Bioinformatic analysis of the inner heterogeneity within MYCN non-amplified pediatric neuroblastoma sub-group

Haiwei Wang (hwwang@sibs.ac.cn)
Fujian Maternity and Child Health Hospital https://orcid.org/0000-0002-9675-4039

Xinrui Wang
Fujian Maternity and Child Health Hospital

Liangpu Xu
Fujian Maternity and Child Health Hospital

Ji Zhang
Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital

Hua Cao
Fujian Maternity and Child Health Hospital

Research article

Keywords: Pediatric neuroblastoma, MYCN, age, DST, TARGET, GEO

Posted Date: July 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-48689/v1

Abstract

Background: Pediatric neuroblastoma is divided into MYCN amplified and MYCN non-amplified sub-groups. However, the extent of heterogeneity within MYCN amplified or non-amplified pediatric neuroblastoma is unclear.

Methods: The prognostic significance of age and MYCN amplification was determined through multivariate cox regression and Kaplan-Meier survival analysis. MYCN non-amplified pediatric neuroblastoma patients were divided into different sub-consensuses using non-negative matrix factorization (NMF) based on the gene expression profiling. Genes particularly expressed in MYCN non-amplified young neuroblastoma patients were identified using Therapeutically Applicable Research to Generate Effective Treatments (TARGET) and Gene Expression Omnibus (GEO) datasets. The prognostic effects of ALCAM, CACNA2D3, DST, EPB41L4A and KIFIB in MYCN non-amplified pediatric neuroblastoma patients were determined by Kaplan-Meier survival.

Results: Age and MYCN amplification were independent prognostic factors in pediatric neuroblastoma. MYCN non-amplified pediatric neuroblastoma comprised young and old two distinct sub-groups. Compared with MYCN non-amplified old neuroblastoma patients, MYCN non-amplified young neuroblastoma patients had better clinical outcomes. MYCN non-amplified pediatric neuroblastoma was divided into three sub-consensuses through NMF assay and each sub-consensus was with significantly different clinical outcomes. However, MYCN amplified pediatric neuroblastoma was relatively homogeneous, and could not divide into sub-groups with different clinical outcomes by age or by NMF assay. ALCAM, CACNA2D3, DST, EPB41L4A and KIFIB were highly expressed in MYCN non-amplified young neuroblastoma patients. Moreover, the high expression levels of ALCAM, CACNA2D3, DST, EPB41L4A or KIFIB were associated with the favorable prognosis of MYCN non-amplified neuroblastoma patients. We also found that DST was an independent prognostic factor in MYCN non-amplified neuroblastoma patients and MYCN non-amplified neuroblastoma young patients with high DST expression levels had the best clinical overall survival.

Conclusions: MYCN non-amplified neuroblastoma was a heterogeneous disease and could be divided into sub-groups based on age or the expression levels of ALCAM, CACNA2D3, DST, EPB41L4A or KIFIB. MYCN non-amplified neuroblastoma young patients with high DST expression levels had the best clinical overall survival.

Background

Initiated from the sympathetic nervous system, pediatric neuroblastoma is a heterogeneous pediatric tumor with variable clinical outcomes [1, 2]. Based on the MYCN amplification status, International Neuroblastoma Risk Group suggests that neuroblastoma comprises at least two distinct groups [3]. About 25% neuroblastoma patients with MYCN amplification are classified into high risk sub-group and are correlated with adverse prognosis [4–6]. Moreover, neuroblastoma patients with high expression

levels of MYCN target genes [7] or MYCN signature [8] are also associated with unfavorable clinical outcomes. On the contrary, the other 75% neuroblastoma patients without MYCN amplification are classified into low risk sub-group and are often had better clinical outcomes [4]. Although this classification provides deep molecular understanding of neuroblastoma, the extent of inner heterogeneity within MYCN amplified or non-amplified sub-group is unknown. Whether the MYCN non-amplified neuroblastoma comprises other sub-consensuses with different prognosis is unclear.

Besides MYCN amplification, age at initial neuroblastoma diagnosis is also a prognostic factor [9]. Compared with young neuroblastoma patients, old neuroblastoma patients usually had worse prognostic outcomes because of the accumulating of somatic alterations with aging [10]. However, the relationship of age at initial neuroblastoma diagnosis and MYCN amplification in the predication of the clinical outcomes of neuroblastoma is unclear. Also, the transcriptional profiling associated with age and MYCN amplification is unknown.

Using the gene expression profiling to identify the molecular inner sub-groups of human cancer is widely accepted [11]. With the availability of published expression and clinical data in TARGET [12], GSE49710 [13] and GSE85047 datasets [14], we analyzed the intra-heterogeneity within MYCN amplified or non-amplified sub-group and determined the relationship of age and MYCN amplification in neuroblastoma patients. Also the transcriptional profiling in MYCN non-amplified young neuroblastoma patients was identified. Our analysis suggested that MYCN non-amplified neuroblastoma was heterogeneous and could be divided into sub-groups based on age at initial neuroblastoma diagnosis or the expression levels of DST. MYCN non-amplified young patients with high DST expression levels had the best clinical overall survival in neuroblastoma.

Methods

Data collection

The clinical files and mRNA-seq datasets of neuroblastoma in TARGET were downloaded from https://ocg.cancer.gov/. The clinical files and microarray expression datasets of neuroblastoma deposited in GEO datasets were downloaded from https://www.ncbi.nlm.nih.gov/geo/, including GSE49710 and GSE85047 datasets.

Prognostic effects of age and MYCN amplification

Neuroblastoma patients were divided into old and young sub-groups based on the mean age at initial diagnosis. Kaplan-Meier estimator in GraphPad Prism software (version 5.0; https://www.graphpad.com/) was applied to determine the different clinical overall survival of neuroblastoma patients with different age or MYCN amplification status using TARGET, GSE49710 and GSE85047 datasets. P values were determined by Log-rank test.

Data processing

FPKM derived from the mRNA-seq dataset was used to measure the gene expression levels in TARGET dataset. The GSE49710 and GSE85047 microarray expression datasets were processed using R software (version 3.5.0; https://www.r-project.org/). When multiple probes were corresponded to the same gene symbol, the averaged expression values were used. The differentially expressed genes in MYCN non-amplified young neuroblastoma patients were determined using Student's t test.

Heatmap

R software "pheatmap" package (version 1.0.12; https://cran.r-project.org/web/packages/pheatmap/index.html) were used to generate the heatmaps. The clustering scale and clustering distance were determined by "average" and 'correlation' methods, respectively.

Venn diagram

The common genes which were differentially expressed in MYCN non-amplified young neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets were generated using VENNY 2.1 software (http://bioinfogp.cnb.csic.es/tools/venny/index.html).

Gene ontology enrichment analysis

Gene ontology analysis was performed using The Database for Annotation, Visualization and Integrated Discovery (DAVID) website (version 6.8; https://david.ncifcrf.gov). Enrichment P-value < 0.05 was considered to be statistical significant.

The Nonnegative Matrix Factorization (NMF) classification

MYCN non-amplified neuroblastoma in TARGET, GSE49710 and GSE85047 datasets were divided into two sub-consensuses or three sub-consensuses using R software "NMF" package (version 0.22.0; https://cran.r-project.org/web/packages/NMF/index.html). 10 number of runs and rank=2:3 were performed to compute the object.

Prognostic effects of ALCAM, CACNA2D3, DST, EPB41L4A and KIFIB

Neuroblastoma patients were divided into high and low expression sub-groups based on the mean expression levels of genes. R software 'Survival' package (version 3.1-8; https://cran.r-project.org/web/packages/survival/ index.html) was used to determine the different clinical outcomes of neuroblastoma patients with high expression levels and low expression levels of ALCAM, CACNA2D3, DST, EPB41L4A and KIFIB. Log-rank test was used to test the P values.

Multivariate cox regression

R software 'survival' package 'coxph' method (version 3.1-8) was used for multivariate cox regression analysis. Log-rank test was used to calculate the P values.

Correlation plot

R software ‘corrplot’ package (version 0.84; https://cran.r-project.org/web/packages/corrplot/index.html) was used to determine the correlation of ALCAM, CACNA2D3, DST, EPB41L4A and KIFIB based on their expression levels in TARGET, GSE49710 and GSE85047 datasets. Spearman’s correlation test was used to determine the correlation coefficients.

Statistical analysis

The box plots were generated from GraphPad Prism software (version 5.0; https://www.graphpad.com/). Statistical analysis was performed using the Student’s t test in R software. P value less than 0.05 was chosen to be statistically significant difference.

Results

Age and MYCN amplification are independent prognostic factors in pediatric neuroblastoma.

Previously, we had shown that neuroblastoma patients with MYCN amplification were associated with adverse prognosis and revealed the prognostic significance of MYCN target genes using TARGET dataset [7]. In the present study, we further tested the prognostic effects of age at initial neuroblastoma diagnosis in TARGET dataset.

According to the TARGET clinical data, the mean age at initial neuroblastoma diagnosis was 3.2 years old. However, the age of neuroblastoma patients were varied significantly. 11.8% (72 out of 608) neuroblastoma patients were younger than one year old, while, 3.3% (20 out of 608) patients were older than eight years old. Previous reports suggested that age was a prognostic factor in neuroblastoma [9]. Consistently, we found that old neuroblastoma patients had worse prognosis than young pediatric neuroblastoma patients in TARGET dataset (Fig. 1a).

The prognostic effects of age in pediatric neuroblastoma were further confirmed using GSE49710 and GSE85047 datasets. The mean age at initial neuroblastoma diagnosis was 2.08 years old in GSE49710 dataset and 2.19 years old in GSE85047 dataset. Similar to the results derived from TARGET dataset, we found that old pediatric neuroblastoma patients had adverse clinical outcomes in both GSE49710 and GSE85047 datasets (Fig. 1a).

MYCN amplification was also associated with the clinical outcomes of pediatric neuroblastoma [5]. So, next, we determined the relationships of age and MYCN amplification in neuroblastoma. We found that there was no significant difference of the mean age in pediatric neuroblastoma patients with or without MYCN amplification in TARGET and GSE85047 datasets (Fig. 1b). Furthermore, the multivariate cox regression assay suggested that age and MYCN amplification were independent prognostic factors in pediatric neuroblastoma in TARGET, GSE49710 and GSE85047 datasets (Fig. 1c).

MYCN non-amplified pediatric neuroblastoma is a heterogeneous disease and MYCN non-amplified young neuroblastoma patients have better prognosis.

Next, we determined the combination of age and MYCN amplification in the predication of the clinical outcomes of pediatric neuroblastoma. Pediatric neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets were divided into four sub-groups based on the mean age and MYCN amplification status. The Kaplan-Meier plots showed that MYCN amplified old patients and MYCN amplified young patients had not different clinical outcomes (Fig. 2a). However, MYCN non-amplified pediatric neuroblastoma was divided into two different sub-groups. MYCN non-amplified young neuroblastoma patients had significant favorable prognosis than MYCN non-amplified old neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets (Fig. 2a). Those results suggested that MYCN non-amplified pediatric neuroblastoma was heterogeneous, and could be divided into old and young pediatric neuroblastoma sub-groups.

Moreover, we showed that all the MYCN amplified pediatric neuroblastoma patients and MYCN non-amplified old pediatric neuroblastoma patients were with histological unfavorable outcomes (Fig. 2b). However, only 55% MYCN non-amplified young pediatric neuroblastoma patients were with histological unfavorable outcomes in TARGET dataset (Fig. 2b). Similarly, 25% MYCN non-amplified young pediatric neuroblastoma patients were in low risk sub-group, while, only one MYCN non-amplified old pediatric neuroblastoma patients were in low risk sub-group in TARGET dataset (Fig. 2b). Those results further confirmed the existing of different sub-groups in MYCN non-amplified pediatric neuroblastoma patients.

Three sub-consensuses of MYCN non-amplified neuroblastoma patients are with different clinical overall survival.

Non-negative matrix factorization (NMF) sub-consensus is a robust cancer classification system based globe gene expression levels [15, 16]. To further address the inner sub-groups of MYCN non-amplified pediatric neuroblastoma patients, we divided the MYCN non-amplified neuroblastoma patients in TARGET dataset into two sub-consensuses (Fig. 3a) or three sub-consensuses (Fig. 3b) using NMF classification. We then tested the clinical overall survival of different sub-consensuses. When the MYCN non-amplified neuroblastoma patients were divided into two sub-consensuses, we found that there was no significant difference in the overall survival between sub-consensus 1 and sub-consensus 2 (Fig. 3a). When divided the MYCN non-amplified neuroblastoma patients into three sub-consensuses, we found that, compared with sub-consensus 1 or sub-consensus 2, MYCN non-amplified neuroblastoma patients in sub-consensus 3 were with significant low overall survival in TAEGET dataset (Fig. 3b).

The three sub-consensuses classification of MYCN non-amplified neuroblastoma was further validated in GSE49710 and GSE85047 datasets. Similarly, MYCN non-amplified neuroblastoma patients in GSE49710 and GSE85047 datasets were divided into three sub-consensuses using NMF (Fig. 3c and 3d). Although not significantly, MYCN non-amplified neuroblastoma patients in sub-consensus 3 were with low overall survival in GSE49710 dataset (P = 0.09) (Fig. 3c). In GSE85047 dataset, MYCN non-amplified

neuroblastoma patients in sub-consensus 1 were with favorable overall survival (Fig. 3d). All those results suggested there were three sub-consensuses in MYCN non-amplified neuroblastoma patients.

MYCN amplified pediatric neuroblastoma is a relatively homogeneous disease.

Previously, we had shown that age alone could not divide the MYCN amplified pediatric neuroblastoma patients into two different sub-groups with different clinical outcomes (Fig. 2a and 2b). Further, we studied the inner sub-groups of MYCN amplified pediatric neuroblastoma patients using NMF assay. MYCN amplified neuroblastoma patients in TARGET dataset were divided into two sub-consensuses (Fig. 4a). We found that, the overall survival between sub-consensus 1 and sub-consensus 2 was no significantly different (Fig. 4a). Even, in the three sub-consensuses classification, the overall survival of each sub-consensus was also not significantly different (Fig. 4b).

Next, we tried to determine whether there were some genes could divide the MYCN non-amplified or MYCN amplified pediatric neuroblastoma into two different sub-groups with different overall survival. To achieve this study, all the prognostic genes in MYCN non-amplified neuroblastoma patients in each TARGET, GSE49710 or GSE85047 dataset were identified, respectively. We found that 226 genes shared common prognostic significance in MYCN non-amplified neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets (Fig. 4c). On the contrary, there was only one gene CSNK1G2 had prognostic significance in MYCN amplified neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets (Fig. 4d). Those results further highlighted that MYCN non-amplified pediatric neuroblastoma was a heterogeneous disease, while, MYCN amplified pediatric neuroblastoma was a relatively homogeneous disease.

Using Kaplan-Meier plots, we demonstrated the prognostic effects of CSNK1G2 in MYCN amplified pediatric neuroblastoma. We found that MYCN amplified neuroblastoma patients with high expression levels of CSNK1G2 were with low overall survival in TARGET, GSE49710 and GSE85047 datasets (Fig. 4e).

Identification of the transcriptional profiling in MYCN non-amplified young neuroblastoma patients

Previously, we had shown that MYCN non-amplified pediatric neuroblastoma was included different sub-groups and MYCN non-amplified young neuroblastoma patients had better prognosis (Fig. 2a). Next, we tried to identify the differentially expressed genes in MYCN non-amplified young neuroblastoma patients in TARGET dataset. Compared with MYCN amplified neuroblastoma patients and MYCN non-amplified old neuroblastoma patients, 64 genes were highly expressed in MYCN non-amplified young neuroblastoma patients in TARGET dataset (Fig. 5a). However, there were only 12 genes were down-regulated in MYCN non-amplified young neuroblastoma patients (Fig. 5a). Those differentially expressed genes were involved in the regulation of neuron projection, regulation of GTPase activity and regulation of cell-cell adhesion (Fig. 5b).

The differentially expressed genes in MYCN non-amplified young neuroblastoma patients in GSE49710 and GSE85047 datasets were also identified. There were 2952 and 612 genes were differentially expressed in MYCN non-amplified young neuroblastoma patients GSE49710 and GSE85047 datasets, respectively (Fig. 5a). Among them, ALCAM, BTBD9, CACNA2D3, DST, EPB41L4A, FGD6, GMEB1, IGSF3 and KIFIB were commonly changed in MYCN non-amplified young neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets (Fig. 5c). Using Heatmaps, we showed that those genes were all highly expressed in MYCN non-amplified young neuroblastoma patients (Fig. 5d).

High expression levels of ALCAM, CACNA2D3, DST, EPB41L4A or KIFIB are associated with the favorable prognosis of MYCN non-amplified neuroblastoma patients.

Next, we determined the prognostic effects of ALCAM, BTBD9, CACNA2D3, DST, EPB41L4A, FGD6, GMEB1, IGSF3 and KIFIB in MYCN non-amplified neuroblastoma patients. Previously, we had identified 226 prognostic genes in MYCN non-amplified neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets (Fig. 4c). ALCAM, CACNA2D3, DST, EPB41L4A and KIFIB were among those 226 prognostic genes and were associated with the good prognosis in MYCN non-amplified neuroblastoma patients. MYCN non-amplified neuroblastoma patients with low expression levels of ALCAM, CACNA2D3, DST, EPB41L4A or KIFIB were with low overall survival in TARGET dataset (Fig. 6a), GSE49710dataset (Fig. 6b) and GSE85047 dataset (Fig. 6c).

Interestingly, high expression levels of ALCAM, CACNA2D3, DST, EPB41L4A or KIFIB were not only associated with the prognosis of MYCN non-amplified neuroblastoma patients, but also were associated with the prognosis of all neuroblastoma patients. Neuroblastoma patients with high expression levels of ALCAM, CACNA2D3, DST, EPB41L4A or KIFIB had favorable clinical overall survival in TARGET dataset (Fig. 7a), GSE49710dataset (Fig. 7b) and GSE85047 dataset (Fig. 7c).

CACNA2D3, DST, EPB41L4A and KIFIB are down-regulated in the sub-consensus 3 of MYCN non-amplified neuroblastoma patients.

Previously, we had shown that there were three sub-consensuses of MYCN non-amplified neuroblastoma patients with different clinical outcomes (Fig. 3). Next, we tested the expression levels of ALCAM, CACNA2D3, DST, EPB41L4A and KIFIB in the three different sub-consensuses in TARGET, GSE49710 and GSE85047 datasets. The expression levels of ALCAM, CACNA2D3, EPB41L4A and KIFIB were not different in the three sub-consensuses of MYCN non-amplified neuroblastoma patients in TARGET dataset (Fig. 8a). Only, compared with sub-consensus 1, DST was lowly expressed in sub-consensus 3 of MYCN non-amplified neuroblastoma patients (Fig. 8a).

Moreover, DST was also lowly expressed in sub-consensus 3 of MYCN non-amplified neuroblastoma patients in GSE49710 (Fig. 8b) and GSE85047 (Fig. 8c) datasets, compared with MYCN non-amplified neuroblastoma patients in sub-consensus 1 or sub-consensus 2. Furthermore, the relative expression levels of CACNA2D3, EPB41L4A and KIFIB were lower in sub-consensus 3 in GSE49710 dataset, compared with MYCN non-amplified neuroblastoma patients in sub-consensus 1 (Fig. 8b). Also,

compared with sub-consensus 2, CACNA2D3, EPB41L4A and KIFIB were lowly expressed in MYCN non-amplified neuroblastoma patients in sub-consensus 3 in GSE85047 dataset (Fig. 8c).

Expression level of DST is an independent prognostic factor in MYCN non-amplified pediatric neuroblastoma.

Next, we tried to determine the associations of ALCAM, CACNA2D3, DST, EPB41L4A and KIFIB in MYCN non-amplified neuroblastoma patients. First, based on their expression levels, we found those genes were highly correlated with each other, as demonstrated the high correlation coefficients of those genes in TARGET, GSE49710 and GSE85047 datasets (Fig. 9a).

Second, using multivariate cox regression, we determined the correlation of age, ALCAM, CACNA2D3, DST, EPB41L4A or KIFIB expressions in the prediction of the clinical overall survival of MYCN non-amplified neuroblastoma patients. In TARGET, GSE49710 and GSE85047 datasets, we found that the expression level of DST was an independent prognostic marker (Fig. 9b). Those results suggested that, although, ALCAM, CACNA2D3, DST, EPB41L4A and KIFIB shared similar expression signature, the prognostic significance of DST was different.

Young patients with high DST expression level have the best prognosis in MYCN non-amplified pediatric neuroblastoma.

Since age associated gene DST was an independent prognostic factor in MYCN non-amplified pediatric neuroblastoma, we wondered if the combination of DST, age and MYCN could achieve best prognostic significance. To test this hypothesis, MYCN non-amplified pediatric neuroblastoma patients in TARGET datasets were divided into old patients with high DST expression, old patients with low DST expression, young patients with high DST expression and young patients with low DST expression four sub-groups. We found that MYCN non-amplified young patients with high DST expression had the better prognosis than other three sub-groups (Fig. 10a). Similarly, in GSE49710 and GSE85047 datasets, MYCN non-amplified young patients with high DST expression also had the best prognosis, while, MYCN non-amplified old patients with low DST expression had the worst prognosis (Fig. 10a).

Furthermore, the contingency graphs showed that among the MYCN non-amplified young pediatric neuroblastoma patients with high DST expression, 73% or 62% patients were in favorable or low risk sub-group, receptively. Contrast with the 16% or 0% patients were in favorable or low risk sub-group in MYCN non-amplified young pediatric neuroblastoma patients with low DST expression (Fig. 10b). Combined all those results suggested that MYCN non-amplified young neuroblastoma patients with high DST expression levels had the best clinical overall survival.

Discussion

Using the combining expression and clinical data derived from TARGET, GSE49710 and GSE85047 datasets, our study delineates the inner heterogeneity within MYCN amplified and non-amplified

neuroblastoma sub-groups. MYCN non-amplified pediatric neuroblastoma is heterogeneous, comprised young and old sub-groups. And MYCN non-amplified young pediatric neuroblastoma patients have better prognosis. NMF assay adds further proofs that MYCN non-amplified pediatric neuroblastoma includes distinct sub-consensuses and each sub-consensus has different clinical outcomes. Furthermore, even young and old MYCN non-amplified pediatric neuroblastoma sub-groups are not biological entity. MYCN non-amplified young patients with high DST expression levels have the best prognosis. On the contrary, MYCN non-amplified old patients with low DST expression levels have unfavorable prognosis. These analyses provide deep understanding of the heterogeneity landscape of neuroblastoma.

Dystonin/BPAG1 (DST) is bullous pemphigoid antigen 1, involving the autoimmune response in the development of bullous pemphigoid [17]. DST is also associated with multiple neurological disorders, including Parkinson's disease [18], Alzheimer's disease [19], epilepsy, dementia and multiple sclerosis [20], through regulation of cytoskeletal dynamics [21] and cell migration [22]. However, the functions of DST in neuroblastoma are barely known. We find that DST is an independent prognostic factor in MYCN non-amplified pediatric neuroblastoma and highly expressed in MYCN non-amplified young neuroblastoma patients. Combination of DST, age and MYCN achieve best prognostic effects in neuroblastoma. ALCAM, CACNA2D3, EPB41L4A and KIFIB share similar expression signature and prognostic effects with DST. Moreover, the prognostic significances of ALCAM, CACNA2D3, EPB41L4A and KIFIB in MYCN non-amplified pediatric neuroblastoma are not previously reported.

In our study, we show that MYCN amplified pediatric neuroblastoma is relatively homogeneous, only one gene CSNK1G2 could divide the MYCN amplified pediatric neuroblastoma into prognostic significant sub-groups. We think that this observation is derived from several reasons. One, MYCN amplified pediatric neuroblastoma is truly biological entity. Another reason is the lack of enough samples to reveal inner complexity and heterogeneity of MYCN amplified pediatric neuroblastoma. There are 124 MYCN non-amplified pediatric neuroblastoma samples with expression data. However, the number of MYCN amplified pediatric neuroblastoma samples with expression data is only 37 in TARGET dataset. Third, in our study, we use only age and NMF assay to distinguish the sub-groups of MYCN amplified pediatric neuroblastoma patients. Using other methods, for example, sub-class mapping (SubMap) [23] or Similarity network fusion (SNF) [24] in large cohort of MYCN amplified patients may achieve better classification of MYCN amplified pediatric neuroblastoma.

The purpose of our study is to determine the molecular inner sub-groups within each of the MYCN amplified and MYCN non-amplified core sub-groups of pediatric neuroblastoma. The present study suggests the homogeneity of MYCN amplified pediatric neuroblastoma. However, MYCN non-amplified pediatric neuroblastoma could be divided into several sub-groups by age and DST expression levels. Our results suggest that MYCN non-amplified neuroblastoma young patients with high DST expression levels have the best clinical overall survival. However, there are limitations for bioinformatic analysis that the results are generated from published TARGET, GSE49710 and GSE85047 datasets and lack of further experimental validations in neuroblastoma cell or neuroblastoma patients. Therefore, functions and

prognosis of DST in MYCN non-amplified neuroblastoma should be further validated. Also the clustering of MYCN amplified neuroblastoma in a large number of patients is further needed.

Conclusions

MYCN non-amplified neuroblastoma is a heterogeneous disease and could be divided into sub-groups based on age or the expression levels of ALCAM, CACNA2D3, DST, EPB41L4A and KIFIB. MYCN non-amplified neuroblastoma young patients with high DST expression levels have the best clinical overall survival.

Abbreviations

TARGET: Therapeutically Applicable Research to Generate Effective Treatments; GEO:Gene Expression Omnibus; DAVID:The Database for Annotation, Visualization and Integrated Discovery; NMF:non-negative matrix factorization; DST:Dystonin/BPAG1; SubMap:sub-class mapping; SNF:Similarity network fusion

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

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

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by grants from Fujian Maternity and Child Health Hospital (No.YCXB 18 – 10 and YCXM 19 – 04). But, the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

HW.W and XR.W designed and performed data analysis. LP.X helped with the data analysis. HW.W wrote the manuscript. JZ and HC reviewed the manuscript and supervised the work. All authors read and approved the final manuscript.

Acknowledgements

We appreciate the generosity of the researches from TARGET and GEO groups for sharing the huge amount of data.

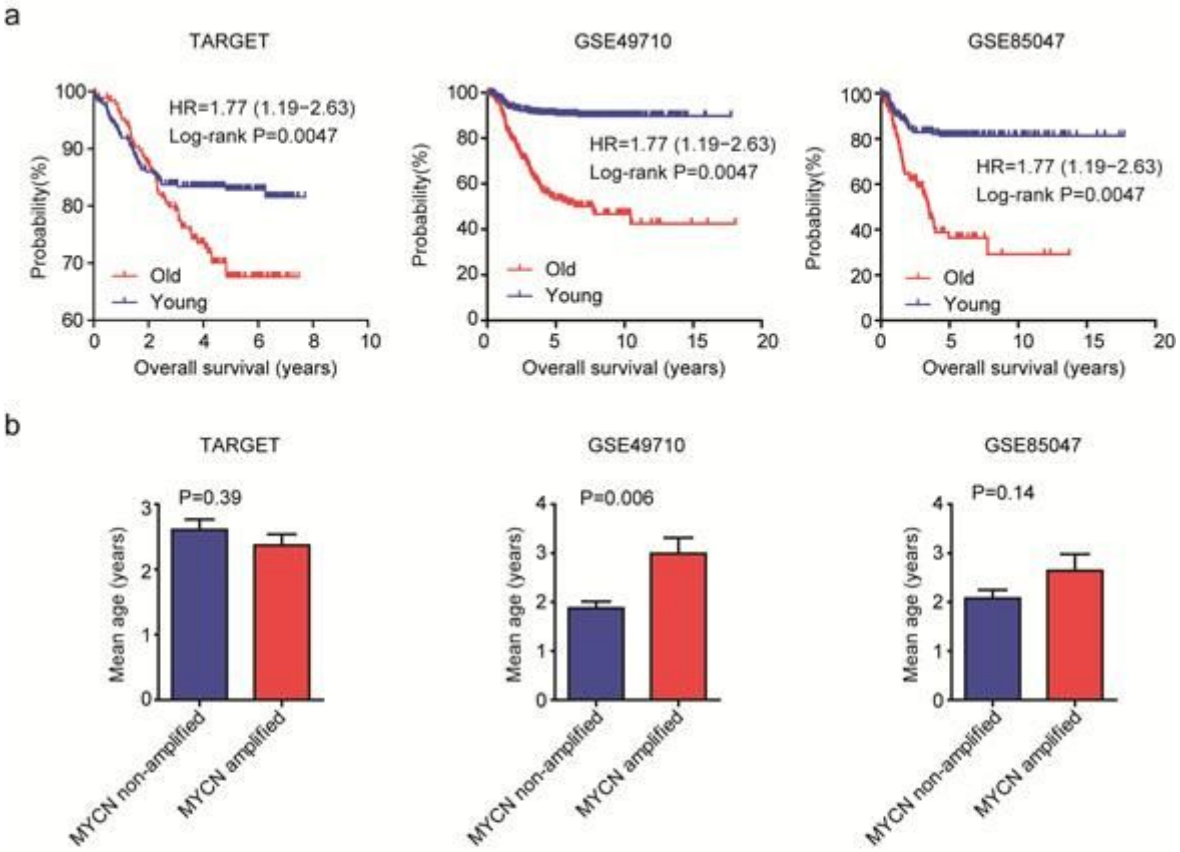
References

1. Maris JM, Hogarty MD, Bagatell R, Cohn SL. Neuroblastoma. Lancet. 2007;369(9579):2106–20.
2. Maris JM. Recent advances in neuroblastoma. N Engl J Med. 2010;362(23):2202–11.
3. 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(2):289–97.
4. Irwin MS, Park JR. Neuroblastoma: paradigm for precision medicine. Pediatr Clin North Am. 2015;62(1):225–56.
5. Campbell K, Gastier-Foster JM, Mann M, Naranjo AH, Van Ryn C, Bagatell R, Matthay KK, London WB, Irwin MS, Shimada H, et al. Association of MYCN copy number with clinical features, tumor biology, and outcomes in neuroblastoma: A report from the Children's Oncology Group. Cancer. 2017;123(21):4224–35.
6. Look AT, Hayes FA, Shuster JJ, Douglass EC, Castleberry RP, Bowman LC, Smith EI, Brodeur GM. Clinical relevance of tumor cell ploidy and N-myc gene amplification in childhood neuroblastoma: a Pediatric Oncology Group study. J Clin Oncol. 1991;9(4):581–91.
7. Wang H, Wang X, Xu L, Zhang J, Cao H. Prognostic significance of MYCN related genes in pediatric neuroblastoma: a study based on TARGET and GEO datasets. BMC Pediatr. 2020;20(1):314.
8. Valentijn LJ, Koster J, Haneveld F, Aissa RA, van Sluis P, Broekmans ME, Molenaar JJ, van Nes J, Versteeg R. Functional MYCN signature predicts outcome of neuroblastoma irrespective of MYCN amplification. Proc Natl Acad Sci U S A. 2012;109(47):19190–5.
9. Park JR, Eggert A, Caron H. Neuroblastoma: biology, prognosis, and treatment. Hematol Oncol Clin North Am. 2010;24(1):65–86.
10. Ben-Aharon I, Elkabets M, Pelossof R, Yu KH, Iacubuzio-Donahue CA, Leach SD, Lowery MA, Goodman KA, O'Reilly EM. Genomic Landscape of Pancreatic Adenocarcinoma in Younger versus Older Patients: Does Age Matter? Clin Cancer Res. 2019;25(7):2185–93.

11. Derry JM, Mangravite LM, Suver C, Furia MD, Henderson D, Schildwachter X, Bot B, Izant J, Sieberts SK, Kellen MR, et al. Developing predictive molecular maps of human disease through community-based modeling. Nat Genet. 2012;44(2):127–30.
12. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, Zhou X, Li Y, Rusch MC, Easton J, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature. 2018;555(7696):371–6.
13. Su Z, Fang H, Hong H, Shi L, Zhang W, Zhang W, Zhang Y, Dong Z, Lancashire LJ, Bessarabova M, et al. An investigation of biomarkers derived from legacy microarray data for their utility in the RNA-seq era. Genome Biol. 2014;15(12):523.
14. Rajbhandari P, Lopez G, Capdevila C, Salvatori B, Yu J, Rodriguez-Barrueco R, Martinez D, Yarmarkovich M, Weichert-Leahey N, Abraham BJ, et al. Cross-Cohort Analysis Identifies a TEAD4-MYCN Positive Feedback Loop as the Core Regulatory Element of High-Risk Neuroblastoma. Cancer Discov. 2018;8(5):582–99.
15. Sadanandam A, Lyssiotis CA, Homicsko K, Collisson EA, Gibb WJ, Wullschleger S, Ostos LC, Lannon WA, Grotzinger C, Del Rio M, et al. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. Nat Med. 2013;19(5):619–25.
16. Wang H, Wang X, Xu L, Zhang J, Cao H. A molecular sub-cluster of colon cancer cells with low VDR expression is sensitive to chemotherapy, BRAF inhibitors and PI3K-mTOR inhibitors treatment. Aging. 2019;11(19):8587–603.
17. Di Zenzo G, Della Torre R, Zambruno G, Borradori L. Bullous pemphigoid: from the clinic to the bench. Clin Dermatol. 2012;30(1):3–16.
18. Foureur N, Descamps V, Lebrun-Vignes B, Picard-Dahan C, Grossin M, Belaich S, Crickx B. Bullous pemphigoid in a leg affected with hemiparesia: a possible relation of neurological diseases with bullous pemphigoid? Eur J Dermatol. 2001;11(3):230–3.
19. Cheng J, Liu HP, Hwang SL, Hsu LF, Lin WY, Tsai FJ. Dystonin/BPAG1 modulates diabetes and Alzheimer's disease cross-talk: a meta-analysis. Neurol Sci. 2019;40(8):1577–82.
20. Brick KE, Weaver CH, Savica R, Lohse CM, Pittelkow MR, Boeve BF, Gibson LE, Camilleri MJ, Wieland CN. A population-based study of the association between bullous pemphigoid and neurologic disorders. J Am Acad Dermatol. 2014;71(6):1191–7.
21. Poliakova K, Adebola A, Leung CL, Favre B, Liem RK, Schepens I, Borradori L. BPAG1a and b associate with EB1 and EB3 and modulate vesicular transport, Golgi apparatus structure, and cell migration in C2.7 myoblasts. PLoS One. 2014;9(9):e107535.
22. Michael M, Begum R, Fong K, Pourreyrone C, South AP, McGrath JA, Parsons M. BPAG1-e restricts keratinocyte migration through control of adhesion stability. J Invest Dermatol. 2014;134(3):773–82.
23. Hoshida Y, Brunet JP, Tamayo P, Golub TR, Mesirov JP. Subclass mapping: identifying common subtypes in independent disease data sets. PLoS One. 2007;2(11):e1195.
24. Cavalli FMG, Remke M, Rampasek L, Peacock J, Shih DJH, Luu B, Garzia L, Torchia J, Nor C, Morrissy AS, et al. Intertumoral Heterogeneity within Medulloblastoma Subgroups. Cancer Cell.

2017;31(6):737–54 e736.

Figures



c

TARGET

Variable	Coefficient	HR	p-value	95% CI
Age	0.08	1.08	0.0001	1.04-1.12
MYCN	-0.56	0.57	2.6e-06	0.45-0.72

GSE49710

Variable	Coefficient	HR	p-value	95% CI
Age	0.34	1	1.98e-12	1-1.1
MYCN	1.97	7.2	2e-16	4.9-10.6

GSE85047

Variable	Coefficient	HR	p-value	95% CI
Age	0.29	1	0.0002	1-1.1
MYCN	1.69	5.4	1.24e-12	3.4-8.6

Figure 1

Age and MYCN amplification are independent prognostic factors in pediatric neuroblastoma. (a) Overall survival was analyzed in old (red) and young (blue) pediatric neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets. P values were generated from Log-rank test. (b) Box plot showed the mean age of MYCN amplified (red) and MYCN non-amplified (blue) pediatric neuroblastoma patients. P values were determined using the Student's t test. (c) Multivariate cox regression was used to test the

prognostic relevance of age and MYCN amplification in neuroblastoma, using TARGET, GSE49710 and GSE85047 datasets.

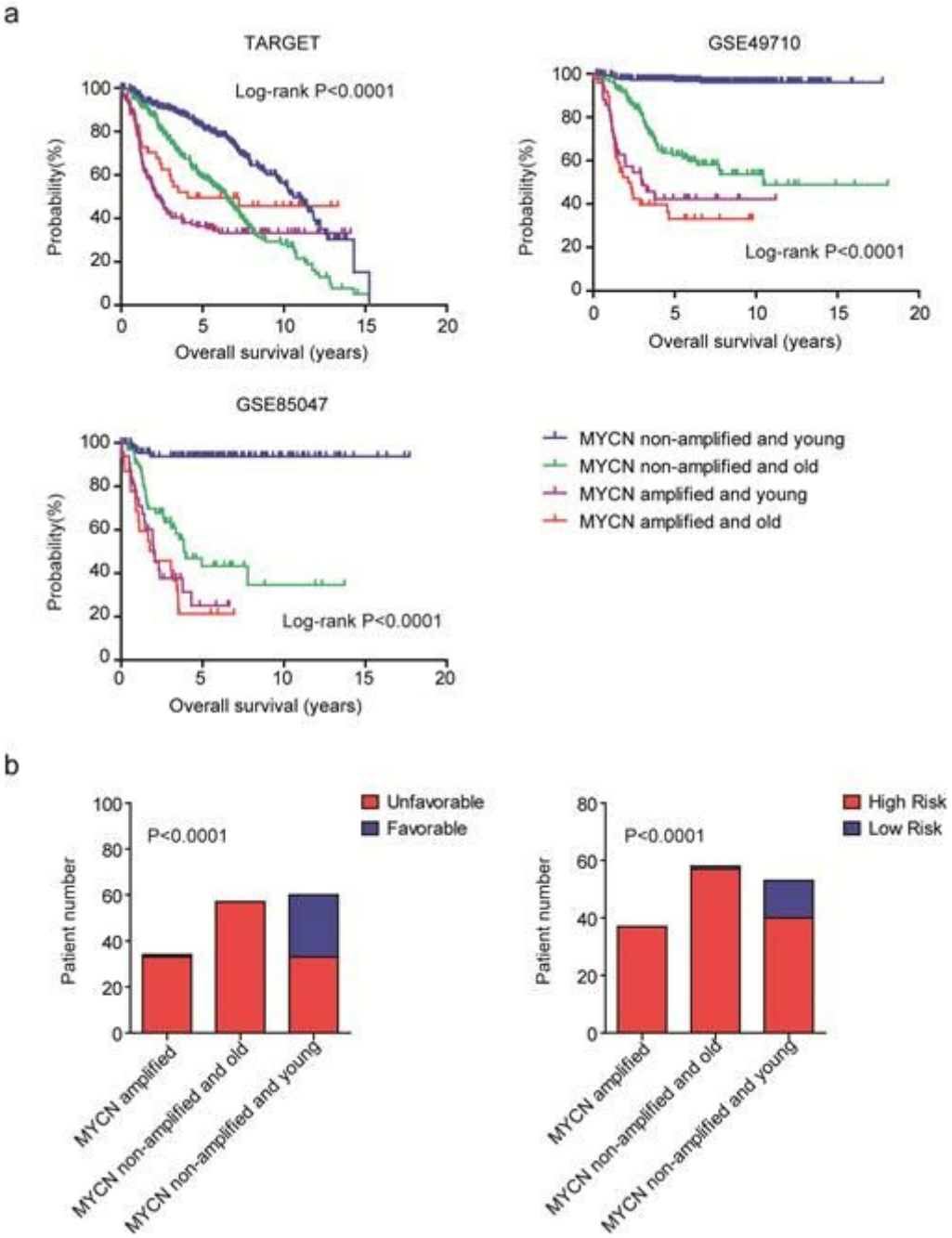


Figure 2

MYCN non-amplified pediatric neuroblastoma is a heterogeneous disease and MYCN non-amplified young neuroblastoma patients have better prognosis. (a) Pediatric neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets were divided into MYCN amplified old, MYCN amplified young, MYCN non-amplified old and MYCN non-amplified young four sub-groups. The Kaplan-Meier plots determined the different overall survival of those four sub-groups. P values were determined using Log-rank test. (b) Contingency graphs showed the number of MYCN amplified, MYCN non-amplified old and MYCN non-

amplified young neuroblastoma patients in each histological sub-group in TARGET dataset. P values were determined by Chi-square test.

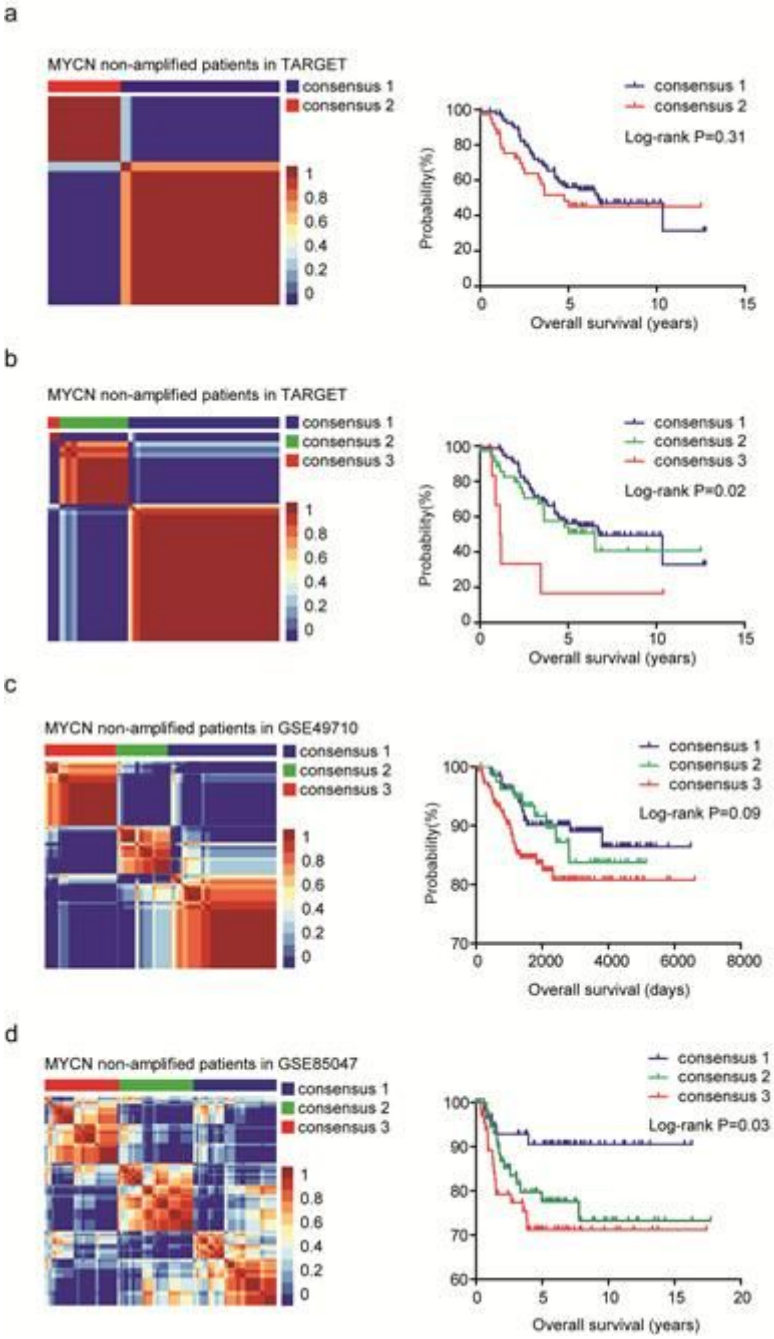


Figure 3

Three sub-consensuses of MYCN non-amplified neuroblastoma patients are with different clinical overall survival. (a) MYCN non-amplified neuroblastoma patients in TARGET dataset were divided into two sub-consensuses based on the globe gene expression signature using NMF assay. Overall survival was analyzed in sub-consensus 1 and sub-consensus 2 MYCN non-amplified pediatric neuroblastoma patients in TARGET datasets. P values were generated from Log-rank test. (b-d) MYCN non-amplified neuroblastoma patients were divided into three sub-consensuses in TARGET (b), GSE49710 (c) and GSE85047 (d) dataset. Overall survival was analyzed in each sub-consensus.

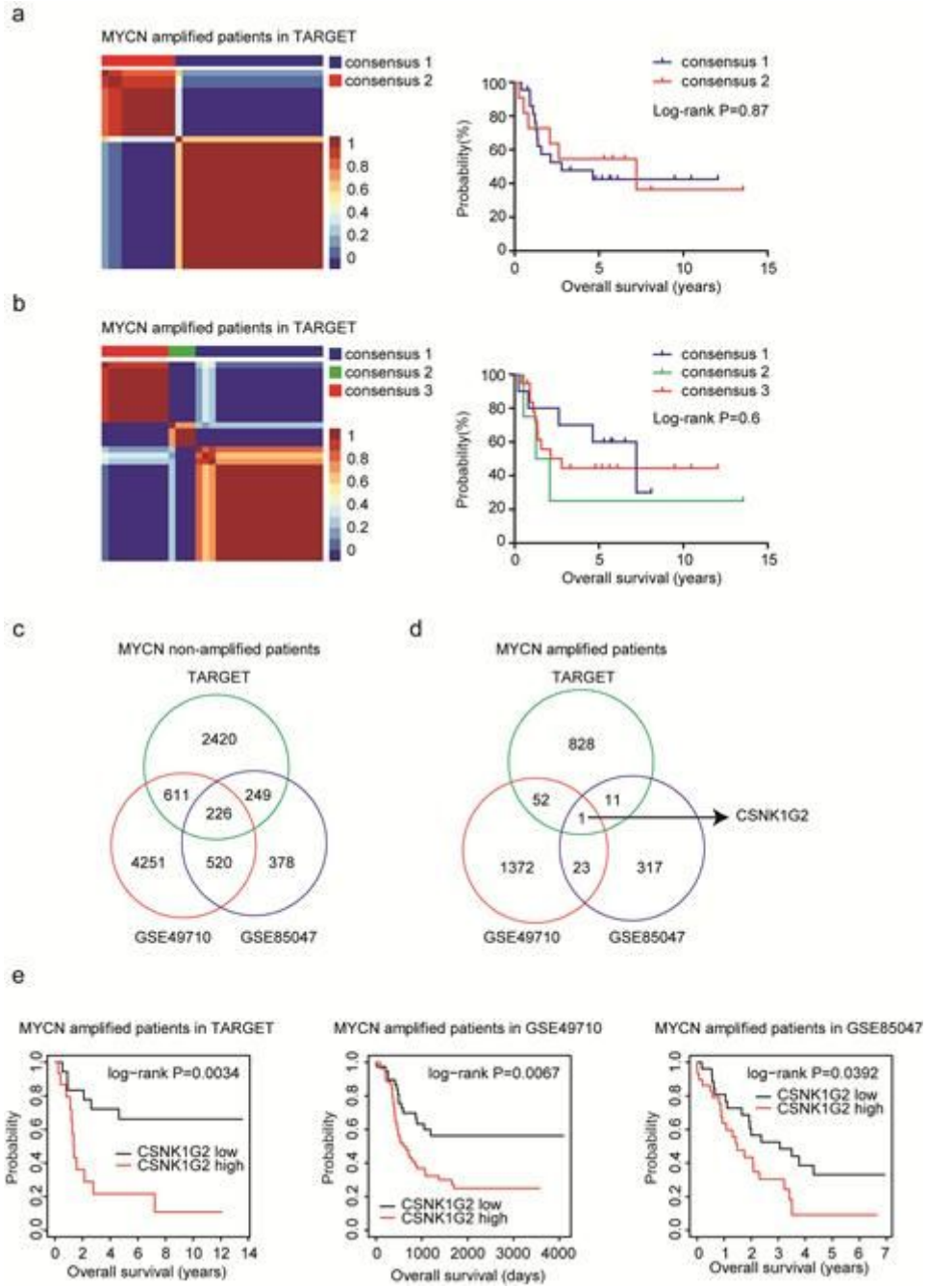


Figure 4

MYCN amplified pediatric neuroblastoma is a relatively homogeneous disease. (a-b) MYCN amplified neuroblastoma patients in TARGET dataset were divided into two sub-consensuses (a) or three sub-consensuses (b) using NMF assay. Overall survival was analyzed in each sub-consensus. (c) Venn diagram depicted the common prognostic genes in MYCN non-amplified neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets. (d) Venn diagram depicted the common prognostic genes in MYCN amplified neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets. (e) The Kaplan-Meier plots demonstrated the prognostic effects of CSNK1G2 in MYCN amplified pediatric neuroblastoma using TARGET, GSE49710 and GSE85047 datasets. Patients were divided into two sub-

groups based on the mean expression levels of CSNK1G2. The log-rank test was used to determine the overall survival P-values

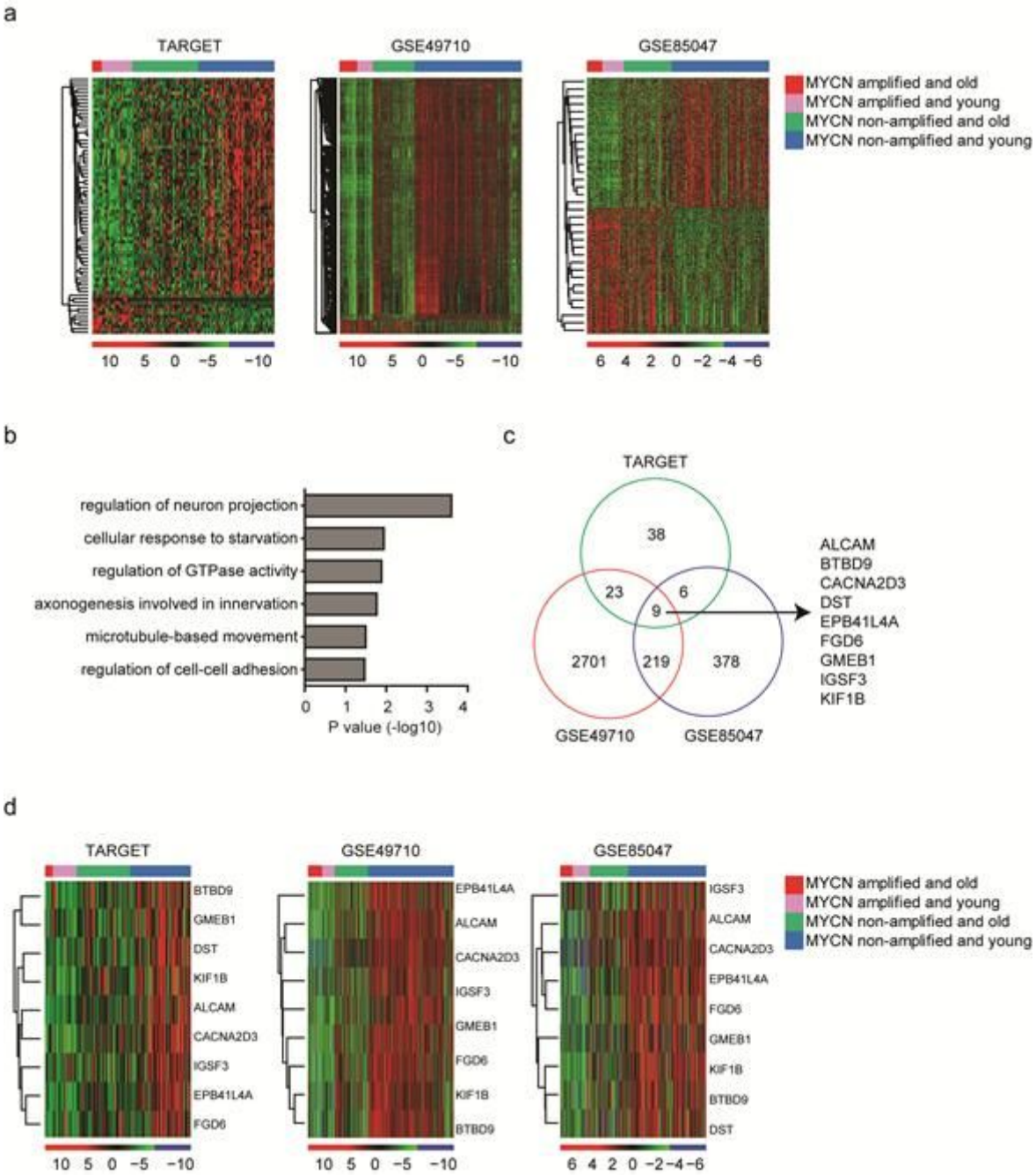


Figure 5

Identification of the transcriptional profiling in MYCN non-amplified young neuroblastoma patients (a) Heatmaps demonstrated the differentially expressed genes in MYCN non-amplified young neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets. (b) The enriched functional gene ontology from the differentially expressed genes in MYCN non-amplified young neuroblastoma patients in TARGET dataset. (c) Venn diagram depicted the common genes which were differentially expressed in MYCN non-amplified young neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets. (d) Heatmaps

demonstrated the expression levels of the common genes in TARGET, GSE49710 and GSE85047 datasets.

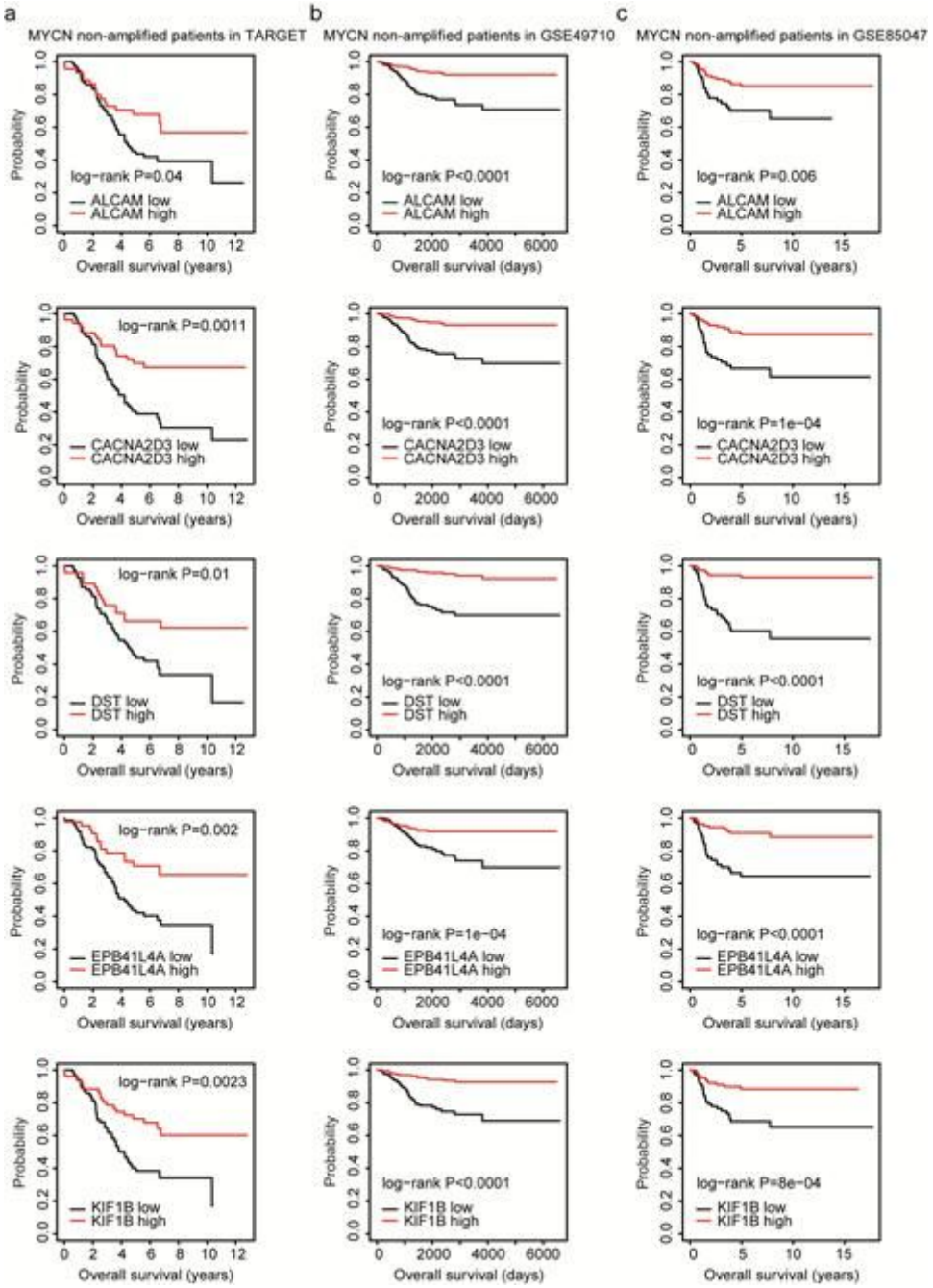


Figure 6

High expression levels of ALCAM, CACNA2D3, DST, EPB41L4A or KIFIB are associated with the favorable prognosis of MYCN non-amplified neuroblastoma patients. (a-c) The Kaplan-Meier plots demonstrated the prognostic effects of ALCAM, CACNA2D3, DST, EPB41L4A and KIFIB in MYCN non-amplified pediatric neuroblastoma using TARGET dataset (a), GSE49710 dataset (b) and GSE85047 dataset (c). Patients were divided into two sub-groups based on the mean expression levels of ALCAM, CACNA2D3, DST, EPB41L4A or KIFIB. The log-rank test was used to determine the overall survival P-values.

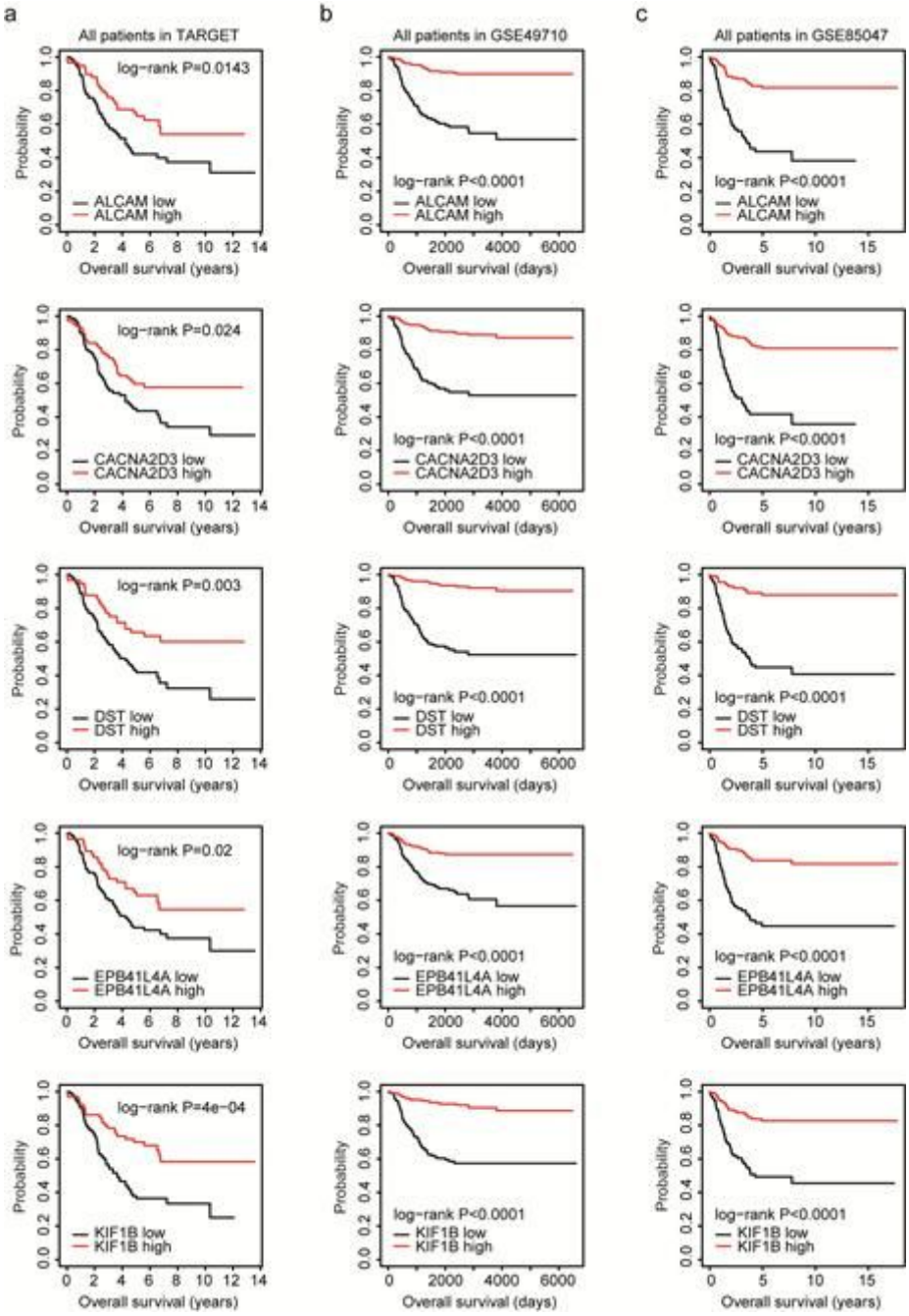


Figure 7

High expression levels of ALCAM, CACNA2D3, DST, EPB41L4A or KIFIB are associated with the favorable prognosis of neuroblastoma patients. The Kaplan-Meier plots demonstrated the prognostic effects of ALCAM, CACNA2D3, DST, EPB41L4A or KIFIB in pediatric neuroblastoma using TARGET dataset (a), GSE49710 dataset (b) and GSE85047 dataset (c).

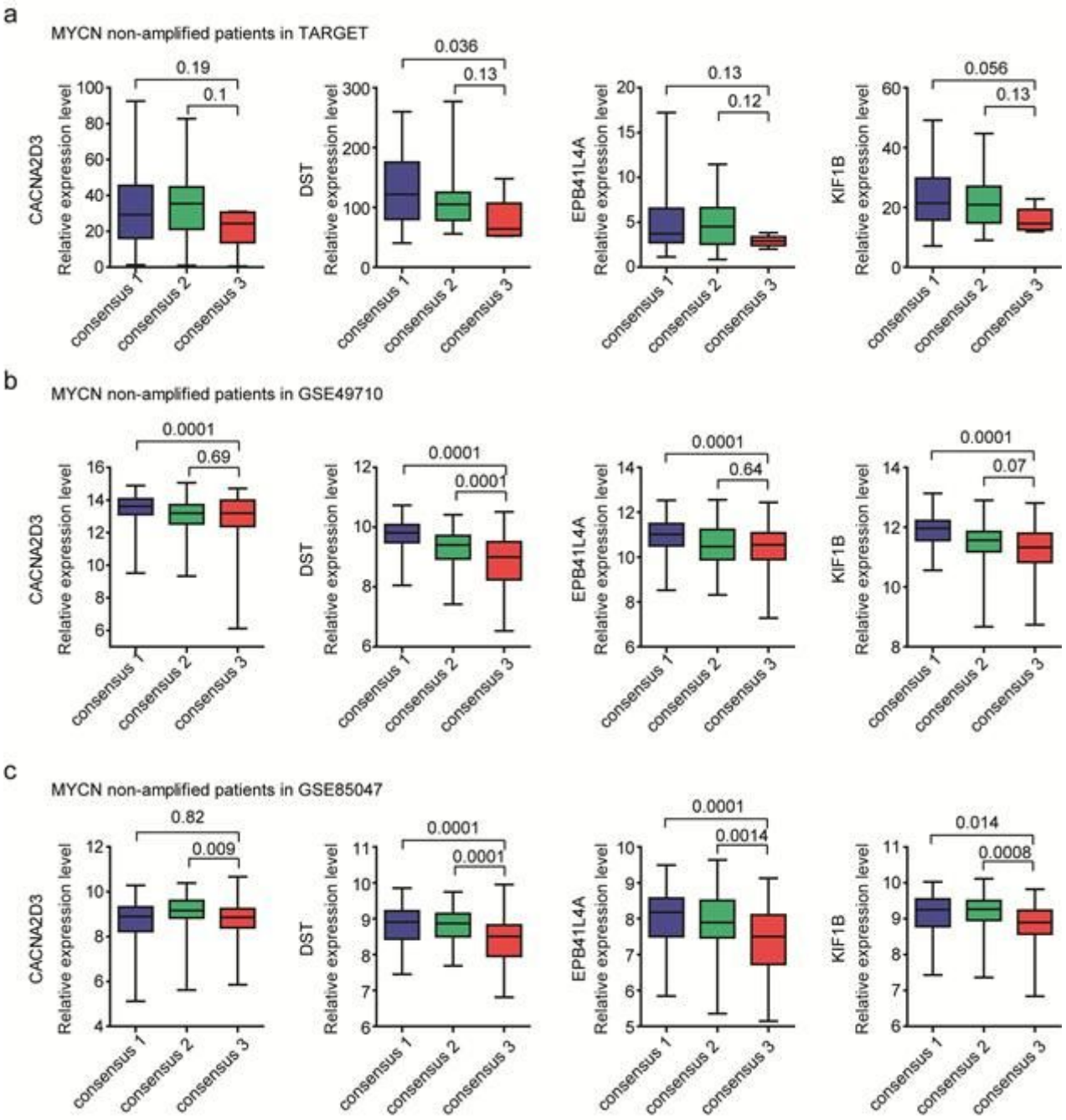
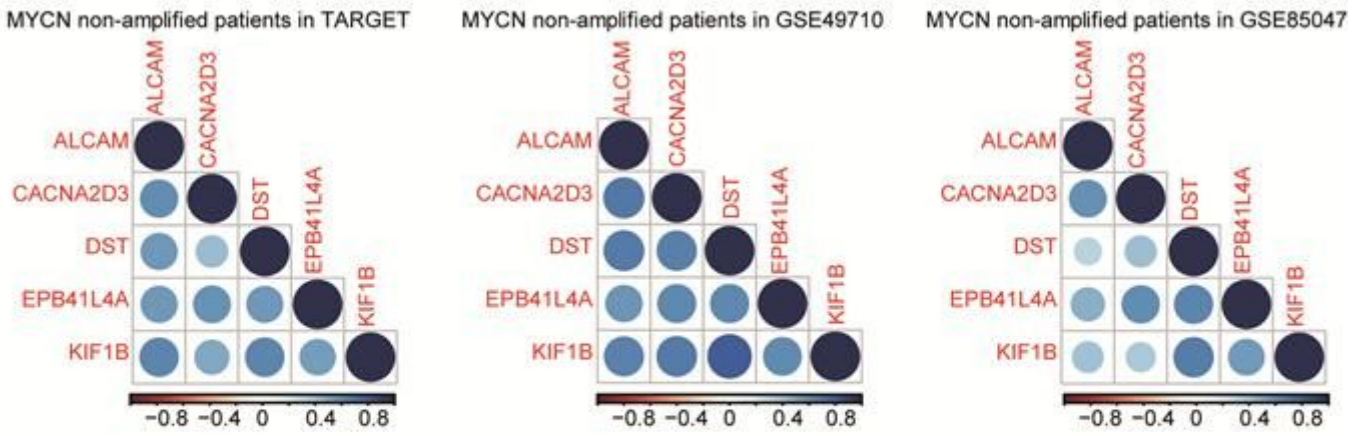


Figure 8

CACNA2D3, DST, EPB41L4A and KIFIB are down-regulated in the sub-consensus 3 of MYCN non-amplified neuroblastoma patients. Box plots showed the relative expression levels of CACNA2D3, DST, EPB41L4A and KIFIB in different sub-consensuses of MYCN non-amplified neuroblastoma patients in TARGET dataset (a), GSE49710 dataset (b) and GSE85047 dataset (c). P values were performed using Student's t test.

a



b

MYCN non-amplified patients in TARGET

Variable	Coefficient	HR	p-value	95% CI
Age	0.018	1.02	0.73	0.92-1.12
ALCAM	0.001	1	0.71	0.99-1
CACNA2D3	-0.012	0.99	0.17	0.97-1
DST	-0.009	0.99	0.013	0.98-1
EPB41L4A	-0.018	0.98	0.78	0.86-1.12
KIF1B	-0.025	0.98	0.24	0.94-1.02

MYCN non-amplified patients in GSE49710

Variable	Coefficient	HR	p-value	95% CI
Age	0.1	1.1	6.5e-06	1.06-1.16
ALCAM	-0.08	0.92	0.63	0.68-1.27
CACNA2D3	-0.18	0.84	0.21	0.64-1.11
DST	-0.62	0.54	0.008	0.34-0.85
EPB41L4A	-0.04	0.96	0.82	0.68-1.53
KIF1B	-0.24	0.78	0.37	0.46-1.33

MYCN non-amplified patients in GSE85047

Variable	Coefficient	HR	p-value	95% CI
Age	0.16	1.17	0.0003	1.08-1.28
ALCAM	0.04	1.04	0.89	0.6-1.8
CACNA2D3	0.12	1.13	0.56	0.75-1.72
DST	-0.87	0.42	0.028	0.19-0.91
EPB41L4A	-0.79	0.45	0.003	0.27-0.77
KIF1B	0.08	1.09	0.82	0.52-2.27

Figure 9

Expression level of DST is an independent prognostic factor in MYCN non-amplified pediatric neuroblastoma. (a) Corrplots demonstrated the association of ALCAM, CACNA2D3, DST, EPB41L4A and KIFIB in MYCN non-amplified neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets. The color of the circle represented the correlation coefficients. (b) Multivariate cox regression was used to determine the correlation of age, ALCAM, CACNA2D3, DST, EPB41L4A or KIFIB expression levels and overall survival in MYCN non-amplified neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets.

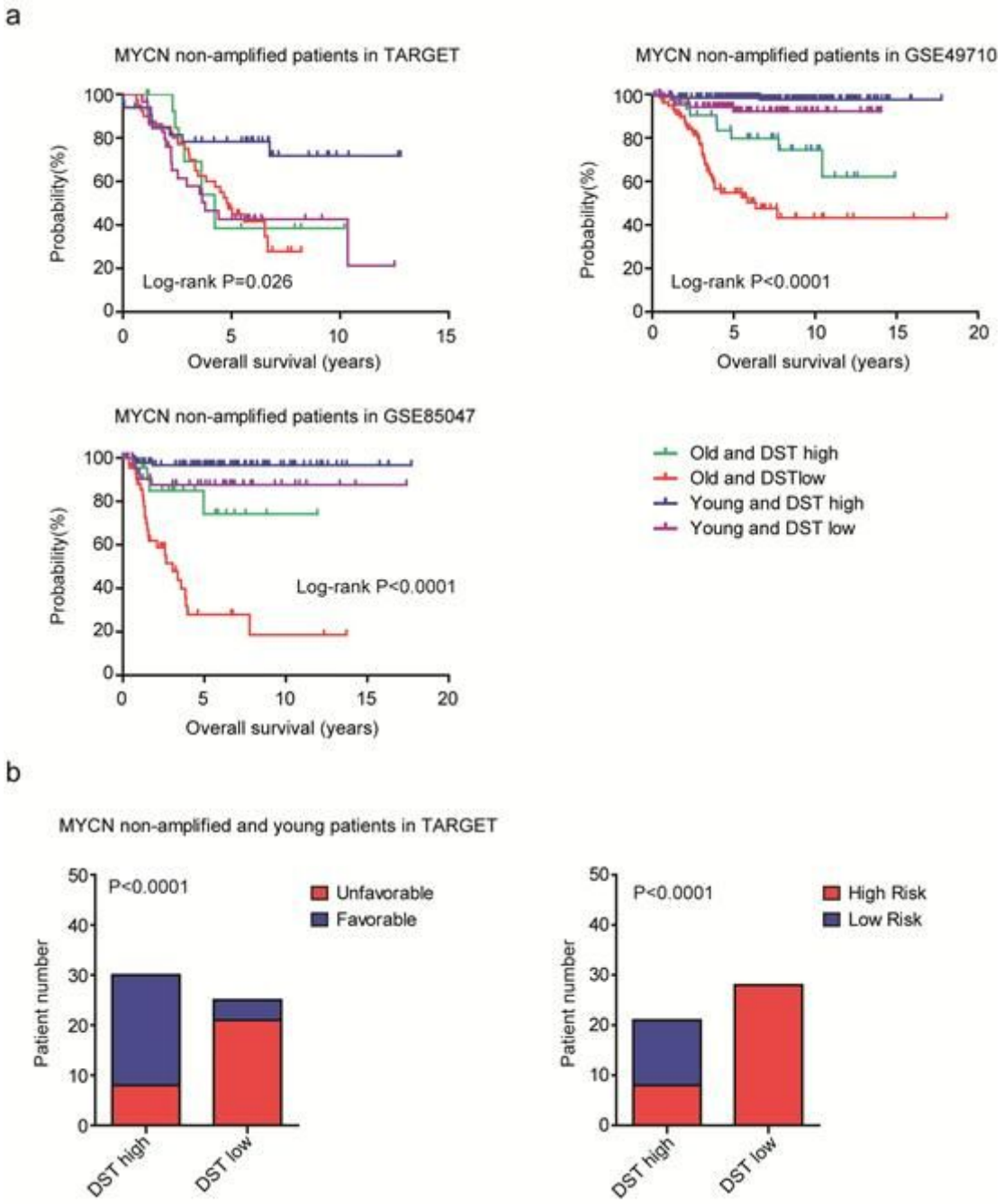


Figure 10

Young patients with high DST expression levels have the best prognosis in MYCN non-amplified pediatric neuroblastoma. (a) MYCN non-amplified pediatric neuroblastoma patients in TARGET, GSE49710 and GSE85047 datasets were divided into old patients with high DST expression, old patients with low DST expression, young patients with high DST expression and young patients with low DST expression four sub-groups. The Kaplan-Meier plots determined the different overall survival of those four sub-groups. (b) Contingency graphs showed the number of MYCN non-amplified young pediatric neuroblastoma patients with high or low DST expression in each histological sub-group. P values were determined by Chi-square test.