Received: 6 January 2020 | Accepted: 13 February 2020

DOI: 10.1002/cac2.12016

CANCER COMMUNICATIONS Open Access

ORIGINAL ARTICLE

Gene signatures associated with genomic aberrations predict prognosis in neuroblastoma

Xiaoyan He[1,3†] | **Chao Qin**[2,3†] | **Yanding Zhao**[3] | **Lin Zou**[1] | **Hui Zhao**[4] | **Chao Cheng**[3,5,6]

[1]Center for Clinical Molecular Medicine, Ministry of Education Key Laboratory of Child Development and Disorders, National Clinical Research Center for Child Health and Disorders, China International Science and Technology Cooperation Base of Child Development and Critical Disorders, Chongqing Key Laboratory of Pediatrics, Children's Hospital of Chongqing Medical University, Chongqing 400014, P. R. China

[2]Beijing Key Lab of Traffic Data Analysis and Mining, School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, P. R. China

[3]Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Lebanon, NH 03766, USA

[4]School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong 999077, P. R. China

[5]Department of Medicine, Baylor College of Medicine, Houston, TX 77030, USA

[6]Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX 77030, USA

Correspondence
Chao Cheng, Ph.D., Department of Medicine, Institute for Clinical and Translational Research, Baylor College of Medicine, Room ICTR 100D, One Baylor Plaza, Baylor College of Medicine, Houston, TX 77030, USA.
Email: chao.cheng@bcm.edu

Funding information
American Cancer Society, Grant/Award Number: IRG-82-003-30; National Natural Science Foundation of China, Grant/Award Number: 81201543; National Center for Advancing Translational Sciences, Grant/Award Number: KL2TR001088

[†]These authors contributed equally to this work.

Abstract

Background: Neuroblastoma (NB) is a heterogeneous disease with respect to genomic abnormalities and clinical behaviors. Despite recent advances in our understanding of the association between the genetic aberrations and clinical features, it remains one of the major challenges to predict prognosis and stratify patients for determining personalized therapy in this disease. The aim of this study was to develop an effective prognosis prediction model for NB patients.

Methods: We integrated diverse computational analyses to define gene signatures that reflect MYCN activity and chromosomal aberrations including deletion of chromosome 1p (Chr1p_del) and chromosome 11q (Chr11q_del) as well as chromosome 11q whole loss (Chr11q_wls). We evaluated the prognostic and predictive values of these signatures in seven NB gene expression datasets (the number of samples ranges from 94 to 498, with a total of 2120) generated from both RNA sequencing and microarray platforms.

Abbreviations: AUC, under the curve; Chr11q_del, chromosome 11q deletion; Chr11q_wls, chromosome 11q whole loss; Chr17q_gain, chromosome 17q gain; Chr17q_wgn, chromosome 17q whole gain; Chr1p_del, chromosome 1p deletion; INRG, International Neuroblastoma Risk Group; INSS, International Neuroblastoma Staging System; LOH, loss of heterozygosity; NB, neuroblastoma.

Results: MYCN signature was a more effective prognostic marker than *MYCN* amplification status and MYCN expression. Similarly, the Chr1p_del score was more prognostic than Chr1p status. The activity scores of MYCN, Chr1p_del and Chr11q_del were associated with poor prognosis, while the Chr11q_wls score was linked to good outcome. We integrated the activity scores of MYCN, Chr1p_del, Chr11q_del, and Chr11q_wls and clinical variables into an integrative prognostic model, which displayed significant performance over the clinical variables or each genomic aberration alone.

Conclusions: Our integrative gene signature model shows a significantly improved forecast performance with prognostic and predictive information, and thereby can be served as a biomarker to stratify NB patients for prognosis evaluation and surveillance programs.

KEYWORDS

chromosome aberration, gene expression profile, *MYCN* gene, neuroblastoma, prognosis, survival analysis

1 | BACKGROUND

Neuroblastoma (NB), the most common solid pediatric tumor, accounts for approximately 7%-10% of pediatric malignancies and 15% of all children cancer death [1–3]. NB displays extensive heterogeneity both biologically and clinically [1,4], ranging from spontaneous regression in neonates to fatal progression in older children despite intensive multimodality treatment. According to the International Neuroblastoma Risk Group (INRG) classification system by incorporating clinical factors such as the patient age at diagnosis and tumor stage (the International Neuroblastoma Staging System, INSS), as well as tumor histopathology, DNA index, and *MYCN* gene status as prognostic factors, patients with NB can be classified into three risk groups (low-risk, intermediate-risk, and high-risk) and be given therapeutic stratification accordingly [5]. Generally, children with NB diagnosed before the age of 1.5 years have a better prognosis than those diagnosed at an older age, and patients at the early stage (stage 1, 2 or 4S) have a better outcome than those at the advanced stage (stage 3, 4) [1,3,6]. NB patients with *MYCN* amplification tend to have a worse prognosis than those without *MYCN* amplification [4,6]. Although the prognosis of NB has been greatly improved, less than 40% of patients with high-risk NB achieved long-term survival even with multiple, aggressive therapies and the recent inclusion of immunotherapy with antibodies targeted at NB-specific antigens, such as GD2 [7–9]. Due to the broadly divergent outcomes, it remains a challenge to develop methods for prognostic prediction so that the most effective therapeutic and management strategies can be designed for diverse subsets of NB.

Accumulated evidence has indicated that many genomic abnormalities, such as loss of chromosome 1p and 11q, gain of chromosome 17q, and *MYCN* gene amplification, are powerful prognostic markers and are strongly associated with clinical outcome [10–12]. However, multiple controversial findings on the prognostic value of gene markers have been observed. For example, *MYCN* amplification does not automatically trigger the increased expression of MYCN [13,14], while the overexpression of MYCN also occurs in the absence of *MYCN* amplification in NB patients [13–15], although *MYCN* amplification is usually considered to be linked to high expression of MYCN; segmental chromosomal aberrations, such as chromosome 11q deletion (Chr11q_del) and chromosome 17q gain (Chr17q_gain), are associated with poor prognoses, while numerical chromosomal aberrations, such as chromosome 11q whole loss (Chr11q_wls) and chromosome 17q whole gain (Chr17q_wgn), tend to be associated with good prognoses [16–18]. This highlights the importance of validating the predictive ability of these existing genetic markers and integrating newly identified prognostic markers into more sophisticated algorithms to further improve the predictive accuracy for clinical applications.

Gene expression profiling has been widely used for biomarker identification. Several well-documented studies have demonstrated that multigene signatures predict the prognosis of NB patients more effectively than traditional individual prognostic marker [19–22]. Gene signatures were usually defined as a set of selected genes identified from statistical analysis. However, many of these signatures are limited by their low reproducibility with poor predictive performance across different datasets; due to interpatient heterogeneity and complex interrelationships between genes. Notably, bioinformatics approaches have been developed to identify significant gene sets or pathway profiles involved in biological processes and diseases, including cancer, based

on gene expression data, which allow for quantitatively measuring the activation of intracellular signaling pathways [23–25]. Our group has previously defined a gene signature to quantify p53 pathway activity and verified that the *p53* gene signature was a better predictor of recurrence-free survival compared with *p53* mutation status in patients with early-stage lung adenocarcinoma [26] based on a rank-based statistical algorithm [27]. Additionally, multiple studies have shown that chromosome aberrations can lead to dramatic expression changes of the genes within the abnormal region and can be detected by gene expression profiles [28–30]. Unfortunately, for NB only a few copy number variation datasets have been generated, which makes it difficult to investigate the prognostic association of chromosomal loss/gain events in NB. In contrast, a large number of gene expression datasets have been published with large sample size and high-quality clinical information including patient outcome information. Therefore, it is desirable to develop computational methods to infer the activity of signaling pathway and the status of chromosome abnormalities for prognosis prediction based on these gene expression data.

In this study, we integrated different computational methods to infer MYCN activity and the statuses of several critical chromosome band aberrations, including chromosome 1p deletion (Chr1p_del), Chr11q_del, and Chr11q_wls, by using gene expression data and attempted to develop a computational model for predicting prognosis of NB patients.

2 | RESOURCES AND METHODS

2.1 | Neuroblastoma gene expression datasets

A summary of the NB gene expression datasets used in this study is provided in Supplementary Table S1. Raw data are available from the Gene Expression Omnibus (GEO) database (www.ncbi.nlm.nih.gov/geo; accession: GSE73517, GSE62564, GSE49710), the ArrayExpress database (www.ebi.ac.uk/arrayexpress/; accession: E-TABM-38, E-MTAB-179), the PREdiction of Clinical Outcomes from Genomic Profiles (PRECOG) portal (https://precog.stanford.edu/; accession: Berwanger_NB), and the Children's Oncology Group (COG) Therapeutically Applicable Research to Generate Effective Treatments (TARGET) data matrix portal (http://target.nci.nih.gov/dataMatrix/TARGET_DataMatrix.html; accession: TARGET_ NB).

The GSE73517 dataset was used to define gene signatures associated with *MYCN* amplification and five chromosome band aberrations (Chr1p_del, Chr11q_del, Chr11q_wls, Chr17q_gain, and Chr17q_wgn). The dataset contain 105 NB samples, providing gene expression profiles, patient clinical information and the statuses of *MYCN* amplification, Chr1p, Chr11q, and Chr17q [31].

The prognostic value of the resulting signatures was assessed in 6 NB gene expression datasets: the GSE62564, GSE49710, E-TABM-38, E-MTAB-179, Berwanger_NB, and TARGET_NB datasets. Among them, the GSE62564 and GSE49710 datasets contain gene expression profiles for the same set of 498 NB patients measured by two different methods with RNA sequencing and microarray analyses, respectively. The other datasets are independent microarray datasets, containing data from 198, 478, 94, and 249 patients, respectively. The data on *MYCN* amplification status are available for GSE49710, E-TABM-38, and E-MTAB-179 datasets. The data on Chr1p status are only available for the E-MTAB-179 dataset. In this article, we labelled datasets according to the first author of the published studies, as followings: Henrich for GSE73517 [31], Su for GSE62564 [32], Wang for GSE49710 [33], Westermann for E-TABM-38 [34], Oberthuer for E-MTAB-179 [35], Berwanger for Berwanger_NB [36], and Pugh for TARGET_NB [37].

2.2 | Defining *MYCN* gene signature and chromosome aberration signatures based on the Henrich dataset

The *MYCN* amplification gene signature was defined based on the Henrich dataset by comparing differential expression of genes between *MYCN* amplification and non-amplification samples while considering confounding variables [26,27]. For each gene, a logistic regression model was constructed using *MYCN* amplification as the response variable. We denote P as the probability of *MYCN* amplification, and the model is formulized as $\log(P/(1-P)) \sim \text{gene} + \text{age} + \text{gender} + \text{stage}$, where the predictor variables are the gene expression level, age at the time of diagnosis (age), gender, and INSS stage (stage 1, 2, 3, 4, or 4S). By applying these models to the Henrich data, we estimated the coefficients (β) and their statistical significance (P) for all genes. Given (β, P) values for all genes; we subsequently defined the MYCN signature, which was represented as a pair of weight profiles, w^+ and w^-, each containing the weights of all genes. For each gene i, w^+ and w^- were assigned in the following way: $w^+ = -\log(P_i)I(\beta_i > 0)$ and $w^- = -\log(P_i)I(\beta_i < 0)$. To avoid extreme values, the weights were trimmed at 10, and then transformed into a value within [0, 1], by subtracting the minimum value and then dividing by the range. If a gene i is more significantly up-regulated in *MYCN* amplification versus non-amplification samples, it will associate with a higher w^+ but w^- of zero. Conversely, a more significantly down-regulated gene will associate with a higher w^- but w^+ of zero. We selected all genes with $w^+ > 0.5$ and $w^- > 0.5$ as up- and down-regulated genes in MYCN amplification versus non-amplification samples, respectively.

According to the statuses of Chr1p, Chr11q, and Chr17q provided by the Henrich dataset, differentially expressed chromosome subbands associated with each chromosome deletion/gain event were obtained ($P < 0.05$). We then merged all genes located in the chromosome subbands as gene sets that can recapitulate the genomic statuses of Chr1p_del, Chr11q_del, Chr11q_wls, Chr17q_gain, and Chr17q_wgn, respectively. After filtering out genes that do not reflect chromosome aberration event (*e.g.*, genes not down-regulated or even up-regulated in Chr11q_del *vs.* Chr11q normal) and shared by two incompatible events for the chromosome band (*e.g.*, down-regulated genes in both Chr11q_del and Chr11q_wls, and up-regulated genes in both Chr17q_gain and Chr17q_wgn), we defined five gene sets for Chr1p_del, Chr11q_del, Chr11q_wls, Chr17q_gain, and Chr17q_wgn. If the number of genes in a set was less than 10, the gene set was halted. Finally, we obtained three final band gene sets to infer the statuses of Chr1p_del, Chr11q_del, and Chr11q_wls.

2.3 | Calculating patient-specific scores for MYCN activity and chromosome aberration events

Given the expression profiles for a group of NB samples, patient-specific *MYCN* activity scores were calculated based on the *MYCN* gene signature by using a previously published method called binding association with sorted expression (BASE) [27].

To calculate chromosome aberration scores, we used a method similar to the one for calculating the MYCN activity score, but with minor modification. Specifically, we replaced the weight profile pair (for *MYCN*) with a single profile containing all genes, with $b_i = 1$ if gene i is include in the gene set associated with a specific chromosome aberration (*e.g.*, Chr1p_del), and otherwise $b_i = 0$. These chromosome band aberration-specific profiles were used as inputs to the BASE algorithm, which output patient-specific scores that indicate the likelihood of the corresponding chromosome aberration event.

2.4 | Predicting patient survival using MYCN, Chr1p_del, Chr11q_del, and Chr11q_wls scores

Cox proportional hazard models were constructed to examine the effectiveness of the activity scores of MYCN, Chr1p_del, Chr11q_del, and Chr11q_wls in predicting patient survival. Patient samples were dichotomized into two groups by using the median of each activity score as a cutoff. Statistical significance was calculated by using Mann-Whitney *U* test. Correlation coefficient between MYCN expression and the MYCN score was calculated by using Spearman correlation, after MYCN expression was log transformed. A univariate Cox regression model was performed to examine difference in survival between the resultant two patient subgroups. A multivariate Cox regression model was used to investigate the effect of each score on survival after adjusting for potential confounding variables such as INSS stage, age, gender, and so on. The Kaplan-Meier method was used to plot survival curves. The difference between the survival curves of different groups was compared with significance being estimated by using a log-rank test. The R package "survival" (https://cran.r-project.org/web/packages/survival/index.html) was used to perform statistical analyses. Specifically, the "coxph" function was used to construct Cox proportional hazard models, the "survfit" function was used to create Kaplan-Meier survival curves, and the "survdiff" function was used to compare the difference between two survival curves. For each analysis, we performed it in all the datasets as long as the variables used for the analysis are available.

2.5 | Comparing prognostic models with different combinations of variables

Clinical variable combination (age, gender, INSS stage), the activity scores of MYCN, Chr1p_del, and Chr11q_del or Chr11q_wls as well as *MYCN* amplification status or Chr1p status were combined as the various models. The C-index was applied to compare the power of the different models. For each model, the C-index was estimated based on 10-fold cross validation. Specifically, the dataset was randomly divided into 10 groups of equal size. Each time, 9 groups were used as the training set, and then survival was predicted in the remaining group based on the trained models. The groups were rotated 10 times so that all samples were predicted exactly once, and then the C-index for each model was calculated. We repeated this procedure for 100 times (each time with a different randomization) to estimate the mean and the standard deviation of C-index for each model.

2.6 | MYCN target genes

MYCN target genes were downloaded from the Molecular Signatures Database (MSigDB, http://www.broadinstitute.org/gsea/msigdb/index.jsp), Section C3 "motif gene sets", subsection "transcription factor targets (TFT)", which provides a total of 271 MYCN predicted target genes (Supplementary Table S2).

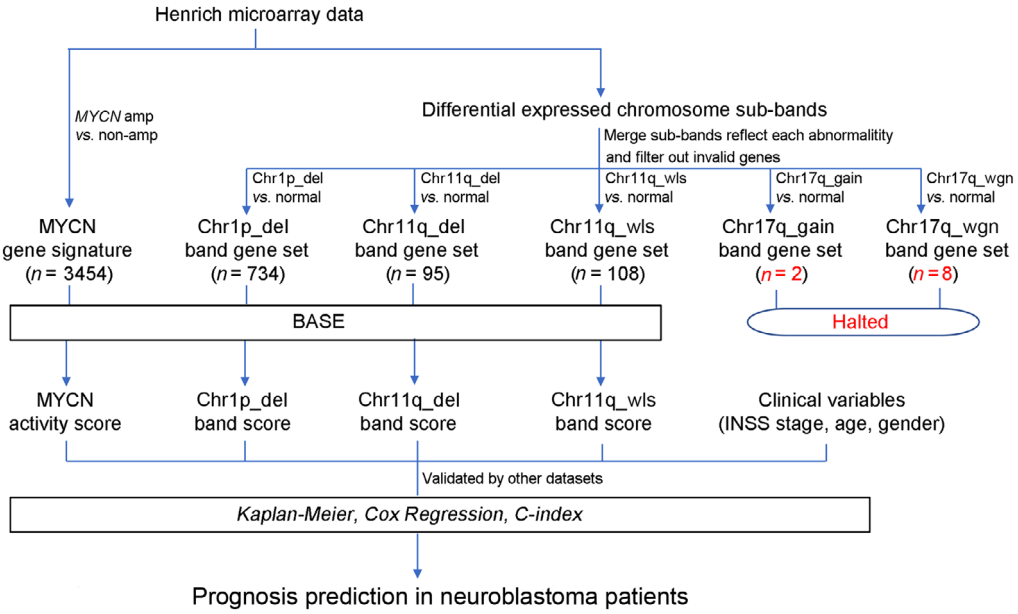

FIGURE 1 Schematic overview of the study. The Henrich microarray data were used to define *MYCN* gene signature and the gene sets for chromosome 1p deletion (Chr1p_del), chromosome 11q deletion (Chr11q_del), chromosome 11q whole loss (Chr11q_wls), chromosome 17q gain (Chr17q_gain), chromosome 17q whole gain (Chr17q_wgn). After removed the gene sets with less than 10 genes, MYCN gene signature and genes sets for Chr1p _del, Chr11q_del, and Chr11q_wls were used to calculate the scores of MYCN activity and corresponding chromosome abnormalities. Then the prognostic values of these scores were determined, and the predictive power of these scores alone or in combination with clinical variables, was compared in all datasets

3 | RESULTS

3.1 | Schematic overview of this study

To establish a powerful outcome prediction model for NB patients, we performed a series of analyses as illustrated in Figure 1. First, we defined gene signatures to recapitulate the genomic statuses of the *MYCN* gene and several chromatin aberration events (Chr1p_del, Chr11q_del, Chr11q_wls, Chr17q_gain, and Chr17q_wgn) by using the Henrich dataset, which provides gene expression profiles for 105 NB patients and the corresponding molecular marker statuses. Since the numbers of genes in the sets for Chr17q_gain and Chr17q_wgn were fewer than 10, these gene sets were discarded.

Next, we used the BASE algorithm to calculate the scores that can infer the status of *MYCN*, Chr1p_del, Chr11q_del, and Chr11q_wls in a patient-specific manner based on the corresponding gene signatures. We then investigated the prognostic value of these scores, specifically, in patients with or without *MYCN* amplification. We also assessed the association of these scores with clinical variables (age, gender, and INSS stage).

Finally, a series of prediction models were established by integrating the clinical variables and the activity score of MYCN, Chr1p_del, Chr11q_del, or Chr11q_wls to stratify patients and predict prognosis in NB.

3.2 | Association of MYCN gene signature with MYCN activity

Statistical analysis of the Henrich dataset identified 671 and 2783 genes that were up- and down-regulated in *MYCN* amplification *versus* non-amplification samples by using $w^+ > 0.5$ and $w^- > 0.5$ as cutoff, respectively (Supplementary Tables S3 and S4). As shown on Figure 2a, we also investigated the overlap of these genes with the 271 MYCN target genes that were available from the MSigDB C3 TFT (Supplementary Table S2). As expected, genes up-regulated in *MYCN* amplification samples were enriched for MYCN regulatory targets.

Next, we investigated whether the MYCN score can inform the *MYCN* status in other datasets. The MYCN scores of *MYCN* amplification samples were significantly higher than those of *MYCN* non-amplification samples in the Wang (Figure 2b), Henrich (Supplementary Figure S1a), Su (Supplementary Figure S1b), Westermann (Supplementary Figure S1c), and Oberthuer datasets (Supplementary Figure S1d). The MYCN score can effectively discriminate *MYCN* amplification from non-amplification samples with high accuracy, as indicated by the area under the curve (AUC) score of 0.99, 0.99, 0.94, and 0.98 in the Wang (Figure 2c), Su (Supplementary Figure S1e), Westermann (Supplementary Figure S1f), and Oberthuer datasets (Supplementary Figure S1g), respectively. These results supported the

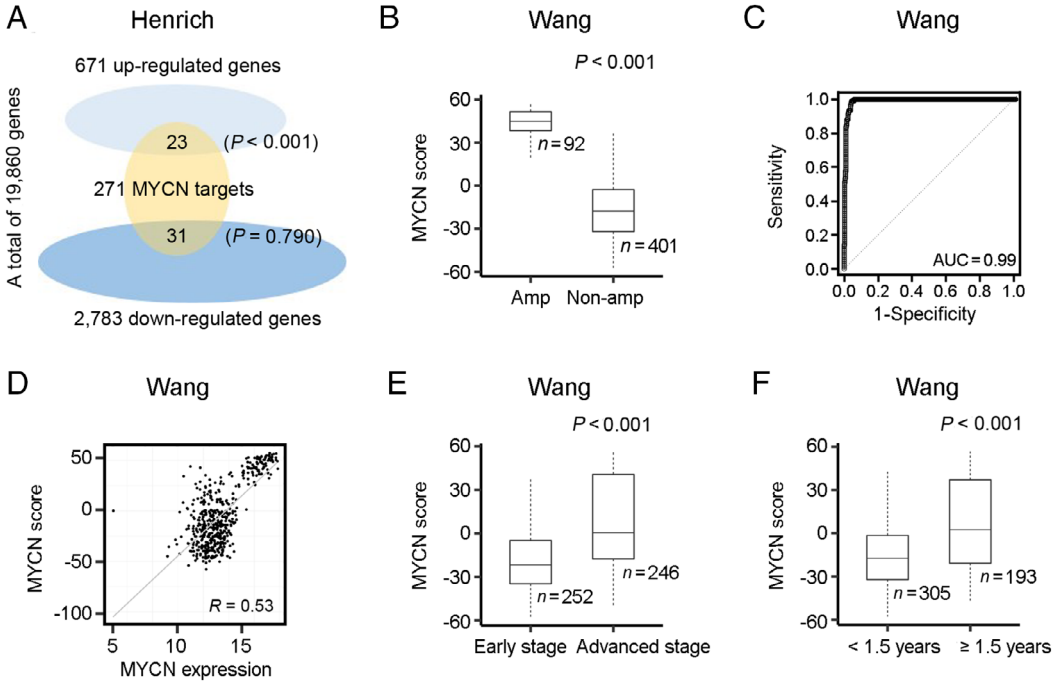

FIGURE 2 Associations of the MYCN activity score with *MYCN* amplification status, MYCN expression, clinical stage, and age. A. Genes up- and down-regulated in NB samples with *MYCN* amplification were enriched in MYCN target genes in the Henrich dataset. B. The MYCN activity score is significantly higher in patients with *MYCN* amplification than in those without *MYCN* amplification in the Wang dataset. C. ROC curve illustrates the diagnostic ability of the MYCN score in the Wang dataset. D. The MYCN score is correlated with MYCN expression level in the Wang dataset. E. The MYCN score is significantly higher in patients at advanced stages than in those at early stages in the Wang dataset. F. The MYCN score is significantly higher in patients older than 1.5 years than in those less than 1.5 years old in the Wang dataset. ROC, receiver operating characteristic curve; AUC, the area under the curve

MYCN score as a biomarker with strong ability to discriminate NB patients with *MYCN* amplification from those without.

We then tested the relationship between the MYCN score and MYCN expression. As shown on Figure 2d, the MYCN score was positively correlated with MYCN expression, although the correlation was moderate ($R = 0.53$) in the Wang dataset. Similar correlations were recognized in the other six datasets (Supplementary Figure S1h-m).

We next examined the association between the MYCN score and the clinical variable combination. In the Wang dataset, we observed a significantly higher MYCN score in patients at advanced stages compared with those at early stages ($P < 0.001$) (Figure 2e), and a higher MYCN score in patients aged ≥ 1.5 years compared with those aged < 1.5 years ($P < 0.001$) (Figure 2f). Similar results were found in other datasets (Supplementary Figure S2a-j). No significant association between the MYCN score and gender was found in all datasets (Supplementary Figure S2k-o).

Overall, these results indicated that the MYCN score was indicative of *MYCN* amplification, and was associated with clinical stage and age, two established prognostic clinical variables.

3.3 | Prognostic values of MYCN score, *MYCN* amplification status, MYCN expression, and clinical variables

Using the Wang dataset, we dichotomized patients into MYCN score-low and -high groups using the median as the cutoff and compared their survival. As shown on Figure 3a, patients with high MYCN score had significantly shorter survival than those with low MYCN score (hazard ratio [HR] = 3.3, 95% confidence interval [CI] = 2.4-4.6, $P < 0.001$). Similar results were observed in other datasets (Supplementary Figure S3a-e). Following that, we separated patients into two groups based on their *MYCN* amplification status and examined the predictive power of the MYCN score in each group. We observed a significant difference in survival between MYCN score-high and -low patients in the *MYCN* non-amplification group (HR = 2.4, 95% CI = 1.7-3.6, $P < 0.001$) (Figure 3b), but not in the *MYCN* amplification group (HR = 0.8, 95% CI = 0.5-1.3, $P = 0.372$, data not shown) in the Wang dataset. These results were further confirmed in other datasets (Supplementary Figure S3f-k). These results suggest that the MYCN activity score contains prognostic value only in patients without *MYCN* amplification.

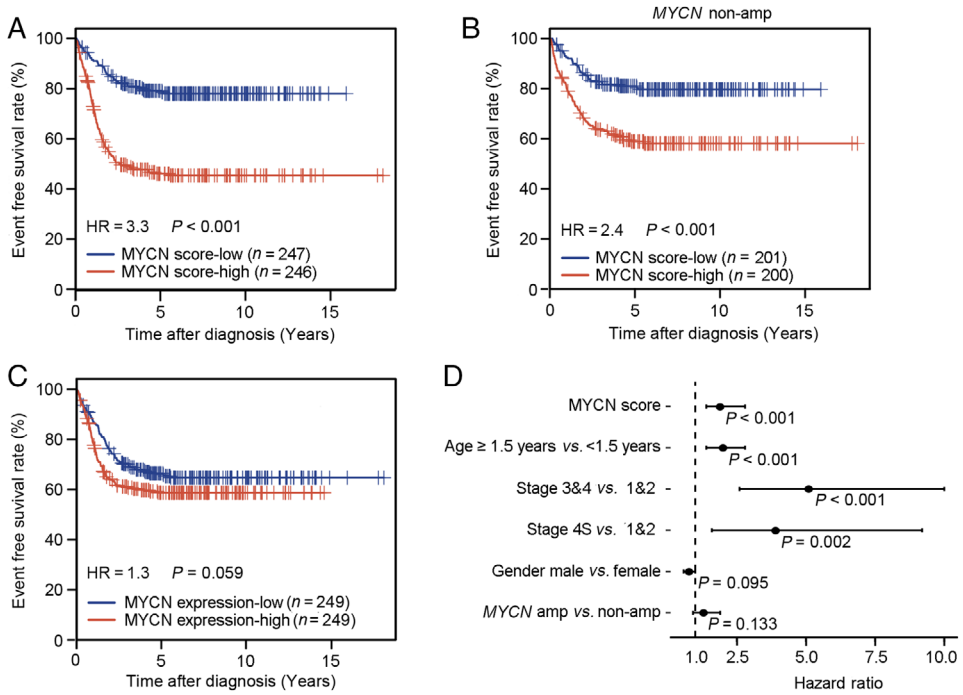

FIGURE 3 Prognostic value of the MYCN score in the Wang dataset after adjusted by clinical variables. A. Patients with high MYCN score had significantly shorter survival than those with low MYCN score. Five patients with MYCN gain were not included in the analysis. B. Survival prediction power of the MYCN score in the *MYCN* non-amplification group. C. MYCN expression was not associated with patients' survival. The median MYCN score (A and B) and the median MYCN expression (C) were used as the cutoff to dichotomize patients into -low and -high groups. HR, hazard ratio. D. A forest plot shows the hazard ratio and *P* value of the MYCN score and several clinical variables estimated by using the multivariate Cox regression model. All analyses were based on the Wang dataset

In contrast, the association between MYCN mRNA level and prognosis was not significant in the Wang dataset (HR = 1.3, 95% CI = 1.0-1.8, $P = 0.059$) (Figure 3c). Similarly, MYCN expression was not associated with survival in the Westermann and Pugh datasets (Supplementary Figure S3l-m), but was significantly associated with survival in the Su, Oberthuer, and Berwanger datasets (Supplementary Figure S3n-p). Thus, MYCN expression cannot be used as a convincing prognostic marker.

We next investigated whether the MYCN score was prognostic after adjusting for clinical variables. Specifically, we constructed a multivariate Cox regression model for the Wang dataset that included dichotomized MYCN score, age, INSS stage, gender, and *MYCN* status. As shown on Figure 3d, the MYCN score remained significant for predicting patient survival even after considering clinical variables and *MYCN* status (HR = 2.4, 95% CI = 1.7-3.6, $P < 0.001$). To further confirm this, the same analysis was performed in the other four datasets. As shown in Supplementary Table S5, the MYCN score was still prognostic in the Su, Oberthuer, and Pugh datasets. In the Westermann dataset, high MYCN score was also associated with short survival (HR = 2.0, 95% CI = 1.0-3.9), but after adjustment it was not significant ($P = 0.056$), presumably due to the relatively small sample size of this dataset.

Taken together, these results indicated that the MYCN score is a more effective prognostic marker than *MYCN* amplification status and MYCN expression.

3.4 | Association of the Chr1p_del score with patient survival

Other than the *MYCN* amplification status, several chromosomal band loss/gain events have been observed with high frequency in NB, for which, however, associations with prognosis have not been investigated systematically. As a proof of concept, we developed gene signatures to capture these chromosome events and examined their prognostic values across different gene expression datasets. Specifically, we defined gene signatures for Chr1p_del, Chr11q_del, and Chr11q_wls (Supplementary Table S6) based on the Henrich dataset. We first investigated whether the Chr1p_del score can inform Chr1p status using the Oberthuer dataset, which provided Chr1p status determined by fluorescence in situ hybridization (FISH) and/or polymerasechain reaction (PCR) [35]. As shown on Figure 4a, we observed a significant increase of the Chr1p_del score in patients with loss of heterozygosity at chromosome 1p (LOH1p) compared with those with normal Chr1p status ($P < 0.001$), with an AUC of 0.95 when

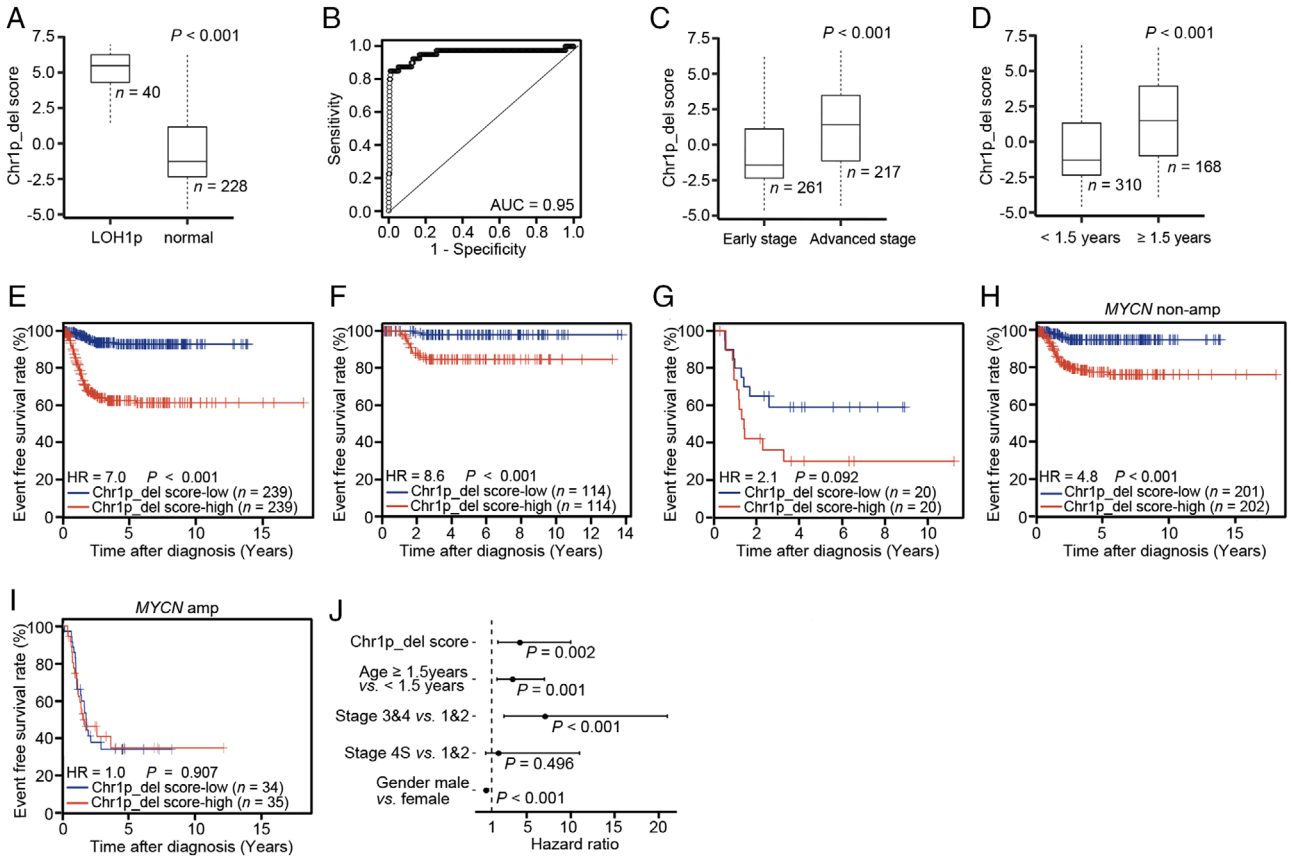

FIGURE 4 Association of the Chr1p_del score with patient prognosis in the Oberthuer dataset. A. The Chr1p_del score is significantly higher in patients with LOH1p than in those without Chr1p aberration. B. ROC curve demonstrates the diagnostic ability of the Chr1p_del score. C. The Chr1p_del score is significantly higher in patients at advanced stages than in those at early stages. D. The Chr1p_del score is significantly higher in patients aged ≥1.5 years than in those aged < 1.5 years. E. High Chr1p_del score was associated with significantly shorter survival as compared with low Chr1p_del score. F and G. survival prediction ability of the Chr1p_del score among patients in the Chr1p normal (F) and LOH1p (G) groups. H and I. Prediction power of the Chr1p_del score in patients without (H) or with *MYCN* amplification (I) groups. The median Chr1p_del score in each group was used as the cutoff to divide patients into Chr1p_del score-low and -high groups. HR, hazard ratio. J. The forest plot exhibits the hazard ratio and *P* value of the Chr1p_del score and the clinical variables evaluated by the multivariate Cox regression model. All analyses were based on the Oberthuer dataset. Chr1p_del, chromosome 1p deletion. ROC, receiver operating characteristic curve; AUC, the area under the curve

the Chr1p_del score was used to discriminate 40 with LOH1p from 268 samples (Figure 4b). The Chr1p_del score was significantly higher in patients at advanced stages than in those at early stages ($P < 0.001$) (Figure 4c), and higher in patients aged ≥ 1.5 years than in those aged < 1.5 years ($P < 0.001$) (Figure 4d). No significant change of the Chr1p_del score was found between male and female patients (data not shown). Similar results were obtained in other datasets (data not shown). These results indicated that the Chr1p_del score derived from the gene signature could correctly reflect Chr1p status in NB samples and were associated with clinical variables such as INSS stage and age.

We further tested whether the Chr1p_del score is predictive of patient survival. In the Oberthuer dataset, patients with high Chr1p_del score had significantly shorter survival than those with low Chr1p_del score (HR = 7.0, 95% CI = 3.9-12.0, $P < 0.001$) (Figure 4e). Similar results were observed in other datasets (Supplementary Figure S4a-e). We then divided patients into two groups on the basis of their reported Chr1p status and examined the association of the Chr1p_del score with patient survival in each group. Interestingly, we observed a significant difference between Chr1p_del score-low and -high patients in the Chr1p normal group (HR = 8.6, 95% CI = 2.0-38.0, $P < 0.001$, Figure 4f), despite no difference in the LOH1p group ($P = 0.092$, Figure 4g). This indicated that the Chr1p_del score could capture not only the Chr1p status but also the downstream effect of Chr1p_del that might also be caused by other mechanisms, *e.g.*, focal deletion of driver genes in Chr1p or epigenetic modifications of this chromosome region.

Because LOH1p usually associates with *MYCN* amplification in NB patients [1,3,6,10,12], we investigated the

association between the Chr1p_del score and *MYCN* status of NB patients. Patients with *MYCN* amplification showed significantly higher Chr1p_del scores than those without *MYCN* amplification (data not shown). Next, we separated patients into *MYCN* amplification and non-amplification groups and performed survival analyses in each group. As shown on Figure 4h-i, high Chr1p_del score was linked to short survival for patients without *MYCN* amplification (HR = 4.8, 95% CI = 2.3-9.9, $P < 0.001$), but no significant association was found for patients with *MYCN* amplification in the Oberthuer dataset. Similar results were observed in other datasets (Supplementary Figure S4f-k), suggesting that the Chr1p_del score provides additional prognostic value only in patients without *MYCN* amplification.

We next evaluated the prognostic value of the Chr1p_del score in the Oberthuer dataset by using a multivariate Cox regression model after considering the following variables: dichotomized Chr1p_del score, age, INSS stage, and gender. As shown on Figure 4j, the Chr1p_del score was significantly associated with patient survival even after adjusting for the clinical variables (HR = 4.2, 95% CI = 1.7-10.0, $P = 0.002$). Similar results were obtained in other datasets (Supplementary Table S7). We also assessed the prognostic values of the Chr1p_del score in *MYCN* amplification and non-amplification -groups separately. As shown in Supplementary Figure S4l and Supplementary Table S7, the Chr1p_del score remained significant for predicting survival even after considering clinical variables and *MYCN* status in the Oberthuer (HR = 2.8, 95% CI = 1.1-7.4, $P = 0.035$) and Wang datasets (HR = 2.1, 95% CI = 1.5-3.1, $P < 0.001$), despite no significance in the Su dataset (HR = 1.4, 95% CI = 1.0-1.9, $P = 0.071$).

Altogether, the Chr1p_del score can be used to infer Chr1p status and is predictive of patient survival. Its prognostic value remains significant after considering the indicated clinical variables and *MYCN* amplification status in NB patients.

3.5 | Inference of Chr11q status based on a combination of two gene signatures

Both Chr11q_del and Chr11q_wls occur in NB patients with high frequencies in the Henrich dataset (Supplementary Table S1), and, interestingly, the former is associated with poor prognosis but the latter with good prognosis, compared to normal Chr11q [16–18]. Notably, loss of Chr11q has been shown to be associated with survival only in patients lacking *MYCN* amplification [3,38,39]. Here, we defined two non-overlapping gene sets to infer Chr11q status and investigated their prognostic values in patients without *MYCN* amplification.

Given that the Chr11q_del and Chr11q_wls scores were derived from the same chromosome region, we tested whether they can reflect their own statuses and distinguish one another. To do this, we divided patients into four groups based on the Chr11q_del and Chr11q_wls scores by using median of the two scores as cutoffs, and then examined whether the two scores can inform reported Chr11q status in the Henrich dataset. As shown on Figure 5a, patients with high Chr11q_del but low Chr11q_wls scores were more likely to have Chr11q_del, whereas those with high Chr11q_wls but low Chr11q_del scores were more likely to have Chr11q_wls. When both scores were low, the patients were more likely to have normal Chr11q. Some patients had high Chr11q_del and Chr11q_wls scores, which indicate abnormal Chr11q status, but it is unclear whether this is associated with a whole loss or a deletion event.

We then examined the prognostic power of the Chr11q_del and Chr11q_wls scores for patients without *MYCN* amplification in the Wang dataset. High Chr11q_del score was associated with significantly shorter survival (HR = 1.8, 95% CI = 1.3-2.6, $P = 0.001$) (Figure 5b), whereas high Chr11q_wls score was related to markedly longer survival (HR = 0.3, 95% CI = 0.2-0.5, $P < 0.001$) (Figure 5c) among these patients. Similar results were observed in other datasets (Supplementary Figure S5a-f). These results indicated that Chr11q_del was linked to poor outcome, whereas Chr11q_wls was associated with good outcome in NB patients without *MYCN* amplification.

By combining the Chr11q_del and Chr11q_wls scores, we divided patients into four groups using the median of the two scores as cutoffs and tested the survival predictive power of the two scores. The survival curves of patients without *MYCN* amplification in the four groups could be distinguished (Figure 5d and Supplementary Table S8) in the Wang dataset. Similar results were found in other datasets (Supplementary Figure S5g-i). As shown in Supplementary Table S8, Chr11q_del was linked to a worse prognosis, while Chr11q_wls was related to a better prognosis in patients without *MYCN* amplification when compared to Chr11q normal. Notably, Chr11q_wls was associated with a better outcome compared to Chr11q_del and Chr11q_normal. Taken together, these results indicated that the scores of Chr11q_del and Chr11q_wls inferred by our algorithm were predictive of patient survival and could be distinguished from one another.

3.6 | An integrated model for prognostic prediction

After validating the contribution of each score of individual genomic alterations to prognosis prediction, we investigated whether the combination of these scores could serve as a more powerful predictor. To this end, we developed a series of predictor models that incorporated different combinations of clinical variables and genomic signature scores (the MYCN,

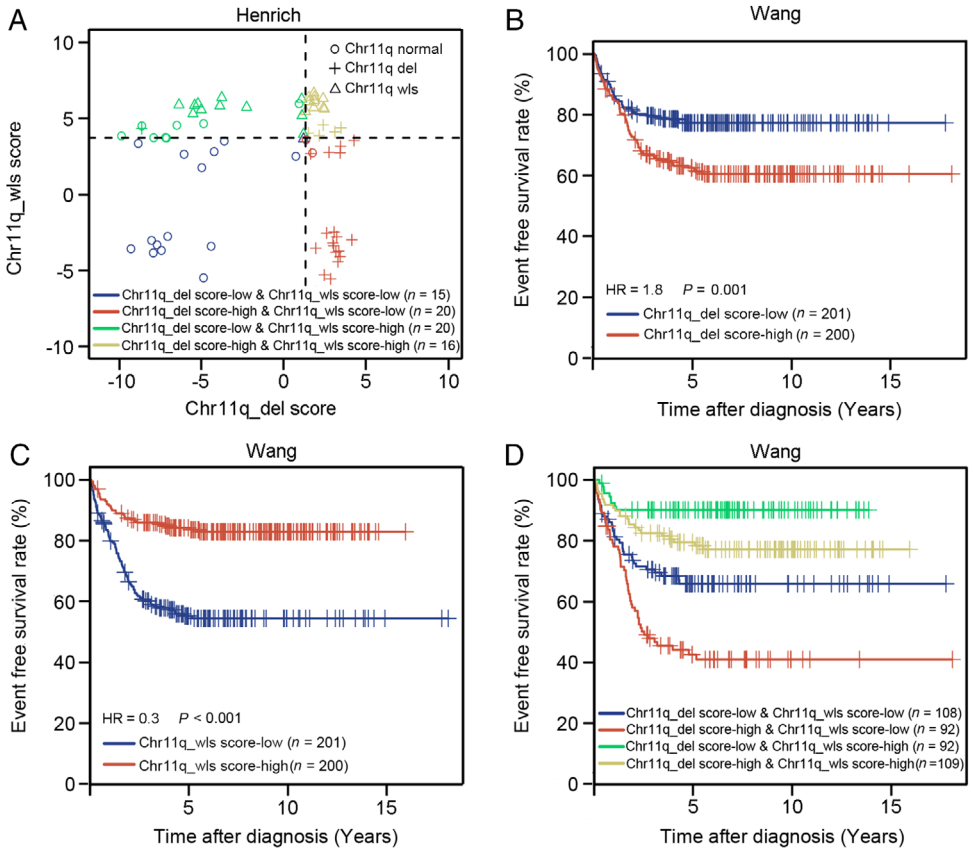

FIGURE 5 Association of the Chr11q_del and Chr11q_wls scores with prognosis of neuroblastoma patients without *MYCN* amplification. A. The scores of Chr11q_del and Chr11q_wls inferred by our algorithm can distinguish from each other in the Henrich dataset. The median scores of Chr11q_del and Chr11q_wls were used as the cutoffs to dichotomize patients into four groups, which are represented by different colors. The patients with low Chr11q_del score and low Chr11q_wls score as the normal group indicated by blue; the patients with high Chr11q_del score and low Chr11q_wls score as the Chr11q_del group indicated by red; the patients with low Chr11q_del score and high Chr11q_wls score as the Chr11q_wls group indicated by green; the patients with high Chr11q_del score and high Chr11q_wls score as the unknown group indicated by yellow. Shapes indicated different Chr11q statuses provided by the Henrich dataset. B. High Chr11q_del score is associated with shorter survival in patients without *MYCN* amplification in the Wang dataset. C. High Chr11q_wls score is linked to longer survival in patients without *MYCN* amplification in the Wang dataset. For each group, the median Chr11q_del or Chr11q_wls score was used as the cutoff to divide patients into score-low and -high groups. HR, hazard ratio. D. The scores of Chr11q_del and Chr11q_wls were predictive of patients' survival and distinguished from each other in patients without *MYCN* amplification in the Wang dataset. The median scores were used as the cutoffs to dichotomize patients into four groups. Chr11q_del, chromosome 11q deletion; Chr11q_wls, chromosome 11q whole loss

Chr1p_del, Chr11q_del, or Chr11q_wls score), with or without the status of genetic aberration (*MYCN* status, Chr_1p status).

Figure 6 and Supplementary Table S9 show a comparison of the discrimination powers of different models by using C-index in the Wang dataset. All models that included clinical variables and each score of genomic aberration were significantly better than either clinical variables alone or each score of genomic aberration alone. Moreover, the model that comprised the MYCN score and clinical variables outperformed the model with *MYCN* status and clinical variables. More importantly, the combination of the MYCN, Chr1p_del, Chr11q_del, or Chr11q_wls score with clinical variables, named here as the final models, were significantly improved over either individual predictor. Notably, the C-index for the two final models were significantly increased by 5.3% and 4.6% respectively compared with the model involving only clinical variables, although they changed only slightly compared with models that contained clinical variables and the MYCN score. Similar results were observed in other datasets (Supplementary Table S9), with a significant improvement in the performance of the Chr1p_del score over that of the Chr1p status in the Oberthuer dataset, in which only the Chr1p status is available.

Overall, the combination of the MYCN, Chr1p_del, and Chr11q_del or Chr11q_wls score with clinical variables exhibits significant improvement over either the clinical variables or genomic aberrations alone.

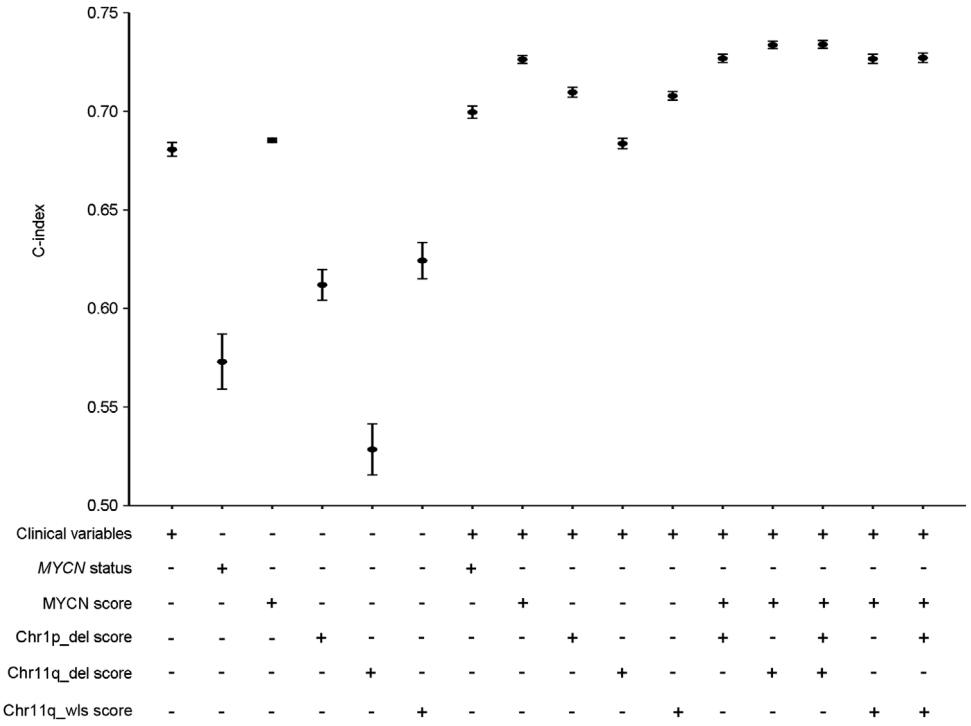

FIGURE 6 The combined gene sets as integrated models for prognostic prediction. The predictive accuracy of various Cox models, which combined the scores of genomic alterations (MYCN, Chr1p_del, Chr11q_del, and Chr11q_wls) with or without clinical variables (age, gender, and INSS stage) and the status of genetic aberration (MYCN status) were compared by using C-index analyses in the Wang dataset. "+" means that the variable was included in the model, and "-" means that the variable was not included in the model

4 | DISCUSSION

In the present study, we developed a computational framework to define gene signatures to infer several clinically important genomic aberrations in NB, including the *MYCN* gene amplification, Chr1p_del, Chr11q_del, and Chr11q_wls. In addition to cytogenetic assays, these signatures can be measured by using traditional molecular biological methods such as PCR and FISH, or by using transcriptomic profiling analyses such as microarray and RNA-seq. Here, we defined a gene signature to calculate the MYCN activity score in NB samples by using our previously described algorithm [26,27]. Our rationale is that the increased regulatory activity of MYCN caused by *MYCN* gene amplification will be eventfully revealed by the alteration of expression of downstream genes. More importantly, the gene signature will capture not only the *MYCN* amplification event but also the changes of MYCN activity caused by other mechanisms (*e.g.*, DNA methylation).

To define gene signatures associated with chromosome aberration events, we applied a different strategy. When a specific chromatin band is deleted, we expect to observe the trend of down-regulation of all genes located in this band with the following exceptions: 1) genes that are intensively regulated by cellular signaling or posttranscriptional events, and 2) genes that are distant from the hot spots associated with the given chromatin band deletion or gain event. Based on this rationale, we defined a gene signature for a chromosome aberration event as the set of genes that are located in the given chromosome regions (i.*e.*, Chr1p_del, Chr11q_del, and Chr11q_wls).

Importantly, our results demonstrated that the MYCN score and chromosome band scores of Chr1p_del, Chr11q_del, and Chr11q_wls inferred by our algorithm reflected the statuses of these genomic features and were predictive of the prognosis of NB patients. Specifically, the MYCN score showed significant superiority in prognostic prediction compared with *MYCN* status and MYCN expression, in consistent with a previous report [15]. The Chr1p_del score showed a similar pattern when compared with Chr1p status. Notably, the Chr11q_del and Chr11q_wls scores were distinguishable from one another, although both were derived from different sets of genes located in Chr11q. In line with previous reports [16–18], the segmental alteration Chr11q_del was associated with short survival, while the numerical abnormality Chr11q_wls was linked to long survival. Importantly, the MYCN score and the Chr1p_del score remain prognostic after adjusting clinical variables. Hence, these genomic feature-based signatures represent effective biomarkers that confer to more accurate prognostic prediction in NB. Although gene expression data were used to define gene signatures associated with chromosome aberration events in the

present study, a recently published study, which contained a large CNV profile of 556 NB cases, gave rise to our concerns [40]. We analyzed the data by using Cox regression analysis and found that 34 chromosomal loss/gain events exhibited significant prognostic values (details shown in the Supplementary Table S10). More high-quality CNV datasets with large sample size will be expected. The complex association between *MYCN* amplification and other genomic alterations also should be noted [41–44]. Particularly, Chr1p_del, Chr11q_del, and *MYCN* amplification are not independent genomic events. Chr1p_del, which presents in 25%-35% of NB patients, was controversial as an independent prognostic indicator because of its close relation with *MYCN* amplification, whereas Chr11q_del, which occurs in 35%-45% of NB patients as an adverse clinical feature with older age of disease onset, late disease stage, and unfavorable histology, is inversely associated with *MYCN* amplification and Chr1p_del [1,3,10,12,45,46]. However, these events are not mutually exclusive, patients with Chr1p_del without *MYCN* amplification [3] and those with concurrent *MYCN* amplification and Chr11q deletion [47] have been reported and were also found in the Henrich dataset (Supplementary Table S11). The presence of Chr1p_del without *MYCN* amplification was also observed in the Oberthuer dataset (Supplementary Table S11). We thus evaluated the prognostic power of the scores of MYCN, Chr1p_del, Chr11q_del, and Chr11q_wls in NB patient groups stratified based on *MYCN* amplification status. We observed that the scores of MYCN, Chr1p_del, Chr11q_del, and Chr11q_wls inferred from our analysis were predictive of patient survival in *MYCN* normal samples. This indicates that they provide additional prognostic information to *MYCN* status and other clinical variables.

Considering that NB is a molecularly defined disease [16,17,19,22,48–51], various genetic features have been integrated to increase the sensitivity and specificity for risk assessment and outcome prediction. Here, we evaluated a series of prediction models that integrated inferred genomic scores of MYCN, Chr1p_del, Chr11q_del, and Chr11q_wls, the status of genetic aberration (*MYCN* status or Chr1p status), and clinical variables. Incorporation of the genetic biomarker, *MYCN* amplification status, achieved more accurate prognostic prediction compared with the model solely based on clinical variables, which can be further improved by using the MYCN score inferred from gene signature analysis. The addition of more inferred genomic scores resulted in slight increase of prediction accuracy. This might be explained by the fact that these genomic events are associated with *MYCN* amplification and that when two genetic aberrations occur together, the prognostic impact of *MYCN* amplification is dominant [3].

Given the poor prognosis of patients with stage 4 or high risk disease, there are still greater challenges in the need for alternative treatment strategies. We investigated whether our proposed model can be used to evaluate the prognosis of stage 4 patients in the Pugh dataset, in which the majority of patients were at stage 4 with a rate of 86.7% (216/249). We found that the improvement, for which the C-index for the two final models were significantly increased by 5.1% and 5.9% respectively compared with the model involving only clinical variables in the Pugh dataset, was similar to that of other independent databases (Supplementary Table S9), suggesting that our proposed models are also suitable for stage 4 patients. In addition, several previously reported gene signatures built for stage 4 or high-risk patients are of great concern. Kinome-27, a specific kinome gene signature, was able to identify a group of high-risk NB patients and could be used as a potential biomarker of targeted therapy [52]. A four-gene signature, in which these genes are direct p53 transcriptional targets, exhibited significant prognostic value in high-risk NB patients [53]. The future work is to extend these findings to pre-clinical or early-phase clinical trials in large independent cohorts of NB patients.

There were several limitations in the present study. Firstly, the effect of treatment strategy on survival could not be analyzed due to the lack of therapeutic information in the selected datasets. In fact, induction and consolidation chemotherapy protocols vary greatly in different countries and studies and may affect patient survival. Secondly, sample selection bias should be considered. For example, the proportions of patients with *MYCN* amplification were 18.5%, 18.5%, 10.6%, and 14.4% in Su, Wang, Westermann, and Oberthuer datasets, which were lower than the widely cited estimates of 25%-33% [4]. Hence, our models need to be further validated in larger and independent patient cohorts.

5 | CONCLUSIONS

Taken together, we developed a bioinformatics algorithm to identify gene signatures that can reflect MYCN activity and several chromosomal aberrations, Chr1p_del, Chr11q_del, and Chr11q_wls, in NB on the basis of gene expression profiles. We also presented a prognosis prediction model by identifying and integrating the MYCN activity score, the scores of Chr1p_del, Chr11q_del, or Chr11q_wls, and clinical variables for NB. Our proposed model displayed significantly improved prognostic prediction and may provide insights for further therapeutic interventions and surveillance programs for NB patients.

ACKNOWLEDGEMENTS

Not applicable

DECLARATIONS

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable

CONSENT FOR PUBLICATION

Not applicable

AVAILABILITY OF DATA AND MATERIAL

All data generated or analyzed during this study are included in this published article and its supplementary information files.

COMPETING INTERESTS

The authors declare that they have no competing interests.

FUNDING

This work was supported by the American Cancer Society (IRG-82-003-30), the National Center for Advancing Translational Sciences of the National Institutes of Health (KL2TR001088) and the National Natural Science Foundation of China (81201543).

AUTHORS' CONTRIBUTIONS

Xiaoyan He analyzed the data and wrote the paper. Chao Qin and Yanding Zhao developed the computational analyses. Chao Cheng designed and supervised the study, and revised the paper. Lin Zou and Hui Zhao provided the biomedical support. All authors contributed to the data interpretation, read and approved the final manuscript.

ORCID

Chao Cheng https://orcid.org/0000-0002-5002-3417

REFERENCES

1. Maris JM, Hogarty MD, Bagatell R, Cohn SL. Neuroblastoma. Lancet. 2007;369:2106–20.
2. Schor NF. Neuroblastoma as a neurobiological disease. J Neurooncol. 1999;41:159–66.
3. Brodeur GM. Neuroblastoma: biological insights into a clinical enigma. Nat Rev Cancer. 2003;3:203–16.
4. Jiang M, Stanke J, Lahti JM. The connections between neural crest development and neuroblastoma. Curr Top Dev Biol. 2011;94:77–127.
5. Cohn SL, Pearson AD, London WB, Monclair T, Ambros PF, Brodeur GM, et al. The International Neuroblastoma Risk Group (INRG) classification system: an INRG Task Force report. J Clin Oncol. 2009;27:289–97.
6. Vo KT, Matthay KK, Neuhaus J, London WB, Hero B, Ambros PF, et al. Clinical, biologic, and prognostic differences on the basis of primary tumor site in neuroblastoma: a report from the international neuroblastoma risk group project. J Clin Oncol. 2014;32:3169–76.
7. Matthay KK, George RE, Yu AL. Promising therapeutic targets in neuroblastoma. Clin Cancer Res. 2012;18:2740–53.
8. Zage PE, Kletzel M, Murray K, Marcus R, Castleberry R, Zhang Y, et al. Outcomes of the POG 9340/9341/9342 trials for children with high-risk neuroblastoma: a report from the Children's Oncology Group. Pediatr Blood Cancer. 2008;51:747–53.
9. Oberthuer A, Juraeva D, Hero B, Volland R, Sterz C, Schmidt R, et al. Revised risk estimation and treatment stratification of low- and intermediate-risk neuroblastoma patients by integrating clinical and molecular prognostic markers. Clin Cancer Res. 2015;21:1904–15.
10. Costa RA, Seuanez HN. Investigation of major genetic alterations in neuroblastoma. Mol Biol Rep. 2018;45:287–95.
11. Whittle SB, Smith V, Doherty E, Zhao S, McCarty S, Zage PE. Overview and recent advances in the treatment of neuroblastoma. Expert Rev Anticancer Ther. 2017;17:369–86.
12. Stigliani S, Coco S, Moretti S, Oberthuer A, Fischer M, Theissen J, et al. High genomic instability predicts survival in metastatic high-risk neuroblastoma. Neoplasia. 2012;14:823–32.
13. He XY, Tan ZL, Mou Q, Liu FJ, Liu S, Yu CW, et al. microRNA-221 Enhances MYCN via Targeting Nemo-like Kinase and Functions as an Oncogene Related to Poor Prognosis in Neuroblastoma. Clin Cancer Res. 2017;23:2905–18.
14. Jacobs JF, van Bokhoven H, van Leeuwen FN, Hulsbergen-van de Kaa CA, de Vries IJ, Adema GJ, et al. Regulation of MYCN expression in human neuroblastoma cells. BMC Cancer. 2009;9:239.
15. Valentijn LJ, Koster J, Haneveld F, Aissa RA, van Sluis P, Broekmans ME, et al. Functional MYCN signature predicts outcome of neuroblastoma irrespective of MYCN amplification. Proc Natl Acad Sci U S A. 2012;109:19190–5.
16. Janoueix-Lerosey I, Schleiermacher G, Michels E, Mosseri V, Ribeiro A, Lequin D, et al. Overall genomic pattern is a predictor of outcome in neuroblastoma. J Clin Oncol. 2009;27:1026–33.
17. Tomioka N, Oba S, Ohira M, Misra A, Fridlyand J, Ishii S, et al. Novel risk stratification of patients with neuroblastoma by genomic signature, which is independent of molecular signature. Oncogene. 2008;27:441–49.
18. Nakazawa A, Haga C, Ohira M, Okita H, Kamijo T, Nakagawara A. Correlation between the International Neuroblastoma Pathology Classification and genomic signature in neuroblastoma. Cancer Sci. 2015;106:766–71.
19. Vermeulen J, De Preter K, Naranjo A, Vercruysse L, Van Roy N, Hellemans J, et al. Predicting outcomes for children with neuroblastoma using a multigene-expression signature: a retrospective SIOPEN/COG/GPOH study. Lancet Oncol. 2009;10:663–71.
20. Shi L, Campbell G, Jones WD, Campagne F, Wen Z, Walker SJ, et al. The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. Nat Biotechnol. 2010;28:827–38.
21. De Preter K, Vermeulen J, Brors B, Delattre O, Eggert A, Fischer M, et al. Accurate outcome prediction in neuroblastoma across independent data sets using a multigene signature. Clin Cancer Res. 2010;16:1532–41.
22. Oberthuer A, Hero B, Berthold F, Juraeva D, Faldum A, Kahlert Y, et al. Prognostic impact of gene expression-based classification for neuroblastoma. J Clin Oncol. 2010;28:3506–15.
23. Fey D, Halasz M, Dreidax D, Kennedy SP, Hastings JF, Rauch N, et al. Signaling pathways models as biomarkers: Patient-specific simulations of JNK activity predict the survival of neuroblastoma patients. Sci Signal. 2015;8:ra130.
24. Borisov NM, Terekhanova NV, Aliper AM, Venkova LS, Smirnov PY, Roumiantsev S, et al. Signaling pathways activation profiles make better markers of cancer than expression of individual genes. Oncotarget. 2014;5:10198–205.
25. Lezhnina K, Kovalchuk O, Zhavoronkov AA, Korzinkin MB, Zabolotneva AA, Shegay PV, et al. Novel robust biomarkers for human bladder cancer based on activation of intracellular signaling pathways. Oncotarget. 2014;5:9022–32.

26. Zhao Y, Varn FS, Cai G, Xiao F, Amos CI, Cheng C. A P53-Deficiency Gene Signature Predicts Recurrence Risk of Patients with Early-Stage Lung Adenocarcinoma. Cancer Epidemiol Biomarkers Prev. 2018;27:86–95.
27. Cheng C, Yan X, Sun F, Li LM. Inferring activity changes of transcription factors by binding association with sorted expression profiles. BMC Bioinformatics. 2007;8:452.
28. Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, et al. Widespread aneuploidy revealed by DNA microarray expression profiling. Nat Genet. 2000;25:333–37.
29. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat Genet. 2006;38:1043–48.
30. Furge KA, Chen J, Koeman J, Swiatek P, Dykema K, Lucin K, et al. Detection of DNA copy number changes and oncogenic signaling abnormalities from gene expression data reveals MYC activation in high-grade papillary renal cell carcinoma. Cancer Res. 2007;67:3171–76.
31. Henrich KO, Bender S, Saadati M, Dreidax D, Gartlgruber M, Shao C, et al. Integrative Genome-Scale Analysis Identifies Epigenetic Mechanisms of Transcriptional Deregulation in Unfavorable Neuroblastomas. Cancer Res. 2016;76:5523–37.
32. Su Z, Fang H, Hong H, Shi L, Zhang W, Zhang W, et al. An investigation of biomarkers derived from legacy microarray data for their utility in the RNA-seq era. Genome Biol. 2014;15:523.
33. Wang C, Gong B, Bushel PR, Thierry-Mieg J, Thierry-Mieg D, Xu J, et al. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. Nat Biotechnol. 2014;32:926–32.
34. Westermann F, Muth D, Benner A, Bauer T, Henrich KO, Oberthuer A, et al. Distinct transcriptional MYCN/c-MYC activities are associated with spontaneous regression or malignant progression in neuroblastomas. Genome Biol. 2008;9:R150.
35. Oberthuer A, Juraeva D, Li L, Kahlert Y, Westermann F, Eils R, et al. Comparison of performance of one-color and two-color gene-expression analyses in predicting clinical endpoints of neuroblastoma patients. Pharmacogenomics J. 2010;10:258–66.
36. Berwanger B, Hartmann O, Bergmann E, Bernard S, Nielsen D, Krause M, et al. Loss of a FYN-regulated differentiation and growth arrest pathway in advanced stage neuroblastoma. Cancer Cell. 2002;2:377–86.
37. Pugh TJ, Morozova O, Attiyeh EF, Asgharzadeh S, Wei JS, Auclair D, et al. The genetic landscape of high-risk neuroblastoma. Nat Genet. 2013;45:279–84.
38. Attiyeh EF, London WB, Mosse YP, Wang Q, Winter C, Khazi D, et al. Chromosome 1p and 11q deletions and outcome in neuroblastoma. N Engl J Med. 2005;353:2243–53.
39. Spitz R, Hero B, Simon T, Berthold F. Loss in chromosome 11q identifies tumors with increased risk for metastatic relapses in localized and 4S neuroblastoma. Clin Cancer Res. 2006;12:3368-3373.
40. Depuydt P, Koster J, Boeva V, Hocking TD, Speleman F, Schleiermacher G, et al. Meta-mining of copy number profiles of high-risk neuroblastoma tumors. Sci Data. 2018;5:180240.
41. Maris JM. Recent advances in neuroblastoma. N Engl J Med. 2010;362:2202–11.
42. Rosswog C, Schmidt R, Oberthuer A, Juraeva D, Brors B, Engesser A, et al. Molecular Classification Substitutes for the Prognostic Variables Stage, Age, and MYCN Status in Neuroblastoma Risk Assessment. Neoplasia. 2017;19:982–90.
43. Berbegall AP, Villamon E, Piqueras M, Tadeo I, Djos A, Ambros PF, et al. Comparative genetic study of intratumoral heterogenous MYCN amplified neuroblastoma versus aggressive genetic profile neuroblastic tumors. Oncogene. 2016;35:1423–32.
44. Thompson D, Vo KT, London WB, Fischer M, Ambros PF, Nakagawara A, et al. Identification of patient subgroups with markedly disparate rates of MYCN amplification in neuroblastoma: A report from the International Neuroblastoma Risk Group project. Cancer. 2016;122:935–45.
45. Caren H, Kryh H, Nethander M, Sjoberg RM, Trager C, Nilsson S, et al. High-risk neuroblastoma tumors with 11q-deletion display a poor prognostic, chromosome instability phenotype with later onset. Proc Natl Acad Sci U S A. 2010;107:4323–28.
46. Mlakar V, Jurkovic Mlakar S, Lopez G, Maris JM, Ansari M, Gumy-Pause F. 11q deletion in neuroblastoma: a review of biological and clinical implications. Mol Cancer. 2017;16:114.
47. Spitz R, Hero B, Simon T, Berthold F. Loss in chromosome 11q identifies tumors with increased risk for metastatic relapses in localized and 4S neuroblastoma. Clin Cancer Res. 2006;12:3368–73.
48. Uryu K, Nishimura R, Kataoka K, Sato Y, Nakazawa A, Suzuki H, et al. Identification of the genetic and clinical characteristics of neuroblastomas using genome-wide analysis. Oncotarget. 2017;8:107513–529.
49. Peifer M, Hertwig F, Roels F, Dreidax D, Gartlgruber M, Menon R, et al. Telomerase activation by genomic rearrangements in high-risk neuroblastoma. Nature. 2015;526:700–704.
50. Molenaar JJ, Koster J, Zwijnenburg DA, van Sluis P, Valentijn LJ, van der Ploeg I, et al. Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. Nature. 2012;483:589–93.
51. Bosse KR, Maris JM. Advances in the translational genomics of neuroblastoma: From improving risk stratification and revealing novel biology to identifying actionable genomic alterations. Cancer. 2016;122:20–33.
52. Rosso R, Cimmino F, Pezone L, Manna F, Avitabile M, Langella C, et al. Kinome expression profiling of human neuroblastoma tumors identifies potential drug targets for ultra high-risk patients. Carcinogenesis. 2017;38:1011–20.
53. Barbieri E, De Preter K, Capasso M, Johansson P, Man TK, Chen Z, et al. A p53 drug response signature identifies prognostic genes in high-risk neuroblastoma. PLoS One. 2013;8:e79843.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: He X, Qin C, Zhao Y, Zou L, Zhao H, Cheng C. Gene signatures associated with genomic aberrations predict prognosis in neuroblastoma. *Cancer Communications*. 2020;40:105–118. https://doi.org/10.1002/cac2.12016