ISSN 2618-1584 Vol. 7, Issue 3: 126-133, DOI: 10.25236/FMSR.2025.070317

Analysis of Exosome-Related Differential Gene Expression Profiles in Acute Myeloid Leukemia and Exploration of Therapeutic Targets

Minxiao Wang, Yanhui Yu, Changsheng Liao*

Graduate Student Department, Changzhi Medical College, Changzhi, Shanxi, China
**Corresponding author: 15603441409@163.com*

Abstract: *This bioinformatics investigation systematically interrogates leukemia stem cell (LSC) transcriptomes from acute myeloid leukemia (AML) patients, focusing on exosome-related biomarkers to delineate their mechanistic involvement and therapeutic potential in AML pathogenesis. We curated two LSC gene expression datasets (GSE17054, GSE24395) from GEO, implementing rigorous normalization and batch effect correction using R-based bioinformatics pipelines (ComBat algorithm via sva package). Harmonized datasets underwent differential expression analysis through empirical Bayes moderated t-tests (limma package), cross-referenced with exosome-associated genes from GeneCards to identify AML-specific candidates. Subsequent DAVID functional annotation revealed critical enrichment in vesicle-mediated transport (FDR<0.05) and ferroptosis pathways ($p=3.2\times10^{-4}$). STRING database analysis coupled with network topology analysis in Cytoscape identified six hub genes (ANXA2, TNFRSF1A, THY1, CD63, IDH1, S100A11) exhibiting significant connectivity (degree centrality ≥8), including membrane trafficking regulators (ANXA2, CD63) and metabolic modulators (IDH1). These findings establish an exosome-related gene signature with therapeutic implications, particularly highlighting IDH1's dual role in 2-hydroxyglutarate metabolism and extracellular vesicle biogenesis. Our multi-omics integration strategy provides a novel framework for understanding AML pathogenesis through the exosome-LSC axis, serving as potential diagnostic biomarkers and therapeutic targets, ultimately advancing therapeutic development for AML.*

Keywords: *AML, Leukemia Stem Cells, Exosomal Genes, Molecular Docking, Bioinformatics Analysis*

1. Introduction

Acute Myeloid Leukemia (AML) is a hematological malignancy characterized by the malignant cloning of myeloid hematopoietic stem cells. Its main pathological features include abnormal proliferation of bone marrow blasts, suppression of normal hematopoiesis, and extramedullary infiltration. AML is clinically characterized by high drug resistance and recurrence rates[1-2], with an increasing incidence rate among the elderly population year by year. Due to poor tolerance to chemotherapy and complex comorbidities in elderly patients, the treatment failure rate is as high as 40%-50%, posing a significant burden on the healthcare system[3-4].

The occurrence of AML involves multilevel regulatory networks such as epigenetic dysregulation, gene mutations (e.g., FLT3, NPM1, DNMT3A), and modulation of the myeloid microenvironment and immune escape[5-7]. Exosomes, originating from various cell types, play a crucial role in intercellular communication by participating in the functional transport of molecules such as proteins, nucleic acids, and lipids across spaces[8-10]. Recent studies have shown that exosomes play a key role in intercellular communication, particularly in the progression of AML, where leukemic stem cells (LSCs) promote myeloid differentiation arrest and immune escape through abnormal exosome secretion[11,12]. Exosome-related genes not only regulate the biosynthesis and secretion of exosomes (e.g., CD63, TSG101, MVB12A) but also participate in the remodeling of the tumor microenvironment[13-15]. LSCs transmit oncogenic proteins (e.g., FLT3-ITD mutants), miRNAs (e.g., miR-17-92 cluster), and signaling molecules (e.g., Wnt/β-catenin pathway ligands) via exosomes, accelerating leukemia progression[16-18] and inducing immune tolerance and treatment resistance[19-21].

In this study, we conducted bioinformatics analysis to systematically dissect the expression profile characteristics of exosome-related genes in LSCs from AML patients, focusing on their dynamic regulatory networks in the pathogenesis of AML. By integrating multi-omics data, we aim to construct

ISSN 2618-1584 Vol. 7, Issue 3: 126-133, DOI: 10.25236/FMSR.2025.070317

a molecular network map mediated by AML-LSC-specific exosomes, providing a theoretical basis for exosome-based therapies targeting LSCs and individualized treatment strategies. With a deeper understanding of the pathological mechanisms of AML exosomes, targeted intervention of exosome secretion or receptor signaling pathways is expected to become a new strategy for improving patient prognosis and promoting the development of precision medicine.

2. Materials and Methods

2.1 Data Acquisition and Processing

This study systematically retrieved transcriptome data from leukemia stem cells (LSCs) and normal hematopoietic stem cells (HSCs) of acute myeloid leukemia (AML) patients using RNA microarray technology from the Gene Expression Omnibus (GEO) database. Core experimental data, including platform description files, series matrix files, TPM files, and count files, were extracted using keywords such as ("acute myeloid leukemia" OR "AML") AND ("leukemia stem cell" OR "LSC") along with "microarray" and "human samples" as filtering criteria. Two independent cohorts were ultimately included: GSE17054 (Institute for Systems Biology, USA, with 9 LSCs vs. 4 HSCs, GPL570-55999) and GSE24395 (Kyushu University, Japan, with 12 LSCs vs. 5 HSCs, GPL6106-11578). Although some samples lacked age/gender information, they were still included to maximize the sample size, with missing variables controlled in subsequent analyses.

2.2 Data Preprocessing for Differentially Expressed Genes (DEGs)

To eliminate the impact of experimental batches and platform differences on the data, both datasets were standardized. The normalizeBetweenArrays function from the limma package was used to perform global scaling and shifting for each sample to ensure consistent data distribution. Subsequently, the avereps function was employed to average duplicate probes, reducing technical noise. The standardized data were then log2(counts + 1) transformed to stabilize variance and compress the data range, making it closer to a normal distribution and enhancing the robustness of subsequent statistical analyses.

2.3 Batch Correction and Differential Analysis After Dataset Merging

Gene identifiers (gene IDs) were extracted from each dataset, and common genes were retained to reduce noise across datasets. Batch correction was performed using the ComBat function from the sva package, estimating batch effects through an empirical Bayes framework. Subsequently, differential expression analysis was conducted using the DESeq function from the DESeq2 package, with P-values adjusted using the FDR method (adjusted P-value < 0.05) to control the false discovery rate. A threshold of $|\log2FC| \geq 0.585$ (corresponding to an original fold change ≥ 1.5) was set to balance sensitivity and specificity.

2.4 Screening of Exosome-Related DEGs

Cross-analysis was conducted between DEGs in the merged dataset and exosome-related genes in the Genecard database to identify AML-LSCs-specific DEGs.

2.5 Enrichment Analysis

GO (Gene Ontology) enrichment analysis, including biological processes (BP), cellular components (CC), and molecular functions (MF), as well as KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis, were performed using the clusterProfiler R package. Additionally, Gene Set Enrichment Analysis (GSEA) was employed to identify gene sets significantly enriched or downregulated in specific biological processes. All analyses were conducted with a significance threshold of $P < 0.05$.

2.6 Construction and Analysis of Protein-Protein Interaction (PPI) Networks and Core Gene Selection

PPI networks were constructed based on the STRING database (confidence score ≥ 0.4) and

ISSN 2618-1584 Vol. 7, Issue 3: 126-133, DOI: 10.25236/FMSR.2025.070317

visualized using Cytoscape software (v3.10.1). The CytoHubba plugin was utilized to comprehensively screen core genes through ten different algorithms, ensuring the integrity and functionality of the network structure.

2.7 Construction of Core Gene Protein Networks

The selected core genes were input into the GENEMANIA platform, with "Homo sapiens" selected as the species, to automatically generate gene interaction networks.

2.8 Statistical Methods, Software, and Tools

All statistical analyses were conducted in the R environment. The Benjamini-Hochberg method was used for multiple testing correction to control the false discovery rate (FDR). R language (version 4.4.2) was employed for data processing and statistical analysis, while Cytoscape software (version 3.10.3) was used for plotting and screening core differential genes.

3. Results

3.1 Data Acquisition and Processing

After preliminary handling, integration, and merging of the two datasets, GSE17054 and GSE24395, a gene expression matrix was successfully obtained. This includes: GSE17054 Dataset: Containing 22,876 gene expression data points. GSE24395 Dataset: Containing 24,981 gene expression data points.

3.2 Data Preprocessing for Differentially Expressed Genes

Through processing, we obtained two normalized gene expression matrix files (Figure 1).

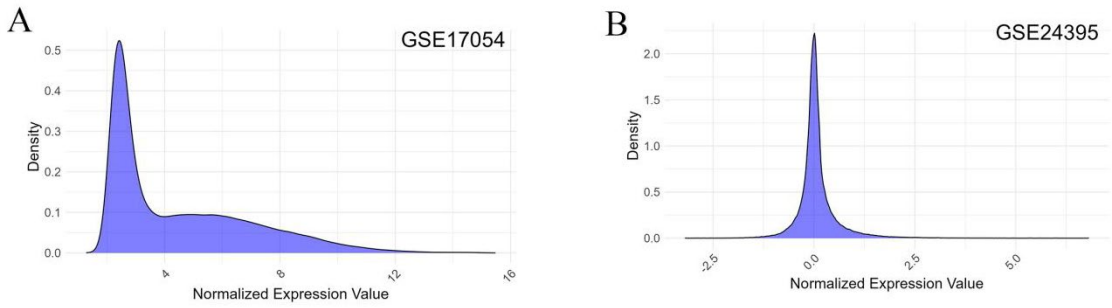

Figure 1 Density Distribution of Normalized Gene Expression.

(A) GSE17054 dataset; (B) GSE24395 dataset. X-axis: log2-transformed normalized expression values (normalized by limma method); Y-axis: Kernel density estimation values. Both datasets exhibit a right-skewed distribution, indicating that lowly expressed genes are predominant, which is consistent with the characteristics of RNA microarray technology.

3.3 Batch Correction and Differential Analysis After Data Integration

After integrating the normalized gene expression data from GSE17054 and GSE24395, we employed a batch correction method to eliminate batch effects between experiments. Subsequently, differential expression analysis was conducted, leading to the successful identification of 550 differentially expressed genes and their expression levels (Figure 2).

(A) PCA Analysis Before Correction: Principal Component 1 (PC1) and Principal Component 2 (PC2) reveal a significant separation between GSE17054 (red) and GSE24395 (green) samples, indicating the presence of technical batch effects. (B) PCA Analysis After Correction: After ComBat correction, there is a significant increase in the overlap between the two groups of samples, suggesting that the technical batch effects have been effectively removed. (C) Heatmap: Hierarchical clustering based on Pearson correlation shows that the normal control samples are grouped under the top blue bar, while the disease samples are grouped under the purple bar. The second row displays different color

ISSN 2618-1584 Vol. 7, Issue 3: 126-133, DOI: 10.25236/FMSR.2025.070317

bars indicating different datasets. In the heatmap, red squares represent high expression, and blue squares represent low expression. Only the top 50 consistently differentially expressed genes with |log2FC| ≥ 0.585 and adj.P.Val < 0.05 across datasets are displayed. (D) Volcano Plot: A total of 550 differentially expressed genes were identified. Genes marked in red are significantly up - regulated in both groups, while genes marked in blue are significantly down - regulated in both groups. Gray dots represent non - differentially expressed genes.

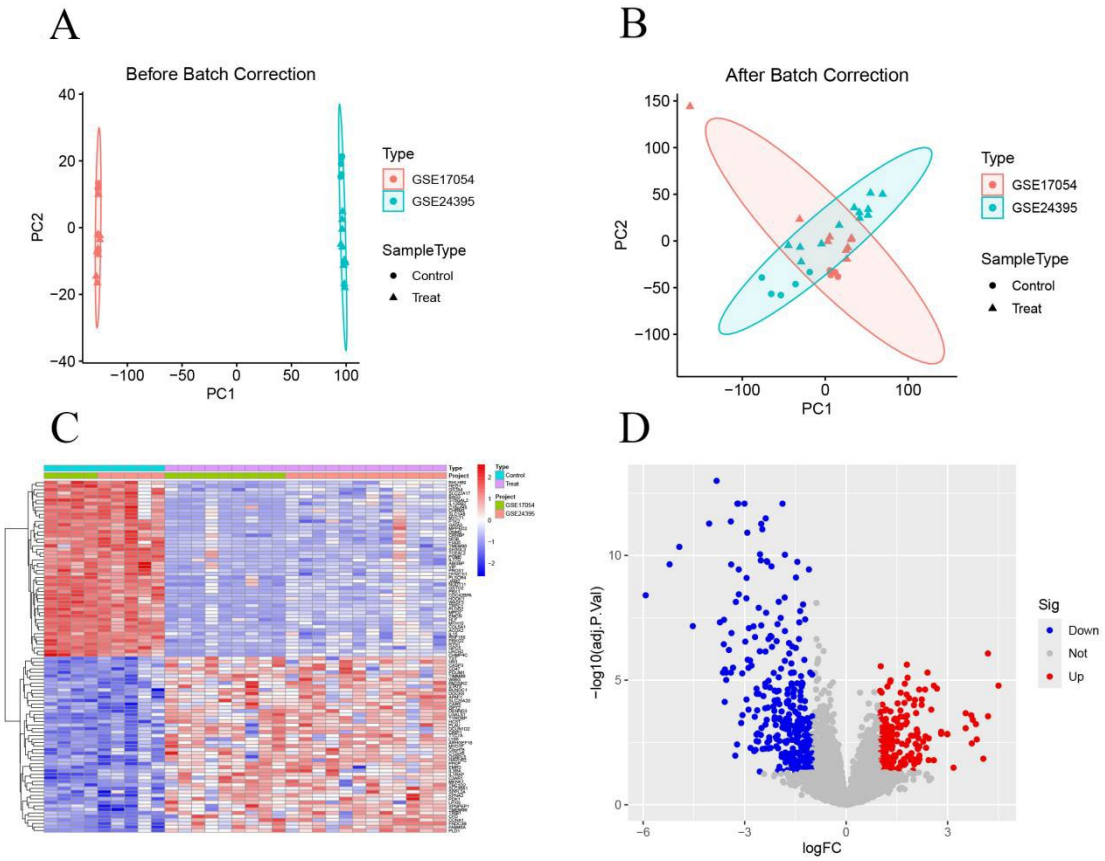

Figure 2 Batch Effect Correction and Differential Analysis

3.4 Screening of Exosome-Related Differentially Expressed Genes

By performing a cross-analysis of the 550 differentially expressed genes with 878 exosome-related genes from the Genecard database, we successfully screened out 41 AML-LSCs exosome-specific differentially expressed genes (DEGs) (Figure 3).

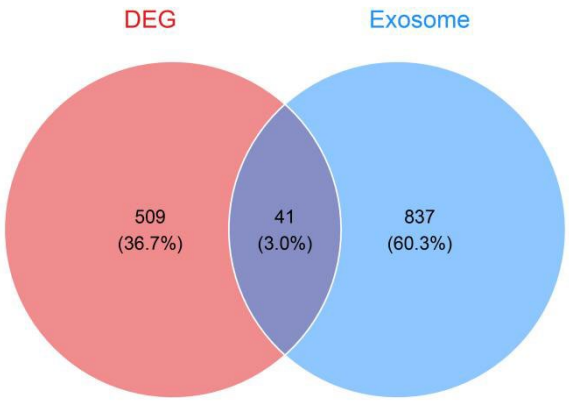

Figure 3 Venn Diagram of Exosome-Related Differential Genes in AML LSCs.

ISSN 2618-1584 Vol. 7, Issue 3: 126-133, DOI: 10.25236/FMSR.2025.070317

Red circle (DEG): 509 differentially expressed genes (36.7%) in leukemia stem cells.Blue circle (Exosome): 837 exosome-related genes (60.3%).Purple overlapping region: 41 co-occurring genes (3.0%) in both AML-LSCs and exosomes.

3.5 Enrichment Analysis

GO, KEGG, and GSEA enrichment analyses were conducted on the 41 exosome-related differentially expressed genes (DEGs), with a significance threshold of $P < 0.05$. Biological Processes (BP): These DEGs were enriched in the regulation of actin polymerization or depolymerization , regulation of actin filament length, and actin filament polymerization. Cellular Components (CC): They were predominantly localized to cytoplasmic vesicle lumens, secretory granule lumens, vesicle lumens, and focal adhesions. Molecular Functions (MF): Significant enrichment was observed in phospholipase inhibitor activity, cell-cell adhesion mediator activity, and cadherin binding, among other functions. KEGG Pathways: The DEGs were involved in pathways such as amino acid biosynthesis, endocytosis, and Salmonella infection.

GSEA Analysis: The analysis revealed significant enrichment of gene sets in pathways related to oxidative phosphorylation, nuclear-cytoplasmic transport, and natural killer cell-mediated cytotoxicity (Figure 4).

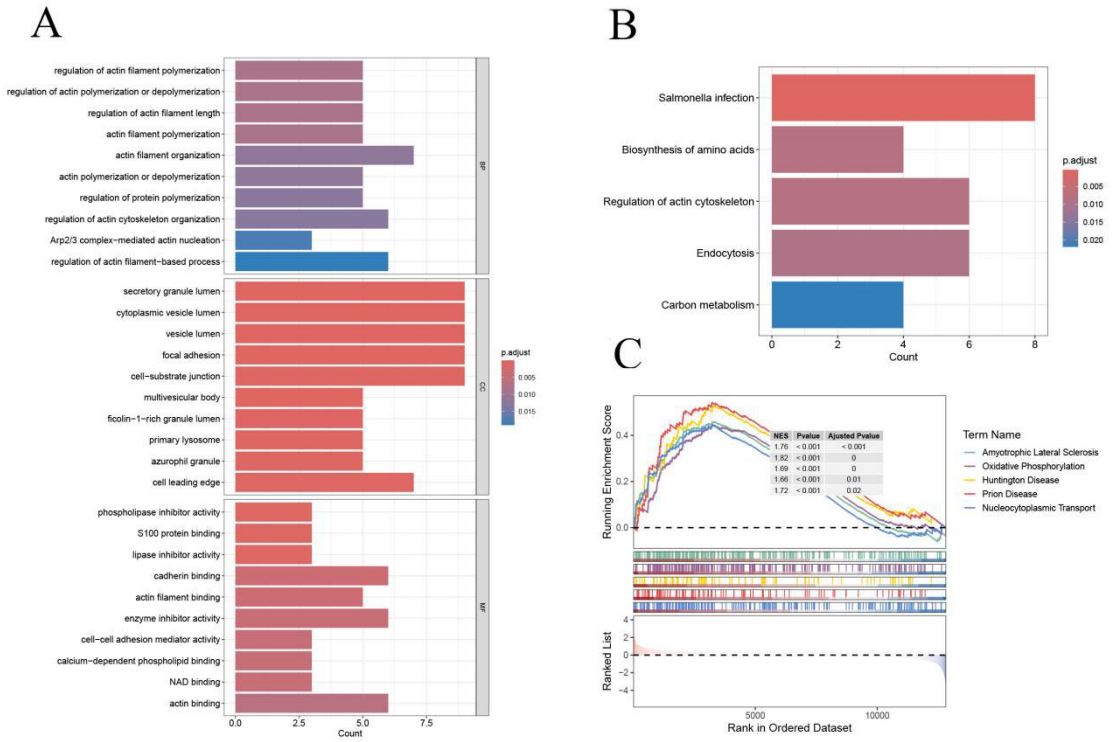

Figure 4 Gene Ontology (GO) and Pathway Enrichment Analysis Results, as well as GSEA Results.

(A) GO Analysis: Biological Process (BP): Differentially expressed genes were significantly enriched in processes such as actin filament polymerization (the length of the bar graph indicates the number of genes, and the color intensity represents the adjusted p-value, with p.adjust < 0.05).Cellular Component (CC): Significant enrichment was observed in structures like secretory granule lumen.Molecular Function (MF): Involved activities such as phosphatidylinositol binding. (B) Pathway Enrichment Analysis: Significantly enriched pathways include Salmonella infection, amino acid biosynthesis, etc. (p.adjust < 0.05). The color coding of the bar graph is the same as above, reflecting the strength of significance. (C) GSEA Analysis: The running enrichment score curve displays the dynamic enrichment trend of gene sets (such as oxidative phosphorylation).

ISSN 2618-1584 Vol. 7, Issue 3: 126-133, DOI: 10.25236/FMSR.2025.070317

3.6 Construction and Analysis of Protein-Protein Interaction (PPI) Network and Screening of Core Genes

Based on the AML exosome-related differentially expressed genes (DEGs), we constructed a PPI network . This network comprises 41 DEGs, forming 30 nodes and 60 edges, where nodes represent proteins encoded by DEGs and edges depict interactions between these proteins. Utilizing Cytoscape software (version 3.10.1), we constructed a differentially expressed PPI network based on these interaction results and analyzed it using 10 algorithms within the CytoHubba plugin. Ultimately, six core hub genes were screened out that met all criteria and had the highest scores: ANXA2, TNFRSF1A, THY1, CD63, IDH1, and S100A11 (Figure 5).

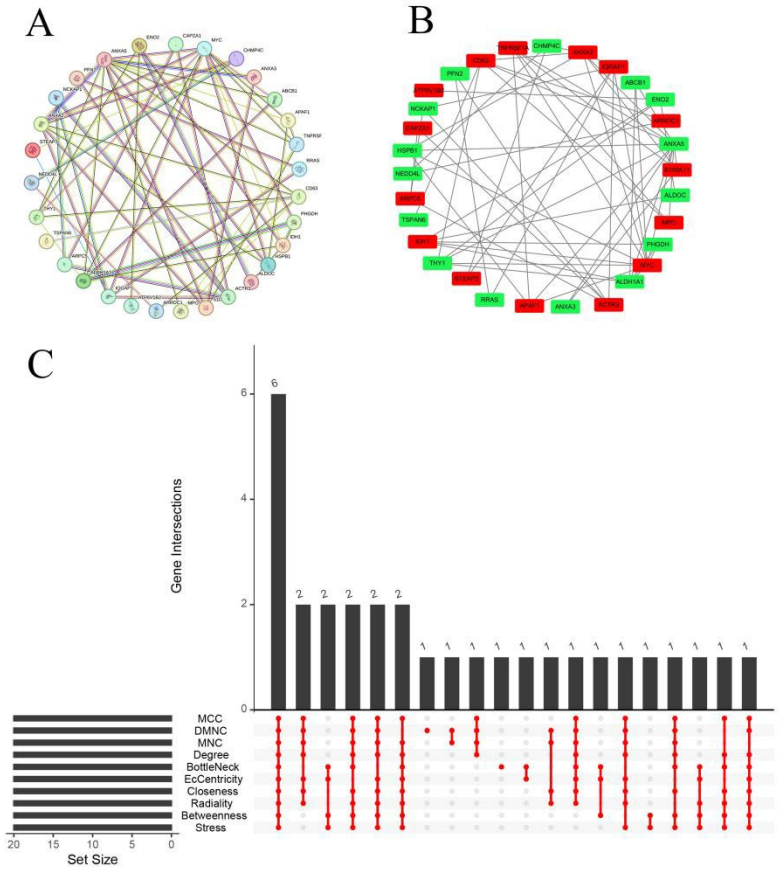

Figure 5 (A) shows the protein-protein interaction (PPI) network diagram of exosome-related differentially expressed genes; (B) presents the PPI network diagram of differentially expressed genes, where red represents upregulated genes and green represents downregulated genes; (C) illustrates the hub genes that meet the requirements based on comprehensive evaluation using 10 algorithms.

3.7 Construction of Protein Interactions Among Core Genes

Analysis of the six core exosome-related DEGs on the GeneMANIA platform unveiled their crucial roles in various biological processes. The results indicated that these genes can directly participate in cell cycle regulation and also indirectly function through interactions with other genes. The primary biological processes involved include membrane microdomains, regulation of endocytosis, receptor catabolism, receptor-mediated endocytosis, positive regulation of endocytosis, phosphatidylinositol binding, and regulation of the I-kappaB kinase/NF-kappaB signaling pathway (Figure 6).

ISSN 2618-1584 Vol. 7, Issue 3: 126-133, DOI: 10.25236/FMSR.2025.070317

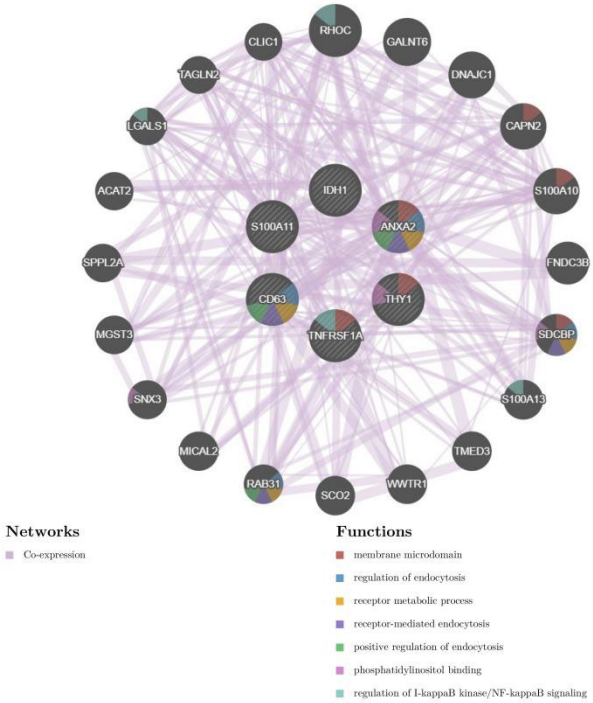

Figure 6 Core gene protein interaction network diagram, the different biological functions represented by node colors.

4. Conclusion

Acute myeloid leukemia (AML) is a malignant hematological disorder originating from hematopoietic stem cells in the bone marrow, characterized by abnormal differentiation and clonal proliferation of myeloid cells. The incidence of AML significantly increases with age, and multidrug resistance, leukemic stem cell (LSC) evasion, and immune microenvironment imbalance are the primary reasons for treatment failure. In recent years, exosomes have emerged as a research hotspot in AML due to their roles in intercellular signaling, regulation of the tumor microenvironment, and the development of drug resistance[22]. In this study, bioinformatics analysis was conducted to screen six exosome-related characteristic genes of LSCs, namely ANXA2, TNFRSF1A, THY1, CD63, IDH1, and S100A11. These findings revealed the crucial roles of these genes in the initiation and progression of AML, providing novel exosome-directed strategies for precision therapy. Furthermore, molecular docking techniques offer new insights into the development of targeted therapeutic drugs.

References

[1] Wachter F, Pikman Y. Pathophysiology of acute myeloid leukemia[J]. Acta Haematologica, 2024, 147(2): 229-246.
[2] Hackl H, Astanina K, Wieser R. Molecular and genetic alterations associated with therapy resistance and relapse of acute myeloid leukemia[J]. Journal of hematology & oncology, 2017, 10: 1-16.
[3] Venugopal S, Sekeres M A. Contemporary Management of Acute myeloid leukemia: a review[J]. JAMA oncology, 2024.
[4] Thol F, Heuser M. Treatment for relapsed/refractory acute myeloid leukemia[J]. Hemasphere, 2021, 5(6): e572.
[5] Conway O'Brien E, Prideaux S, Chevassut T. The epigenetic landscape of acute myeloid leukemia[J]. Advances in hematology, 2014, 2014(1): 103175.
[6] Daver N, Venugopal S, Ravandi F. FLT3 mutated acute myeloid leukemia: 2021 treatment algorithm. Blood Cancer J 11: 104[EB/OL].(2021).
[7] Krause D S, Fulzele K, Catic A, et al. Differential regulation of myeloid leukemias by the bone marrow microenvironment[J]. Nature medicine, 2013, 19(11): 1513-1517.
[8] Tkach M, Théry C. Communication by extracellular vesicles: where we are and where we need to go[J]. Cell, 2016, 164(6): 1226-1232.

ISSN 2618-1584 Vol. 7, Issue 3: 126-133, DOI: 10.25236/FMSR.2025.070317

[9] Maas S L N, Breakefield X O, Weaver A M. Extracellular vesicles: unique intercellular delivery vehicles[J]. Trends in cell biology, 2017, 27(3): 172-188.
[10] Bobrie A, Colombo M, Raposo G, et al. Exosome secretion: molecular mechanisms and roles in immune responses[J]. Traffic, 2011, 12(12): 1659-1668.
[11] Stelmach P, Trumpp A. Leukemic stem cells and therapy resistance in acute myeloid leukemia[J]. Haematologica, 2023, 108(2): 353.
[12] Wang H, Hu H, Zhang Q, et al. Dynamic transcriptomes of human myeloid leukemia cells[J]. Genomics, 2013, 102(4): 250-256.
[13] Kumar B, Garcia M, Weng L, et al. Acute myeloid leukemia transforms the bone marrow niche into a leukemia-permissive microenvironment through exosome secretion[J]. Leukemia, 2018, 32(3): 575-587.
[14] Hong C S, Muller L, Boyiadzis M, et al. Isolation and characterization of CD34+ blast-derived exosomes in acute myeloid leukemia[J]. PloS one, 2014, 9(8): e103310.
[15] Amin A H, Al Sharifi L M, Kakhharov A J, et al. Role of Acute Myeloid Leukemia (AML)-Derived exosomes in tumor progression and survival[J]. Biomedicine & Pharmacotherapy, 2022, 150: 113009.
[16] Trino S, Laurenzana I, Lamorte D, et al. Acute myeloid leukemia cells functionally compromise hematopoietic stem/progenitor cells inhibiting normal hematopoiesis through the release of extracellular vesicles[J]. Frontiers in Oncology, 2022, 12: 824562.
[17] Huan J, Hornick N I, Shurtleff M J, et al. RNA trafficking by acute myelogenous leukemia exosomes[J]. Cancer research, 2013, 73(2): 918-929.
[18] Li H, Xie C, Lu Y, et al. Exosomal mir92a promotes cytarabine resistance in myelodysplastic syndromes by activating Wnt/β-catenin signal pathway[J]. Biomolecules, 2022, 12(10): 1448.
[19] Saadh M J, K. Abdulsahib W, Ashurova D, et al. FLT3-mutated AML: immune evasion through exosome-mediated mechanisms and innovative combination therapies targeting immune escape[J]. Expert Review of Anticancer Therapy, 2025, 25(2): 143-150.
[20] Can C, Yang X, Jia H, et al. Exosomal circ_0006896 promotes AML progression via interaction with HDAC1 and restriction of antitumor immunity[J]. Molecular Cancer, 2025, 24(1): 4.
[21] Wang C, Song R, Yuan J, et al. Exosome-Shuttled METTL14 From AML-Derived Mesenchymal Stem Cells Promotes the Proliferation and Radioresistance in AML Cells by Stabilizing ROCK1 Expression via an m6A-IGF2BP3-Dependent Mechanism[J]. Drug Development Research, 2025, 86(1): e70025.
[22] Yang C, Yang H, Liu J, et al. Focus on exosomes: novel pathogenic components of leukemia[J]. American journal of cancer research, 2019, 9(8): 1815.