NCAPG is a Potential Biomarker for Prognosis of Treatment in Children with High-risk Neuroblastoma

Meng Fu
  Children's Hospital of Fudan University

Wei Wu
  Children's Hospital of Fudan University

Yunxia Sun
  Zhejiang Chinese Medical University

Rui Zhao
  Children's Hospital of Fudan University

Hao Li
  Children's Hospital of Fudan University

Zhiping Li (✉ zpli@fudan.edu.cn)
  Children's Hospital of Fudan University   https://orcid.org/0000-0001-6194-023X

NCAPG is a Potential Biomarker for Prognosis of Treatment in Children with High-risk Neuroblastoma

Meng Fu[1], Wei Wu[1], Yunxia Sun[2], Rui Zhao[3], Hao Li[3]*, Zhiping Li[1]*

[1]Department of Clinical Pharmacy, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai, China.
[2]Academy of Chinese Medical Sciences, Zhejiang Chinese Medical University, Hangzhou, China.
[3]Department of Neurosurgery, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai, China.
Correspondence should be addressed to Zhiping Li (E-mail: zpli@fudan.edu.cn)

Abstract

Background

High-risk neuroblastoma, as a lethal pediatric tumor, has unfavorable clinical outcomes, and accounts for ~15% of childhood cancer-related mortality. However, the biomarkers for treatment prognosis have remained largely elusive. The present study sought to identify potential genes associated with the progression and prognosis of high-risk neuroblastoma.

Methods

The gene expression profiles of neuroblastoma were downloaded from the Gene Expression omnibus (GEO) database, and samples from GSE62564 and GSE16476 were applied as the training set and validation set, respectively. Bioinformatic analysis were performed by R. Subsequently, the biological functions of key gene were investigated by transwell, western blot, and flow cytometry analysis. miRNA that regulated the expression of key gene were predicted and the bind site were validated by luciferase reporter analysis and Q-PCR in children with high-risk neuroblastoma.

Results

Analysis shown that 120 differentially expressed genes (DEGs) in the blue module were identified to be significantly related to high-risk neuroblastoma by weighted correlation network analysis (WGCNA). After MCODE and Cytohubba algorithm analyses, 26 genes were recognized as hub genes. Subsequently, univariate COX and LASSO regression analyses identified a six-gene signature (NCAPG, UBE2C, MELK, CCNA2, KIF15, and BIRC5) and constructed a prognostic model. The expression of these six genes in high-risk neuroblastoma children was significantly higher than that in non-high-risk neuroblastoma children. These signature genes and the prognostic model demonstrated strong prognostic capability using Kaplan-Meier (KM) curves. Subsequently, NCAPG was selected as key gene and the results indicated that knockdown NCAPG suppressed the migration and invasion, increased the apoptosis rate and decreased the Bcl-2 level in neuroblastoma cells. In addition, the luciferase reporter analysis method showed that miR-495-3p can negatively regulate the expression of NCAPG. Q-PCR indicated that miR-495-3p highly expressed in children with high-risk neuroblastoma.

Conclusion

Our findings identified that miR-495-3p/ NCAPG could serve as prognostic biomarkers of high-risk neuroblastoma and may advance the understanding of the molecular pathogenesis.

Keywords: High-risk Neuroblastoma; Prognosis; Biomarker; NCAPG; MicroRNAs

Introduction

Neuroblastoma is a lethal pediatric tumor originating from neurons of the sympathetic nervous system that is responsible for approximately 8% of all childhood cancers and 15% of childhood malignant cancer mortality[1]. It arises in early childhood with a median age at diagnosis of 18 months[2]. A high frequency of metastatic disease at diagnosis, such as lymph node metastasis, bone metastasis, bone marrow metastasis, and liver metastasis, is a hallmark of neuroblastoma. Neuroblastoma is highly heterogeneous with diverse clinical presentations that include spontaneous regression, therapy-resistant progression and metastatic disease with a poor prognosis even after intensive multimodality treatment[3]. Given this wide range of outcomes, patients with neuroblastoma are risk-stratified based on a combination of tumor stage, age at the time of diagnosis, and pathology (histology, differentiation, amplification of MYCN, diploidy, and 11q aberration) to guide the selection of treatment strategies[4].

Risk-based treatment as a tailored therapy approach has effectively reduced therapy for low-risk patients to avoid long-term complications while augmenting and targeting treatments for high-risk patients to improve overall survival and quality of life[5, 6]. Despite these efforts, While the survival for patients with low-and intermediate-risk disease approaches 100%, the 5-year survival rate for high-risk NB patients is less than 50%. 60% of patients have high-risk NB, and the prognosis of treatment in such patients remains poor. Although Significant progress has been made to comprehend the molecular mechanisms involved in the etiology and pathogenesis of NB, the factors affecting the prognosis of patients with high-risk neuroblastoma have remained largely elusive.

The rapid technological breakthrough of microarray and sequencing technologies has provided an excellent tool and platform for cancer research[7, 8]. Gene expression profiles have been used to identify genes associated with the progression of neuroblastoma by screening differentially expressed genes[9]. However, the high degree of interconnectivity between genes and the specific factors affecting the prognosis of high-risk patients have been ignored. Weighted correlation network analysis (WGCNA) is an exploratory data tool widely used to find modules of highly correlated genes and to explore the correlation patterns between genes and clinical features.

Here, in order to effectively find prognostic related genes, we constructed a co-expression gene network based on gene expression data from neuroblastoma patients obtained from the Gene Expression Omnibus (GEO) database, and the correlation between modules and clinical traits was further analyzed. Pathway analyses were performed to investigate pathways significantly associated with high-risk neuroblastoma. The hub genes associated with high-risk neuroblastoma were screened using a protein-protein interaction (PPI) network. The hub genes affecting prognosis were identified by univariate Cox and least absolute shrinkage and selection operator (LASSO) regression analyses. The biological functions of uninvestigated hub genes have been further explored by transwell, western blot, and flow cytometry.

Meanwhile, miRNAs that regulated signature gene expression were predicted and measured in the blood circulation of children with high-risk neuroblastoma (Figure 1).
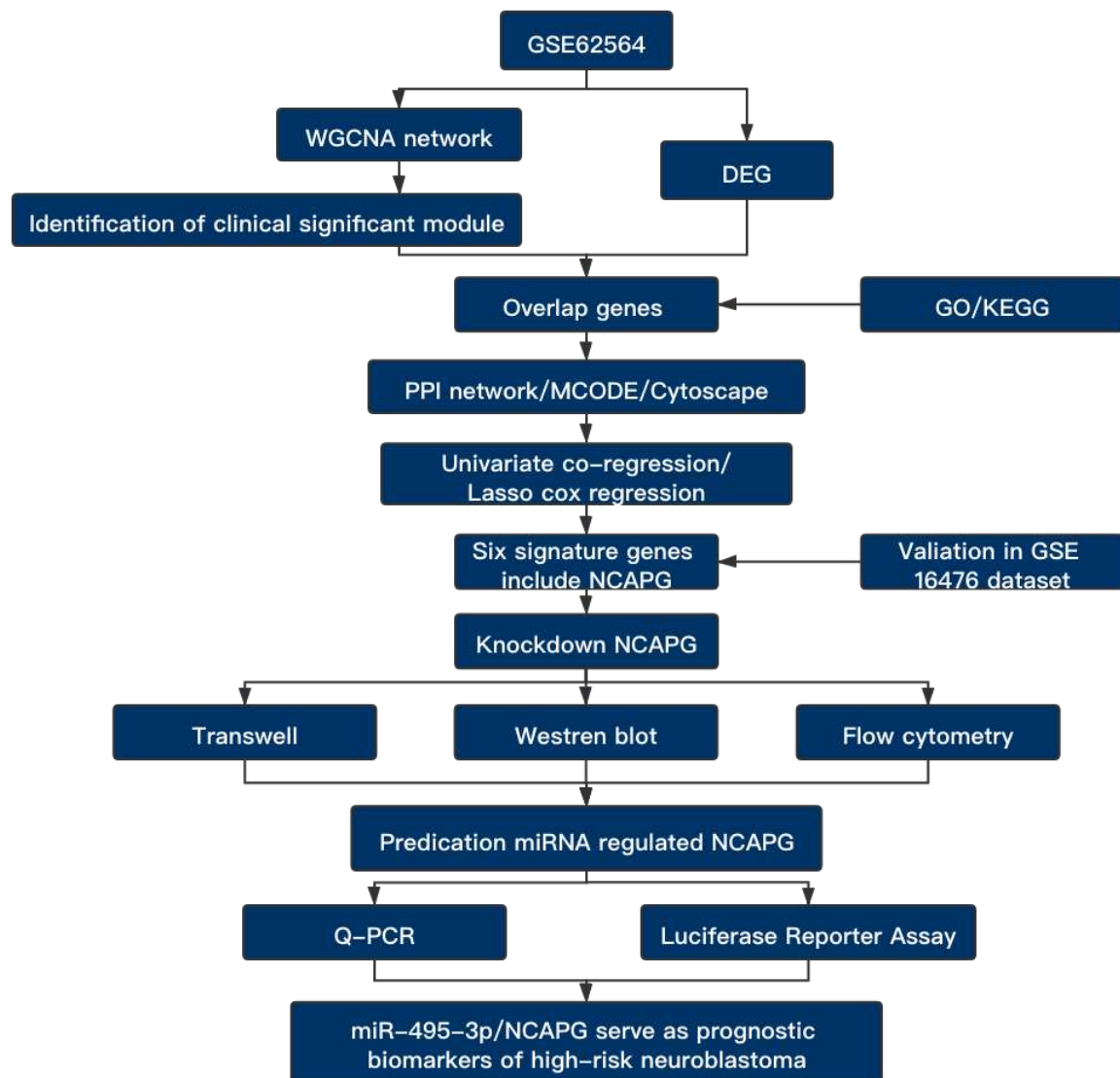

Figure 1: Overall flowchart of the study design.

Materials and Methods

Patient Samples

Blood samples from 36 subjects aged 0 to 18 years were collected from the Children's Hospital of Fudan University. Tumor extent was evaluated using CT and/or MRI. Tumor staging was determined according to the International Neuroblastoma Staging System (INSS)[10]. Patients whose clinical data were not available were excluded from the analysis. According to criteria published by the International Neuroblastoma Risk Group (INRG), 24 children were diagnosed with high-risk neuroblastoma (HR-NB), and 12 children were diagnosed with non-high-risk neuroblastoma (NHR-NB) (Table S1). Approximately 2 ml of peripheral blood from all patients was collected. Then, the serum samples were obtained by centrifugation at 1200 ×g and stored at -80°C. Hemolytic serum samples were excluded. The protocol was approved by the Institutional Ethics Committee of the Children's Hospital of Fudan University. The

procedures were carried out according to the ethical principles of the Declaration of Helsinki. Informed consent from each enrolled individual was obtained.

Cell Lines and Small interfering RNA (siRNA) transfection

Human neuroblastoma SH-SY5Y cell line was purchased from ATCC. The cells were cultured in 1640 medium (Thermo Fisher Scientific) containing 15% fetal bovine serum (FBS, (Thermo Fisher Scientific)) and 100 U/mL penicillin/ streptomycin ((Thermo Fisher Scientific) at 37 °C in a humidifed atmosphere with 5% CO2. SH-SY5Y (5×10^4 cells/well) were reverse transfected with 100 nM of each of the siRNAs using Lipofectamine RNAiMAX reagent (Invitrogen) in 24 wells plate at the indicated concentrations according to the manufacturer's instructions. At 6 h post transfection, the culture medium was replaced with fresh medium containing 15% FBS. The cells were harvested after 24 h of transfection for the following assays. The siRNA sequences were as follows: Si-h-NCAPG: sense, 5'-GGAGUUC AUUCAUUACCUUT T -3' and antisense, 5'-AAGGU AAUGAAUGAACUCCTT-3'.

Data Acquisition and Screening

The GSE62545 and GSE16476 datasets were retrieved from the GEO (www.ncbi.nlm.nih.gov/geo/). In this study, patients with unclear risk classification or older than 18 years were excluded. A total of 419 samples (GSE62564) and 76 samples (GSE16476) were included for subsequent analysis (Table S1). Then, to ensure the practicability and accuracy of the prognostic genes for neuroblastoma, samples from GSE62564 and GSE16476 were applied as the training set and validation set, respectively. For RNAs with multiple probes, the mean expression values were determined as the final gene expression value of the gene.

Construction of Dynamic Weighted Gene Co-expression Network

The WGCNA was performed based on previous protocols using the "WGCNA" R package to explore the correlation between the clinical features and gene modules[11, 12]. All the included samples of GSE62564 were analyzed, and outliers in the clustering results were removed. The revised expression profiles containing 422 samples were used to test the degree of independence and average connectivity of the modules by the gradient test. The appropriate power value was determined when the scale independence value was equal to 0.9. The WGCNA algorithm was then used to construct the co-expression modules and extract the gene information in each module. In the process, a module size=30 was set as the cut-off threshold, and modules with similar expression profiles were merged.

Clinically Significant Relevant Module Identification and Hub Gene Screening

After obtaining the gene modules, the correlation between the clinical traits and the module eigengenes was evaluated by Pearson's correlation test. The module with the highest correlation coefficient and a p value < 0.05 for high-risk clinical features was selected for subsequent analysis. The gene significance (GS) and module membership (MM) were evaluated. Based on the criteria of MM>0.8 and GS>0.2, the significant gene sets in the selected module were screened. Meanwhile, DEGs between HR-NB and NHR-NB were obtained using the "limma" R package with a cut-off value of |log2fold change (FC)| > 1 and

an adjusted P value < 0.05. The hub genes were identified by determining the intersection of the DEGs and the significant gene sets obtained from the WGCNA analysis.

Protein-Protein Interaction Network Construction and Analysis

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; http://string-db.org/) database was used to analyze the interaction of hub genes and construct a PPI network[13]. The PPI network was visualized by Cytoscape[14], and the MCODE plugin was used to select the hub clustered subnetworks of highly intraconnected nodes with the default parameters (degree cut-off ≥ 2, node score cut-off ≥ 0.2, K-core ≥ 2, and max depth = 100). Furthermore, the plugin CytoHubba was utilized to explore the significant nodes in the clustered hub.

Construction of the Gene-Related Prognostic Model

To determine the survival associated genes, univariate Cox regression analysis was performed by using the "survival" and "survminer" R packages. Genes for which P < 0.01 were selected as survival-related genes. LASSO analysis of the survival-related genes was performed to select the most valuable prognostic markers. The risk score (RS) of each sample was calculated based on a linear combination of the coefficient (β) multiplied by its expression level. KM curve analysis and receiving operating characteristic (ROC) analysis were further conducted to investigate the relationship between the RS and overall survival in GSE62564, and the prognostic model was validated based on GSE16476.

Transswell Assay

Transwell assay was used to detect the capacity of invasion and migration. Briefly, 5 × 104 in 200 μL serum-free DMEM were reseeded into the top of the insert of a Boyden chamber (Corning Inc., USA) with 300 μg/mL Matrigel (BD Bioscience, USA). The culture medium with 15% FBS was used as the chemical attractant and added to the well below. After 12h (migration) or 48h (invasion) incubation, cells that passed through the filter were fixed with 0.1% paraformaldehyde, and stained with 0.1% crystal violet solution. The cells were imaged at 200× magnification in six random fields, and measured using the ImageJ software.

Flow Cytometry

Cells were harvested by trypsinization and suspended in Annexin V binding solution. Then, Annexin V-FITC staining solution was added to the cell suspension and gently mixed at 4°C for 15 minutes in the dark, after which 5 mL propidium iodide (PI) staining solution was added to the cell suspension. Subsequently, the cell-apoptosis rate was detected using flow cytometry (BD FACSCalibur, USA).

Western blot analysis

After cell lysates were prepared, the protein concentration was measured using the Enhanced BCA Protein Assay Kit (Beyotime). Protein samples were separated by 10% SDS-PAGE electrophoresis at 120V for 80min and transferred onto PVDF membranes at 80V for 60min. The membranes were blocked by 5% nonfat milk and incubated overnight at 4 °C with one of the following primary antibodies: anti-Bcl-2 (1:1000, Abcam, UK), anti-NCAPG (1:1000, Abcam, UK), anti-beta-actin antibody (1:1000, Abcam, UK). The next day, membranes were incubated with the appropriate secondary antibody Goat Anti-Rabbit IgG H&L (HRP) for 1

hour at room temperature. Signals were detected using ECL Western Blotting Substrate (Beyotime, China), and protein bands were quantified by measuring the signal intensity using Image J.

Prediction and Detection miRNAs that Regulate Signature Gene Expression

To further identify the miRNAs that regulate signature gene expression, the prediction analysis was conducted with the starBase database (http://starbase.sysu.edu.cn/). The intersecting miRNAs were obtained by determining the intersection of the signature gene-related miRNA set. Afterwards, the expression of overlapping miRNAs was analyzed using RT-PCR. Total RNA was extracted using TRIzol reagent (TaKaRa, China) according to the manufacturer's protocol. Reverse transcription reactions were performed using a FastQuant RT Kit (Tiangen, China). The primers were designed and then synthesized by Sangon Biotechnology Ltd. (China). RT-PCR was performed according to the instructions of the PCR Kit (Tiangen, China). U6 was used as an internal reference. The fold changes were calculated by means of relative quantification (2-ΔΔCt method).

Luciferase Reporter Assay

The predicted binding site of the miRNA in the target gene was verified using a luciferase reporter gene assay performed in HEK-293T. HEK-293T cells in 24-well plates and after 24 h incubation the confuence reaches to 60–70%. NCAPG-3′-UTR wt and NCAPG-3′-UTR mutant reporter plasmids were constructed in advanced. Then, miRNA mimics and a mimic negative control (NC) were cotransfected with the luciferase reporter vector into human embryonic kidney (HEK)-293 T cells. A dual luciferase assay system (Promega, USA) was used to measure luciferase activity.

Statistical Analysis

The R-studio platform (v. 3.6.2) was used for the WGCNA analysis, differential expression analysis, GO enrichment analysis, KEGG pathway analysis, univariate Cox regression, LASSO regression, K-M curve and ROC curve generation and Pearson's correlation analysis. An adjusted P value or a P value < 0.05 indicated a statistically significant difference.

Results

Identification of Key Modules Associated with High-risk Neuroblastomas

After data quality evaluation and preprocessing, an expression matrix with 419 samples and corresponding clinical information from GSE62564 was used for gene co-expression network construction (Figure S1A). To ensure the network scale-free topology distribution, a soft threshold power (β) equal to 7 (R2=0.9) was selected for further analysis (Figure 2A, B). After similar modules were merged, a total of 19 distinct gene co-expression modules were identified based on the average linkage hierarchical clustering and β values (Figure S1B and 2C). Each module was designated by unique colors to distinguish the modules, and the genes in the gray module were excluded from further analysis. The relationships between the modules and clinical traits were assessed (Figure 2D). The blue module was found to be significantly positively correlated with high-risk neuroblastomas and was considered to be a key module for subsequent research.
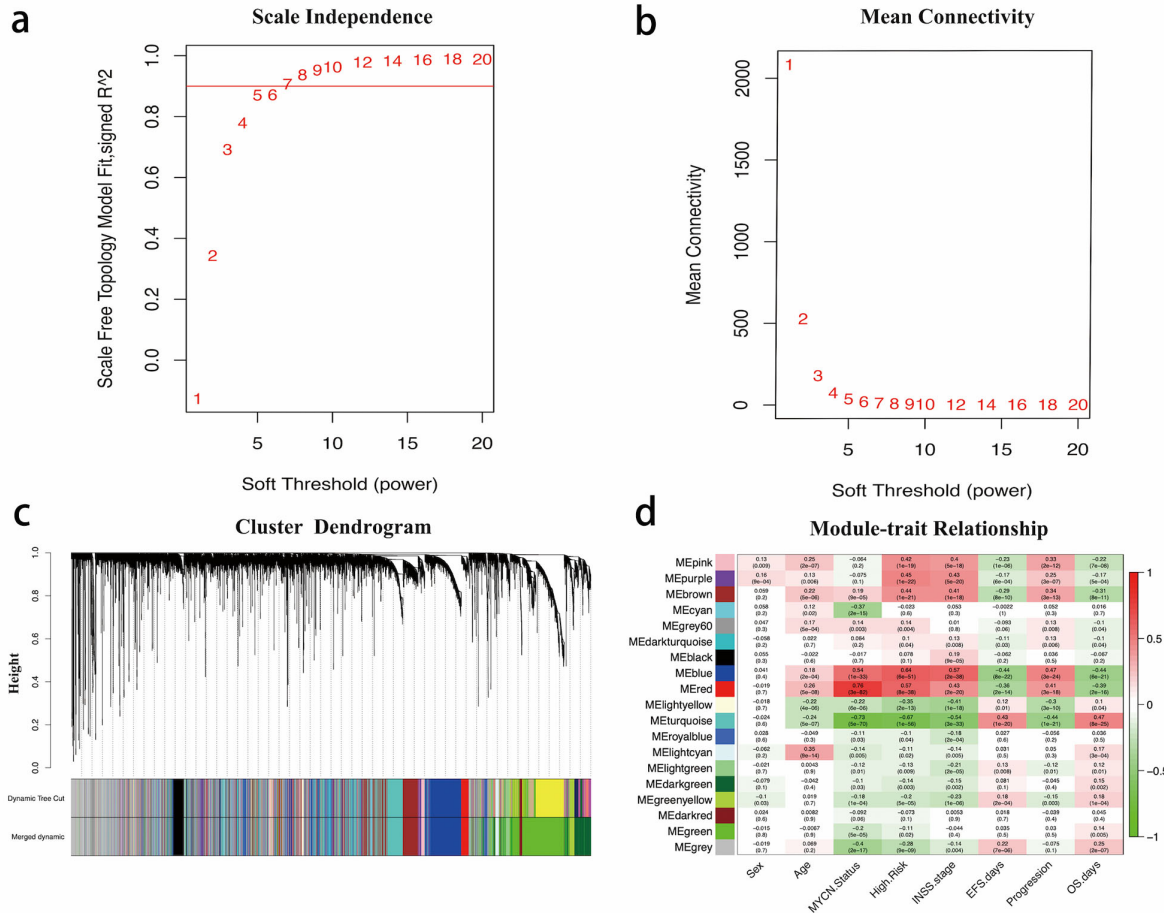
Figure 2: Construction and identification of modules associated with clinical traits. (A) Analysis of the scale-free fit index for various soft-thresholding powers (β). (B) The effect of β values on the mean connectivity of the co-expression network. (C) Identification of co-expressed gene modules in neuroblastoma patients. (D) Heatmap plot of the correlation between module eigengenes and clinical traits. Red indicates a positive correlation, and green indicates a negative correlation.

Identification of the Hub Gene Set

The genes that met the criteria of MM>0.8 and GS>0.2 in the blue module were screened (Figure 3A). Meanwhile, the DEGs between HR-NB and NHR-NB were obtained by differential expression analysis (Figure 3B). A total of 120 genes were obtained by taking the intersection of the DEGs and significant genes in the blue module (Figure 3C). The PPI network was constructed using the STRING database (Figure S2). After calculation with the MCODE algorithm, 89 genes in the cluster with the highest score were selected to better characterize the blue module (Figure 3D). To further identify the vital candidate genes associated with HR-NB, the topological features of the cluster were described according to the following four parameters: MCC, degree, betweenness and closeness. After determining the intersection of the top 30 genes according to the four algorithms, 26 genes were recognized as hub genes for subsequent analysis (Figure 3E, Table S2).
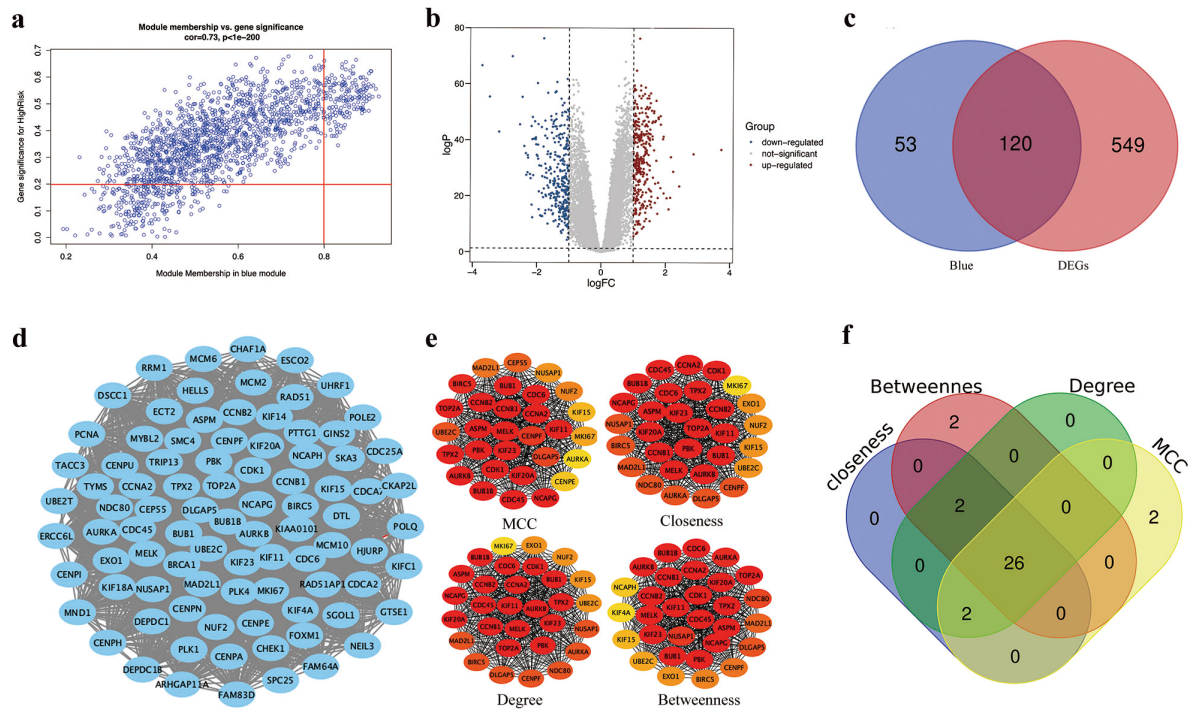
Figure 3. Identification of hub genes based on the training set. (A) A scatterplot showing the correlation between Gene Significance (GS) and Module Membership (MM) in the blue module; a gene was considered a significant gene in this module if MM > 0.8 and GS >2. (B) Volcano plot of the significance of the gene expression differences between high-risk and non-high-risk patients; the cut-offs were |log(FC)| > 1 and a p-value < 0.05. (C) Venn diagram analysis of the significant gene set in the blue module and the evidently differentially expressed gene set. (D) The highest scoring cluster in the protein-protein network of overlapping genes according to the MCODE algorithm. (E) The top 30 genes in the highest scoring cluster based on the MCC, degree, betweenness and closeness algorithms. (F) Venn diagram of the four gene sets.

Functional Annotation

The 120 intersecting genes in the DEG sets and the significant genes in the blue module were subjected to GO function and KEGG pathway analyses to examine their potential function. The biological process (BP) analysis indicated that 120 genes were particularly enriched in chromosome segregation, DNA replication, and nuclear division. The cellular component (CC) of these genes mainly included chromosomal regions, condensed chromosomes, and spindles. The molecular function (MF) were chiefly involved in catalytic activity, ATPase activity, and DNA helicase activity (Figure 4A). According to KEGG pathway analysis, the significant pathways included cell cycle, p53 signaling pathway, and DNA replication (Figure 4B).
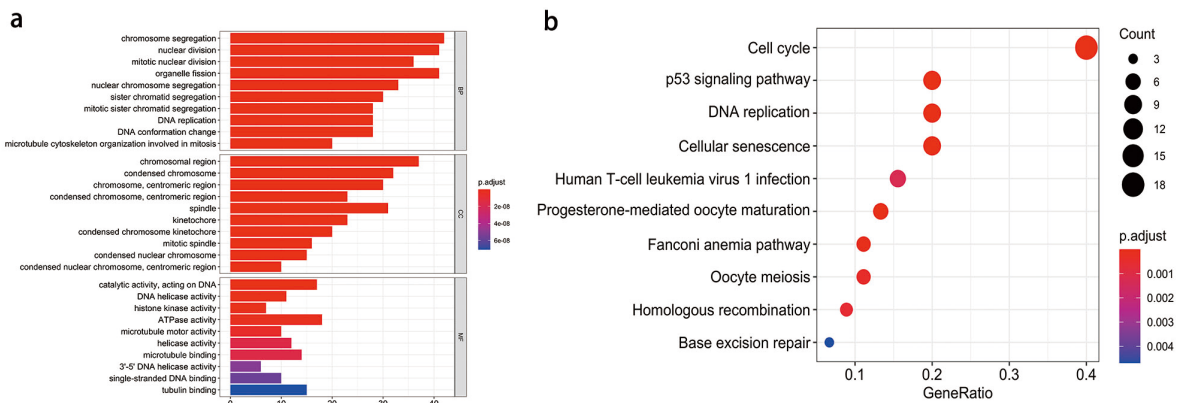

8

Figure 4. Functional enrichment analysis of the 120 intersecting genes. (A) GO enrichment analysis results. (B) KEGG pathway enrichment analysis results.

Construction of the Prognostic Model

After univariate Cox regression analysis, 22 genes were identified as survival related genes according to the criterion of P < 0.01 (Table S3). In order to further search for genes related to the survival of NBL patients, LASSO logistic regression was performed, and six genes (NCAPG, UBE2C, MELK, CCNA2, KIF15, and BIRC5) were identified as the final prognostic model (Figure 5A, B). Subsequently, the expression levels of six genes in NBL patients were verified. The results shown that the level of these six genes were outstandingly increased in HR-NB compared to NHR-NB (Figure 5C). Kaplan-Meier analysis showed that the high expression of these six genes was Closely related with a shorter overall survival time in neuroblastoma patients (Figure 5D). Based on the RS calculated by the prognostic model formula, patients were divided into a high-RS group and a low-RS group using the median RS as the cutoff (Figure 5E). The KM curves revealed that the high-RS group showed a poorer overall survival rate compared to the low-RS group (Figure 5F-H).
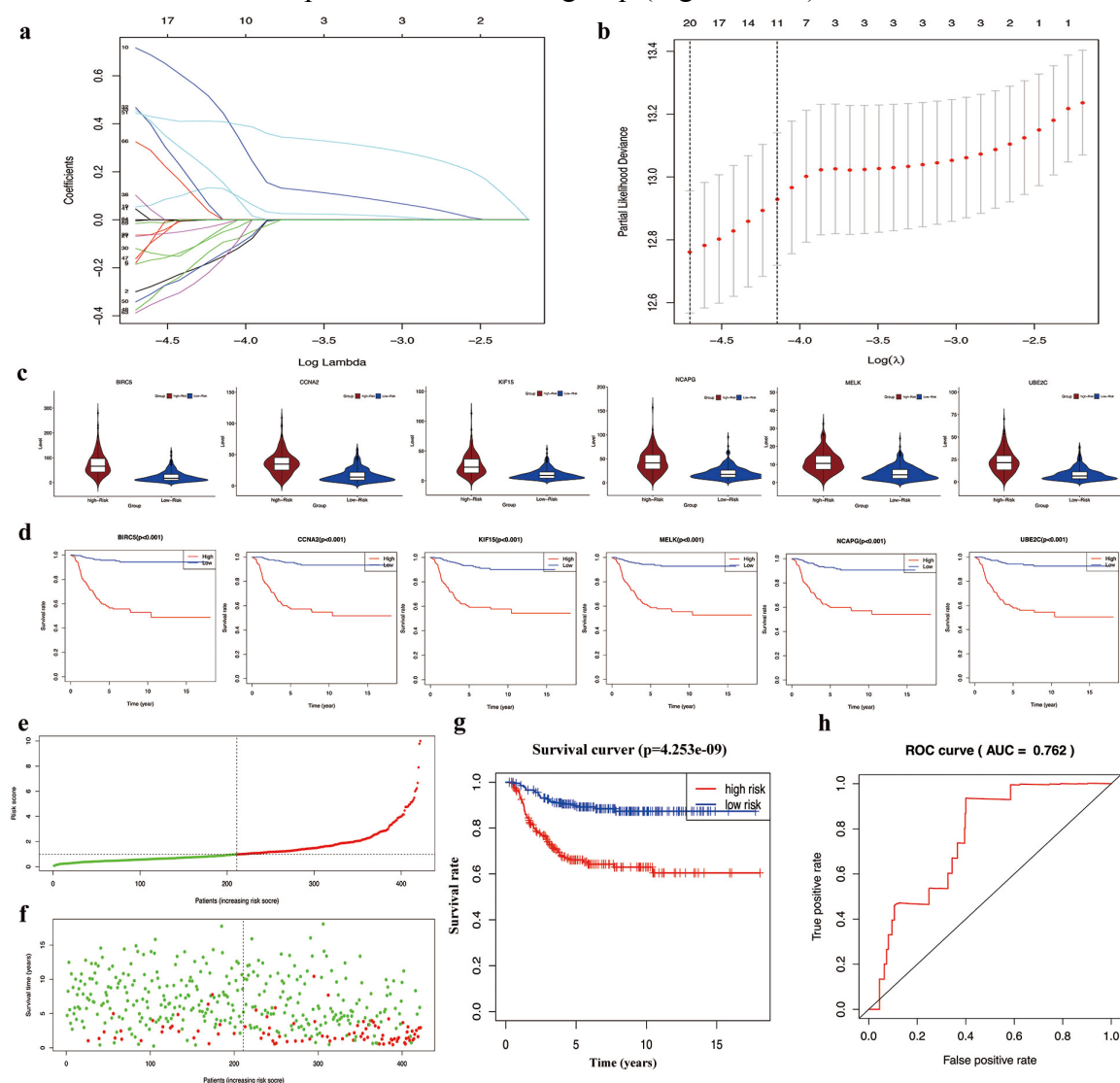

Figure 5. Construction of the prognostic model of high-risk neuroblastoma based on GSE62564. (A) Least absolute shrinkage and selection operator (LASSO) coefficient profiles of the fractions of 22 genes. (B) Tenfold cross-validation of the tuning parameter selection in the LASSO model. (C) Boxplot of the significance of the gene expression levels of 6 signature genes. (D) Overall survival analysis of 6 signature genes. (E) Risk score

distribution of neuroblastoma patients. (F) Survival status of neuroblastoma patients. (G). Survival analysis of the association between the risk score and the overall survival time of neuroblastoma patients. (H) Receiver operating characteristic (ROC) curves and area under the curve (AUC) statistics used to evaluate the predictive power of the prognostic model.

Verification of the Prognostic Model

The above findings are verified in GSE16476. The results indicated that the six signature genes were significantly higher in patients with HR-NB compared to patients with NHR-NB (Figure 6A). The expression level of the six signature genes was negatively associated with overall survival (Figure 6B). In addition, KM analysis showed that the survival time of patients in the high-RS group was shorter than that of patients in the low-RS group (Figure 6C). This result suggests that NCAPG upregulation of NCAPG, UBE2C, MELK, CCNA2, KIF15, and BIRC5 were effectively predictor of the poor prognosis of NBL patients.
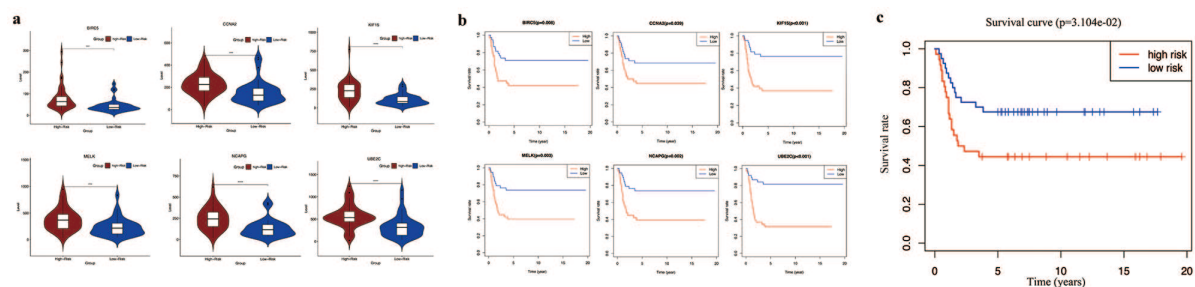

Figure 6. Verification of the prognostic model by using GSE16476. (A) Boxplot of the significance of the gene expression levels of 6 signature genes. (B) Overall survival analysis of 6 signature genes. (C) Kaplan–Meier curves for overall survival.

Knockdown of NCAPG Suppressed NBL cell migration and invasion

Among the six genes mentioned above, except for NCAPG, the others have been proven to play an important role in NBL. Therefore, this study focused on the effect of NCAPG on NBL. To investigate the biological function of NCAPG in NBL, the NCAPG siRNA intervention is implemented in NBL cell lines SH-SY5Y. Westren blot analisis indicated that NCAPG siRNA intervention knockdown the protein level of NCAPG (Figure 7A). The transwell assays were conducted to evaluate the effects of NCAPG on migration and invasion. As shown in Figure 7B, C, downregulation of NCAPG significantly inhibited the migration and invasion of SH-SY5Y cells.
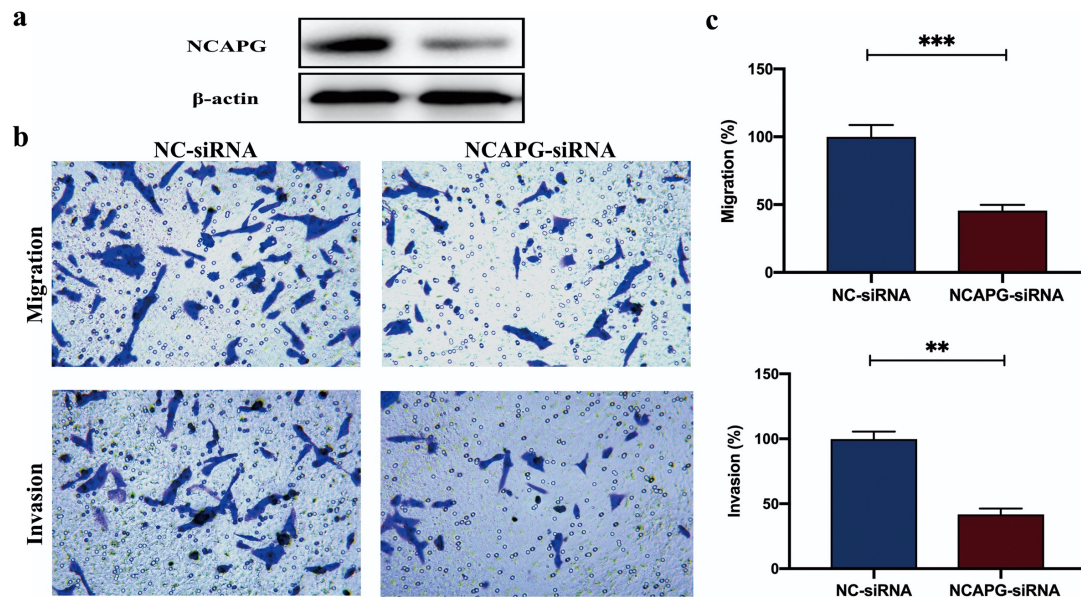
Figure 7. Transwell analysis of SH-SY5Y cells. Downregulation of NCAPG significantly inhibited the migration and invasion of SH-SY5Y cells.

Knockdown of NCAPG promote NBL cell apoptosis

The effects of NCAPG on NBL cell apoptosis were explored by flow cytometry and western blot. Flow cytometry analysis was performed after FITC Annexin V/ PI staining, which showed NCAPG siRNA increased the degree of apoptosis of SH-SY5Y cells（Figure 8A, B）. In addition, the expression level of Bcl-2, an apoptosis-related protein, is detected by the WB method. The results that NCAPG siRNA transfection greatly decreased expression of Bcl-2 (Figure 8C). These data indicate that knockdown of NCAPG can effectively reduce apoptosis in NBL cells.
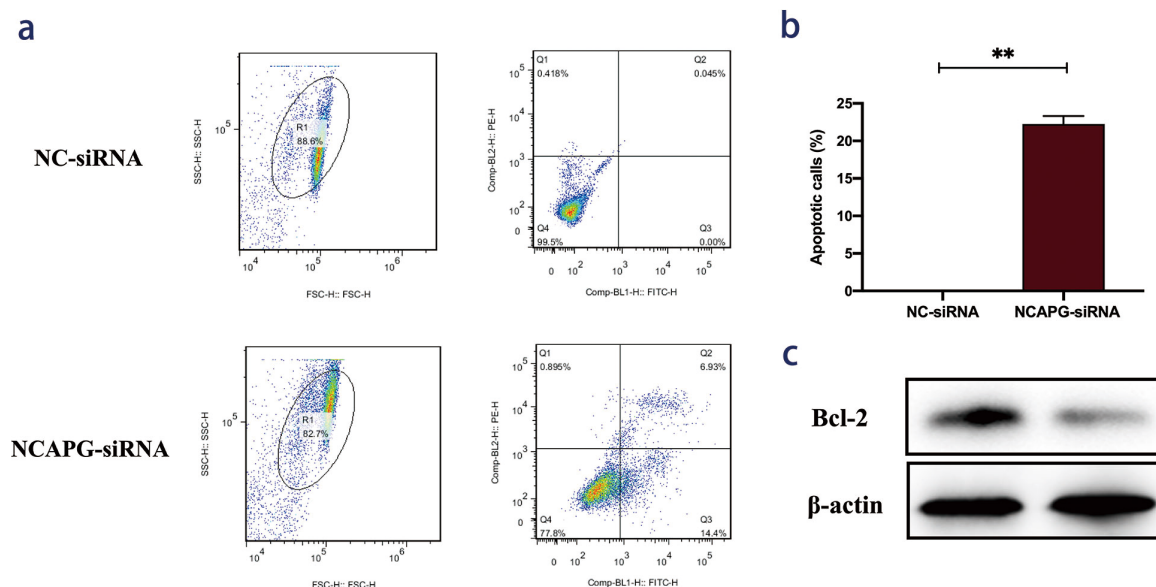

Figure 8. Flow cytometry analysis and western blot of SH-SY5Y cells. Downregulation of NCAPG significantly induced the apoptosis rate of SH-SY5Y cells (A, B). Downregulation of NCAPG remarkly decreased expression of Bcl-2.

11

Expression of Signature Gene-related miRNAs in High-risk Neuroblastoma

In order to find prognostic biomarkers in blood circulation, miRNAs that affect the expression of signature genes were predicted. Four overlapping miRNAs (miR-381-3p, miR-641, miR-300, and miR-495-3p) were obtained by determining the intersection of six gene-related miRNA sets (Figure 9A and Table S4). The circulating levels of these four miRNAs in children with HS-NB and NHS-NB were detected. As Figure 9B~E shows, the expression of miR-495-3p in children with HS-NB were evidently lower than those in children with NHS-NB. There was no significant difference in the miR-300, miR-641 and miR-381-3p expression levels between HS-NB patients and NHS-NB patients. The predicted binding site of the miR-495-3p and NCAPG was verified using luciferase reporter gene assay. The results showed that the luciferase activity of NCAPG-Wt in the miR-495-3p mimic group was greatly decreased compared with NC group. In NCAPG-Mut cells, there was no obvious difference between the miR-495-3p mimic group and the NC group (Figure 10). This result illustrated that the down-regulated miR-495-3p in NHS-NB may lead to up-regulation of its negative regulatory target genes NCAPG.
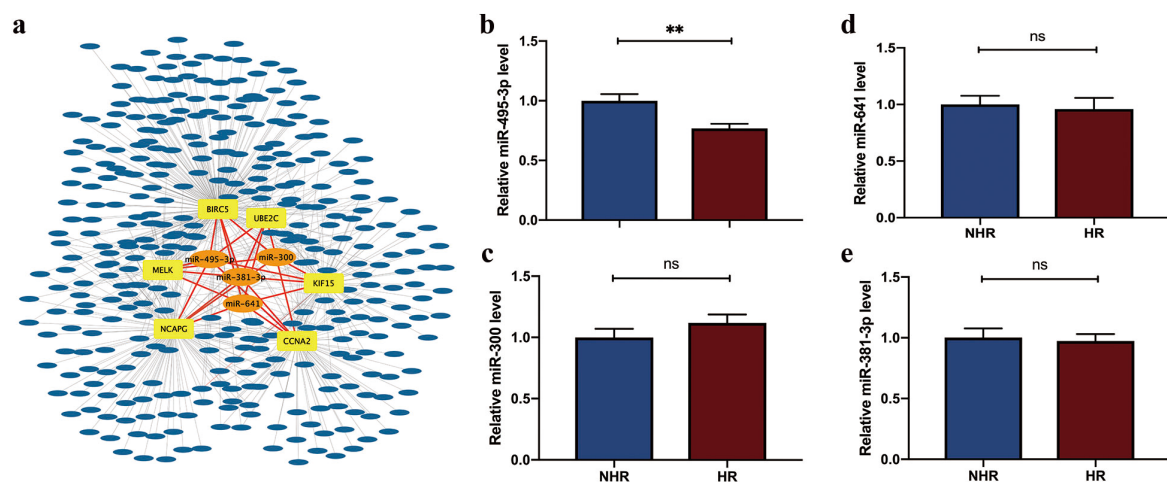

Figure 9. Expression levels of miRNAs that regulated six signature genes expression. MiRNAs prediction results that regulate six mRNA expression (A). The expression of miR-495-3p (B), miR-300 (C), miR-641 (D), and miR-381-3p (E) in children with HS-NB and NHS-NB.
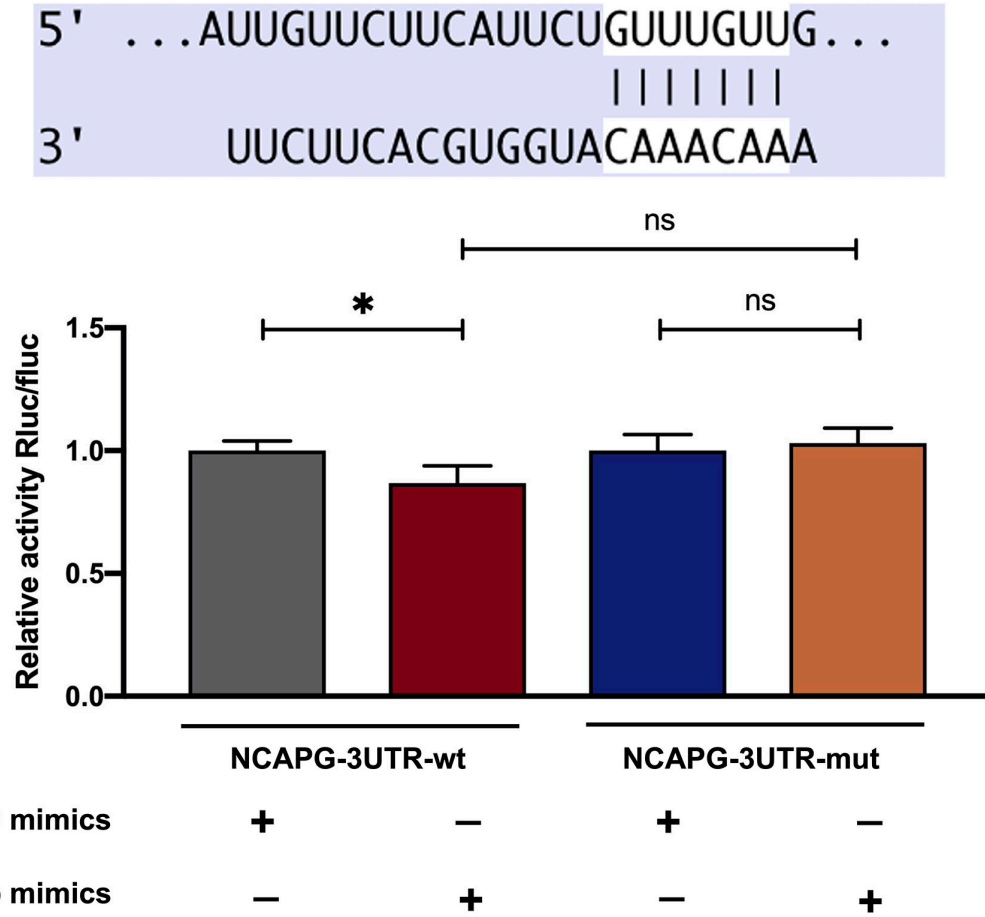
Figure 10. The binding sites of miR-495-3p in the 3′-UTR of NCAPG and the relative fluorescence unit of co-transfected cells detected by dual luciferase reporter gene assay.

Discussion

High-risk neuroblastoma as a fatal neoplasm has a higher incidence among children[1]. Current high-risk treatment regimens consist of multiagent induction chemotherapy, surgical tumor resection, autologous stem-cell rescue as a consolidation treatment, posttransplant radiotherapy, and postconsolidation treatment with biological agents and immunotherapy to treat minimal residual disease[3, 15, 16], which has improved 5-year overall survival to 50%[6, 16, 17]. Albeit comprehensive therapy and advances in the molecular exploration of neuroblastoma, current treatments are ineffective in approximately half of children with high-risk neuroblastoma, and further major advances in the development of prognostic of treatments are imperative[18].

In this study, prognostic-related genes in high-risk NB are explored. The GSE62564 dataset, with 419 neuroblastoma samples, was employed to obtain robust results. The WGCNA network of neuroblastoma was constructed, and the blue module was identified to be closely related to the clinical features of high-risk patients. Meanwhile, the DEGs of patients with HR-NB and patients with NHR-NB were obtained. Subsequently, 120 overlapping genes were obtained by determining the intersection of the blue module and the DEGs. To further screen genes related to the prognosis of HR-NB, the PPI network of 120 intersecting genes were built. After calculation by using the MCODE and CytoHubba algorithms, 26 hub genes were identified. Through univariate regression and LASSO logistic regression, the prognostic model was generated, which consisted of NCAPG, MELK, CCNA2, UBE2C, KIF15, and BIRC5. To

explore the relevance of the prognostic models to survival, survival analyses were performed, and the results indicated that the six signature genes have powerful predictive capacity for prognosis of HR-NB. Furthermore, the results were validated in the GSE16476 dataset.

According to the above six genes, most of them have been proved to be related to the prognosis of neuroblastoma. MELK is maternal embryonic leucine zipper kinase. It is involved in many tumor-related processes, such as tumor invasion, chemotherapeutic drug resistance, and stem cell renewal, by regulating the inflammatory pathway and affecting tumor mitotic activity[19-21]. Studies have shown that MELK is overexpressed in NB patients and is a potential therapeutic target for NB patients [22, 23]. CCNA2 is an important member of the highly conserved cyclin family and plays critical roles in the regulation of the cell cycle[24]. Overexpression of CCNA2 has been demonstrated in neuroblastoma patients and MYCN-amplified neuroblastoma cells. UBE2C is involved in the ubiquitin-proteasome pathway, which regulates various cancer-related cellular processes [25-28]. Study indicated that UBE2C plays a vital role in mitosis and is a promising therapeutic target for neuroblastoma. KIF15 is involved in various cellular processes, including proliferation, apoptosis, differentiation and development, that lead to poor prognosis in tumor patients[29, 30]. Patients with MYCN-amplified high-risk neuroblastoma have single nucleotide polymorphisms in the KIF15 gene[31]. BIRC5 is a protein involved in the intrinsic apoptotic pathway and is stage-dependently overexpressed in tumor tissue. Multiple studies have shown that BIRC5 is gained in 49% of NB tumors, and is correlated with a poor prognosis of high-risk NB. In addition, studies have indicated that abnormal BIRC5 expression induced by MYCN in neuroblastoma could be a target for therapy[32-35]. NCAPG, as a mitotic-associated chromosomal condensing protein, could promote cell proliferation, invasion, and migration. NCAPG overexpression cause poor prognosis in patients with many kinds of tumors, including prostate cancer, pediatric glioma, renal cell carcinoma, hepatocellular carcinoma, multiple myeloma, and melanoma[36-41]. However, the correlation between the high expression of NCAPG and the prognosis of high-risk neuroblastoma has not been explored. Therefore, this study investigated the biological function of NCAPG in neuroblastoma.

In human neuroblastoma SH-SY5Y cell line, we evaluated the effect of NCAPG on neuroblastoma migration, invasion and apoptosis. The results confirmed that knockdown of NCAPG could significantly suppress SH-SY5Y cell migration and induced apoptosis and invasion. Infer from this, NCAPG has the potential to be a new biomarker for prognosis of treatment in neuroblastoma.

On this basis, to further search for biomarkers in the blood circulation of HR-NB patients, the miRNAs that negatively regulated the six signature genes were predicted. Four miRNAs were obtained, and the expression of miR-495-3p in the circulation in children with HR-NB were significantly downregulated compared with that in children with NHR-NB. Meanwhile, the luciferase reporter assay indicated that miR-495-3p negatively regulated the expression of NACPG. Studies have indicated that miR-495-3p participate in the proliferation, invasion, and migration of various cancer cells and affect tumor progression[42-44]. Our results demonstrated that miR-495-3p/NCAPG could serve as circulating biomarkers to predict the prognosis of HR-NB.

There are still some limitations in our work. Although the results have been well validated based on independent data sets, neuroblastoma cells and blood samples from children with HR-NB, we have not further validated our conclusions through the analysis of tumor tissues.

In-depth examination of the mechanism underlying how these biomarkers affect high-risk neuroblastoma needs further study.

Conclusions

In summary, we have presented a systematic and comprehensive analysis by using WGCNA and LASSO. We identified six-gene signature which was of great value to predict progression and poor prognosis of HR-NB. We further determined that inhibiting NCAPG in the six genes can effectively inhibit the migration of neuroblastoma cells and induce neuroblastoma cell invasion and apoptosis. Meanwhile, miR-495-3p negatively regulate the expression of NCAPG, and were down-regulated in the serum of HR-NB children. These findings may shed light on advanced understanding of the prognostic factors of high-risk neuroblastoma.

Abbreviations

WGCNA: weighted gene co-expression network analysis; GEO: Gene Expression omnibus; DEGs: differentially expressed genes; KM: Kaplan-Meier; PPI: protein-protein interaction; LASSO: least absolute shrinkage and selection operator; HR-NB: high-risk neuroblastoma; NHR-NB: non-high-risk neuroblastoma; ROC: receiving operating characteristic; RS: risk score; BP: biological process; CC: cellular component; MF: molecular function.

Acknowledgements

Not applicable.

Author's Contributions

ZL, HL, and MF designed the research. MF and WW performed the data analysis work. MF, YS, and RZ conducted transwell, western blot, RT-PCR, and flow cytometry experiments. MF, HL, and ZL wrote and revised the manuscript. All authors read and approved the final manuscript.

Funding

Availability of data and materials

We declare that the materials described in the manuscript will be freely avail- able to all scientists for non-commercial purposes.

Ethics approval and consent to participate

Ethics approval was approved by the Institutional Ethics Committee of the Children's Hospital of Fudan University and all patients consent to participate into this study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Reference

1. Matthay KK, Maris JM, Schleiermacher G, Nakagawara A, Mackall CL, Diller L, Weiss WA: **Neuroblastoma**. *Nat Rev Dis Primers* 2016, 2:16078.
2. London WB, Castleberry RP, Matthay KK, Look AT, Seeger RC, Shimada H, Thorner P, Brodeur G, Maris JM, Reynolds CP *et al*: **Evidence for an age cutoff greater than 365 days for neuroblastoma risk group stratification in the Children's Oncology Group**. *J Clin Oncol* 2005, 23:6459-6465.
3. Brodeur GM: **Spontaneous regression of neuroblastoma**. *Cell Tissue Res* 2018, 372:277-286.
4. Cohn SL, Pearson AD, London WB, Monclair T, Ambros PF, Brodeur GM, Faldum A, Hero B, Iehara T, Machin D *et al*: **The International Neuroblastoma Risk Group (INRG) classification system: an INRG Task Force report**. *J Clin Oncol* 2009, 27:289-297.
5. Tolbert VP, Matthay KK: **Neuroblastoma: clinical and biological approach to risk stratification and treatment**. *Cell Tissue Res* 2018, 372:195-209.
6. Smith V, Foster J: **High-Risk Neuroblastoma Treatment Review**. *Children* 2018, 5.
7. 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.
8. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA *et al*: **Mutational landscape and significance across 12 major cancer types**. *Nature* 2013, 502:333-339.
9. Zhong X, Liu Y, Liu H, Zhang Y, Wang L, Zhang H: **Identification of Potential Prognostic Genes for Neuroblastoma**. *Front Genet* 2018, 9:589.
10. Brodeur GM, Pritchard J, Berthold F, Carlsen NL, Castel V, Castelberry RP, De Bernardi B, Evans AE, Favrot M, Hedborg F *et al*: **Revisions of the international criteria for neuroblastoma diagnosis, staging and response to treatment**. *Prog Clin Biol Res* 1994, 385:363-369.
11. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis**. *BMC Bioinformatics* 2008, 9:559.
12. Langfelder P, Horvath S: **Fast R Functions for Robust Correlations and Hierarchical Clustering**. *J Stat Softw* 2012, 46.
13. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP *et al*: **STRING v10: protein-protein interaction networks, integrated over the tree of life**. *Nucleic Acids Res* 2015, 43:D447-452.
14. Su G, Morris JH, Demchak B, Bader GD: **Biological network exploration with Cytoscape 3**. *Curr Protoc Bioinformatics* 2014, 47:8.13.11-24.
15. Pinto NR, Applebaum MA, Volchenboum SL, Matthay KK, London WB, Ambros PF, Nakagawara A, Berthold F, Schleiermacher G, Park JR *et al*: **Advances in Risk Classification and Treatment Strategies for Neuroblastoma**. *J Clin Oncol* 2015, 33:3008-3017.
16. Almstedt E, Elgendy R, Hekmati N, Rosen E, Warn C, Olsen TK, Dyberg C, Doroszko M, Larsson I, Sundstrom A *et al*: **Integrative discovery of treatments for high-risk neuroblastoma**. *Nat Commun* 2020, 11:71.

17. Simon T, Berthold F, Borkhardt A, Kremens B, De Carolis B, Hero B: **Treatment and outcomes of patients with relapsed, high-risk neuroblastoma: results of German trials**. *Pediatr Blood Cancer* 2011, 56:578-583.

18. Tas ML, Reedijk AMJ, Karim-Kos HE, Kremer LCM, van de Ven CP, Dierselhuis MP, van Eijkelenburg NKA, van Grotel M, Kraal K, Peek AML *et al*: **Neuroblastoma between 1990 and 2014 in the Netherlands: Increased incidence and improved survival of high-risk neuroblastoma**. *Eur J Cancer* 2020, 124:47-55.

19. Giuliano CJ, Lin A, Smith JC, Palladino AC, Sheltzer JM: **MELK expression correlates with tumor mitotic activity but is not required for cancer growth**. *Elife* 2018, 7.

20. Janostiak R, Rauniyar N, Lam TT, Ou J, Zhu LJ, Green MR, Wajapeyee N: **MELK Promotes Melanoma Growth by Stimulating the NF-κB Pathway**. *Cell Rep* 2017, 21:2829-2841.

21. Li B, Yan J, Phyu T, Fan S, Chung TH, Mustafa N, Lin B, Wang L, Eichhorn PJA, Goh BC *et al*: **MELK mediates the stability of EZH2 through site-specific phosphorylation in extranodal natural killer/T-cell lymphoma**. *Blood* 2019, 134:2046-2058.

22. Pitner MK, Taliaferro JM, Dalby KN, Bartholomeusz C: **MELK: a potential novel therapeutic target for TNBC and other aggressive malignancies**. *Expert Opin Ther Targets* 2017, 21:849-859.

23. Chlenski A, Park C, Dobratic M, Salwen HR, Budke B, Park JH, Miller R, Applebaum MA, Wilkinson E, Nakamura Y *et al*: **Maternal Embryonic Leucine Zipper Kinase (MELK), a Potential Therapeutic Target for Neuroblastoma**. *Mol Cancer Ther* 2019, 18:507-516.

24. Ko E, Kim Y, Cho EY, Han J, Shim YM, Park J, Kim DH: **Synergistic effect of Bcl-2 and cyclin A2 on adverse recurrence-free survival in stage I non-small cell lung cancer**. *Ann Surg Oncol* 2013, 20:1005-1012.

25. Dastsooz H, Cereda M, Donna D, Oliviero S: **A Comprehensive Bioinformatics Analysis of UBE2C in Cancers**. *Int J Mol Sci* 2019, 20.

26. Guo J, Wu Y, Du J, Yang L, Chen W, Gong K, Dai J, Miao S, Jin D, Xi S: **Deregulation of UBE2C-mediated autophagy repression aggravates NSCLC progression**. *Oncogenesis* 2018, 7:49.

27. Liu G, Zhao J, Pan B, Ma G, Liu L: **UBE2C overexpression in melanoma and its essential role in G2/M transition**. *J Cancer* 2019, 10:2176-2184.

28. Guo J, Jin D, Wu Y, Yang L, Du J, Gong K, Chen W, Dai J, Miao S, Xi S: **The miR 495-UBE2C-ABCG2/ERCC1 axis reverses cisplatin resistance by downregulating drug resistance genes in cisplatin-resistant non-small cell lung cancer cells**. *EBioMedicine* 2018, 35:204-221.

29. Yu X, He X, Heindl LM, Song X, Fan J, Jia R: **KIF15 plays a role in promoting the tumorigenicity of melanoma**. *Exp Eye Res* 2019, 185:107598.

30. Wang J, Guo X, Xie C, Jiang J: **KIF15 promotes pancreatic cancer proliferation via the MEK-ERK signalling pathway**. *Br J Cancer* 2017, 117:245-255.

31. Hungate EA, Applebaum MA, Skol AD, Vaksman Z, Diamond M, McDaniel L, Volchenboum SL, Stranger BE, Maris JM, Diskin SJ *et al*: **Evaluation of Genetic Predisposition for MYCN-Amplified Neuroblastoma**. *J Natl Cancer Inst* 2017, 109.

32. Eckerle I, Muth D, Batzler J, Henrich KO, Lutz W, Fischer M, Witt O, Schwab M, Westermann F: **Regulation of BIRC5 and its isoform BIRC5-2B in neuroblastoma**. *Cancer Lett* 2009, 285:99-107.

33. Lamers F, van der Ploeg I, Schild L, Ebus ME, Koster J, Hansen BR, Koch T, Versteeg R, Caron HN, Molenaar JJ: **Knockdown of survivin (BIRC5) causes apoptosis in neuroblastoma via mitotic catastrophe**. *Endocr Relat Cancer* 2011, 18:657-668.

34. Hagenbuchner J, Kiechl-Kohlendorfer U, Obexer P, Ausserlechner MJ: **BIRC5/Survivin as a target for glycolysis inhibition in high-stage neuroblastoma**. *Oncogene* 2016, 35:2052-2061.

35. Wang B, Li X, Zhao G, Yan H, Dong P, Watari H, Sims M, Li W, Pfeffer LM, Guo Y *et al*: **miR-203 inhibits ovarian tumor metastasis by targeting BIRC5 and attenuating the TGFβ pathway**. *J Exp Clin Cancer Res* 2018, 37:235.

36. Eberlein A, Takasuga A, Setoguchi K, Pfuhl R, Flisikowski K, Fries R, Klopp N, Fürbass R, Weikard R, Kühn C: **Dissection of genetic factors modulating fetal growth in cattle indicates a substantial role of the non-SMC condensin I complex, subunit G (NCAPG) gene**. *Genetics* 2009, 183:951-964.

37. Goto Y, Kurozumi A, Arai T, Nohata N, Kojima S, Okato A, Kato M, Yamazaki K, Ishida Y, Naya Y *et al*: **Impact of novel miR-145-3p regulatory networks on survival in patients with castration-resistant prostate cancer**. *Br J Cancer* 2017, 117:409-420.

38. Liang ML, Hsieh TH, Ng KH, Tsai YN, Tsai CF, Chao ME, Liu DJ, Chu SS, Chen W, Liu YR *et al*: **Downregulation of miR-137 and miR-6500-3p promotes cell proliferation in pediatric high-grade gliomas**. *Oncotarget* 2016, 7:19723-19737.

39. Yamada Y, Arai T, Kojima S, Sugawara S, Kato M, Okato A, Yamazaki K, Naya Y, Ichikawa T, Seki N: **Regulation of antitumor miR-144-5p targets oncogenes: Direct regulation of syndecan-3 and its clinical significance**. *Cancer Sci* 2018, 109:2919-2936.

40. Cohen Y, Gutwein O, Garach-Jehoshua O, Bar-Haim A, Kornberg A: **The proliferation arrest of primary tumor cells out-of-niche is associated with widespread downregulation of mitotic and transcriptional genes**. *Hematology* 2014, 19:286-292.

41. Gong C, Ai J, Fan Y, Gao J, Liu W, Feng Q, Liao W, Wu L: **NCAPG Promotes The Proliferation Of Hepatocellular Carcinoma Through PI3K/AKT Signaling**. *Onco Targets Ther* 2019, 12:8537-8552.

42. Guan YX, Zhang MZ, Chen XZ, Zhang Q, Liu SZ, Zhang YL: **Lnc RNA SNHG20 participated in proliferation, invasion, and migration of breast cancer cells via miR-495**. *J Cell Biochem* 2018, 119:7971-7981.

43. Sun L, Liu L, Yang J, Li H, Zhang C: **SATB1 3'-UTR and lncRNA-UCA1 competitively bind to miR-495-3p and together regulate the proliferation and invasion of gastric cancer**. *J Cell Biochem* 2019, 120:6671-6682.

44. He Z, Dang J, Song A, Cui X, Ma Z, Zhang Z: **NEAT1 promotes colon cancer progression through sponging miR-495-3p and activating CDK6 in vitro and in vivo**. *J Cell Physiol* 2019, 234:19582-19591.
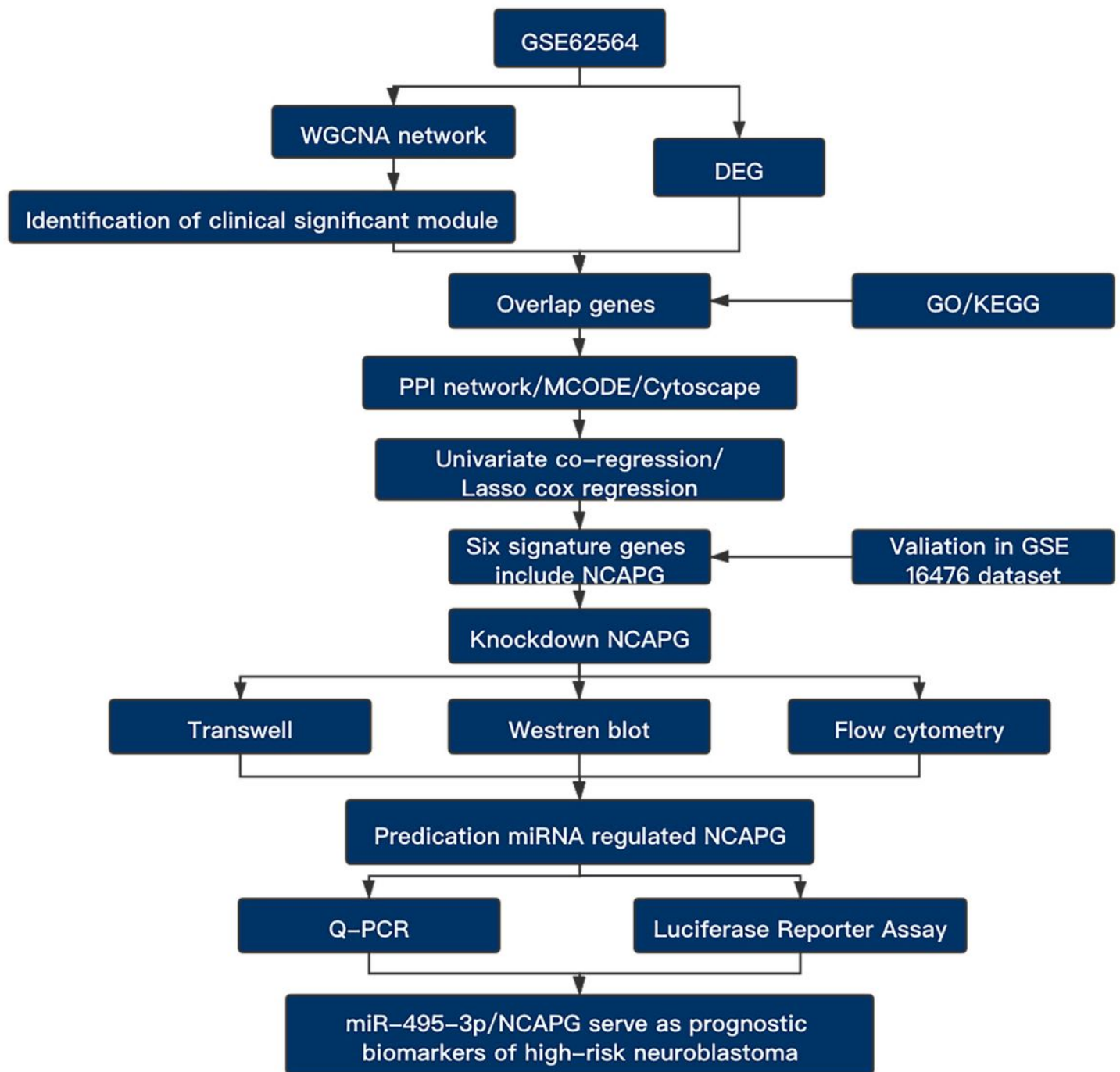
Figures

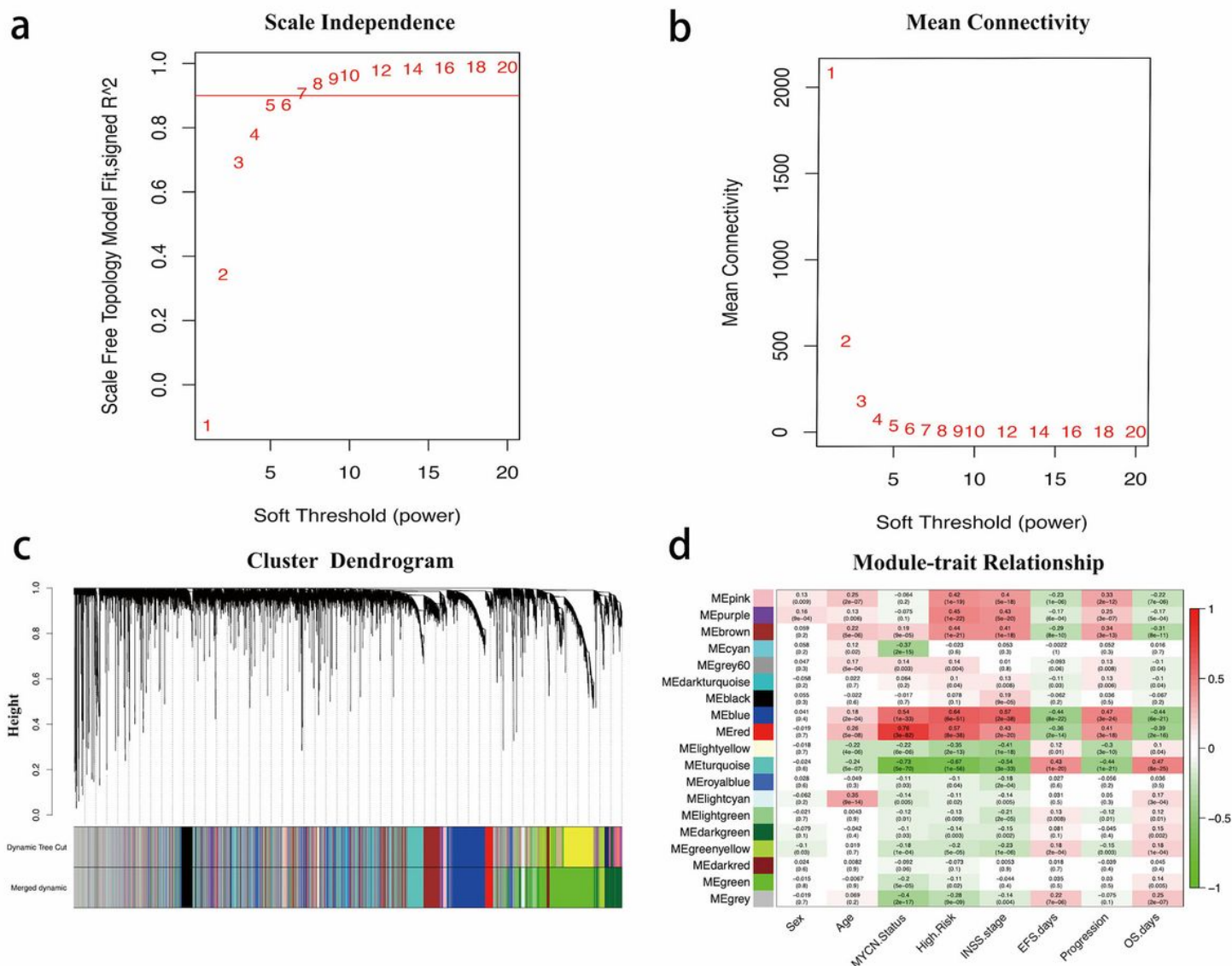**Figure 1**

Overall flowchart of the study design.

Figure 2

Construction and identification of modules associated with clinical traits. (A) Analysis of the scale-free fit index for various soft-thresholding powers (β). (B) The effect of β values on the mean connectivity of the co-expression network. (C) Identification of co-expressed gene modules in neuroblastoma patients. (D) Heatmap plot of the correlation between module eigengenes and clinical traits. Red indicates a positive correlation, and green indicates a negative correlation.

Figure 3

Identification of hub genes based on the training set. (A) A scatterplot showing the correlation between Gene Significance (GS) and Module Membership (MM) in the blue module; a gene was considered a significant gene in this module if MM > 0.8 and GS >2. (B) Volcano plot of the significance of the gene expression differences between high-risk and non-high-risk patients; the cut-offs were |log(FC)| > 1 and a p-value < 0.05. (C) Venn diagram analysis of the significant gene set in the blue module and the evidently differentially expressed gene set. (D) The highest scoring cluster in the protein-protein network of overlapping genes according to the MCODE algorithm. (E) The top 30 genes in the highest scoring cluster based on the MCC, degree, betweenness and closeness algorithms. (F) Venn diagram of the four gene sets.
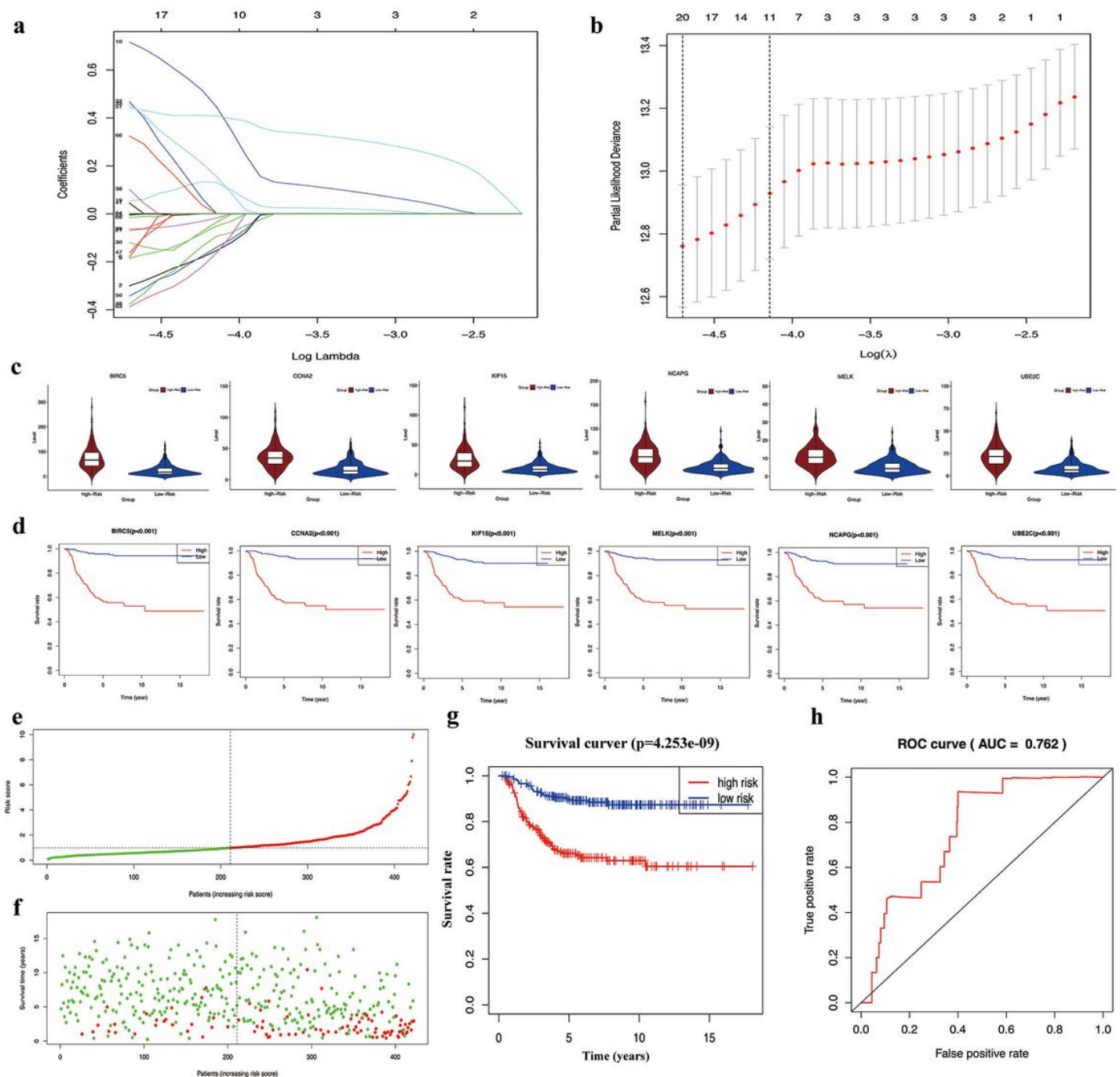
Figure 4

Functional enrichment analysis of the 120 intersecting genes. (A) GO enrichment analysis results. (B) KEGG pathway enrichment analysis results.

Figure 5

Construction of the prognostic model of high-risk neuroblastoma based on GSE62564. (A) Least absolute shrinkage and selection operator (LASSO) coefficient profiles of the fractions of 22 genes. (B) Tenfold cross-validation of the tuning parameter selection in the LASSO model. (C) Boxplot of the significance of the gene expression levels of 6 signature genes. (D) Overall survival analysis of 6 signature genes. (E) Risk score distribution of neuroblastoma patients. (F) Survival status of neuroblastoma patients. (G). Survival analysis of the association between the risk score and the overall survival time of neuroblastoma patients. (H) Receiver operating characteristic (ROC) curves and area under the curve (AUC) statistics used to evaluate the predictive power of the prognostic model.

Figure 6

Verification of the prognostic model by using GSE16476. (A) Boxplot of the significance of the gene expression levels of 6 signature genes. (B) Overall survival analysis of 6 signature genes. (C) Kaplan–Meier curves for overall survival.
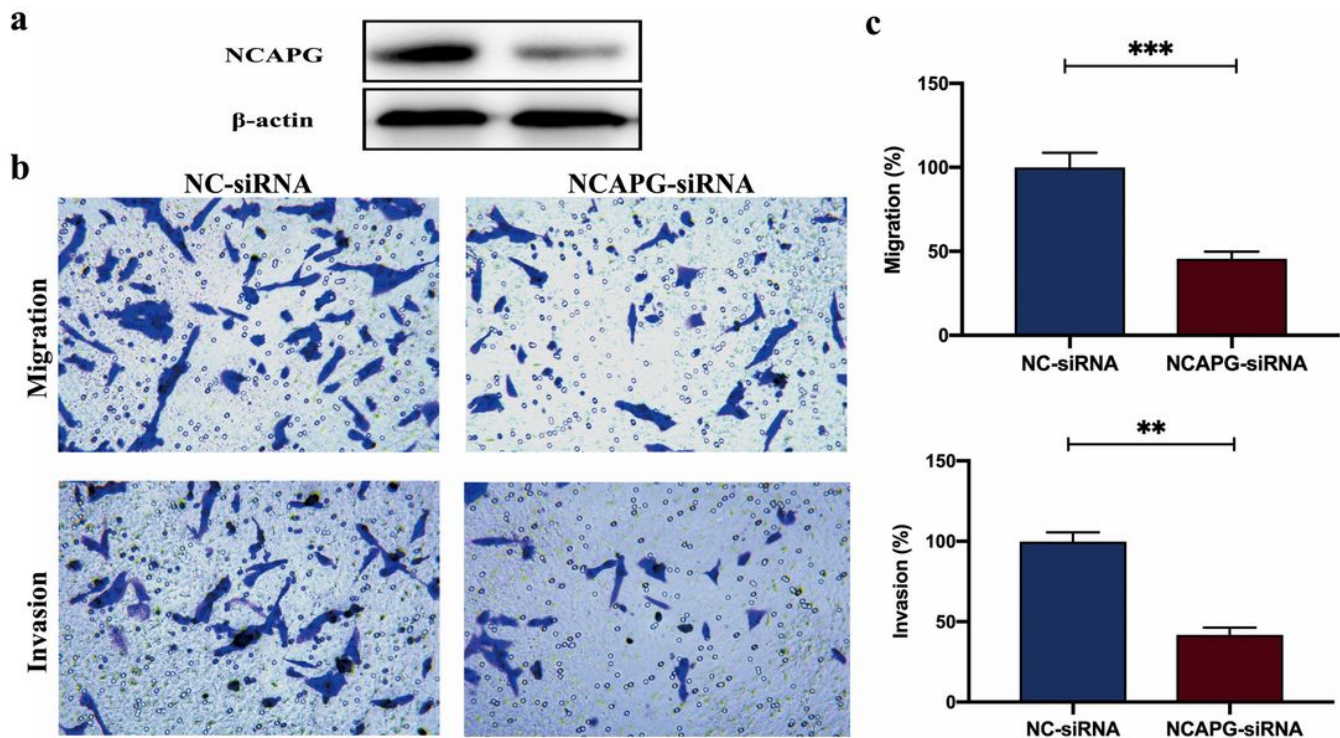

Figure 7

Transwell analysis of SH-SY5Y cells. Downregulation of NCAPG significantly inhibited the migration and invasion of SH-SY5Y cells.
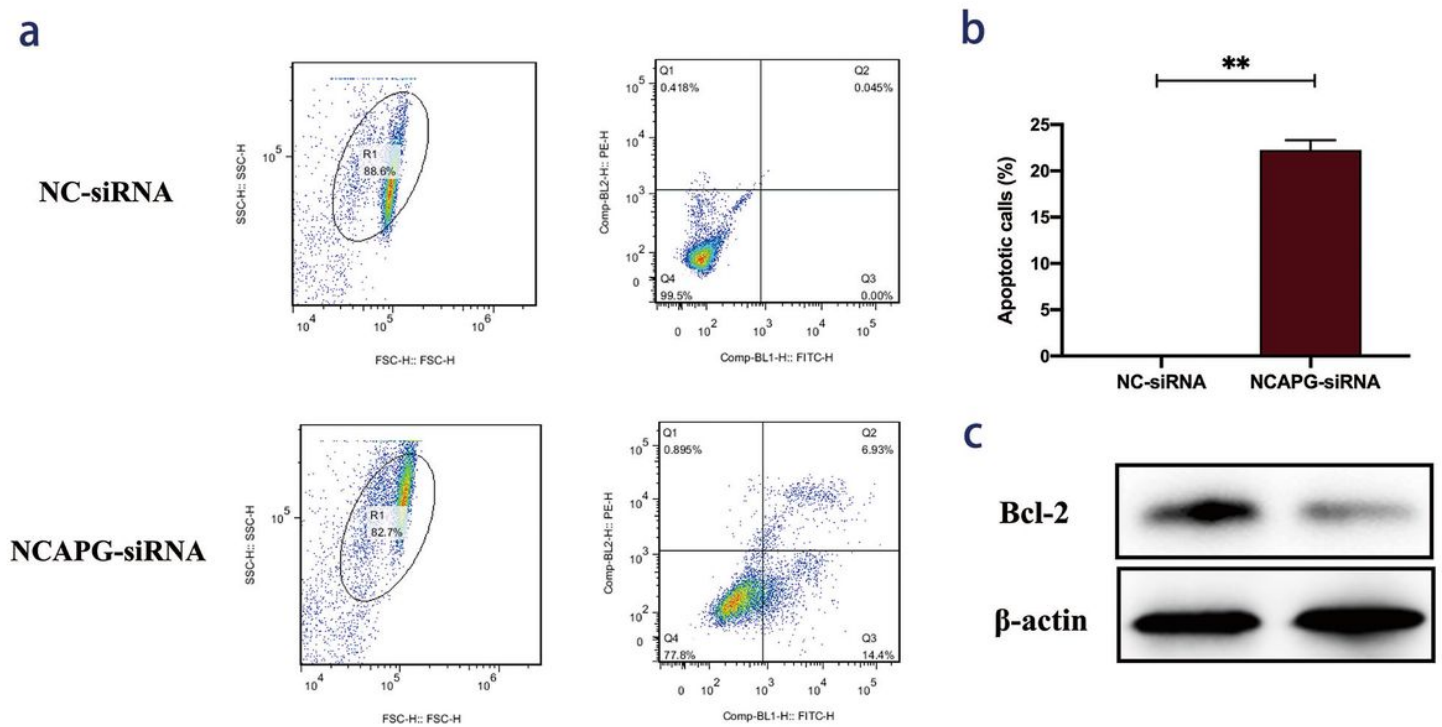
Figure 8

Flow cytometry analysis and western blot of SH-SY5Y cells. Downregulation of NCAPG significantly induced the apoptosis rate of SH-SY5Y cells (A, B). Downregulation of NCAPG remarkly decreased expression of Bcl-2.
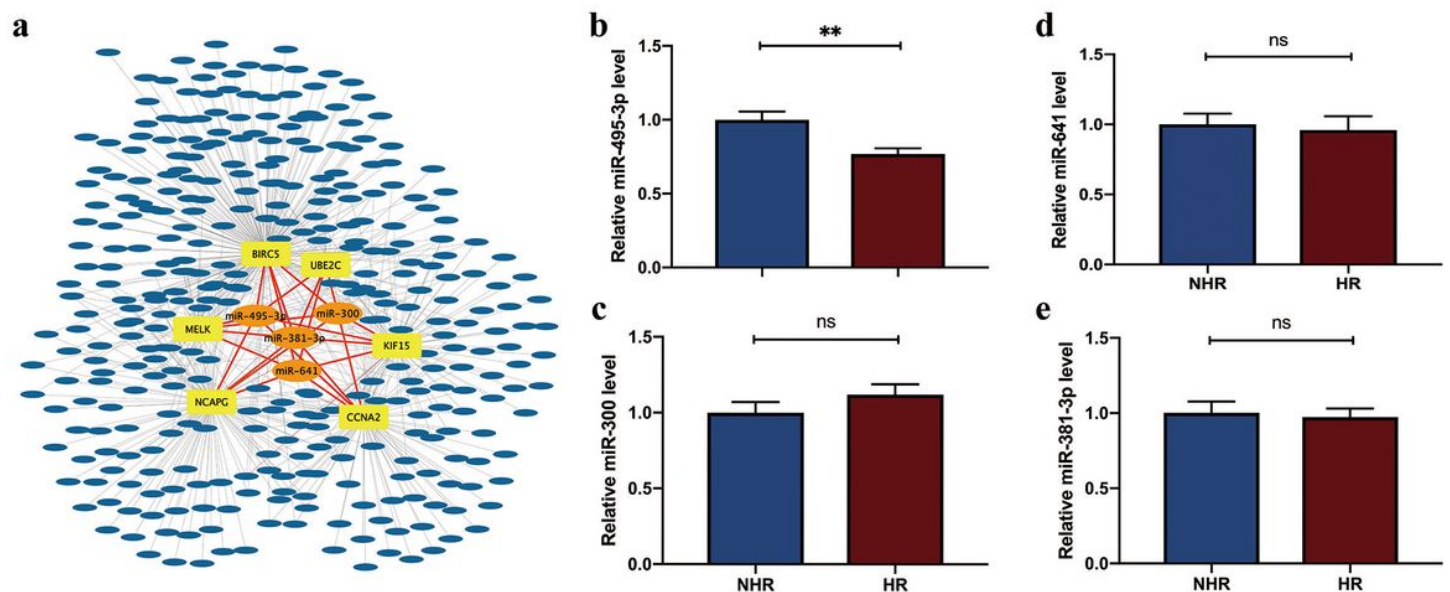

Figure 9

Expression levels of miRNAs that regulated six signature genes expression. MiRNAs prediction results that regulate six mRNA expression (A). The expression of miR-495-3p (B), miR-300 (C), miR-641 (D), and
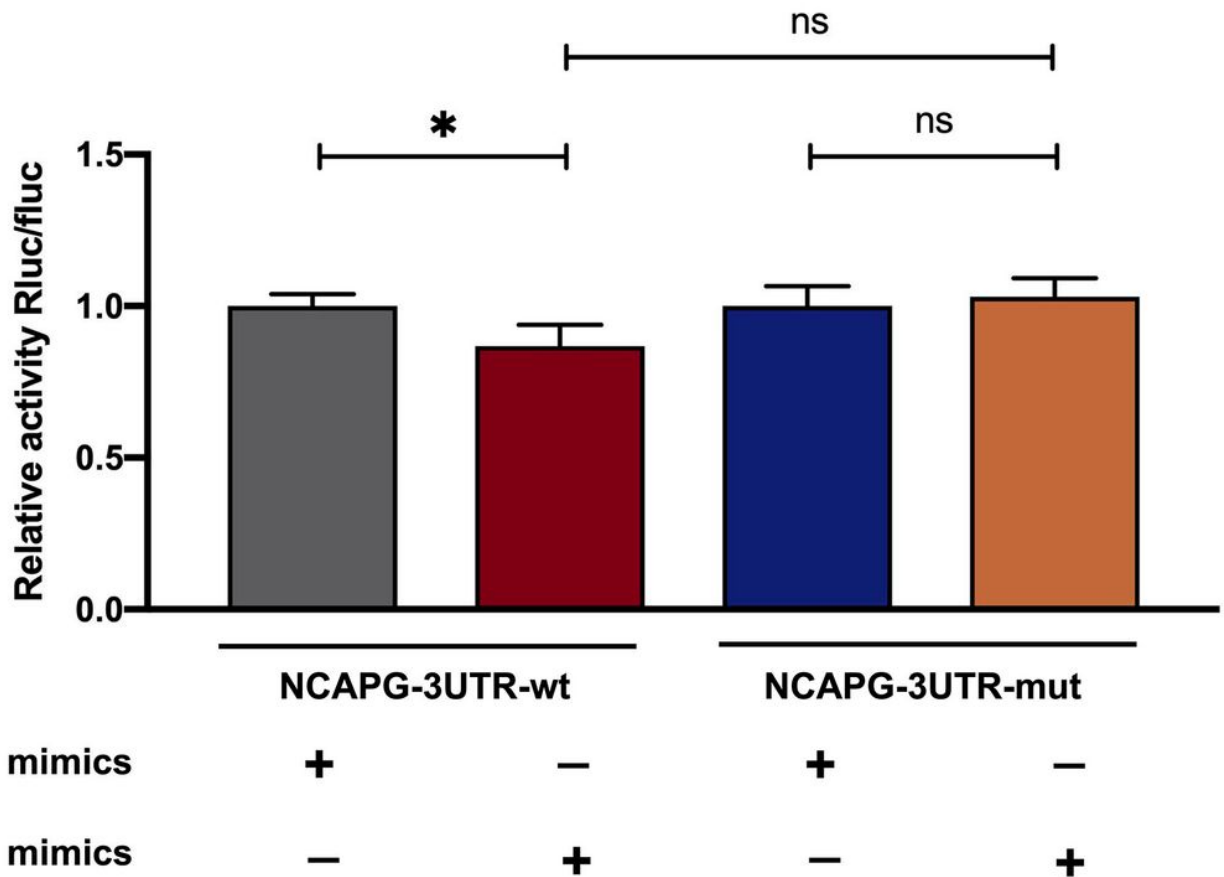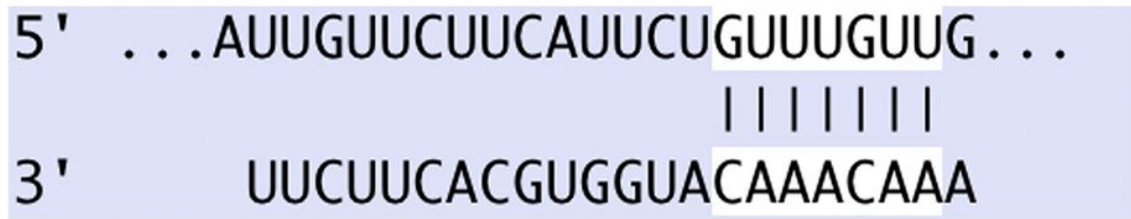
miR-381-3p (E) in children with HS-NB and NHS-NB.

Figure 10

The binding sites of miR-495-3p in the 3ꞋUTR of NCAPG and the relative fluorescence unit of co-transfected cells detected by dual luciferase reporter gene assay.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.jpg
- FigureS2.jpg
- TableS1.docx
- TableS2.docx

- TableS3.docx
- TableS4.xlsx