31. Sato K, Mano H, Ariyama T, Inazawa J, Yazaki Y and Hirai H: Molecular cloning and analysis of the human Tec protein-tyrosine kinase. Leukemia 8: 1663-1672, 1994.
32. Goncalves AC, Alves R, Baldeiras I, Cortesão E, Carda JP, Branco CC, Oliveiros B, Loureiro L, Pereira A, Nascimento Costa JM, *et al*: Genetic variants involved in oxidative stress, base excision repair, DNA methylation, and folate metabolism pathways influence myeloid neoplasias susceptibility and prognosis. Mol Carcinog 56: 130-148, 2017.
33. Pei S, Minhajuddin M, Callahan KP, Balys M, Ashton JM, Neering SJ, Lagadinou ED, Corbett C, Ye H, Liesveld JL, *et al*: Targeting aberrant glutathione metabolism to eradicate human acute myelogenous leukemia cells. J Biol Chem 288: 33542-33558, 2013.
34. Miraki-Moud F, Ghazaly E, Ariza-Mcnaughton L, Hodby KA, Clear A, Anjos-Afonso F, Liapis K, Grantham M, Sohrabi F, Cavenagh J, *et al*: Arginine deprivation using pegylated arginine deiminase has activity against primary acute myeloid leukemia cells in vivo. Blood 125: 4060-4068, 2015.
35. Li G, Song Y, Zhang Y, Wang H and Xie J: miR-34b targets HSF1 to suppress cell survival in acute myeloid leukemia. Oncol Res 24: 109-116, 2016.
36. Miwa H, Beran M and Saunders GF: Expression of the Wilms' tumor gene (WT1) in human leukemias. Leukemia 6: 405-409, 1992.
37. Feldhahn N, Arutyunyan A, Stoddart S, Zhang B, Schmidhuber S, Yi SJ, Kim YM, Groffen J and Heisterkamp N: Environment-mediated drug resistance in Bcr/Abl-positive acute lymphoblastic leukemia. Oncoimmunology 1: 618-629, 2012.
38. Boultwood J, Pellagatti A, Watkins F, Campbell LJ, Esoof N, Cross NC, Eagleton H, Littlewood TJ, Fidler C and Wainscoat JS: Low expression of the putative tumour suppressor gene gravin in chronic myeloid leukaemia, myelodysplastic syndromes and acute myeloid leukaemia. Br J Haematol 126: 508-511, 2004.
39. Armengol G, Canellas A, Alvarez Y, Bastida P, Toledo JS, Pérez-Iribarne Mdel M, Camós M, Tuset E, Estella J, Coll MD, *et al*: Genetic changes including gene copy number alterations and their relation to prognosis in childhood acute myeloid leukemia. Leuk Lymphoma 51: 114-124, 2010.
40. Seedhouse CH, Mills KI, Ahluwalia S, Grundy M, Shang S, Burnett AK, Russell NH and Pallis M: Distinct poor prognostic subgroups of acute myeloid leukaemia, FLT3-ITD and P-glycoprotein-positive, have contrasting levels of FOXO1. Leuk Res 38: 131-137, 2014.
41. Gasparetto M, Pei S, Minhajuddin M, Khan N, Pollyea DA, Myers JR, Ashton JM, Becker MW, Vasiliou V, Humphries KR, *et al*: Targeted therapy for a subset of acute myeloid leukemias that lack expression of aldehyde dehydrogenase 1A1. Haematologica 102: 1054-1065, 2017.
42. Barabási AL, Gulbahce N and Loscalzo J: Network medicine: A network-based approach to human disease. Nat Rev Genet 12: 56-68, 2011.
43. Taipale M, Jarosz DF and Lindquist S: HSP90 at the hub of protein homeostasis: Emerging mechanistic insights. Nat Rev Mol Cell Biol 11: 515-528, 2010.
44. Flandrin-Gresta P, Solly F, Aanei CM, Cornillon J, Tavernier E, Nadal N, Morteux F, Guyotat D, Wattel E and Campos L: Heat Shock Protein 90 is overexpressed in high-risk myelodysplastic syndromes and associated with higher expression and activation of focal adhesion kinase. Oncotarget 3: 1158-1168, 2012.
45. Tian WL, He F, Fu X, Lin JT, Tang P, Huang YM, Guo R and Sun L: High expression of heat shock protein 90 alpha and its significance in human acute leukemia cells. Gene 542: 122-128, 2014.
46. Flandrin P, Guyotat D, Duval A, Cornillon J, Tavernier E, Nadal N and Campos L: Significance of heat-shock protein (HSP) 90 expression in acute myeloid leukemia cells. Cell Stress Chaperones 13: 357-364, 2008.
47. Lin S, Li J, Zhou W, Qian W, Wang B and Chen Z: BIIB021, an Hsp90 inhibitor, effectively kills a myelodysplastic syndrome cell line via the activation of caspases and inhibition of PI3K/Akt and NF-κB pathway proteins. Exp Ther Med 7: 1539-1544, 2014.

48. Lazenby M, Hills R, Burnett AK and Zabkiewicz J: The HSP90 inhibitor ganetespib: A potential effective agent for Acute Myeloid Leukemia in combination with cytarabine. Leuk Res 39: 617-624, 2015.
49. Zheng N, Schulman BA, Song L, Miller JJ, Jeffrey PD, Wang P, Chu C, Koepp DM, Elledge SJ, Pagano M, *et al*: Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. Nature 416: 703-709, 2002.
50. Nakayama KI and Nakayama K: Ubiquitin ligases: Cell-cycle control and cancer. Nat Rev Cancer 6: 369-381, 2006.
51. Ishino R, Minami K, Tanaka S, Nagai M, Matsui K, Hasegawa N, Roeder RG, Asano S and Ito M: FGF7 supports hematopoietic stem and progenitor cells and niche-dependent myeloblastoma cells via autocrine action on bone marrow stromal cells in vitro. Biochem Biophys Res Commun 440: 125-131, 2013.
52. Zhang J and Li Y: Therapeutic uses of FGFs. Semin Cell Dev Biol 53: 144-154, 2016.
53. Hart KC, Robertson SC and Donoghue DJ: Identification of tyrosine residues in constitutively activated fibroblast growth factor receptor 3 involved in mitogenesis, Stat activation, and phosphatidylinositol 3-kinase activation. Mol Biol Cell 12: 931-942, 2001.
54. Webster MK and Donoghue DJ: Constitutive activation of fibroblast growth factor receptor 3 by the transmembrane domain point mutation found in achondroplasia. EMBO J 15: 520-527, 1996.
55. Kang S, Elf S, Dong S, Hitosugi T, Lythgoe K, Guo A, Ruan H, Lonial S, Khoury HJ, Williams IR, *et al*: Fibroblast growth factor receptor 3 associates with and tyrosine phosphorylates p90 RSK2, leading to RSK2 activation that mediates hematopoietic transformation. Mol Cell Biol 29: 2105-2117, 2009.
56. Airoldi I, Di Carlo E, Banelli B, Moserle L, Cocco C, Pezzolo A, Sorrentino C, Rossi E, Romani M, Amadori A and Pistoia V: The IL-12Rbeta2 gene functions as a tumor suppressor in human B cell malignancies. J Clin Invest 113: 1651-1659, 2004.
57. Jo SH, Schatz JH, Acquaviva J, Singh H and Ren R: Cooperation between deficiencies of IRF-4 and IRF-8 promotes both myeloid and lymphoid tumorigenesis. Blood 116: 2759-2767, 2010.
58. Mangiavacchi A, Sorci M, Masciarelli S, Larivera S, Legnini I, Iosue I, Bozzoni I, Fazi F and Fatica A: The miR-223 host non-coding transcript linc-223 induces IRF4 expression in acute myeloid leukemia by acting as a competing endogenous RNA. Oncotarget 7: 60155-60168, 2016.
59. Otto N, Manukjan G, Göhring G, Hofmann W, Scherer R, Luna JC, Lehmann U, Ganser A, Welte K, Schlegelberger B and Steinemann D: ICSBP promoter methylation in myelodysplastic syndromes and acute myeloid leukaemia. Leukemia 25: 1202-1207, 2011.
60. Pogosova-Agadjanyan EL, Kopecky KJ, Ostronoff F, Appelbaum FR, Godwin J, Lee H, List AF, May JJ, Oehler VG, Petersdorf S, *et al*: The prognostic significance of IRF8 transcripts in adult patients with acute myeloid leukemia. PLoS One 8: e70812, 2013.
61. Xu L, Gu ZH, Li Y, Zhang JL, Chang CK, Pan CM, Shi JY, Shen Y, Chen B, Wang YY, *et al*: Genomic landscape of CD34+ hematopoietic cells in myelodysplastic syndrome and gene mutation profiles as prognostic markers. Proc Natl Acad Sci USA 111: 8589-8594, 2014.
62. Cancer Genome Atlas Research Network, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson A, Hoadley K, Triche TJ Jr, Laird PW, *et al*: Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 368: 2059-2074, 2013.