

# Bioinformatics-Based Identification of Key Prognostic Genes in Neuroblastoma with a Focus on Immune Cell Infiltration and Diagnostic Potential of VGF

Qiang Guo<sup>1,\*</sup>, Yang Xiao<sup>2,\*</sup>, Jing Chu<sup>3</sup>, Yu Sun<sup>2</sup>, Shaomei Li<sup>3</sup>, Shihai Zhang<sup>1</sup>

<sup>1</sup>Department of Clinical Laboratory, Anhui Children's Hospital, Hefei, People's Republic of China; <sup>2</sup>Xuzhou Medical University, Jiangsu, People's Republic of China; <sup>3</sup>Department of Pathology, Anhui Children's Hospital, Hefei, People's Republic of China

\*These authors contributed equally to this work

Correspondence: Shihai Zhang, Department of Clinical Laboratory, Anhui Children's Hospital, Wangjiang Road 39, Hefei, People's Republic of China, Tel +86-13856048692, Email zhangshihai888@126.com

**Objective:** This study aims to identify differentially expressed genes (DEGs) in neuroblastoma (NB) through comprehensive bioinformatics analysis and machine learning techniques. We seek to elucidate these DEGs' biological functions and associated signaling pathways. Furthermore, our objective extends to predicting upstream microRNAs (miRNAs) and relevant transcription factors of pivotal genes, with the ultimate goal of guiding clinical diagnostics and informing future treatment strategies for Neuroblastoma.

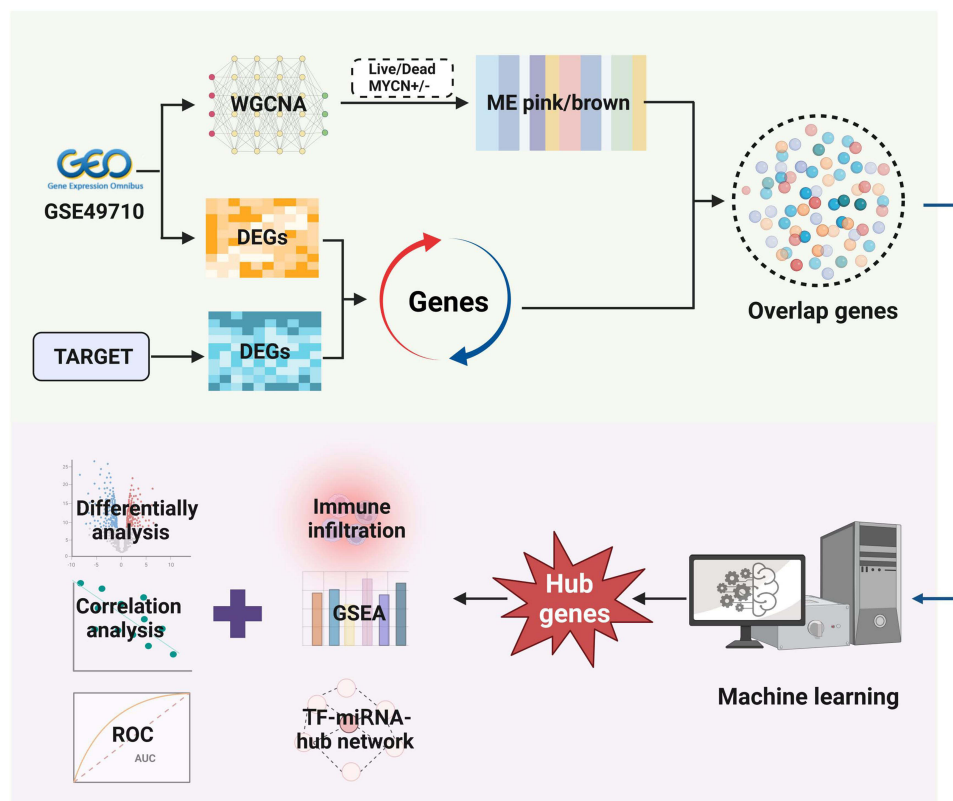
**Methods:** In this study, we sourced datasets GSE49710 and TARGET from the GEO and UCSC-XENA databases, respectively. Differentially expressed genes (DEGs) were identified using the R language "limma" package. Subsequent Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of these DEGs were conducted using the "clusterProfiler" package. We employed Weighted Gene Co-expression Network Analysis (WGCNA) to isolate the most significant modules associated with death and MYCN amplification, specifically MEpink and MEbrown modules. These modules were then cross-referenced with the DEGs for further GO and KEGG pathway analyses. LASSO regression analysis, facilitated by the "glmnet" package, was utilized to pinpoint three hub genes. We performed differential analysis on these genes and constructed Receiver Operating Characteristic (ROC) curves for disease diagnosis purposes. Immune infiltration analysis was conducted using the "GSVA" package's ssGSEA function. Additionally, single-gene Gene Set Enrichment Analysis (GSEA) on the hub gene was carried out based on Reactome and KEGG databases. Upstream miRNA and transcription factors associated with the hub gene were predicted using RegNetwork, with visual representations created in Cytoscape. Furthermore, to validate the three identified markers in neuroblastoma tissues, quantitative Real-Time Polymerase Chain Reaction (qRT-PCR) analysis was conducted.

**Results:** We identified 483 differentially expressed genes (DEGs) in neuroblastoma. These genes predominantly function in protein translation, membrane composition, and RNA transcription regulation, and are implicated in multiple signaling pathways relevant to neurodegenerative diseases. Utilizing LASSO regression analysis, we pinpointed three hub genes: *VGF*, *DGKD*, and *C19orf52*. The Receiver Operating Characteristic (ROC) curve analysis yielded Area Under Curve (AUC) values of 0.751 and 0.722 for *VGF*, 0.79 and 0.656 for *DGKD*, and 0.8 and 0.753 for *C19orf52*, respectively. Our immune infiltration analysis revealed significant correlations among monocytes, follicular helper T cells, and CD4+ T cells. Notably, in the death group, we observed heightened infiltration levels of activated CD4+ T cells, macrophages, and Th2 cells. *C19orf52* exhibited a close association with the infiltration of monocytes, CD4+ T cells, and Th2 cells, with P-values less than 0.05. Furthermore, qRT-PCR analysis corroborated the upregulation of *VGF* in neuroblastoma tissues, further validating our findings.

**Conclusion:** The hub genes (*VGF*, *DGKD*, and *C19orf52*) of neuroblastoma are screened. *VGF*, one of the hub genes, may have a high diagnostic value and is involved in the immune cell infiltration in neuroblastoma tissue, which may be used as a biomarker for the diagnosis of neuroblastoma and provides a new direction for clinical prognosis prediction and management improvement.

**Keywords:** neuroblastoma, WGCNA, LASSO regression, bioinformatics analysis

## Graphical Abstract



## Introduction

Neuroblastoma, predominantly affecting children under the age of 5, is the most common form of cancer in infants, accounting for about 50% of all cancer cases in this demographic and 7–10% of pediatric cancers.<sup>1</sup> Annually, approximately 650 cases are diagnosed in the United States. Notably, in China, neuroblastoma and ganglioneuroblastoma are the primary cancers in infants under one year, accounting for 14.35% of cases in girls and 15.84% in boys.<sup>2</sup> This neuroendocrine tumor, primarily arising from neural crests of the sympathetic nervous system, commonly occurs in the adrenal glands but can also manifest in the pelvis, abdomen, chest, and neck. Despite representing only 8% of childhood cancers, neuroblastoma is responsible for 15% of pediatric cancer-related deaths.<sup>3</sup> The clinical presentation of individuals with neuroblastoma is varied. Some affected individuals undergo spontaneous degeneration or differentiation into benign ganglioma, while others can present with an aggressive disease with metastatic dissemination leading to death.<sup>4,5</sup> Neuroblastoma patients can be categorized into low, medium, and high-risk groups based on a variety of factors, such as the biological features and clinical presentation of the tumor, including age, histopathological features, disease stage, presence of MYCN amplification, and others.<sup>3,6</sup> The surgical and chemotherapeutic response of the individuals in the medium and low-risk groups is better, leading to an above 90% long-term survival rate.<sup>7</sup> On the other hand, high-risk patients often present with widespread metastatic lesions. Despite receiving aggressive treatment, including high-dose chemotherapy in combination with surgery, autologous bone marrow stem cell transplantation, and radiotherapy, they still present a below 50% long-term survival rate.<sup>8–10</sup> Boosting the rate of long-term survival and cure rate of high-risk neuroblastoma patients is crucial for the improvement of the overall prognosis and represents a pressing challenge in both basic research and clinical therapy. As such, the identification of new and effective targets for neuroblastoma diagnosis and treatment is a vital step toward addressing this issue.

While traditional research methodologies have provided insights into the disease, there remains a need for more comprehensive approaches. Prediction of gene connections and molecular pathways can be facilitated by bioinformatics analysis.<sup>11,12</sup> This methodology is particularly suited to address the intricate nature of neuroblastoma due to its ability to analyze and interpret large-scale biological data. We aim to identify key molecular and genetic factors contributing to the disease. This includes the analysis of differentially expressed genes (DEGs), signaling pathways, and potential biomarkers that are crucial in the progression of neuroblastoma. The Gene Expression Ontology (GEO) serves as an international public repository for archiving and freely distributing various forms of high-throughput functional genomic data, including microarrays and next-generation sequencing.<sup>13</sup> UCSC Xena (<http://xena.ucsc.edu/>) is a comprehensive genome-related database that offers numerous functionalities for tumor research and provides visual data analysis tools for public data centers. The Weighted Correlation Network Analysis (WGCNA) is an R package designed for conducting weighted correlation network analysis. It can serve both as a data exploration tool and a method for ranking genes to identify clusters or modules of highly correlated genes.<sup>14</sup> In recent years, machine learning (ML)-based techniques have gained popularity in addressing a significant challenge in genetic data analysis: the extraction of meaningful genes. Since different studies have identified varying sets of key genes, there is still room to improve the confidence of gene identification using ML and statistics-based bioinformatics models. Once the hub gene is determined, pathway enrichment and immune cell infiltration analyses are carried out. The regulatory mechanism is further explored by examining the interaction mode of miRNA to assess whether miRNA plays a mediating role in real gene expression behaviors. While there have been numerous studies exploring the genetic and molecular aspects of neuroblastoma, our research represents a pioneering effort in applying a combined approach of bioinformatics, and machine learning to thoroughly analyze gene expression and immune cell infiltration in neuroblastoma. To the best of our knowledge, this comprehensive methodology has not been previously employed in neuroblastoma.

In summary, this study was based on the analysis of detailed information from large data sets. Hence, it is expected to find some biomarkers on the prognosis of neuroblastoma. These can be further utilized in the diagnosis and prognosis of NB patients, thus helping them to overcome the clinical heterogeneous disorder and providing a new direction for the research concerning clinical prognosis, diagnosis, and treatment of neuroblastoma.

## Materials and Methods

### Data Acquisition and Differential Gene Screening

The gene chip data set GSE49710 containing 498 neuroblastoma samples was accessed at the GEO database, and the gene chip data set TARGET with 150 neuroblastoma samples was retrieved from the UCSC-XENA database. The two groups of samples were homogenized and standardized using the R language “preprocessCore” software package, which is capable of adjusting for batch effects in the data analysis. The differentially expressed genes (DEGs) were filtered out through the R language “limma” package<sup>15</sup> to correct the screening conditions of  $p < 0.05$  and  $|\log_2\text{FC}| > 0.5$ . The volcano and heat maps were drawn through R to visualize differential genes. “Venny” (<https://bioinfogp.cnb.csic.es/tools/venny/index.html>), a network analysis tool, was utilized for analyzing DEG intersections.

### Gene Function Enrichment and Pathway Analysis

The R language “clusterprofiler” package<sup>16</sup> was utilized to execute GO (<http://www.geneontology.org>), KEGG (<http://www.genome.jp/kegg/pathway.html>), and Reactome (<http://reactome.org/>) enrichment analysis on DEGs. GO is a widely used annotation tool for genes and gene products, including cellular components (CC), biological processes (BP), and molecular function (MF). Genome sequencing and other high-throughput experimental technologies produce large-scale molecular data sets that are integrated and interpreted by KEGG. KEGG pathway<sup>17</sup> is a collection database of metabolic pathways, enriched to correct  $p < 0.05$ . The Reactome<sup>18</sup> was employed to correct  $p < 0.05$  as an enrichment term.

### Weighted Gene Co-Expression Network Analysis (WGCNA)

“WGCNA” package<sup>14</sup> in R language was employed to identify modules significantly associated with mortality and MYCN amplification within the GSE49710 dataset. Cluster analysis was performed using the package’s dedicated

functions, with the “pickSoft Threshold” function being particularly instrumental in determining an optimal soft threshold value of 5. This analysis also involved the use of the ‘blockwise Modules’ function to generate gene modules, each containing a minimum of 50 genes and based on a merging threshold of 0.25. We specifically focused on modules showing a high correlation with mortality and MYCN amplification for subsequent Module Membership (MM) and Gene Significance (GS) analysis. Furthermore, to identify overlapping genes, differential genes, and those within the most pertinent modules related to death and MYCN amplification were mapped using R language and visualized in “Venny”.

## LASSO Regression Analysis

The least absolute shrinkage and selection operator (LASSO) regression is a simplified model obtained by constructing a penalty function, which can be used to process a large number of collinear sample data. The variable can be reduced by adjusting the penalty coefficient  $\lambda$ . When  $\lambda$  is bigger, the punishment is stronger. Thereby, the independent variable with a large influence on the dependent variable can be found. This research utilized the R language “glmnet” package<sup>19</sup> to carry out regression analysis on GSE49710 and TARGET data sets to further narrow the gene range and obtain 29 and 36 variables, respectively.

## ROC Diagnostic Analysis of Hub Gene

The receiver operating characteristic curve (ROC) can be utilized to directly determine the diagnostic efficacy of each diagnostic index. The larger the area under the curve (AUC), the more diagnostic value it has. The R language “pROC” package<sup>20</sup> was employed to draw the ROC curve of the obtained hub gene and screen the biomarkers with diagnostic potential according to the AUC value.

## Immune Infiltration Analysis

The GSE49710 data set was analyzed through the function ssGSEA of the R language GSVA (version 3.6) package,<sup>21</sup> and the “ggplot2” package<sup>22</sup> was used for mapping. To analyze the infiltration of immune cells in neuroblastoma, compare the variation in immune cell infiltration between the survival group and the death group, and assess the link between the selected neuroblastoma hub gene and the rate of immune infiltration in neuroblastoma tissue.

## Construction of TF-miRNA-Hub Co-Regulatory Network

RegNetwork (<https://regnetworkweb.org/>) was utilized for the prediction of the miRNA and transcription factors upstream of the hub gene. Cytoscape<sup>23</sup> software was employed for visualization of the TF-miRNA-hub co-regulatory network to further verify the relationship between the hub gene and neuroblastoma at the molecular level.

## Clinical Specimens

The Formalin-fixed paraffin-embedding samples of 4 NB tissues and 4 non-NB tissues (lymph nodes) were collected from Anhui Children’s Hospital (Hefei, China) without chemotherapy or radiotherapy before surgical excision and all patients gave informed consents. The study has been approved by the Ethics Committee of the Anhui Provincial Children’s Hospital and was conducted in compliance with the Helsinki Declaration.

## RNA Isolation and Determination of Target Gene Expression Using qRT-PCR

To confirm our bioinformatics results, qRT-PCR was conducted on NB and non-NB tissues. Total RNA was extracted using a kit (AmoyDx, China). Then RNA was converted into cDNA via reverse transcription by a Reverse transcription tool (Novoprotein, China). qRT-PCR analysis was completed via a normal protocol from Power SYBR Green (Novoprotein, China). Every protocol was completed as per the supplier’s specifications. The  $\Delta\text{ct}$  results were normalized to the values of glyceraldehyde-3-phosphate dehydrogenase (GAPDH). qRT-PCR analysis and data acquisition were completed via an ABI Q5 apparatus. All specimens were studied three times. The primers were as follows:

*VGF*

Forward(F): CGCTGACCCGAGTGAATCTG

Reverse (R): CATA CGCGCTGGAATTGA



*C19orf52*

Forward (F): CGCTGTGTATGTGGGTCTG

Reverse (R): AGAGCCCCAGGTTGACGTA

*DGKD*

Forward (F): CTTGAGGGCGAACGCTTTA

Reverse (R): TTTTGGTACTGGATTGAGCTACG

*GAPDH*

Forward (F): GGAGCGAGATCCCTCCAAAAT

Reverse (R): GGCTGTTGTCATACTTCTCATGG

## Statistical Analysis

Statistical analyses were performed by GraphPad Prism 7.0 software. Data were presented as the mean  $\pm$  SD. The *t*-test was used for comparison between the two groups.

## Results

### Flow Chart

Figure 1 Depicts the analysis flow chart of this research.

### DEGs Recognition of Neuroblastoma

In the GSE49710 dataset, there were 6132 upregulated and 5529 downregulated differentially expressed genes (DEGs) (Figure 2A). A heat map illustrating the differential gene expression levels in the GSE49710 dataset is shown in Figure 2B. In the TARGET dataset, 420 DEGs were upregulated and 272 were downregulated (Figure 2C), with Figure 2D displaying a heat map of the differential gene expression levels in the TARGET dataset.

### GO and KEGG Analysis of Neuroblastoma DEGs

A Venn diagram was utilized to delineate the intersection of differentially expressed genes (DEGs) across two datasets. We identified 265 upregulated genes (Figure 3A) and 188 downregulated genes (Figure 3B). Subsequent analysis of co-differential

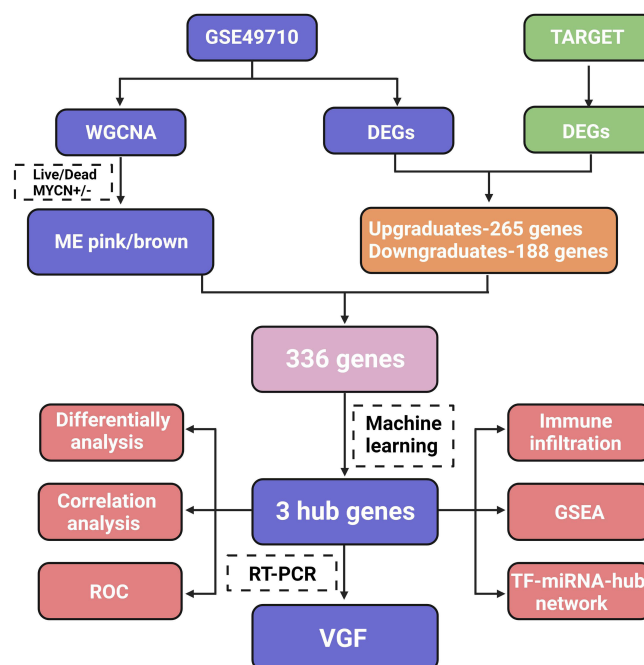
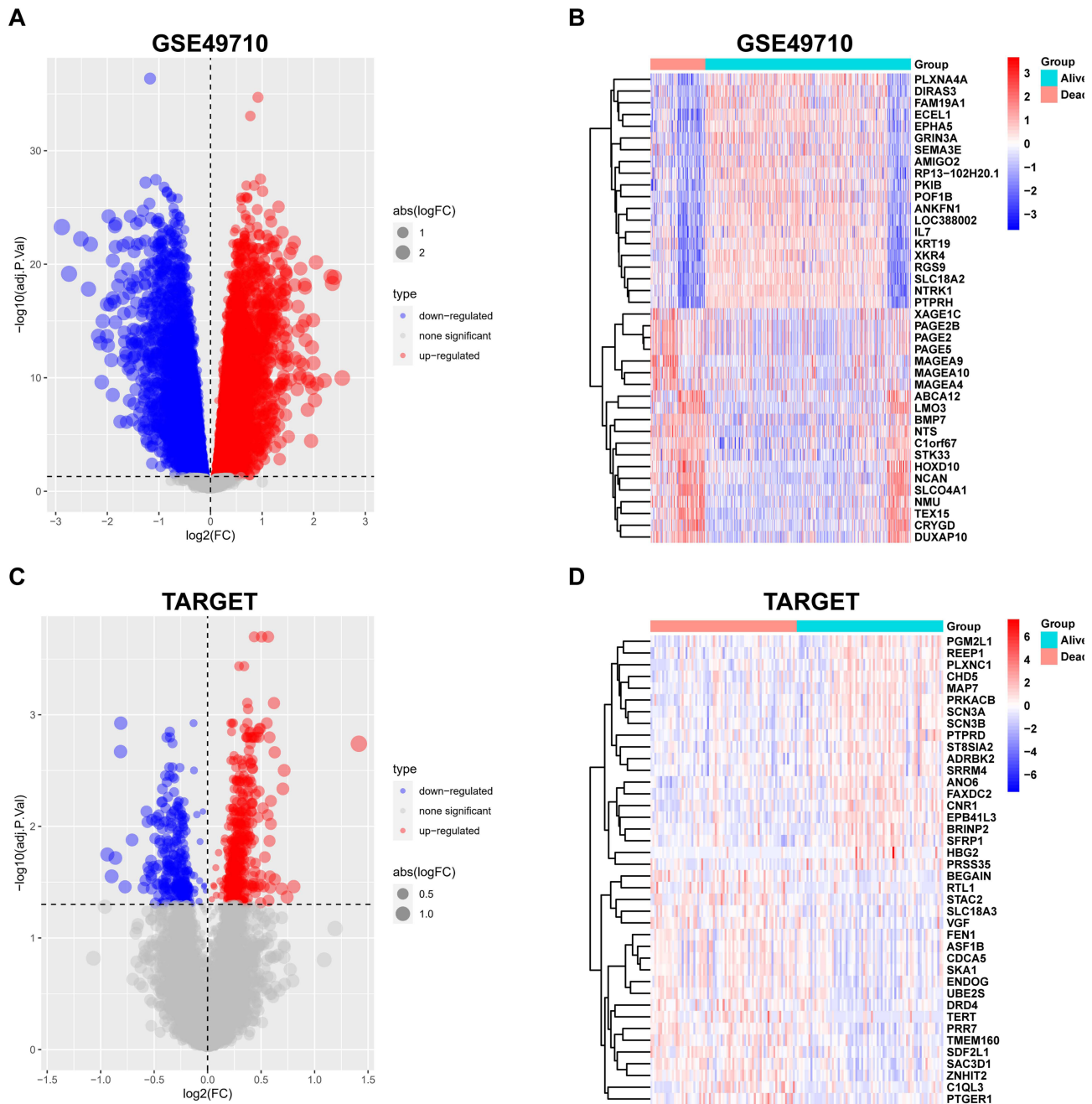
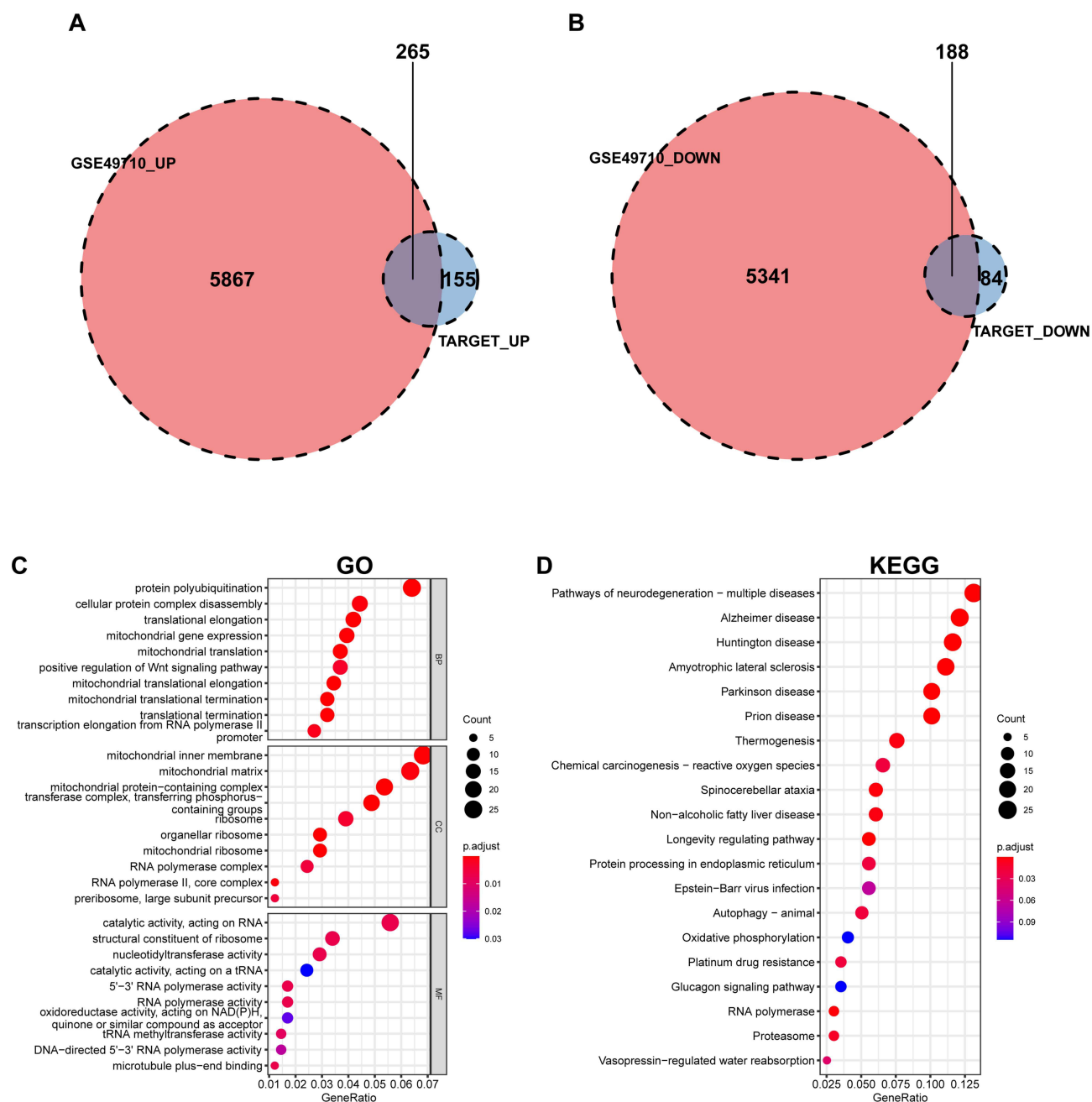


Figure 1 Flowchart of data preparation, processing, analysis, and validation.



**Figure 2** DEGs recognition of neuroblastoma. **(A)** Difference analysis volcano map in GSE49710 dataset **(B)** Difference analysis heat map in GSE49710 dataset. **(C)** Difference analysis volcano map in TARGET dataset **(D)** Difference analysis heat map of TARGET dataset. **(A and C)** Difference analysis volcano map **(B and D)** Difference analysis heat map (GSE49710 up-regulated 6132, down-regulated 5529. TARGET up-regulated 420, down-regulated 272).

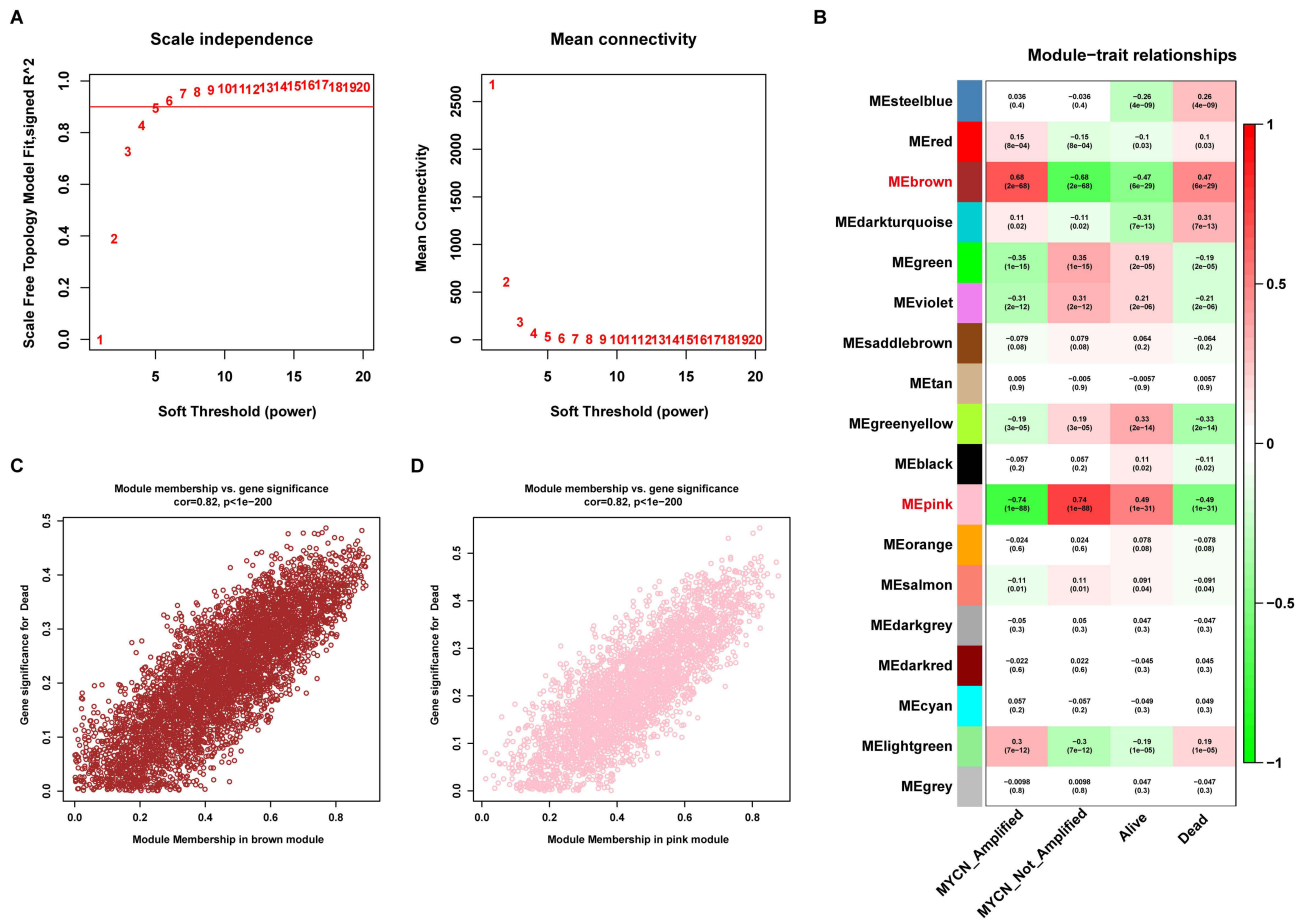
genes revealed that the DEGs primarily influenced biological processes such as protein polyubiquitination, cell protein complex disassembly, translation elongation, and mitochondrial gene expression. Notably, the cell components significantly impacted by these DEGs were enriched in areas like the mitochondrial inner membrane, mitochondrial matrix, mitochondrial protein complex, and transferase complexes involved in phosphorus transfer. Furthermore, the DEGs exhibited a pronounced effect on molecular functions including catalytic RNA activity, ribosomal components, and nucleotide transferase activity, as illustrated in Figure 3C. Additionally, KEGG enrichment analysis indicated a significant association of neuroblastoma DEGs with neurodegenerative diseases, encompassing conditions such as amyotrophic lateral sclerosis, Alzheimer's disease, Parkinson's disease, and Huntington's disease (Figure 3D).



**Figure 3** GO and KEGG analysis of neuroblastoma DEGs. (A) Differential analysis up-regulated gene intersection (265 genes). (B) Differential genes down-regulated gene intersection, 188 genes were obtained. (C) GO functional enrichment analysis (including BP MF CC). (D) KEGG pathway enrichment analysis results.

## Weighted Gene Co-Expression Network Analysis

WGCNA was conducted for screening the genes with a stronger correlation with death and MYCN amplification from the GSE49710 gene in the data set. The results showed that when the soft threshold of the selection adjacency matrix was five, the links among genes in the network were most consistent with the scale-free network distribution (Figure 4A). The power value was set to five to divide 18 gene modules. Among these modules, the modules named MEbrown and MEpink had the strongest correlation with Gene Set Expression (GSE) death and MYCN amplification (Figure 4B). It suggested that the genes in MEbrown and MEpink modules had a stronger correlation with GSE death and MYCN amplification. Afterward, further study was carried out for MEbrown and MEpink modules and death, whose resulting data depicted a positive association between Module Membership (MM) and Gene Significance (GS) (Figure 4C and 4D).



**Figure 4** Weighted gene co-expression network analysis. **(A)** Screening soft threshold powers value, screening results for 5 when the most appropriate. **(B)** Association of gene modules with clinical traits. **(C)** The positive correlation between gene significance for death and module membership in the brown module. **(D)** The strong positive correlation between gene significance for death and module membership in the pink module.

## Overlapping Gene Screening and GO and KEGG Analysis

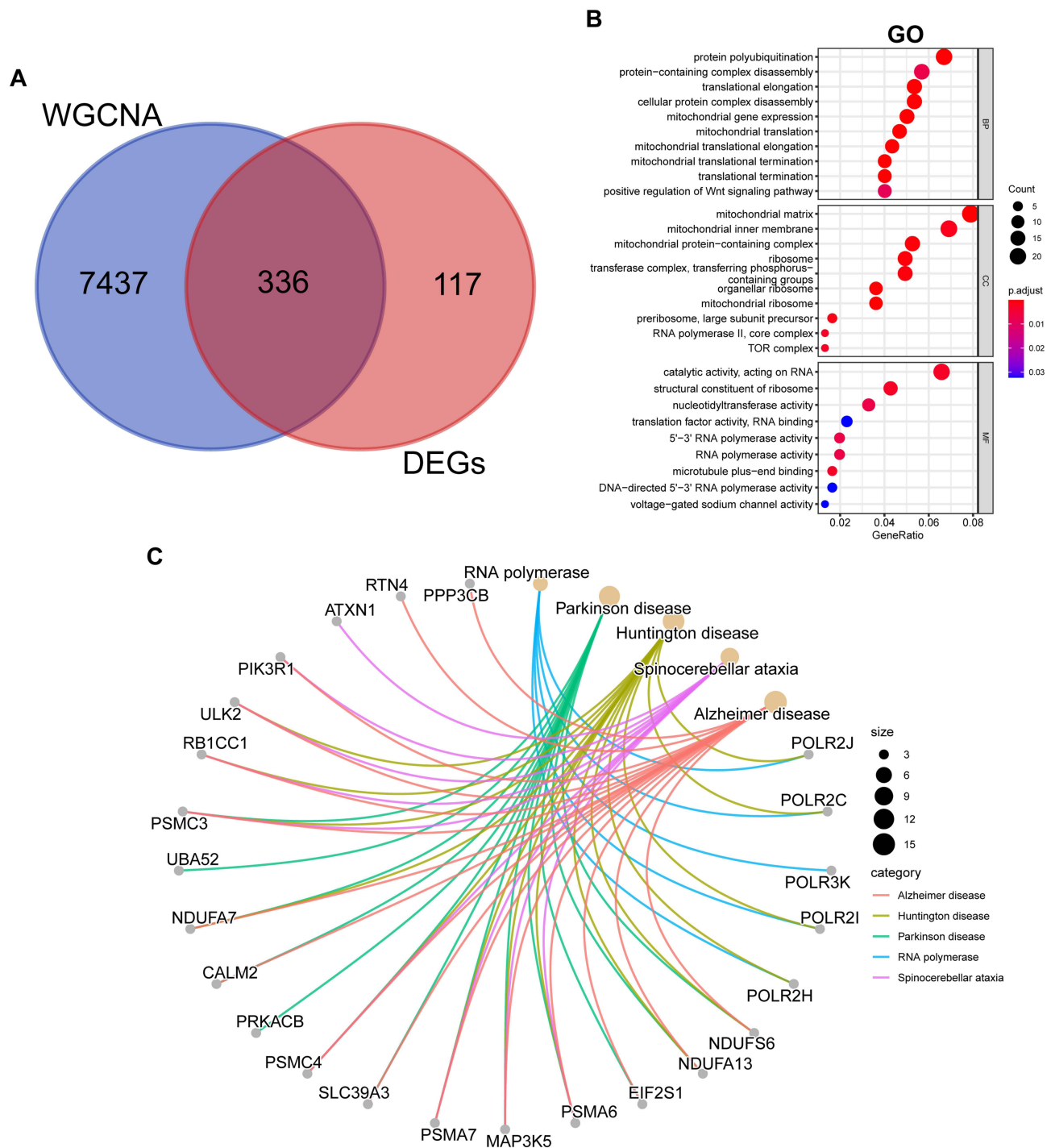
The intersection of MEbrown and MEPink module genes and co-differential genes was obtained by using the Venn diagram, and a total of 336 overlapping genes were obtained (Figure 5A). GO analysis and TOP5 KEGG were carried out to draw an enrichment network diagram, aiming to further clarify the association between overlapping genes and functions and mechanisms of action. GO analysis results showed that its function was mainly related to protein polyubiquitination and catalytic RNA synthesis (Figure 5B). The KEGG enrichment network diagram (Figure 5C) depicts the relationship between the top 5 KEGG result genes and pathways.

## LASSO Regression Analysis Results

The regression analysis of GSE49710 and TARGET data sets was carried out using LASSO regression to further narrow the gene range. Through cross-validation and at the lowest mean square error, 29 genes were screened by GSE49710 analysis, 36 by TARGET analysis, and three hub genes were obtained by using the intersection of the Venn diagram (Figure 6).

## Hub Gene Difference, Correlation Analysis, and ROC Diagnosis Analysis

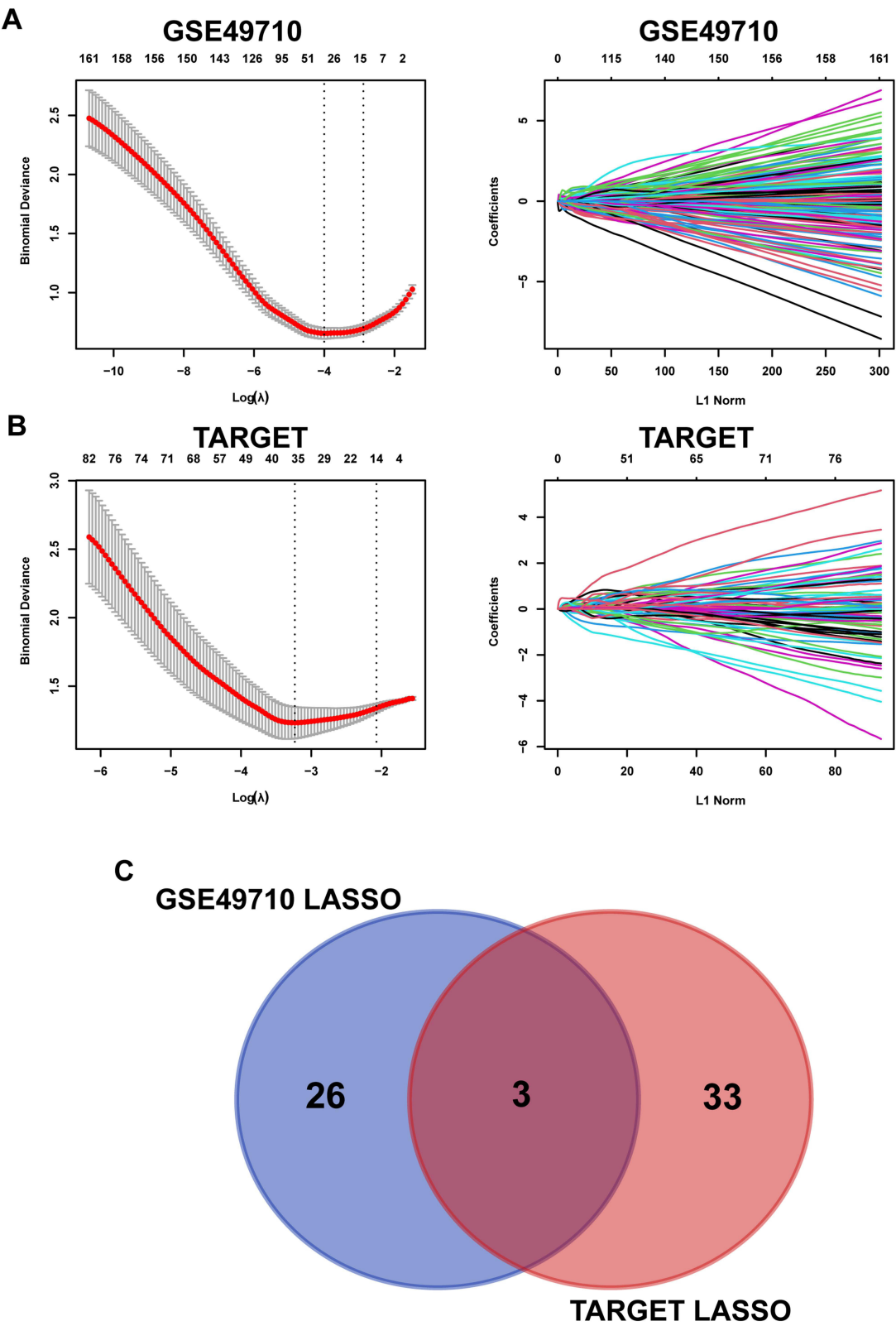
The difference analysis of the three hub genes obtained after Lasso analysis was carried out in the two data sets. The value was log FC, showing two genes with high expression and one gene with low expression (Figure 7A). At the same time, the correlation of three hub genes was analyzed. *VGF* was positively associated with *C19orf52* and depicted



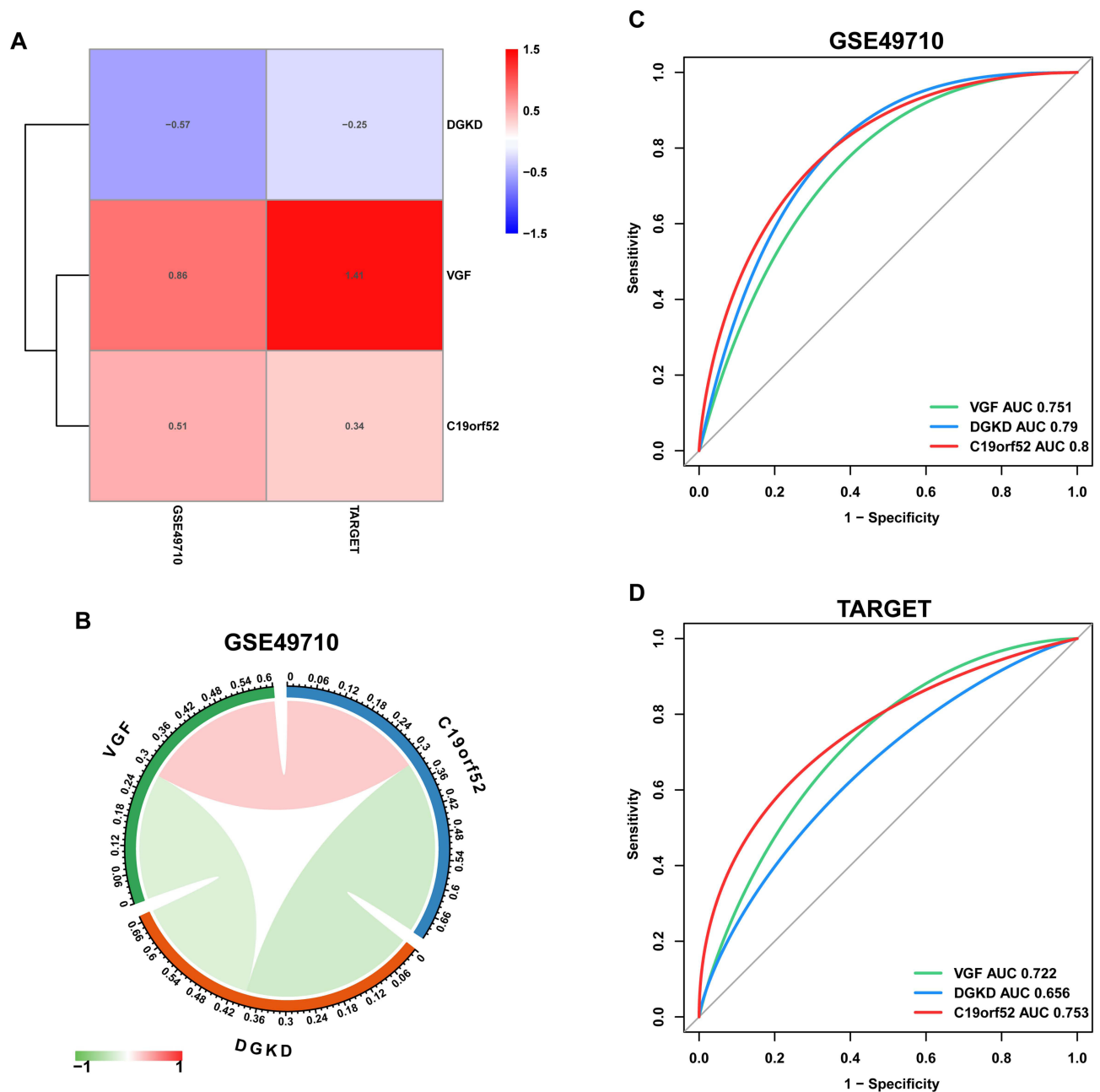
**Figure 5** Overlapping gene screening and GO and KEGG analysis. **(A)** Two module genes intersected with differential genes to obtain 336 genes. **(B)** GO enrichment analysis of 336 genes. **(C)** Corresponding relationship between genes and pathways.

a negative association with *DGKD*. *C19orf52* was also negatively associated with *DGKD* (Figure 7B). In addition, ROC diagnostic analysis was performed on three hub genes. When  $AUC > 0.5$ , the nearer the AUC value is to 1, the more favorable the diagnostic effect of the diagnostic object. Based on the two data sets of GSE49710 and TARGET, the ROC curves of the three hub genes were drawn utilizing the R language pROC software package. The results indicated that the AUC value of the hub gene was basically between 0.6 and 0.8, and the diagnostic value of *C19orf52* ( $AUC = 0.8, 0.753$ ) was relatively high (Figure 7C and D).





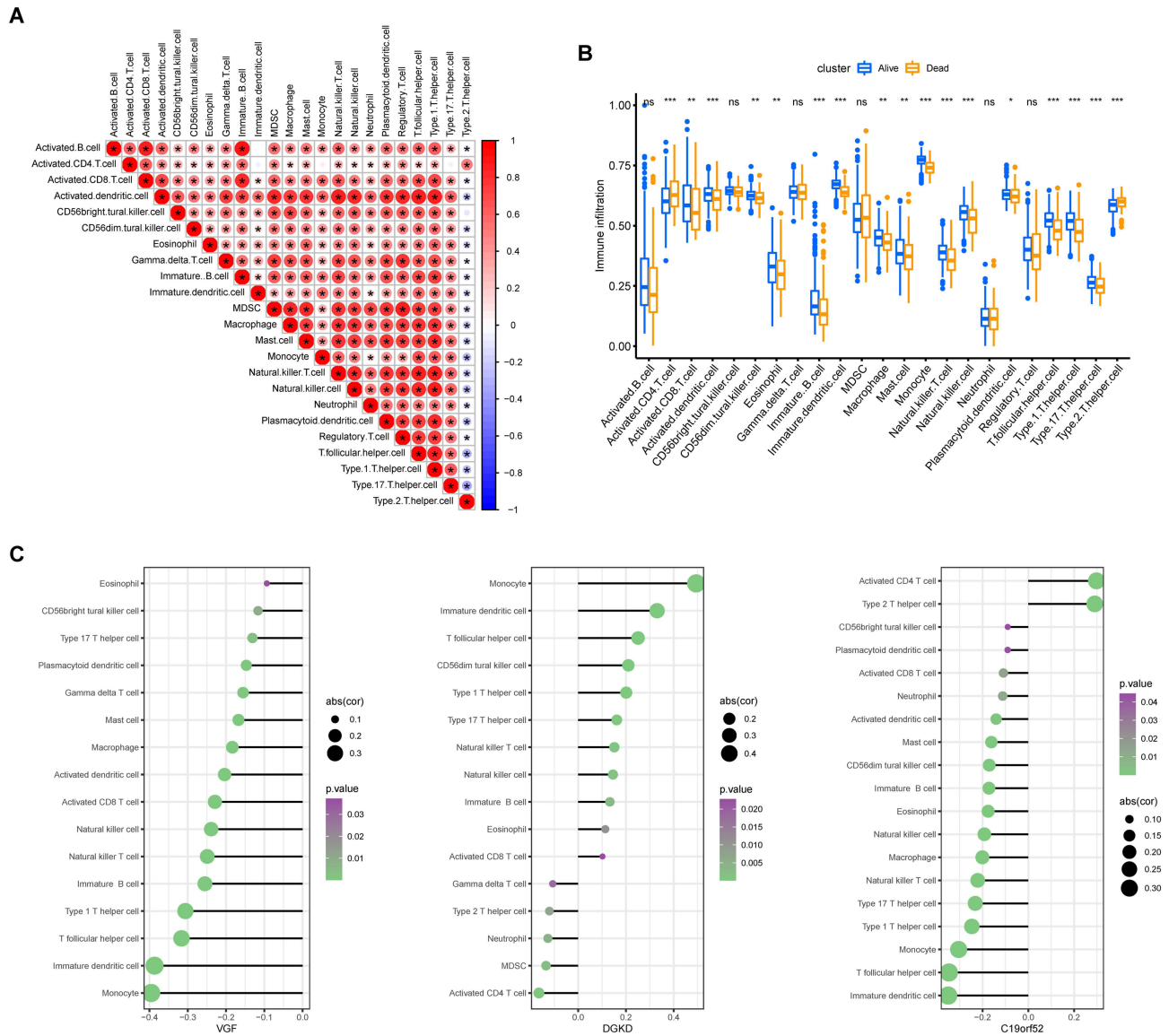
**Figure 6** LASSO regression analysis results. **(A)** GSE49710 lasso regression analysis (screened 29 genes). **(B)** TARGET lasso regression analysis (36 genes screened). **(C)** The results of the two methods intersected to obtain three hub genes.



**Figure 7** Hub gene difference, correlation analysis, and ROC diagnosis analysis. **(A)** Hub gene difference analysis (value represents logFC value). **(B)** Correlation analysis of core genes (red lines represent positive correlation, green lines represent negative correlation, the deeper the color, the stronger the correlation). **(C and D)** Receiver-operating characteristic curve.

## Immune Infiltration Analysis

The function ssGSEA of the R language GSVA package was utilized for the evaluation of the proportion of immune infiltrates on the GSE49710 data set. Correlation analysis of the infiltration of 23 different immune cells was performed, of which monocytes, follicular helper T cells, and CD4+ T cells were strongly correlated (Figure 8A). The GSE49710 data set was categorized into the death and the survival groups, wherein it was determined that the rate of infiltration of activated CD4+ T cells, macrophages, mast cells, plasmacytoid dendritic cells, and Th2 cells was high in the death group. Whereas the degree of infiltration of the activated dendritic cells, immature B cells, immature dendritic cells, natural killer T cells, monocytes, follicular helper T cells, natural killer cells, Th1 and Th17 cells were high in the survival group (Figure 8B). The ggplot2 was used to assess the association between hub genes and immune cells. *VGF* expression

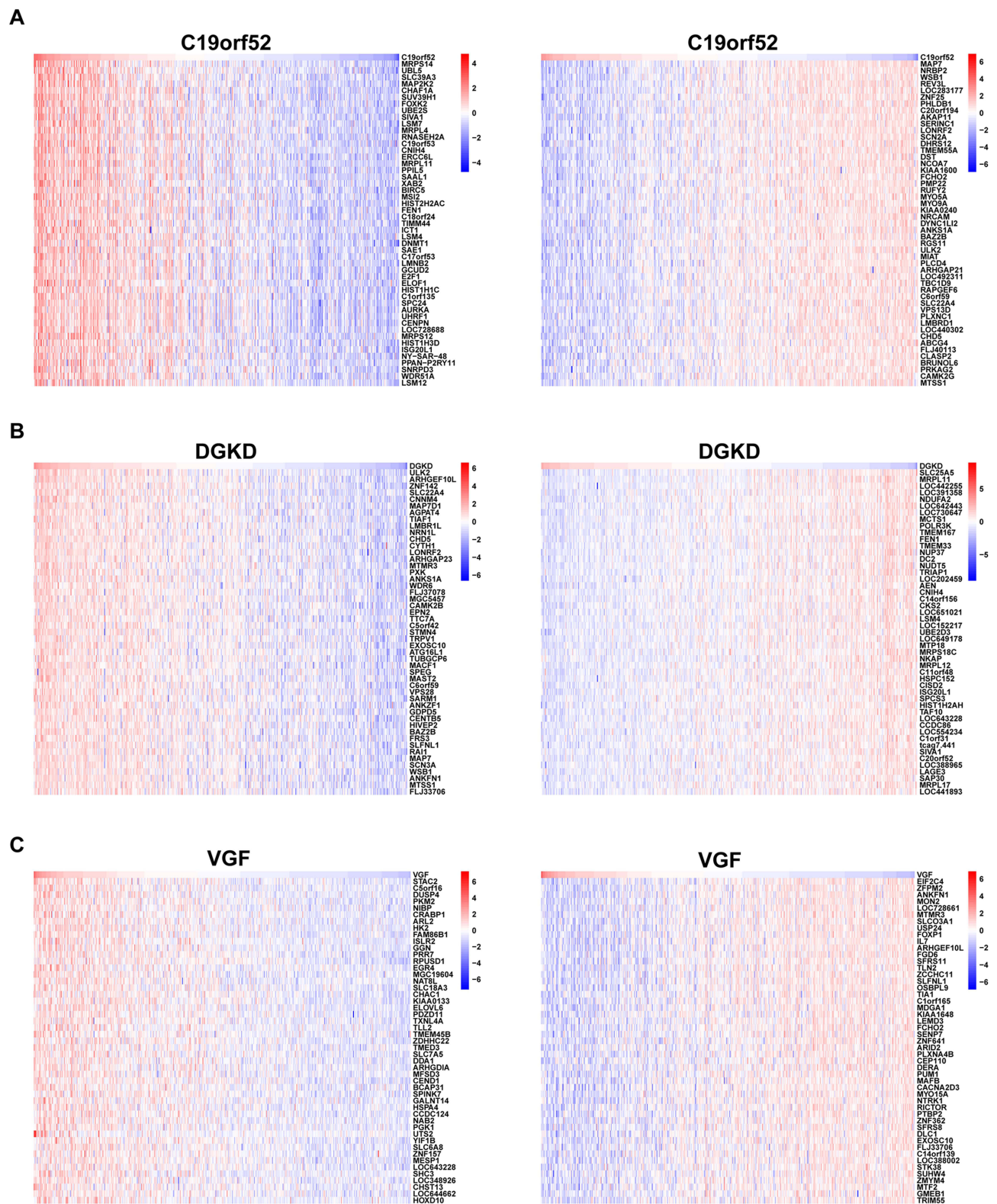


**Figure 8** Immune infiltration analysis. **(A)** Correlation of different immune cell infiltration. **(B)** Difference of immune infiltration between death group and survival group (ns: Not significant ( $p \geq 0.05$ ), \*:  $p < 0.05$ , \*\*:  $p < 0.01$ , \*\*\*:  $p < 0.001$ ). **(C)** Correlation between the hub gene and immune cell infiltration (only  $p < 0.05$  immune cells).

negatively correlates with the presence of monocytes and immature dendritic cells. In contrast, higher levels of *DGKD* are positively linked to increased infiltration of activated CD4+ T cells and T follicular helper cells. *C19orf52* shows a complex relationship with immune cells, having both positive and negative correlations; for instance, a positive correlation with monocyte infiltration and a negative one with activated CD4+ T cells and T follicular helper cells. (Figure 8C).

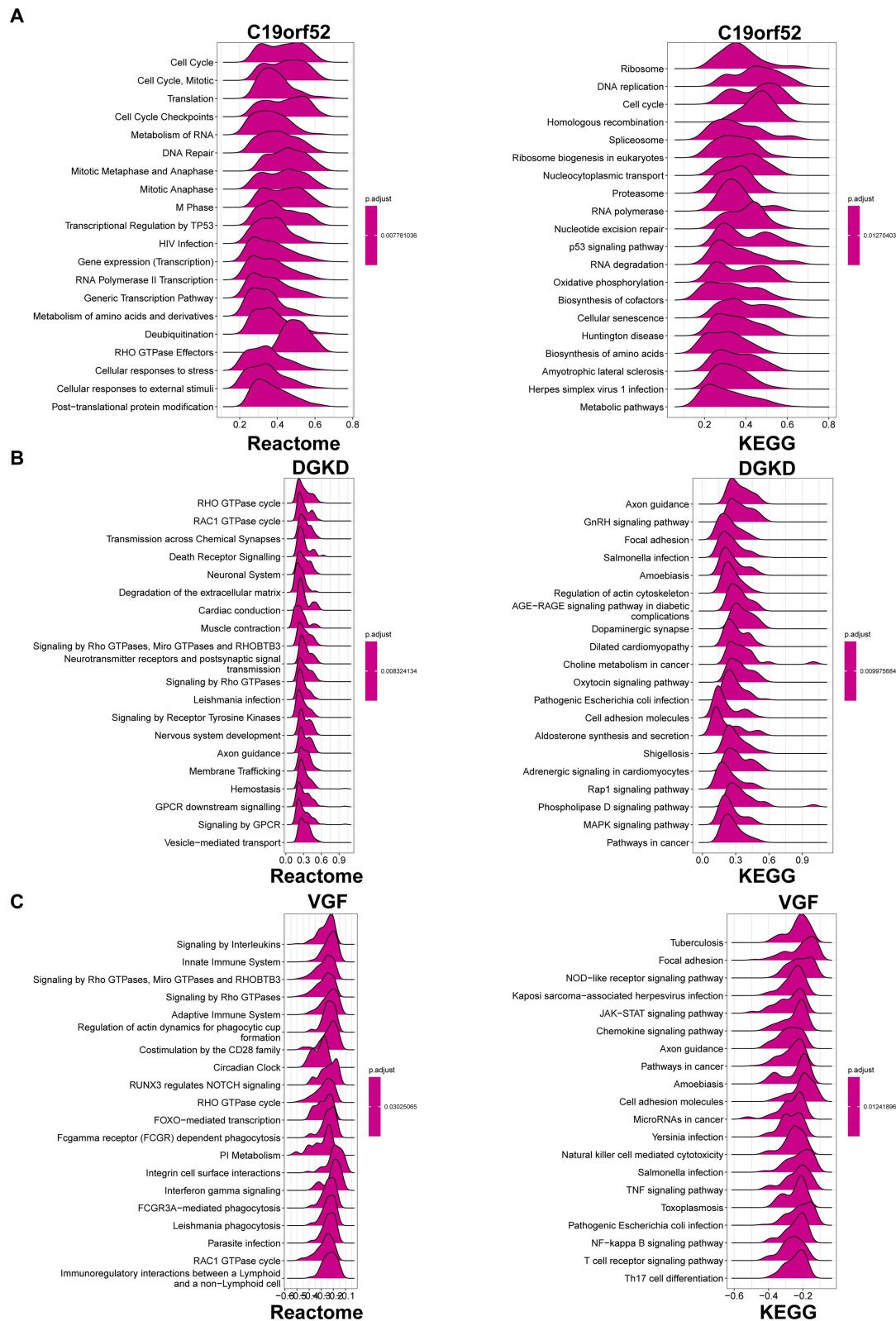
# Single-Gene GSEA Analysis and Construction of TF-miRNA-Hub Co-Regulatory Network

Correlation analysis was performed between the three hub genes and all genes in the data set GSE49710, and a heat map was used to display the expression of positively correlated genes (top 50) and the expression of negatively correlated genes (first 50) (Figure 9). According to the data of the correlation analysis, the single-gene GSEA analysis based on Reactome and KEGG was executed, and the first 20 results of the three hub genes were displayed respectively (Figure 10). To further verify the link between the hub genes and neuroblastoma, the RegNetwork was utilized for the



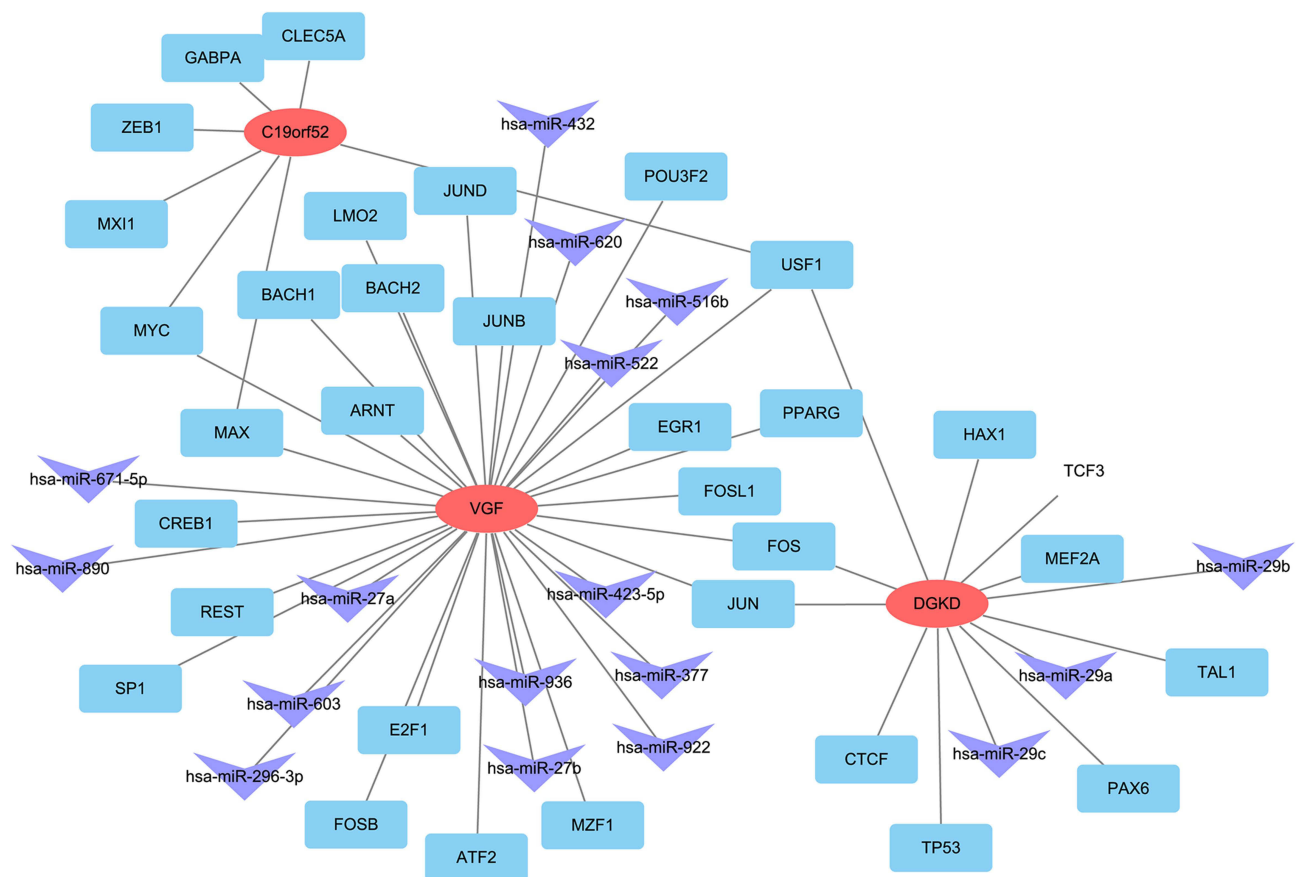
**Figure 9** Heatmap of the gene correlation. Gene expression analysis of positive correlation top50 and negative correlation top50 of Hub gene. **(A)** *C19orf52* Correlation Analysis Heatmaps. **(B)** *DGKD* Correlation Analysis Heatmaps. **(C)** *VGF* Correlation Analysis Heatmaps.





**Figure 10** Single-gene GSEA analysis. Based on the results of correlation analysis in Figure 8, GSEA analysis of single gene based on Reactome and KEGG was performed. (A) GSEA Analysis for *C19orf52*. (B) GSEA Analysis for *DGKD*. (C) GSEA Analysis for *VGF*.





**Figure 11** Upstream prediction. Prediction of miRNAs and transcription factors upstream of three Hub genes.

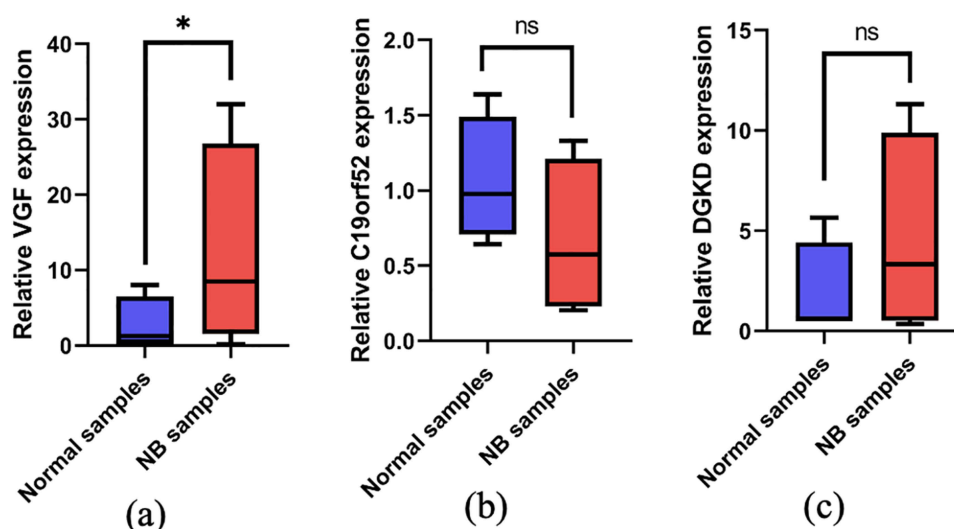
prediction of the miRNAs and related transcription factors upstream of the hub gene, while the Cytoscape software was utilized for mapping analysis. The data depicted that *CLEC5A*, *GABPA*, *ZEB1*, *MXI1*, *MYC*, *JUND*, *LMO2*, *BACH1*, *BACH2*, *JUNB*, *MAX*, *ARNT*, *CREB1*, *REST*, *SP1*, *POU3F2*, *USF1*, *EGR1*, *PPARG*, *FOSL1*, *FOS*, *JUN*, *HAX1*, *MEF2A*, *CTCF*, *TP53*, *TAL1*, *PAX6*, *E2F1*, *FOSB*, *ATF2*, and *MZF1* may be potential TFs that regulate the hub gene expression. The potential upstream miRNAs of the hub gene included has-miR-432, has-miR-620, has-miR-516b, has-miR-522, has-miR-671-5p, has-miR-890, has-miR-27a, has-miR-603, has-miR-296-3p, has-miR-936, has-miR-27b, has-miR-423-5p, has-miR-377, has-miR-922, has-miR-29a, has-miR-29b, and has-miR-29c (Figure 11).

## Validation of Key Genes by Real-Time qPCR

To confirm the involvement of key genes in the pathogenesis of neuroblastoma (NB), quantitative Real-Time Polymerase Chain Reaction (qRT-PCR) was utilized to measure the expression levels of three key genes in four NB tissue samples and four non-NB lymph gland samples. Our findings revealed that *VGF* is significantly upregulated in neuroblastoma tissues (Figure 12), corroborating our bioinformatics analysis. However, the expression levels of *DGKD* and *C19orf52* did not show a statistically significant difference when comparing NB tissues with normal tissues. These observations indicate that the hub gene *VGF* might play a role in the etiology of neuroblastoma and could potentially serve as a biomarker for this disease.

## Discussion

NB is a significant cause of childhood malignancy-related death, and early diagnosis and treatment are essential for extending patients' survival time.<sup>24</sup> Therefore, it is necessary to investigate prognostic indicators for NB further. With the development of bioinformatics, sequencing technology is increasingly used to explore the early diagnosis, treatment, and prognosis of



**Figure 12** The relative mRNA expression of VGF (a), *C19orf52* (b) and *DGKD* (c) in NB tissues and non-NB tissues. The experiments were repeated three times, and each experiment was triplicated. (ns: Not significant ( $p \geq 0.05$ ), \*:  $p < 0.05$ ).

cancer.<sup>25</sup> In the field of pediatric neuroblastoma, studies utilizing bioinformatics to identify prognostic markers remain relatively scarce. There has been only a report highlighting *GDPD5*, related to lipid metabolism, as a potential prognostic biomarker.<sup>26</sup> Additionally, variations in the expression of genes linked to ferroptosis have been observed between stages 1 and 4 of neuroblastoma.<sup>27</sup> There is a pressing need for continued investigation into these biomarkers and their validation in clinical samples to enhance our understanding of the disease and improve prognostic accuracy. This study aims to seek hub genes (DEGs) between NB with a favorable prognosis and NB with an unfavorable prognosis to understand the mechanism of NB further and potentially provide diagnostic biomarkers and therapeutic targets.

The data sets GSE49710 and TARGET were accessed at the databases GEO and UCSC-XENA, respectively, and the DEGs were screened by using R language. The respective upregulated and downregulated DEGs reached almost 265 and 188 in number. The GO analysis and KEGG enrichment analysis were performed with the data depicting the elevated concentration of neuroblastoma DEGs in the pathways of neurodegenerative disorders such as amyotrophic lateral sclerosis, Parkinson's disease, Huntington's disease, and Alzheimer's disease. Among them, protein polyubiquitination, translation elongation, catalytic RNA activity, and mitochondrial gene expression can lead to the incidence and development of tumors.<sup>28–31</sup> Existing research has indicated that ubiquitination performs a vital function in cancer and neurodegenerative disorders, so the regulation of ubiquitination pathways is considered to be a promising therapeutic strategy for tumors and neurodegenerative diseases.

To screen out DEGs with stronger correlations with death traits and MYCN amplification from the GSE49710 data set, WGCNA analysis was conducted. The high expression of the MYCN gene was positively correlated with death traits and decreased the survival rate of neuroblastoma, which was consistent with the conclusion described in the literature, such as the high expression of MYCN can promote the proliferation of neuroblastoma cells and increase invasion and metastasis, leading to poor prognosis.<sup>7</sup> In addition, WGCNA analysis was performed on the TARGET data set, but the results were not very significant, so they were not displayed in the figures.

To further narrow the range of DEGs, the machine learning LASSO regression analysis was performed on the two data sets GSE49710 and TARGET, and finally, three hub genes were obtained, ie *C19orf52*, *DGKD*, and *VGF*. The difference and correlation analysis of the hub genes were conducted, and the results depicted two genes that were highly expressed and one lesser expressed gene, of which *VGF* was positively associated with *C19orf52* and negatively associated with *DGKD*. *C19orf52* depicted a negative association with *DGKD*. Further ROC diagnostic analysis depicted that the AUC (AUC=0.8, 0.753) of *C19orf52* was relatively high, with a high diagnostic value.

Neuroendocrine factor *VGF* is a polypeptide precursor composed of 617 amino acids, mainly expressed in neurons.<sup>32</sup> Adult *VGF* has been detected in several parts of the brain, including the adrenal medulla, cerebral cortex, hippocampus, hypothalamus, cerebral olfactory system, as well as motor neurons of the spinal cord.<sup>33</sup> Existing studies have shown that

*VGF* may be a key regulator of Alzheimer's disease.<sup>34</sup> In addition, clinical studies have reported that *VGF* is related to lung cancer,<sup>35,36</sup> pancreatic cancer,<sup>37</sup> and glioma,<sup>38</sup> but its role in neuroblastoma has not been reported, and its mechanism needs to be further studied.

The human TIM22 complex considers the *C19orf52* as its novel metazoan-specific subunit. *C19orf52*, essential for the TIM22 complex stability, is integrated into the inner membrane of the mitochondria. Additionally, it plays a role in the assembly of hTIM22. There have been few reports on *C19orf52* in recent years.

Diacylglycerol kinase delta (*DGKD*), an enzyme of the phosphatidylinositol (PI) cycle, performs a vital function in signal transduction, mainly by regulating the balance between these signaling lipids.<sup>39</sup> Recent research has implicated *DGKD* in the development and function of the central nervous system.<sup>40</sup> Mizuno et al proposed that at the initial/early stage of neuroblastoma cell differentiation, *DGKD* upregulated neurite outgrowth.<sup>41</sup>

After the hub genes were identified, the possible mechanism was explored. Immune infiltration is an important part of the tumor microenvironment. The types and methods of infiltrating cells are correlated with clinical outcomes and are expected to become therapeutic targets to improve prognosis. In this research, the degree of immune cell infiltration was evaluated, and survival analysis was performed. The resulting data depicted that the progression of neuroblastoma was promoted through the downregulation of monocytes by *VGF* and *C19orf52* as well as through the upregulation of monocytes by *DGKD* and through the upregulation of CD4+T cells and Th2 cells by *C19orf52*, increasing the mortality. Therefore, monocytes, CD4+T cells, and Th2 cells can be regarded as therapeutic targets, providing new ideas for the treatment of neuroblastoma. Those hypotheses needed more research to reveal the intricate interplay between genes and immunocytes.

To gain an in-depth understanding of the hub genes, a single-gene GSEA analysis was executed based on Reactome and KEGG, and the first 20 results of the three hub genes were displayed, respectively. *C19orf52* is mainly involved in the cell cycle and mitosis to enhance the proliferation of neuroblastoma cells. The combination of *VGF* mediates the interaction between cells and extracellular matrix through integrin cell surface interaction to regulate cell movement, proliferation, and apoptosis, resulting in tumor growth, infiltration, metastasis, and other effects. *DGKD* is mainly involved in choline metabolism in cancer, induces neuropathy in the tumor microenvironment, and leads to tumor growth and metastasis, resulting in perineural infiltration. This research verified the mechanism of hub gene-induced neuroblastoma by utilizing RegNetwork to predict the upstream miRNA and related transcription factors of the hub gene. Among these, *USF1* is directly linked with the three hub genes. Many studies have linked *USF1* to the proliferation, metastasis, invasion, and drug resistance of many tumors, which indicates the potential of the *USF1* gene as a marker for diagnosis, targeted therapy, and prognosis of neuroblastoma.

Our study identified several differentially expressed genes (DEGs) potentially involved in neuroblastoma pathogenesis. To provide a deeper understanding, we explored the functional roles, regulatory networks, and interactions among these DEGs. Gene Ontology (GO) and KEGG enrichment analyses showed that the DEGs are primarily involved in critical processes such as protein translation, membrane composition, and RNA transcription regulation. These findings suggest that the DEGs may influence neuroblastoma development by modulating essential cellular functions.<sup>42</sup> The Weighted Gene Co-expression Network Analysis (WGCNA) identified significant modules related to key outcomes, such as MYCN amplification and patient survival. These modules suggest that the identified hub genes: *VGF*, *DGKD*, and *C19orf52* play crucial roles in the molecular networks driving neuroblastoma progression. Furthermore, our analysis of upstream miRNAs and transcription factors provided insights into the regulatory mechanisms controlling these DEGs, highlighting potential targets for therapeutic intervention.

However, the current experiment still has some limitations. Firstly, the clinical sample size for validating hub genes is relatively small, and large-scale clinical trials are needed. Secondly, further validation of hub genes at the protein level is also required.

## Conclusion

In conclusion, the DEGs of neuroblastoma and their biological functions were preliminarily studied through bioinformatics analysis, and three hub genes were obtained, *C19orf52*, *DGKD*, and *VGF*. *VGF* is upregulated in neuroblastoma tissues, consistent with the findings from bioinformatics analysis. The miRNAs and related transcription factors upstream

of the hub genes were predicted. These genes which require more clinical specimens to validate their findings may have the potential to become targets for neuroblastoma diagnosis, targeted therapy, and prediction of prognosis. However, there is no research or report on the correlation between *C19orf52* and *USF1* gene and neuroblastoma. Therefore, this study can provide new research areas and references for subsequent clinical and basic research.

## Data Sharing Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Ethics Approval and Consent to Participate

The study involving human material were performed in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Anhui Children's Hospital, Informed consent was obtained from all legal guardian of the patients included in this study.

## Acknowledgments

We thank Bullet Edits Limited for the linguistic editing and proofreading of the manuscript.

## Author Contributions

ZSH was responsible for the idea and concept of the paper. All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work. All authors contributed to the article and approved the submitted version.

## Funding

There is no funding to report.

## Disclosure

All the authors declare that they have no conflict of interest.

## References

1. Linabery AM, Ross JA. Childhood and adolescent cancer survival in the US by race and ethnicity for the diagnostic period 1975-1999. *Cancer*. 2008;113:2575-2596. doi:10.1002/cncr.23866
2. Ni X, Li Z, Li X, et al. Socioeconomic inequalities in cancer incidence and access to health services among children and adolescents in China: a cross-sectional study. *Lancet*. 2022;400:1020-1032. doi:10.1016/S0140-6736(22)01541-0
3. Fonseka P, Liem M, Ozcitti C, Adda CG, Ang CS, Mathivanan S. Exosomes from N-Myc amplified neuroblastoma cells induce migration and confer chemoresistance to non-N-Myc amplified cells: implications of intra-tumour heterogeneity. *J Extracell Vesicles*. 2019;8:1597614. doi:10.1080/20013078.2019.1597614
4. Scheer M, Bork K, Simon F, Nagasundaram M, Horstkorte R, Gnanapragassam VS. Glycation leads to increased polysialylation and promotes the metastatic potential of neuroblastoma cells. *Cells*. 2020;9:868. doi:10.3390/cells9040868
5. Depuydt P, Boeva V, Hocking TD, et al. Genomic amplifications and distal 6q loss: novel markers for poor survival in high-risk neuroblastoma patients. *J Natl Cancer Inst*. 2018;110:1084-1093. doi:10.1093/jnci/djy022
6. Koneru B, Lopez G, Farooqi A, et al. Telomere maintenance mechanisms define clinical outcome in high-risk neuroblastoma. *Cancer Res*. 2020;80:2663-2675. doi:10.1158/0008-5472.CAN-19-3068
7. Pinto NR, Applebaum MA, Volchenboum SL, et al. Advances in risk classification and treatment strategies for neuroblastoma. *J Clin Oncol*. 2015;33:3008-3017. doi:10.1200/JCO.2014.59.4648
8. Upton K, Modi A, Patel K, et al. Epigenomic profiling of neuroblastoma cell lines. *Sci Data*. 2020;7:116. doi:10.1038/s41597-020-0458-y
9. Almstedt E, Elgendy R, Hekmati N, et al. Integrative discovery of treatments for high-risk neuroblastoma. *Nat Commun*. 2020;11:71. doi:10.1038/s41467-019-13817-8
10. Berthold F, Falduum A, Ernst A, et al. Extended induction chemotherapy does not improve the outcome for high-risk neuroblastoma patients: results of the randomized open-label GPOH trial NB2004-HR. *Ann Oncol*. 2020;31:422-429. doi:10.1016/j.annonc.2019.11.011
11. Srivastava P, Mangal M, Agarwal SM. Understanding the transcriptional regulation of cervix cancer using microarray gene expression data and promoter sequence analysis of a curated gene set. *Gene*. 2014;535:233-238. doi:10.1016/j.gene.2013.11.028

12. Mancini-DiNardo D, Judkins T, Woolstenhulme N, et al. Design and validation of an oligonucleotide microarray for the detection of genomic rearrangements associated with common hereditary cancer syndromes. *J Exp Clin Cancer Res.* **2014**;33:74. doi:10.1186/s13046-014-0074-9
13. Toro-Dominguez D, Martorell-Marugan J, Lopez-Dominguez R, et al. ImaGEO: integrative gene expression meta-analysis from GEO database. *Bioinformatics.* **2019**;35:880–882. doi:10.1093/bioinformatics/bty721
14. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* **2008**;9:559. doi:10.1186/1471-2105-9-559
15. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**;43:e47. doi:10.1093/nar/gkv007
16. Wu T, Hu E, Xu S, et al. ClusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation.* **2021**;2:100141. doi:10.1016/j.xinn.2021.100141
17. Du J, Yuan Z, Ma Z, Song J, Xie X, Chen Y. KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. *Mol Biosyst.* **2014**;10:2441–2447. doi:10.1039/C4MB00287C
18. Gillespie M, Jassal B, Stephan R, et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **2022**;50:D687–D692. doi:10.1093/nar/gkab1028
19. Engebretsen S, Bohlin J. Statistical predictions with glmnet. *Clin Clin Epigenet.* **2019**;11:123. doi:10.1186/s13148-019-0730-1
20. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf.* **2011**;12:77. doi:10.1186/1471-2105-12-77
21. Hanzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf.* **2013**;14:7. doi:10.1186/1471-2105-14-7
22. Villanueva RAM, Chen ZJ. ggplot2: elegant graphics for data analysis. 2nd ed. Measurement. **2019**;17: 160–167.
23. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape stringApp: network analysis and visualization of proteomics data. *J Proteome Res.* **2019**;18:623–632. doi:10.1021/acs.jproteome.8b00702
24. Park JA, Cheung NV. Targets and antibody formats for immunotherapy of neuroblastoma. *J Clin Oncol.* **2020**;38:1836–1848. doi:10.1200/JCO.19.01410
25. Xia XQ, Jia Z, Porwollik S, et al. Evaluating oligonucleotide properties for DNA microarray probe design. *Nucleic Acids Res.* **2010**;38:e121. doi:10.1093/nar/gkq039
26. Luo T, Peng J, Li Q, et al. GDPD5 related to lipid metabolism is a potential prognostic biomarker in neuroblastoma. *Int J Mol Sci.* **2022**;23:13740.
27. Chen Y, Li Z, Cao Q, Guan H, Mao L, Zhao M. Ferroptosis-related gene signatures in neuroblastoma associated with prognosis. *Front Cell Dev Biol.* **2022**;10:871512. doi:10.3389/fcell.2022.871512
28. Dai H, Yan M, Li Y. The zinc-finger protein ZCCHC2 suppresses retinoblastoma tumorigenesis by inhibiting HectH9-mediated K63-linked polyubiquitination and activation of c-Myc. *Biochem Biophys Res Commun.* **2020**;521:533–538. doi:10.1016/j.bbrc.2019.10.163
29. Kobayashi D, Tokuda T, Sato K, et al. Identification of a specific translational machinery via TCTP-EF1A2 interaction regulating NF1-associated tumor growth by affinity purification and data-independent mass spectrometry acquisition (AP-DIA). *Mol Cell Proteomics.* **2019**;18:245–262. doi:10.1074/mcp.RA118.001014
30. Jin Q, Liu G, Wang B, et al. High methionyl-tRNA synthetase expression predicts poor prognosis in patients with breast cancer. *J Clin Pathol.* **2020**;73:803–812. doi:10.1136/jclinpath-2019-206175
31. Gillingwater TH, Wishart TM. Mechanisms underlying synaptic vulnerability and degeneration in neurodegenerative disease. *Neuropathol Appl Neurobiol.* **2013**;39:320–334. doi:10.1111/nan.12014
32. Alder J, Thakker-Varia S, Bangasser DA, et al. Brain-derived neurotrophic factor-induced gene expression reveals novel actions of VGF in hippocampal synaptic plasticity. *J Neurosci.* **2003**;23:10800–10808. doi:10.1523/JNEUROSCI.23-34-10800.2003
33. Russo-Neustadt A, Beard RC, Cotman CW. Exercise, antidepressant medications, and enhanced brain derived neurotrophic factor expression. *Neuropsychopharmacology.* **1999**;21:679–682. doi:10.1016/S0893-133X(99)00059-7
34. Beckmann ND, Lin WJ, Wang M, et al. Multiscale causal networks identify VGF as a key regulator of Alzheimer's disease. *Nat Commun.* **2020**;11:3942. doi:10.1038/s41467-020-17405-z
35. Hwang W, Chiu YF, Kuo MH, et al. Expression of neuroendocrine factor VGF in lung cancer cells confers resistance to EGFR kinase inhibitors and triggers epithelial-to-mesenchymal transition. *Cancer Res.* **2017**;77:3013–3026. doi:10.1158/0008-5472.CAN-16-3168
36. Yang LH, Lee RK, Kuo MH, et al. Neuronal survival factor VGF promotes chemoresistance and predicts poor prognosis in lung cancers with neuroendocrine feature. *Int J Cancer.* **2022**;151:1611–1625. doi:10.1002/ijc.34193
37. Alrawashdeh W, Jones R, Dumartin L, et al. Perineural invasion in pancreatic cancer: proteomic analysis and in vitro modelling. *Mol Oncol.* **2019**;13:1075–1091. doi:10.1002/1878-0261.12463
38. Wang X, Prager BC, Wu Q, et al. Reciprocal signaling between glioblastoma stem cells and differentiated tumor cells promotes malignant progression. *Cell Stem Cell.* **2018**;22:514–528. doi:10.1016/j.stem.2018.03.011
39. Bozelli JC Jr, Yune J, Aulakh SS, et al. Human diacylglycerol kinase epsilon N-terminal segment regulates the phosphatidylinositol cycle, controlling the rate but not the acyl chain composition of its lipid intermediates. *ACS Chem Biol.* **2022**;17:2495–2506. doi:10.1021/acscchembio.2c00387
40. Leach NT, Sun Y, Michaud S, et al. Disruption of diacylglycerol kinase delta (DGKD) associated with seizures in humans and mice. *Am J Hum Genet.* **2007**;80:792–799. doi:10.1086/513019
41. Mizuno S, Kado S, Goto K, Takahashi D, Sakane F. Diacylglycerol kinase zeta generates dipalmitoyl-phosphatidic acid species during neuroblastoma cell differentiation. *Biochem Biophys Res.* **2016**;8:352–359. doi:10.1016/j.bbrep.2016.10.004
42. Ma K, Zhang P, Xia Y, et al. A signature based on five immune-related genes to predict the survival and immune characteristics of neuroblastoma. *BMC Med Genomics.* **2022**;15:242. doi:10.1186/s12920-022-01400-y



**Pharmacogenomics and Personalized Medicine****Dovepress****Publish your work in this journal**

Pharmacogenomics and Personalized Medicine is an international, peer-reviewed, open access journal characterizing the influence of genotype on pharmacology leading to the development of personalized treatment programs and individualized drug selection for improved safety, efficacy and sustainability. This journal is indexed on the American Chemical Society's Chemical Abstracts Service (CAS). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/pharmacogenomics-and-personalized-medicine-journal>