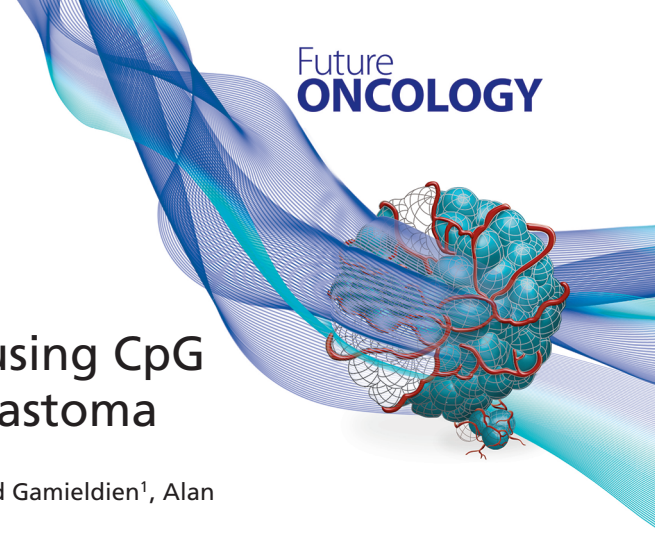
Research Article

*For reprint orders, please contact: reprints@futuremedicine.com*

Future ONCOLOGY



# Predicting amplification of *MYCN* using CpG methylation biomarkers in neuroblastoma

Abdulazeez Giwa[1], Sophia Catherine Rossouw[1], Azeez Fatai[2], Junaid Gamieldien[1], Alan Christoffels[1] & Hocine Bendou*,[1]

[1]SAMRC Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Bellville, 7535, South Africa

[2]Department of Biochemistry, Lagos State University, Nigeria

*Author for correspondence: hocine@sanbi.ac.za

**Background:** Neuroblastoma is the most common extracranial solid tumor in childhood. Amplification of *MYCN* in neuroblastoma is a predictor of poor prognosis. **Materials and methods:** DNA methylation data from the TARGET data matrix were stratified into *MYCN* amplified and non-amplified groups. Differential methylation analysis, clustering, recursive feature elimination (RFE), machine learning (ML), Cox regression analysis and Kaplan–Meier estimates were performed. **Results and Conclusion:** 663 CpGs were differentially methylated between the two groups. A total of 25 CpGs were selected by RFE for clustering and ML, and a 100% clustering accuracy was obtained. ML validation on three external datasets produced high accuracy scores of 100%, 97% and 93%. Eight survival-associated CpGs were also identified. Therapeutic interventions may need to be targeted to patient subgroups.

**Lay abstract:** Neuroblastoma is the most common extracranial solid tumor in childhood. Elevated levels of the MYCN protein in neuroblastoma is a predictor of poor prognosis. It is the most relevant prognostic factor in neuroblastoma and predicting *MYCN* gene amplification (which leads to increased gene expression and more protein) from epigenetic data rather than genetic testing might be useful in the oncology clinic. This study was designed to identify a DNA methylation (epigenetic) signature that can be used to diagnose *MYCN* amplification without actually testing for the gene. The authors also aimed to correlate this DNA methylation signature with patient survival and poorer prognosis. Based on statistical and computational methods applied to DNA methylation data for neuroblastoma, signatures that are predictive of *MYCN* amplification and poor prognosis were found, which clinicians can use for early patient diagnosis and selection of the best therapies for patients at high risk.

First draft submitted: 28 April 2021; Accepted for publication: 21 September 2021; Published online: 9 November 2021

**Keywords:** differential methylation analysis • machine learning • *MYCN* amplification • neuroblastoma • prognostic markers

Neuroblastoma is the most commonly occurring extracranial solid tumor in childhood and accounts for approximately 15% of pediatric cancer-related deaths [1–3]. It can develop anywhere along the sympathetic nervous system, with 60% of tumors occurring in the abdominal region, of which approximately half are located in the medulla of the adrenal glands [4,5]. Tumors in high-risk neuroblastoma patients are often metastatic, resulting in survival rates of less than 50% [1]. Characteristics for high-risk neuroblastoma include age, loss of chromosome 1p or 11q and amplification of *MYCN* [6,7]. Amplification of *MYCN* is a well studied genomic alteration found in approximately 22% of cases [8] and is a predictor of poor prognosis [9,10], although patients without *MYCN* amplification may also have a poor outcome [11]. Neuroblastoma is a heterogeneous disease with outcomes ranging from spontaneous regression, as seen in some tumors, to relentless progression despite extensive and varied therapies [11].

Some alterations identified in neuroblastoma include mutations in *ALK*, *ATRX* and *TERT* [12–15]. Through genomic sequencing, it is now known that pediatric cancers, including neuroblastoma, have fewer mutations compared with adult cancers [16,17], with many primary neuroblastomas not containing recognizable driver mutations. This suggests the involvement of epigenetic alterations. It has been shown that epigenetic factors, especially alter-

Future Medicine

ations in DNA methylation, play a role in the pathogenesis of neuroblastoma [18,19]. For example, hypermethylation of *TERT* was proposed as a biomarker for poor prognosis in neuroblastoma [20]. In addition, DNA methylation of *CASP8* and *RASSF1A* was linked to the development and progression of neuroblastoma [19] and was associated with poor prognosis [21].

Another indicator of poor prognosis and survival in cancer is the CpG island methylator phenotype (CIMP) [22], first established in colorectal cancer (CRC) and then adopted in neuroblastoma [23]. In a genome-wide study of 140 neuroblastomas, CIMP was defined by methylation of 5 CpG islands (CGIs) in the PCDHB family and was associated with the methylation of promoter CGIs of several tumor suppressor genes, such as *RASSF1A* and *BLU* [24]. Most *MYCN* amplified cases exhibited CIMP. However, the presence of CIMP was also detected in many cases without *MYCN* amplification [24], emphasizing the need for an accurate predictor of *MYCN* amplification.

While studies have proposed different methods of diagnosing *MYCN* amplification in neuroblastoma (such as PCR-based and hybridization laboratory methods) [25–28], the authors are unaware of any study assessing the use of CpG methylation biomarkers. Therefore, this study was designed to identify a CpG methylation signature that is diagnostic of *MYCN* amplification. In addition, the authors aim to identify CpGs that are associated with survival and poor prognosis. The use of methylation biomarkers for the diagnosis of *MYCN* amplification and prognosis of survival has the potential to be clinically useful in deciding treatment strategy and could also be cost effective.

## Materials & methods

### Dataset

The Therapeutically Applicable Research to Generate Effective Treatments (TARGET; https://ocg.cancer.gov/programs/target) initiative uses comprehensive genomic approaches for the molecular characterization of hard-to-treat childhood cancers. Neuroblastoma methylation data is accessible via the TARGET data matrix portal and is mainly composed of high-risk samples with available clinical information. The level 3 methylation dataset includes beta values from 235 samples (https://target-data.nci.nih.gov/Public/NBL/methylation_array/L3/). Based on the clinical information contained in the metadata file (https://target-data.nci.nih.gov/Public/NBL/clinical/harmonized/TARGET_NBL_ClinicalData_Discovery_20170525.xlsx), International Neuroblastoma Staging System (INSS) stage 4 samples with known *MYCN* status were considered for inclusion. A total of 126 samples (45 samples with *MYCN* amplification and 81 samples without *MYCN* amplification) were downloaded from the TARGET data matrix portal. Missing values were imputed using methyLImp package [29], which applies a computationally efficient imputation method based on linear regression. The R-package ChAMP [30] was used for data normalization and elimination of variability and batch effects between groups. A total of 396,065 CpGs were used for downstream analysis.

### Differential methylation analysis

Differential methylation analysis was done using the ChAMP package [30]. In finding biologically relevant differentially methylated probes, the ChAMP package uses the limma R-package [31] to compare two groups. Differential methylation analysis between *MYCN* amplified and non-amplified groups was performed as illustrated in Figure 1. Differentially methylated CpGs with a p-value $< 0.05$ and an absolute value of delta-beta $>0.4$ were considered statistically significant. Heatmaps were plotted using the gplots R package. Gene ontology (GO) and disease enrichment analysis was carried out to functionally annotate the differentially methylated genes (DMGs) using Enrichr [32,33].

### Machine learning

The factoextra R package [34] was used to create and visualize the clustering of samples based on all significant CpGs into their respective groups. A dendrogram plotting function was applied to the data objects produced from the application of an ensemble of hierarchical and k-means clustering algorithms to the significant CpGs [34]. Prior to the classification tasks, recursive feature elimination (RFE) was done to extract the most important features and eliminate those that create bias and negatively contribute to model performance. In the RFE, a linear support vector machine (SVM) algorithm was used to build machine learning (ML) models that were evaluated with repeated stratified ten-fold cross-validation (three repeats) to determine the best parameters and features to use in the ML classification. The CpG features selected by RFE were then used to create a training set from the 126 TARGET samples. LibSVM [35] was used to build an SVM model with the following parameters: kernel = linear, cost = 10. The SVM model was then tested on tumor methylation samples from two independent Gene Expression Omnibus

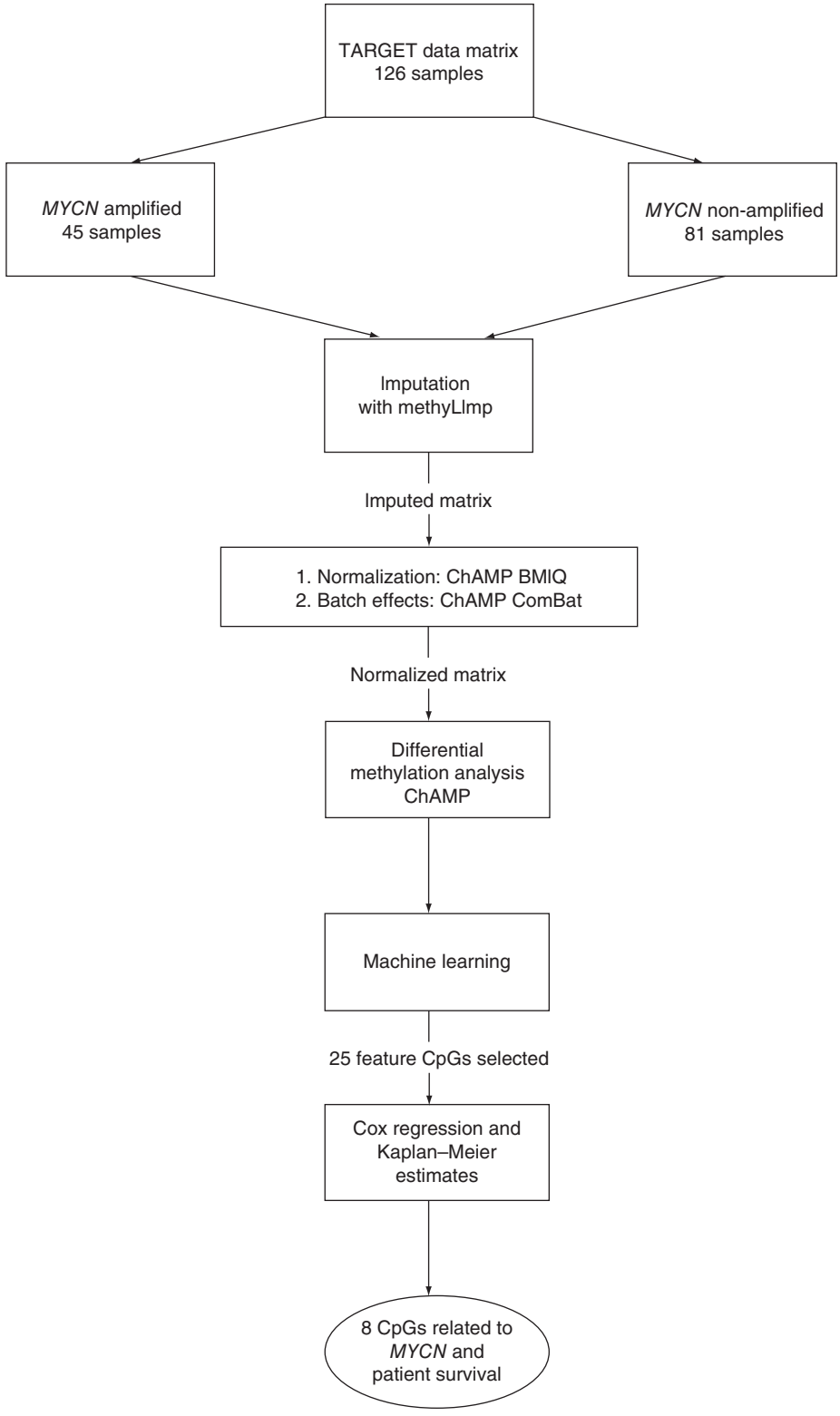



**Figure 1. Workflow depicting the steps and tools used in this analysis.** The workflow has five principal steps: data retrieval from data matrix portal, data preprocessing using methyLImp and ChAMP R packages, differential methylation analysis with ChAMP, machine learning training and testing using a support vector machine and survival analysis using glmnet R package and Kaplan–Meier curves.

Table 1. Fourteen highly methylated genes between the *MYCN* amplified and *MYCN* non-amplified groups.

| Gene | Gene name | Chr | Significant CpGs (n) |
|---|---|---|---|
| *NXPH1* | Neurexophilin 1 | 7 | 14 |
| *SOX2-OT* | SOX2 overlapping transcript | 3 | 12 |
| *DLX5* | Distal-less homeobox 5 | 7 | 10 |
| *TFAP2D* | Transcription factor AP-2 delta | 6 | 10 |
| *CAVIN3* | Caveolae associated protein 3 | 11 | 8 |
| *VAX2* | Ventral anterior homeobox 2 | 2 | 8 |
| *TERT* | Telomerase reverse transcriptase | 5 | 7 |
| *HHEX* | Hematopoietically expressed homeobox | 10 | 7 |
| *KRT19* | Keratin 19 | 17 | 7 |
| *RNF207* | Ring finger protein 207 | 1 | 7 |
| *MIRLET7BHG* | MIRLET7B host gene | 22 | 7 |
| *CHRNE* | Cholinergic receptor nicotinic epsilon subunit | 17 | 6 |
| *DLX6-AS1* | DLX6 antisense RNA 1 | 7 | 6 |
| *TMCO3* | Transmembrane and coiled-coil domains 3 | 13 | 6 |

(GEO) datasets, GSE54719 [36] and GSE120650 [37], as well as on matched primary tumor and relapse samples from another independent GEO dataset, GSE65306 [38]. These three independent datasets are composed of samples of different INSS stages (Stages 1, 2, 3, 4 and 4s) and with known *MYCN* amplification statuses. The evaluation metrics for the SVM model were precision, recall and accuracy.

### Cox regression analysis & Kaplan–Meier estimates

To determine which of the most highly significant CpGs that best correlated with patient survival, a Cox regression model based on the lasso algorithm of the glmnet R package [39–41], and which was evaluated by leave-one-out cross-validation, was used. The model assigns each CpG a regression coefficient value. CpGs with a zero coefficient were considered to have no effect on survival and were therefore eliminated. The method described by Ng *et al.* [42] was followed, whereby a CpG score value is calculated for each patient as a linear combination of beta values of the top significant CpGs weighted by their corresponding coefficients obtained from the Cox regression model. A median value was inferred from the patient scores. Each score was then compared with the median and patients were assigned a status value of 1 or 0 depending on whether the score was above or below the median. Kaplan–Meier (K–M) estimates and hazard ratios (HRs) were then calculated for the overall survival (OS) and event-free survival (EFS) according to patient status information. Additionally, K–M analysis was performed using the CpGs identified by the Cox regression models on one of the independent test datasets, GSE65306, as this is the only dataset that provides survival information. K–M curves were generated using the *ggsurvplot* function from the survminer R package. A complete workflow describing the steps followed in this study is shown in (Figure 1).

## Results

### Differential methylation analysis

Differential methylation analysis between the *MYCN* amplified and *MYCN* non-amplified groups identified 663 differentially methylated CpGs of 369 DMGs. Of these 369 DMGs, 238 and 131 DMGs had high and low methylation, respectively, in the *MYCN* amplified group compared with the non-amplified group. A total of 14 genes were highly methylated between the *MYCN* amplified and non-amplified groups. Table 1 shows information about the highly methylated genes, defined as having at least six differentially methylated CpGs with *NXPH1* having the highest number of differentially methylated CpGs.

Pathway enrichment analysis indicated that the DMGs were enriched in pathways of extracellular matrix (ECM) organization, cardiac hypertrophic response and neural crest differentiation (Table 2). With regard to molecular function and biological processes, GO analysis revealed that the DMGs were mainly enriched in the regulation of transcription and cell differentiation. In addition, for the disease enrichment analysis, the DMGs were associated with different cancers (kidney, liver and nasopharynx) as well as heart conduction disease (Table 2).

Table 2. Enrichment analysis of the 369 differentially methylated genes.

| Pathways, disease, ontology | adj. p-value | Overlapped genes (n) | Genes |
|---|---|---|---|
| **WikiPathways** | | | |
| Cardiac hypertrophic response WP2795 | 0.030335 | 6 | *HDAC4, PPP3CA, NPPA, GUCA1A, AKT1, PRKCA* |
| Neovascularization processes WP4331 | 0.030335 | 4 | *FLT4, CXCR4, AKT1, EPHB4* |
| Neural crest differentiation WP2064 | 0.030335 | 8 | *HDAC4, TFAP2B, DLX5, TLX2, MPZ, ETS1, BMP7, SOX5* |
| **Reactome** | | | |
| Extracellular matrix organization Homo sapiens R-HSA-1474244 | 0.014525 | 17 | *PTPRS, COL22A1, COL23A1, LTBP4, PDGFA, PRKCA, PCOLCE, LTBP2, BMP7, COL5A1, COL4A1, COL4A4, TIMP2, NCAM1, CD44, DDR2, ITGA9* |
| **Jensen diseases** | | | |
| Kidney cancer | 0.01632 | 74 | *SAMD9L, TRIO, PTPRS, FLT4, CHD5, KNDC1, FRY, AFF3, SIPA1L3, CELSR3, PTPRG, DOCK10, CDH4, HEPHL1, GRM6, AKT1, RNF150, ERC2, MCF2L2, EPHB4, SOX5, ARHGEF10, MEF2C, UNC13A, RIPK4, TET1, EBF3, WDR72, FRMD4A, NAV1, TYK2, DNM2, ADCY9, BANK1, COL4A1, PARD3, COL4A4, ANGPTL2, DDR2, ANKRD11, LTBP2, LRP2, THBS2, ACACA, ARNTL, KIAA0556, GRIN2A, CUX1, PCBP1, CDH22, G3BP1, MAN1C1, TRPM8, NCAM2, RGS22, XDH, DYNC1I1, TMEM132D, FARP1, TFAP2B, DNAH10, TFAP2D, COL22A1, PTCH1, DAB2IP, ATP2B4, NFATC1, TMCO3, SDK1, COL5A1, KCNS2, ZNF536, RGS12, DZIP3* |
| Heart conduction disease | 0.02058 | 8 | *SFRP2, RNF207, CXCR4, FRMD4A, C9ORF3, PTPRG, FOXP1, ITGA9* |
| Liver cancer | 0.03701 | 24 | *RSPH6A, TMEM132D, SAMD9L, TRIO, DNAH10, PTPRS, COL22A1, AGAP2, KNDC1, LRP2, FRY, THBS2, ACACA, SDK1, SNTG2, CUX1, COL5A1, COL4A1, GRM6, PARD3, ZNF536, FAM65B, NCAM1, ERC2* |
| Nasopharynx carcinoma | 0.04403 | 3 | *HLA-A, HS3ST4, ITGA9* |
| **GO: biological process** | | | |
| Positive regulation of cell differentiation (GO: 0045597) | 0.0224 | 14 | *IFITM1, MEF2C, ZBTB16, SOX11, BMP7, ARNTL, SFRP2, PDPN, MYF6, AKT1, ASB4, CMKLR1, SOX5, DDR2* |
| Negative regulation of cell differentiation (GO: 0045596) | 0.0224 | 12 | *TBX1, TRIO, SFRP2, COL5A1, ZBTB16, PTCH1, DAB2IP, ANP32B, XDH, MEIS2, SMAD7, ARNTL* |
| Positive regulation of ossification (GO: 0045778) | 0.0224 | 7 | *IFITM1, MEF2C, ZBTB16, FZD9, SOX11, BMP7, DDR2* |
| Regulation of transcription from RNA polymerase II promoter (GO: 0006357) | 0.0224 | 49 | *ZCCHC12, CD40, CHD5, ENO1, ETS1, PPP3CA, HHEX, CHP2, TEAD4, MEF2C, SMARCC1, ZHX2, TET1, EBF3, SOX11, FOXP1, SFRP2, GAL, ETV3L, EZR, SKAP1, HDAC4, DLX5, DOT1L, GATA4, ARNTL, RXRA, CUX1, DEAF1, HSF5, TBX1, WWOX, TFAP2B, TFAP2D, BCL11B, PTCH1, ZBTB16, DAB2IP, ATP2B4, NFATC1, BMP7, MEIS2, FLI1, SMAD7, NFIB, TLX2, NFIC, MYF6, ZNF536* |
| Cellular response to growth factor stimulus (GO: 0071363) | 0.0224 | 11 | *TBX1, WWOX, MEF2C, DLX5, FLT4, DAB2IP, AKT1, BMP7, CD44, SOX5, SMAD7* |
| Positive regulation of transcription DNA-templated (GO: 0045893) | 0.02774 | 39 | *HDAC4, CD40, DLX5, DOT1L, GATA4, ETS1, ARNTL, PPP3CA, HHEX, RXRA, DEAF1, ATOH8, CHP2, AKT1, TEAD4, TBX1, WWOX, TFAP2B, MEF2C, SMARCC1, BCL11B, ZBTB16, DAB2IP, TET1, EBF3, SOX11, NFATC1, BMP7, MEIS2, FLI1, SMAD7, DNM2, SFRP2, GAL, NFIB, TLX2, NFIC, MYF6, SKAP1* |
| Positive regulation of transcription from RNA polymerase II promoter (GO: 0045944) | 0.02774 | 32 | *HDAC4, CD40, DLX5, DOT1L, GATA4, ETS1, ARNTL, PPP3CA, HHEX, RXRA, TERT, CHP2, TEAD4, TBX1, WWOX, TFAP2B, MEF2C, BCL11B, DAB2IP, TET1, EBF3, SOX11, NFATC1, MEIS2, SMAD7, SFRP2, GAL, NFIB, TLX2, NFIC, MYF6, SKAP1* |
| Regulation of transmembrane receptor protein seine/threonine kinase signaling pathway (GO: 0090092) | 0.03248 | 5 | *TFAP2B, LTBP4, SOX11, BMP7, SMAD7* |
| **GO: molecular function** | | | |
| RNA polymerase II regulatory region sequence-specific DNA binding (GO: 0000977) | 0.01956 | 22 | *TFAP2B, MEF2C, SMARCC1, TFAP2D, DLX5, ZBTB16, SOX11, GATA4, NFATC1, ENO1, ETS1, MEIS2, HHEX, RXRA, CUX1, NFIB, DEAF1, NFIC, ZFP42, MYF6, ZNF536, HSF5* |
| Transcription factor activity RNA polymerase II core promoter proximal region sequence-specific binding (GO: 0000982) | 0.04652 | 15 | *MEF2C, BCL11B, DLX5, EBF3, SOX11, ENO1, ETS1, MEIS2, ARNTL, HHEX, NFIB, NFIC, ZFP42, ZNF536, ZNF467* |

GO: Gene ontology.

Table 3. Twenty-five differentially methylated CpGs between the *MYCN* amplified and *MYCN* non-amplified groups selected by recursive feature elimination.

| CpG | adj. p-value |
|---|---|
| cg00540828 | 4.276078e-29 |
| cg01710189 | 2.392558e-24 |
| cg13558971 | 5.777789e-23 |
| cg23930334 | 2.474014e-22 |
| cg03364683 | 7.279483e-22 |
| cg19944656 | 6.866947e-20 |
| cg23186333 | 1.801227e-19 |
| cg25310824 | 1.964955e-19 |
| cg20818806 | 3.380732e-19 |
| cg09973986 | 1.051204e-18 |
| cg06484432 | 4.085595e-18 |
| cg07476617 | 2.075058e-17 |
| cg22865905 | 2.403524e-17 |
| cg09175843 | 4.275436e-17 |
| cg22886575 | 1.358736e-16 |
| cg25841625 | 1.564289e-16 |
| cg26487157 | 4.648216e-16 |
| cg22871253 | 1.528965e-15 |
| cg12595667 | 5.146262e-15 |
| cg17939889 | 9.408868e-15 |
| cg14020052 | 1.592441e-14 |
| cg02658690 | 1.720897e-14 |
| cg15455864 | 4.358962e-14 |
| cg16047279 | 1.709752e-13 |
| cg22076311 | 9.729690e-12 |

Table 4. Machine learning prediction of *MYCN* amplification of samples from GSE54719, GSE120650 and GSE65306 test sets.

| *MYCN* amplification | GSE54719 | | GSE120650 | | GSE65306 | |
|---|---|---|---|---|---|---|
| | Precision | Recall | Precision | Recall | Precision | Recall |
| Yes | 1.0 | 1.0 | 1.0 | 0.87 | 0.83 | 1.0 |
| No | 1.0 | 1.0 | 0.96 | 1.0 | 1.0 | 0.89 |
| Accuracy | 100% (35/35) | | 97% (56/58) | | 93% (26/28) | |

## Machine learning

RFE was performed and selected 25 CpGs as the most important features for classifying samples by their amplification groups (Table 3). These 25 CpGs accurately clustered (100% precision) the 126 TARGET samples by their *MYCN* amplification group, with no sample misclassified (Figure 2). The dendrogram plot of hierarchical and k-means applied to the 663 significant CpGs resulted in the correct clustering of 122 (96%) of the 126 TARGET samples (Figure 3). The 25 CpGs selected by RFE from the 663 CpGs, as the most informative features in the data, were then used in ML training and test set construction (Table 3). The training set was built on the 126 TARGET data matrix samples, while the test sets comprised 35, 58 and 28 samples from the GSE54719, GSE120650 and GSE65306 datasets, respectively. Repeated stratified ten-fold cross-validation yielded an accuracy of 96%. Evaluation of the SVM model using the GSE54719, GSE120650 and GSE65306 test sets resulted in high accuracies of 100%, 97% and 93%, respectively, for correctly predicted samples (Table 4).

## Cox regression analysis & Kaplan–Meier estimates

A minimal subset of eight CpGs with a nonzero coefficient was selected by the Cox regression model (Table 5).

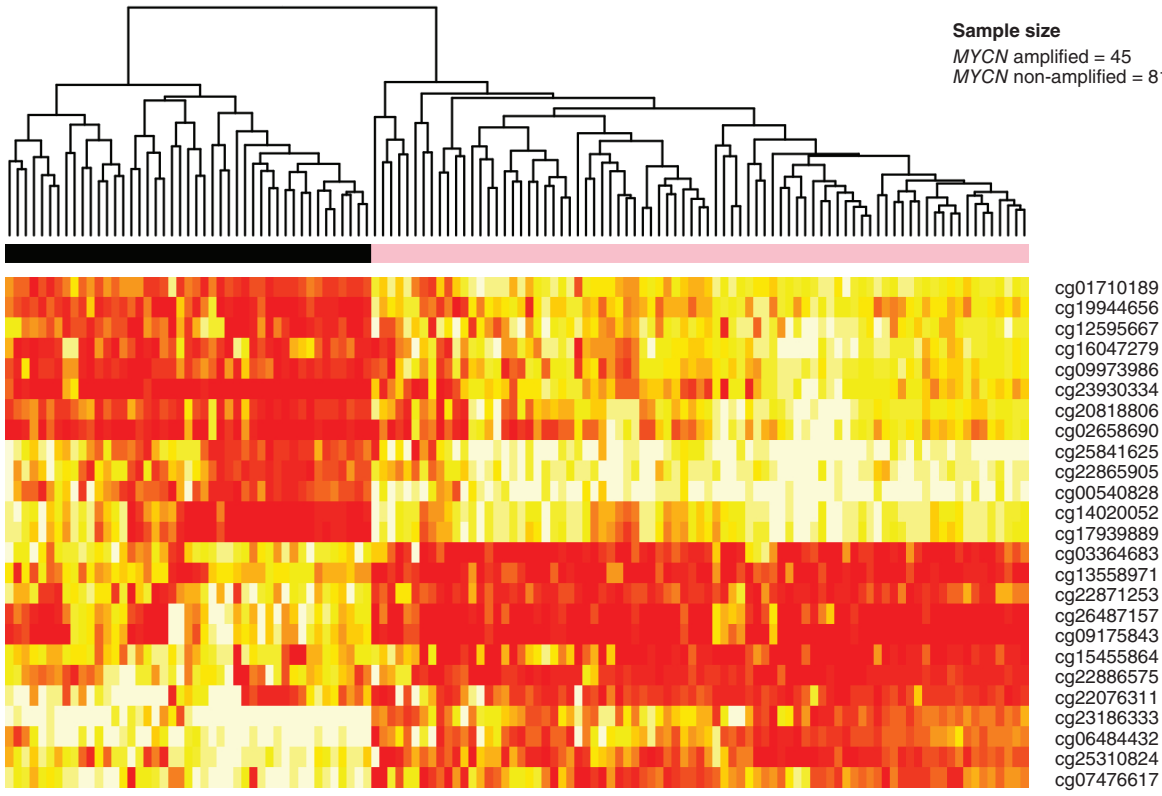



**Figure 2. Heatmap of the methylation level of the 25 most significant CpG sites (selected by recursive feature elimination) in the *MYCN* amplification groups.** A total of 126 samples were classified including 45 *MYCN* amplified samples and 81 *MYCN* non-amplified samples. Two main clusters were identified with the *MYCN* amplified samples principally clustered in the left cluster while the *MYCN* non-amplified samples were principally clustered in the right cluster. *MYCN* amplified samples are coded black while *MYCN* non-amplified samples are coded pink. No sample was misclassified. The heatmap colors represent intensity ranging from a lower intensity of red to a higher intensity of yellow.

**Table 5. CpGs selected by the Cox regression model and their respective regression coefficients for overall survival and event-free survival.**

| CpG† | Coefficient OS | DMG | CpG | Coefficient EFS | DMG |
|---|---|---|---|---|---|
| cg00540828 | -0.01461534 | *CUX1* | cg01710189 | -1.25426856 | *PDLIM2* |
| cg01710189 | -1.86558620 | *PDLIM2* | cg13558971 | 0.56167125 | *ATP2B4* |
| cg13558971 | 0.36211640 | *ATP2B4* | cg25310824 | 0.46332724 | *SEPP1* |
| cg25310824 | 0.53756846 | *SEPP1* | cg07476617 | 0.44985062 | *CFLAR* |
| cg07476617 | 0.51575895 | *CFLAR* | cg12595667 | -0.06647858 | *CXCR4* |
| cg22886575 | 0.18633969 | *HMX2* | | | |
| cg12595667 | -0.21609796 | *CXCR4* | | | |
| cg15455864 | 0.13284453 | *SYMPK* | | | |

†Eight CpGs and 5 CpGs were found to be significantly associated with overall survival and event-free survival respectively.
DMG: Differentially methylated genes; EFS: Event-free survival; OS: Overall survival.

These CpGs are believed to have a role in patient survival. All eight CpGs were associated with OS, while only five were associated with EFS. The positive coefficient CpGs were methylated in the *MYCN* amplified group, while the negative coefficient CpGs were methylated in the *MYCN* non-amplified group. The CpGs and their coefficients were used to calculate a CpG score and assign a status value of 1 or 0 for each patient. K–M estimates for overall survival (OS) and EFS based on patient statuses are shown in Figure 4. *MYCN* status correlated with CpG score and survival (OS: HR = 5.11; $p < 0.0001$; EFS: HR = 4.845; $p < 0.0001$). Patients with a CpG score above the

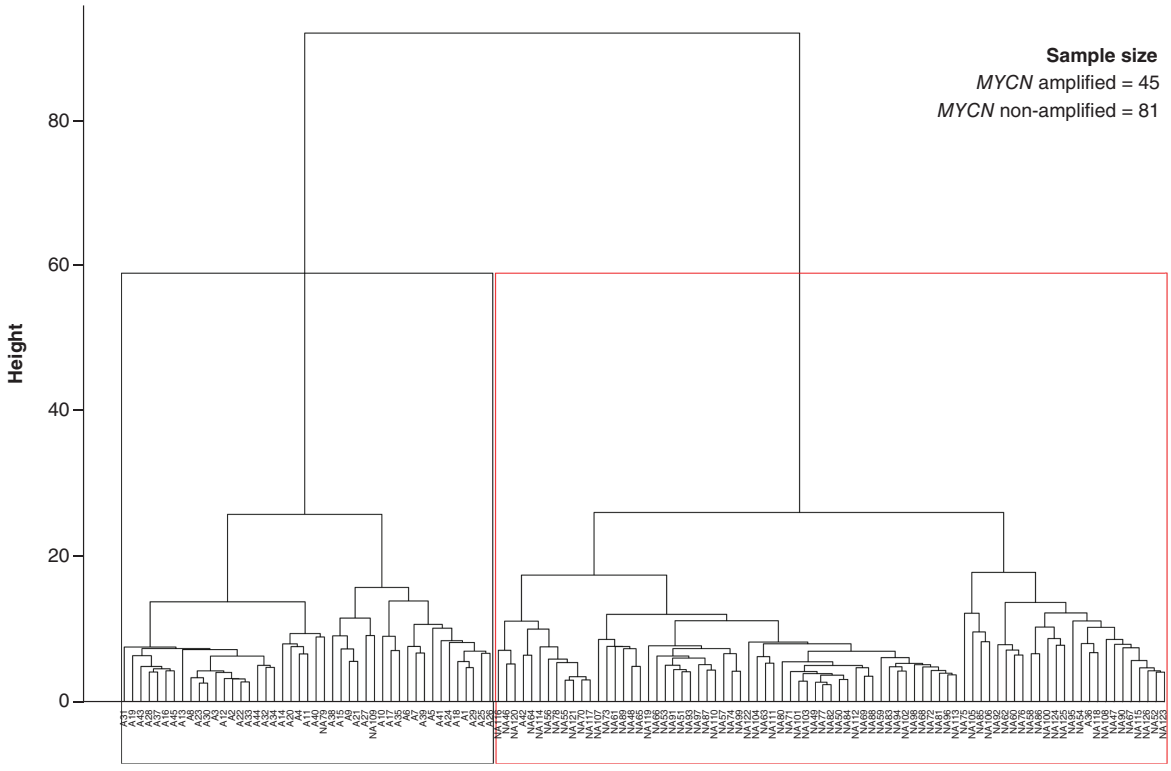



**Figure 3. Clustering of *MYCN* amplification samples using all 663 significant CpGs.** The samples are on the x-axis, with *MYCN* amplified samples labeled A and *MYCN* non-amplified samples labeled NA. Two main clusters were identified with the *MYCN* amplified samples principally clustered in the left cluster (black box) while the *MYCN* non-amplified samples were principally clustered in the right cluster (red box). A total of 126 samples were clustered including 45 *MYCN* amplified samples and 81 *MYCN* non-amplified samples. Four (2 *MYCN* amplified and 2 *MYCN* non-amplified) samples were misclassified.

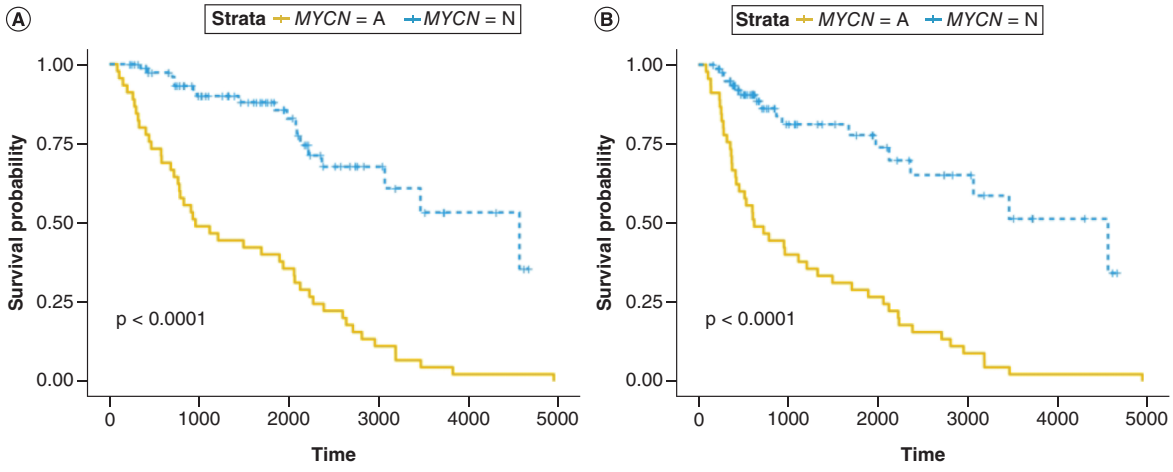



**Figure 4. Kaplan–Meier estimates of (A) overall survival and (B) event-free survival according to *MYCN* amplification and patient status values.** 1 or 0 depending on whether CpG scores were above or below the median. Estimates for *MYCN* amplified patients are in yellow and non-amplified are in blue.

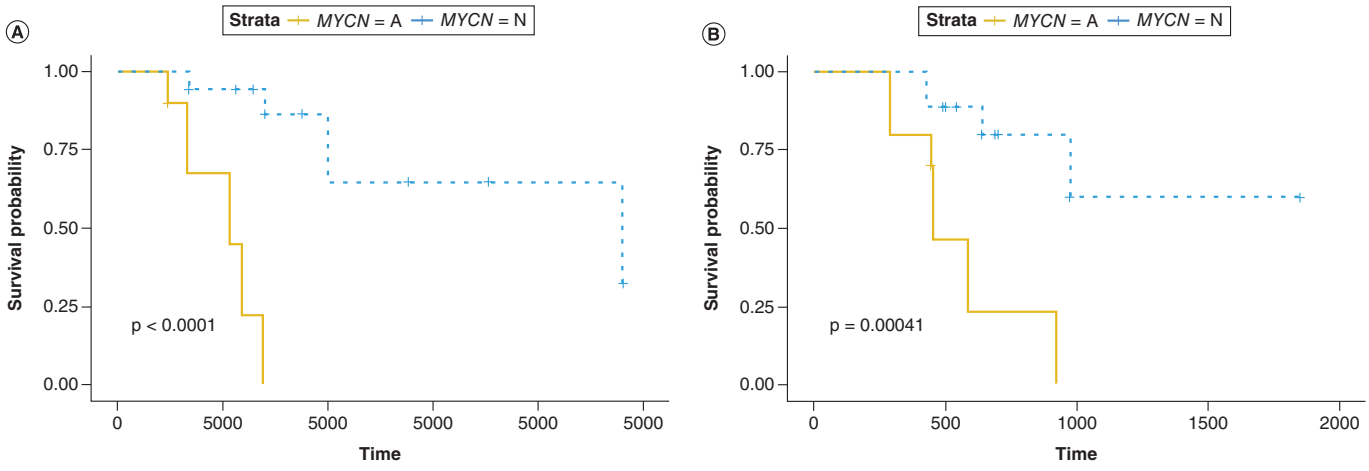



**Figure 5. Kaplan–Meier estimates for the test dataset GSE65306 of (A) overall survival and (B) event-free survival according to *MYCN* amplification and patient status values.** 1 or 0 depending on whether CpG scores were above or below the median. Estimates for *MYCN* amplified patients are in yellow and non-amplified are in blue.

median were predominantly *MYCN* amplified and those below were mostly non-amplified. Similarly, significant results were also observed when applying the previous CpGs and their corresponding coefficients identified by the Cox regression models to the methylation data and survival information from the GSE65306 test dataset (OS: HR = 35.87; $p < 0.0001$; EFS: HR = 7.99; $p < 0.00041$). The K–M curves for the GSE65306 test dataset are shown in Figure 5.

## Discussion

The authors aimed to identify predictive methylation biomarkers of the amplification of *MYCN* in neuroblastoma. The differential methylation analysis identified genes that have also been associated with neuroblastoma [20,22,43] and other cancers [44,45]. The differential methylation analysis and clustering results (Figures 2 & 3) also demonstrate that *MYCN* amplification alters the methylation landscape in neuroblastoma and that this methylation landscape differs from that in *MYCN* non-amplified neuroblastoma. The accuracy of the clustering, based on all significant CpGs (Figure 3), validated the differential methylation results and demonstrated their possible utility as features for ML prediction.

Transcriptional regulation, which was enriched by both molecular function and biological process ontologies, is an activity that promotes tumor cell proliferation. Therefore, an increase in global transcription activity suggests the rapid proliferation of tumor cells. We have also suggested that cardiac disorders may be a cause of mortality in neuroblastoma patients [46] and that this may be a treatment-related late effect [47]. In addition, ECM organization was also enriched in the GO analysis, and this correlates with the fact that the ECM is a major structural component of the tumor microenvironment, and also plays an important role in tumor progression and cell signaling [48–51].

### Machine learning & sample classification outcomes

Regardless of the stage of neuroblastoma cancer (i.e., INSS stages 1, 2, 3, 4 and 4s), the high accuracies obtained by the SVM model (Table 4), constructed on the basis of the 25 significant CpGs selected by RFE, demonstrate the validity of these CpGs for the diagnostic identification of *MYCN* amplification in neuroblastoma. The accuracy obtained from the prediction of *MYCN* amplification in both tumor and relapse samples of the GSE65306 dataset indicates the specificity and accuracy of the CpGs in identifying *MYCN* amplification in patients with neuroblastoma. The two misclassified samples were from the paired primary tumor and relapse sample of the same patient. The CpG sites found here can serve as methylation biomarkers of *MYCN* amplification and assist our understanding of tumorigenesis in *MYCN* amplified and *MYCN* non-amplified neuroblastoma. This knowledge may allow for the development of reliable and rapid diagnostic assessment of *MYCN* amplification with potential cost- and time-saving advantages. In addition, the methylated genes may serve as therapeutic targets, as the amplification of *MYCN* is a known indicator of poor prognosis in neuroblastoma.

CIMP has also been used to assess the prognosis of patients with neuroblastoma. However, as noted in the introduction, the CIMP phenotype can be observed in patients without *MYCN* amplification, which may limit its utility as a biomarker for *MYCN* amplification. In addition, conflicting findings have been reported for the prognostic role of CIMP in CRC due to differences in CIMP definitions [52]. This difficulty in defining CIMP can also occur in neuroblastoma.

Loss of *CASP8* expression by methylation was believed to be a predictor of *MYCN* amplification [53]. However, in a large-scale study, no correlation was observed between the expression of *CASP8* and the amplification of *MYCN* [54]. This may explain why *CASP8* was not among the identified DMGs in this study, although we found that its paralog, *CFLAR*, was differentially methylated and associated with both OS and EFS. Being a biomarker for poor outcome (for example methylation of *RASSF1A* and *CASP8*) does not necessarily mean being a biomarker for *MYCN* amplification, as poor outcomes can also be observed in non-amplified cases [11]. The current method established a CpG methylation signature specific for predicting *MYCN* amplification phenotype and poor prognosis in neuroblastoma.

### Cox regression analysis & Kaplan–Meier estimates

Reducing the list of CpGs obtained from the ML analysis, the Cox regression model selected eight CpGs related to patient OS (Table 5). The K–M curves (Figures 4 & 5) show the ability of these eight CpGs to distinguish between shorter and longer OS (i.e., *MYCN* amplified and non-amplified groups respectively, with a significant p-value $< 0.0001$ and HR $> 1$). Additionally, the results of the survival analysis (Figures 4 & 5) validated the poor prognosis in patients with *MYCN* amplified [9,10], and suggested eight CpGs for the prognostic diagnosis of *MYCN* amplification.

Since a CpG score above the median is associated with shorter survival, higher methylation of the three CpGs with positive coefficients, cg13558971, cg25310824, cg07476617 from *ATP2B4*, *SEPP1* and *CFLAR* genes, respectively (Table 5), predicts *MYCN* amplification and indicates poor prognostic outcome. The *CASP8* paralog, *CFLAR*, is an apoptosis regulator, with significantly higher expression in lung cancer tissues [55]. Apoptosis resistance is one of the hallmarks of cancer initiation and progression [56] and the role of apoptosis in cancer and its potential as a cancer therapy target has been reviewed [57,58]. *SEPP1* is involved in selenium transport and has been associated with neuroblastoma [59] and some other cancers, including prostate cancer [60], gastric adenocarcinoma [61] and renal cell cancer [62]. Similar to the current results, Wang *et al.* also found *SEPP1* and 13 other genes to be prognostic for OS in neuroblastoma [59]. *ATP2B4* belongs to a family of plasma membrane pumps ($Ca^{2+}$-ATPases) involved in calcium homeostasis. Cellular processes important for tumorigenesis, such as proliferation, apoptosis and angiogenesis are influenced by calcium ions [63]. The association of cg13558971 (*ATP2B4*) methylation with survival is suggestive of the importance of calcium transport in neuroblastoma. Satheesh and Busselberg [64] reviewed the role of intracellular calcium in the development and treatment of neuroblastoma and Florea *et al.* confirmed the importance of calcium signaling in neuroblastoma cells in response to chemotherapy [65]. The specific roles and mechanisms of these genes in neuroblastoma still need to be fully clarified; however, this study shows that their methylation is associated with amplification of *MYCN* and poor outcomes. This knowledge may be used to develop a predictive biomarker panel to assess patient survival in *MYCN* amplified groups.

### Highly methylated genes between the *MYCN* amplified & *MYCN* non-amplified groups

Among the 14 highly methylated genes (Table 1) were genes related to neuroblastoma as well as other cancers. *NXPH1* is a neuronal glycoprotein involved in neuronal differentiation that forms a complex with alpha neurexin proteins that promote adhesion between dendrites and axons. Decock *et al.* proposed *NXPH1* as a prognostic methylation biomarker for EFS in neuroblastoma [22]. *NXPH1* may promote the growth of tumors by stimulating the proliferation of neuroblastoma stem cells. *SOX2-OT* is a long noncoding RNA with restricted expression toward the brain. It regulates *SOX2*, a key regulator of pluripotency. It has been associated with several cancers, including gastric [66,67], pancreatic [68], cholangiocarcinoma [69] and hepatocellular carcinoma [70]. *DLX5* is a homeobox-containing transcription factor that promotes neuronal differentiation, neural crest development and is essential for osteogenesis [71,72]. In acute myeloid leukemia, promoter hypermethylation of *DLX5* led to reduced expression and correlated with shorter OS [44]. Higher methylation of *DLX5* has also been reported in colorectal cancer tissue [73] and breast cancer [74]. *DLX6-AS1* is a long noncoding RNA that partially enhances the proliferation, migration and invasive abilities of neuroblastoma cells [75]. Olsson *et al.* [20] reported similar results with hypermethylation

of *DLX5* and *DLX6-AS1* in aggressive International Neuroblastoma Risk Group (INRG) stage M neuroblastoma tumors.

*TFAP2D* belongs to the activator protein-2 transcription factor family that is essential in cellular processes such as apoptosis, migration and differentiation, and have been implicated in cancer [76,77]. Hypermethylation of a member of its protein family, *TFAP2E,* is associated with clinical nonresponsiveness to chemotherapy in colorectal cancer [78] and acts as a tumor suppressor in neuroblastoma [79]. *TFAP2D* may also be acting the same role and further studies are required to substantiate this claim. *TERT* rearrangements were reported to be the most frequent genetic alterations after *MYCN* amplification in high-risk neuroblastoma [14,15]. Consistent with the current results, Olsson *et al.* also reported hypermethylation of *TERT* in aggressive tumors of INRG stage M neuroblastoma [20]. Cancer cells activate *TERT* to maintain their telomeres [14]. *TMCO3* belongs to a family of transporter proteins involved in the coupling of export of monovalent cations such as potassium or sodium to import protons across the cellular membrane. Mutations in *TMCO3* alongside other genes have been suggested as markers for evaluating the effects of chemotherapy in patients with neuroblastoma [43]. Methylation of *TMCO3* may likely be associated with a more favorable prognosis.

*CAVIN3* plays an important role in the protein kinase c-delta tumor suppression pathway [80]. The loss of *CAVIN3* has been observed in different cancers, such as breast [81], cervix [82], bladder [83], lung [84] and stomach [85] cancers. It is commonly altered in colorectal cancer by promoter hypermethylation [86]. It is also downregulated in breast cancer, possibly due to methylation [87]. *KRT19* is highly expressed in multiple cancers, serving as a diagnostic marker [88]. High *KRT19* expression is associated with clinical progression in lung cancer [89] and correlates with poor prognosis in breast cancer [90,91]. Further studies are necessary to uncover the precise roles of these genes in neuroblastoma.

In this study, we aimed to propose a CpG signature that accurately predicts *MYCN* amplification and is associated with patient survival. This method, summarized in (Figure 1), generated a CpG signature capable of predicting *MYCN* amplification with high precision. We used the ChAMP R package [30] to discover DMGs between *MYCN* amplified and non-amplified neuroblastoma tumors and the ChAMP.DMP() function was designed to compare methylation between two phenotypes. The high prediction accuracy scores obtained from the SVM model demonstrate the correctness of the obtained CpG signature, with the number of the significant CpGs used in the prediction selected by RFE. DNA methylation-based diagnostic methods are increasingly used in clinical practice [92–95]. Aberrant DNA methylation patterns have been observed and reported to play a role in neuroblastoma pathogenesis [24]. It is expected that epigenetic prognostics and therapeutics will be commonplace in the clinic in the near future.

It should be noted that this was a purely computational-based study. Although some of the DMGs identified here have already been associated with neuroblastoma, functional studies are needed to discover the role of certain DMGs in the pathogenesis of neuroblastoma.

## Conclusion

In this study, the authors demonstrated the utility of CpG methylation profiling for subgrouping neuroblastoma tumors and for predicting clinical events. Twenty-five CpGs capable of stratifying neuroblastoma samples were identified on the basis of *MYCN* amplification status, thereby demonstrating their utility as diagnostic indicators of *MYCN* amplification in neuroblastoma. The impact of the 25 CpGs methylation on the survival of patients with neuroblastoma was also evaluated using Cox regression analysis and eight CpGs associated with survival in neuroblastoma were identified. The DMGs reported in this study include some known genes associated with neuroblastoma as well as novel ones. This study furthers our understanding of the mechanisms of tumor progression in neuroblastoma. Therapeutic interventions may need to be targeted to patient subgroups to optimize treatment outcomes and improve survival.

### Financial & competing interests disclosure

The work reported herein was made possible through funding by the South African Medical Research Council through its Division of Research Capacity Development under the Mid-Career Scientist Programme from funding received from the South African National Treasury. This work was also supported by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation of South Africa (Grant ID: 64751). The content hereof is the sole responsibility of the authors and does not necessarily represent the official views of the SAMRC. The authors have no other relevant affiliations

## Summary points

- Amplification of *MYCN* in neuroblastoma is a predictor of poor prognosis.
- A total of 663 CpGs were differentially methylated between *MYCN* amplified and non-amplified groups.
- Fourteen highly methylated genes were identified.
- High clustering accuracy based on the 25 CpGs selected by recursive feature elimination was observed.
- High accuracy scores for *MYCN* amplification or non-amplification prediction were also found.
- Eight CpGs were associated with overall survival.
- Five CpGs were associated with event-free survival.
- Therapeutic interventions may need to be targeted to patient subgroups.

or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

### Data sharing statement

R and python scripts were implemented for this study and are available in https://github.com/SANBI-SA/NBMethyl.git.

### Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/

## References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Maris JM. Recent advances in neuroblastoma. *N. Engl. J. Med.* 362(23), 2202–2211 (2010).
2. Smith MA, Seibel NL, Altekruse SF *et al.* Outcomes for children and adolescents with cancer: challenges for the twenty-first century. *J. Clin. Oncol.* 28(15), 2625–2634 (2010).
3. Ward E, DeSantis C, Robbins A, Kohler B, Jemal A. Childhood and adolescent cancer statistics, 2014. *CA Cancer J. Clin.* 64(2), 83–103 (2014).
4. Zhang L, Lv C, Jin Y *et al.* Deep learning-based multi-omics data integration reveals two prognostic subtypes in high-risk neuroblastoma. *Front. Genet.* 9, 477 (2018).
5. Johnsen JI, Dyberg C, Wickström M. Neuroblastoma – a neural crest derived embryonal malignancy. *Front. Mol. Neurosci.* 12, 9 (2019).
6. Mueller S, Matthay KK. Neuroblastoma: biology and staging. *Curr. Oncol. Rep.* 11(6), 431–438 (2009).
7. Tonini G, Nakagawara A, Berthold F. Towards a turning point of neuroblastoma therapy. *Cancer Lett.* 326(2), 128–134 (2012).
8. Newman E, Nuchtern J. Recent biologic and genetic advances in neuroblastoma: implications for diagnostic, risk stratification, and treatment strategies. *Semin. Pediatr. Surg.* 25(5), 257–264 (2016).
9. Brodeur GM, Seeger RC, Schwab M, Varmus HE, Bishop JM. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* 224(4653), 1121–1124 (1984).

•• **In this report, amplification of MYCN was found to be a common event in untreated neuroblastoma, and correlates with advanced-stage disease.**

10. Seeger RC, Brodeur GM, Sather H *et al.* Association of multiple copies of the N-myc oncogene with rapid progression of neuroblastomas. *N. Engl. J. Med.* 313(18), 1111–1116 (1985).

•• ***MYCN* amplification was associated with rapid progression and worse prognosis in neuroblastoma.**

11. Baali I, Acar AE, Aderinwale TW, HafezQorani S, Kazan H. Predicting clinical outcomes in neuroblastoma with genomic data integration. *Biol. Direct* 13(1), 20 (2018).
12. Mossé YP, Laudenslager M, Longo L *et al.* Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature* 455(7215), 930–935 (2008).
13. Cheung NV, Zhang J, Lu C *et al.* Association of age at diagnosis and genetic mutations in patients with neuroblastoma. *JAMA* 307(10), 1062–1071 (2012).
14. Valentijn LJ, Koster J, Zwijnenburg DA *et al.* TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors. *Nat. Genet.* 47(12), 1411–1414 (2015).
15. Peifer M, Hertwig F, Roels F *et al.* Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* 526(7575), 700–704 (2015).

16. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. *Science* 339(6127), 1546–1558 (2013).

17. Sweet-Cordero EA, Biegel JA. The genomic landscape of pediatric cancers: implications for diagnosis and treatment. *Science* 363(6432), 1170–1175 (2019).

• **This report demonstrates the need for multi-omic and integrative studies to address the knowledge gaps in pediatric cancers.**

18. Ratner N, Brodeur GM, Dale RC, Schor NF. The "neuro" of neuroblastoma: neuroblastoma as a neurodevelopmental disorder. *Ann. Neurol.* 80(1), 13–23 (2016).

19. Decock A, Ongenaert M, Vandesompele J, Speleman F. Neuroblastoma epigenetics: from candidate gene approaches to genome-wide screenings. *Epigenetics* 6(8), 962–970 (2011).

20. Olsson M, Beck S, Kogner P, Martinsson T, Carén H. Genome-wide methylation profiling identifies novel methylated genes in neuroblastoma tumors. *Epigenetics* 11(1), 74–84 (2016).

•• **This study demonstrates the use of methylation profiling for patient stratification and reports hypermethylated genes found in advanced neuroblastoma tumors.**

21. Hassan W, Bakry M, Siepmann T, Illigens B. Association of RASSF1A, DCR2, and CASP8 methylation with survival in neuroblastoma: a pooled analysis using reconstructed individual patient data. *Biomed. Res. Int.* 2020, 7390473 (2020).

22. Decock A, Ongenaert M, Cannoodt R *et al.* Methyl-CpG-binding domain sequencing reveals a prognostic methylation signature in neuroblastoma. *Oncotarget* 7(2), 1960–1972 (2016).

23. Asada K, Abe M, Ushijima T. Clinical application of the CpG island methylator phenotype to prognostic diagnosis in neuroblastomas. *J. Hum. Genet.* 58(7), 428–433 (2013).

24. Abe M, Ohira M, Kaneda A *et al.* CpG island methylator phenotype is a strong determinant of poor prognosis in neuroblastomas. *Cancer Res.* 65(3), 828–834 (2005).

25. Hiyama E, Hiyama K, Yokoyama T *et al.* Rapid detection of MYCN gene amplification and telomerase expression in neuroblastoma. *Clin. Cancer Res.* 5(3), 601–609 (1999).

26. Oude Luttikhuis M, Iyer VK, Dyer S, Ramani P, McConville CM. Detection of MYCN amplification in neuroblastoma using competitive PCR quantitation. *Lab. Invest.* 80(2), 271–273 (2000).

27. Mathew P, Valentine MB, Bowman LC *et al.* Detection of MYCN gene amplification in neuroblastoma by fluorescence *in situ* hybridization: a pediatric oncology group study. *Neoplasia* 3(2), 105–109 (2001).

•• **This study demonstrates the detection of amplification of MYCN using fluorescence *in situ* hybridization technique.**

28. Yagyu S, Iehara T, Tanaka S *et al.* Serum-based quantification of MYCN gene amplification in young patients with neuroblastoma: potential utility as a surrogate biomarker for neuroblastoma. *PLoS ONE* 11(8), e0161039 (2016).

29. Di Lena P, Sala C, Prodi A, Nardini C. Missing value estimation methods for DNA methylation data. *Bioinformatics* 35(19), 3786–3793 (2019).

30. Morris TJ, Butcher LM, Feber A *et al.* ChAMP: 450 k chip analysis methylation pipeline. *Bioinformatics* 30(3), 428–430 (2014).

31. Ritchie ME, Phipson B, Wu D *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43(7), e47 (2015).

32. Chen EY, Tan CM, Kou Y *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14, 128 (2013).

33. Kuleshov MV, Jones MR, Rouillard AD *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44(W1), W90–W97 (2016).

34. Kassambara A, Mundt F. Factoextra R package (2020). https://cloud.r-project.org/package=factoextra

35. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2(3), 1–27 (2011).

36. Gómez S, Castellano G, Mayol G, Queiros A, Martín-Subero JI, Lavarino C. DNA methylation fingerprint of neuroblastoma reveals new biological and clinical insights. *Genom. Data* 5, 360–363 (2015).

• **This report details the methods and quality control of the GSE54719 dataset.**

37. Ackermann S, Cartolano M, Hero B *et al.* A mechanistic classification of clinical phenotypes in neuroblastoma. *Science* 362(6419), 1165–1170 (2018).

38. Schramm A, Koster J, Assenov Y *et al.* Mutational dynamics between primary and relapse neuroblastomas. *Nat. Genet.* 47(8), 872–877 (2015).

39. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33(1), 1–22 (2010).

40. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* 39(5), 1–13 (2011).

41. Tibshirani R, Bien J, Friedman J *et al.* Strong rules for discarding predictors in lasso-type problems. *J. R. Stat. Soc. Series B Stat. Methodol.* 74(2), 245–266 (2012).
42. Ng SW, Mitchell A, Kennedy JA *et al.* A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* 540(7633), 433–437 (2016).
43. Duan C, Wang H, Chen Y *et al.* Whole exome sequencing reveals novel somatic alterations in neuroblastoma patients with chemotherapy. *Cancer Cell Int.* 18, 21 (2018).
44. Zhang TJ, Xu ZJ, Gu Y *et al.* Identification and validation of prognosis-related DLX5 methylation as an epigenetic driver in myeloid neoplasms. *Clin. Transl. Med.* 10(2), e29 (2020).
45. Li Z, Li J, Ji D *et al.* Overexpressed long noncoding RNA Sox2ot predicts poor prognosis for cholangiocarcinoma and promotes cell proliferation and invasion. *Gene* 645, 131–136 (2018).
46. Giwa A, Fatai A, Gamieldien J, Christoffels A, Bendou H. Identification of novel prognostic markers of survival time in high-risk neuroblastoma using gene expression profiles. *Oncotarget* 11(46), 4293–4305 (2020).
47. Friedman DN, Henderson TO. Late effects and survivorship issues in patients with neuroblastoma. *Children (Basel)* 5(8), 107 (2018).
48. Walker C, Mojares E, Del Rio Hernandez A. Role of extracellular matrix in development and cancer progression. *Int. J. Mol. Sci.* 19(10), 3028 (2018).
49. Poltavets V, Kochetkova M, Pitson SM, Samuel MS. The role of the extracellular matrix and its molecular and cellular regulators in cancer cell plasticity. *Front. Oncol.* 8, 431 (2018).
50. Eble JA, Niland S. The extracellular matrix in tumor progression and metastasis. *Clin. Exp. Metastas.* 36(3), 171–198 (2019).
51. Nallanthigal S, Heiserman JP, Cheon D. The role of the extracellular matrix in cancer stemness. *Front. Cell. Dev. Biol.* 7, 86 (2019).
52. Jia M, Gao X, Zhang Y, Hoffmeister M, Brenner H. Different definitions of CpG island methylator phenotype and outcomes of colorectal cancer: a systematic review. *Clin. Epigenetics* 8, 25 (2016).
53. Teitz T, Wei T, Valentine MB *et al.* Caspase 8 is deleted or silenced preferentially in childhood neuroblastomas with amplification of MYCN. *Nat. Med.* 6(5), 529–535 (2000).
54. Fulda S, Poremba C, Berwanger B *et al.* Loss of caspase-8 expression does not correlate with MYCN amplification, aggressive disease, or prognosis in neuroblastoma. *Cancer Res.* 66(20), 10016–10023 (2006).
55. Zheng H, Zhang Y, Zhan Y *et al.* Expression of DR5 and c-FLIP proteins as novel prognostic biomarkers for non-small cell lung cancer patients treated with surgical resection and chemotherapy. *Oncol. Rep.* 42(6), 2363–2370 (2019).
56. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 144(5), 646–674 (2011).
57. Lowe SW, Lin AW. Apoptosis in cancer. *Carcinogenesis* 21(3), 485–495 (2000).
58. Pfeffer CM, Singh ATK. Apoptosis: a target for anticancer therapy. *Int. J. Mol. Sci.* 19(2), 448 (2018).
59. Wang Y, Luo H, Cao J, Ma C. Bioinformatic identification of neuroblastoma microenvironment-associated biomarkers with prognostic value. *J. Oncol.* 2020, 5943014 (2020).
60. Gonzalez-Moreno O, Boque N, Redrado M *et al.* Selenoprotein-P is down-regulated in prostate cancer, which results in lack of protection against oxidative damage. *Prostate* 71(8), 824–834 (2011).
61. Wang Q, Gong L, Dong R *et al.* Tissue microarray assessment of selenoprotein P expression in gastric adenocarcinoma. *J. Int. Med. Res.* 37(1), 169–174 (2009).
62. Meyer HA, Endermann T, Stephan C *et al.* Selenoprotein P status correlates to cancer-specific mortality in renal cancer patients. *PLoS ONE* 7(10), e46644 (2012).
63. Monteith GR, McAndrew D, Faddy HM, Roberts-Thomson SJ. Calcium and cancer: targeting $Ca^{2+}$ transport. *Nat. Rev. Cancer* 7(7), 519–530 (2007).
64. Satheesh NJ, Büsselberg D. The role of intracellular calcium for the development and treatment of neuroblastoma. *Cancers (Basel)* 7(2), 823–848 (2015).
65. Florea AM, Varghese E, McCallum JE *et al.* Calcium-regulatory proteins as modulators of chemotherapy in human neuroblastoma. *Oncotarget* 8(14), 22876–22893 (2017).
66. Farhangian P, Jahandoost S, Mowla SJ, Khalili M. Differential expression of long non-coding RNA SOX2OT in gastric adenocarcinoma. *Cancer Biomark.* 23(2), 221–225 (2018).
67. Qu F, Cao P. Long noncoding RNA SOX2OT contributes to gastric cancer progression by sponging miR-194-5p from AKT2. *Exp. Cell Res.* 369(2), 187–196 (2018).
68. Li Z, Jiang P, Li J *et al.* Tumor-derived exosomal lnc-Sox2ot promotes EMT and stemness by acting as a ceRNA in pancreatic ductal adenocarcinoma. *Oncogene* 37(28), 3822–3838 (2018).
69. Li Z, Li J, Ji D *et al.* Overexpressed long noncoding RNA Sox2ot predicts poor prognosis for cholangiocarcinoma and promotes cell proliferation and invasion. *Gene* 645, 131–136 (2018).

70. Shi XM, Teng F. Up-regulation of long non-coding RNA Sox2ot promotes hepatocellular carcinoma cell metastasis and correlates with poor prognosis. *Int. J. Clin. Exp. Pathol.* 8(4), 4008–4014 (2015).
71. Merlo GR, Zerega B, Paleari L, Trombino S, Mantero S, Levi G. Multiple functions of Dlx genes. *Int. J. Dev. Biol.* 44(6), 619–626 (2000).
72. Heo JS, Lee SG, Kim HO. Distal-less homeobox 5 is a master regulator of the osteogenesis of human mesenchymal stem cells. *Int. J. Mol. Med.* 40(5), 1486–1494 (2017).
73. Mitchell SM, Ross JP, Drew HR *et al.* A panel of genes methylated with high frequency in colorectal cancer. *BMC Cancer* 14, 54 (2014).
74. Karsli-Ceppioglu S, Dagdemir A, Judes G *et al.* The epigenetic landscape of promoter genome-wide analysis in breast cancer. *Sci. Rep.* 7(1), 6597 (2017).
75. Li C, Wang S, Yang C. Long non-coding RNA DLX6-AS1 regulates neuroblastoma progression by targeting YAP1 via miR-497-5p. *Life Sci.* 252, 117657 (2020).
76. Sun L, Zhao Y, Gu S, Mao Y, Ji C, Xin X. Regulation of the HMOX1 gene by the transcription factor AP-2δ with unique DNA binding site. *Mol. Med. Rep.* 10(1), 423–428 (2014).
77. Li Z, Xu X, Luo M *et al.* Activator protein-2β promotes tumor growth and predicts poor prognosis in breast cancer. *Cell. Physiol. Biochem.* 47(5), 1925–1935 (2018).
78. Ebert MPA, Tänzer M, Balluff B *et al.* TFAP2E–DKK4 and chemoresistance in colorectal cancer. *N. Engl. J. Med.* 366(1), 44–53 (2012).
79. Hoshi R, Watanabe Y, Ishizuka Y *et al.* Depletion of TFAP2E attenuates adriamycin-mediated apoptosis in human neuroblastoma cells. *Oncol. Rep.* 37(4), 2459–2464 (2017).
80. Lee JH, Byun DS, Lee MG *et al.* Frequent epigenetic inactivation of hSRBC in gastric cancer and its implication in attenuated p53 response to stresses. *Int. J. Cancer* 122(7), 1573–1584 (2008).
81. Roy D, Calaf G, Hei TK. Allelic imbalance at 11p15.5–15.4 correlated with c-Ha-ras mutation during radiation-induced neoplas-tic transformation of human breast epithelial cells. *Int. J. Cancer* 103(6), 730–737 (2003).
82. Kozlowski L, Filipowski T, Rucinska M *et al.* Loss of heterozy-gosity on chromosomes 2p, 3p, 18q21.3 and 11p15.5 as a poor prognostic factor in stage II and III (FIGO) cervical cancer treated by radiotherapy. *Neoplasma* 53(5), 440–443 (2006).
83. Panani AD, Ferti AD, Raptis SA, Roussos C. Novel recurrent structural chromosomal aberrations in primary bladder cancer. *Anticancer Res.* 24(5A), 2967–2974 (2004).
84. Zhao B, Bepler G. Transcript map and complete genomic sequence for the 310 kb region of minimal allele loss on chromosome segment 11p15.5 in non-small-cell lung cancer. *Oncogene* 20(56), 8154–8164 (2001).
85. Moskaluk CA, Rumpel CA. Allelic deletion in 11p15 is a common occurrence in esophageal and gastric adenocarcinoma. *Cancer* 83(2), 232–239 (1998).
86. Lee JH, Kang MJ, Han HY *et al.* Epigenetic alteration of PRKCDBP in colorectal cancers and its implication in tumor cell resistance to TNFα-induced apoptosis. *Clin. Cancer Res.* 17(24), 7551–7562 (2011).
87. Wikman H, Sielaff-Frimpong B, Kropidlowski J *et al.* Clinical relevance of loss of 11p15 in primary and metastatic breast cancer: association with loss of PRKCDBP expression in brain metastases. *PLoS One* 7(10), e47537 (2012).
88. Sharma P, Alsharif S, Bursch K *et al.* Keratin 19 regulates cell cycle pathway and sensitivity of breast cancer cells to CDK inhibitors. *Sci. Rep.* 9(1), 14650 (2019).
89. Yuan X, Yi M, Dong B, Chu Q, Wu K. Prognostic significance of KRT19 in lung squamous cancer. *J. Cancer* 12(4), 1240–1248 (2021).
90. Saloustros E, Perraki N, Apostolaki S *et al.* Cytokeratin-19 mRNA-positive circulating tumor cells during follow-up of patients with operable breast cancer: prognostic relevance for late relapse. *Breast Cancer Res.* 13(3), R60 (2011).
91. Kabir NN, Rönnstrand L, Kazi JU. Keratin 19 expression correlates with poor prognosis in breast cancer. *Mol. Biol. Rep.* 41(12), 7729–7735 (2014).
92. Sahm F, Schrimpf D, Stichel D *et al.* DNA methylation-based classification and grading system for meningioma: a multicentre, retrospective analysis. *Lancet Oncol.* 18(5), 682–694 (2017).
93. Capper D, Jones DTW, Sill M *et al.* DNA methylation-based classification of central nervous system tumours. *Nature* 555, 469–474 (2018).
94. Capper D, Stichel D, Sahm F *et al.* Practical implementation of DNA methylation and copy-number-based CNS tumor diagnostics: the Heidelberg experience. *Acta Neuropathol.* 136(2), 181–210 (2018).
95. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV. on behalf of the CCGA Consortium. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann. Oncol.* 31(6), 745–759 (2020).