Research Article

Epigenome-wide association study reveals CpG sites related to COG of neuroblastoma

Hao Zhao¹,², Xiaojun Zhou³, Hu Sun², Dongyun Zhao³, Hongfei Liu², Bin Huang¹,⁴, Xingang Li¹,⁴ and Yinghao Gu²

¹Department of Neurosurgery, Qilu Hospital of Shandong University and Institute of Brain and Brain-Inspired Science, Shandong University, Jinan 250012, Shandong, China; ²Department of Neurosurgery, Zibo Central Hospital Affiliated to Shandong University, Zibo 255036, Shandong, China; ³Department of Paediatric Neurology, Zibo Central Hospital Affiliated to Shandong University, Zibo 255036, Shandong, China; ⁴Shandong Key Laboratory of Brain Function Remodeling, Qilu Hospital of Shandong University, Jinan 250012, Shandong, China

Correspondence: Xingang Li (lxg@sdu.edu.cn) and Yinghao Gu (guyinghaosd@yeah.net)

Background. Neuroblastoma (NB) is the most common extracranial solid tumor in infants and children. Its variable location and complex pathogenesis make NB hard for early diagnosis and risk classification. Methodology. We analyzed the methylation data of 236 samples from patients with NB in Therapeutically Applicable Research to Generate Effective Treatments (TARGET) database. Kaplan–Meier survival analysis was used for comparing overall survival of NB patients in different groups. Epigenome-wide association study (EWAS) was conducted to screen CpGs significantly associated with NB patients’ Children’s Oncology Group (COG). Logistic regression method was used for constructing a model to predict NB patients’ COG. Results. NB patients in low COG showed significantly superior prognosis than those in high COG. A total of seven CpG sites were found closely related to COG. Logistic regression model based on those CpGs showed superior performance in separating NB patients in different COGs. Conclusions. The present study highlights the important role of DNA methylation in NB development, which might provide evidence for treatment decisions for children NB.

Introduction
Neuroblastoma (NB) is one of the most common malignant tumors in infants and children. The incidence of NB in extracranial solid tumor in childhood. Approximately 7% of the malignant tumors occur in children born to 14 years old, with a fatality rate of up to 15% [1]. NB usually originates from the sympathic nervous system in abdomen or chest, most commonly from adrenal gland. Patients with different origins and spread of tumor cells exhibit different symptoms, such as abdominal distension, constipation, dyspnea, skin mass, bone pain and anemia [2]. NB is a self-limiting disease, which is possible for spontaneously cured and recovery when the age of diagnosis is less than 18 months [3]. Children with NB are usually treated with surgical resection, chemotherapy, radiotherapy, and autologous hematopoietic stem cell transplantation. Appropriate treatments based on risk classification could reduce the treatment-related toxicities and improve the prognosis. Patients with low or intermediate risk have satisfactory prognosis, while the recurrence rate of patients with high risk is more than 50%, with 5-year survival rate of 40–50% [4]. However, the occult primary site and diverse phenotypes make NB hard for early diagnosis and most of the patients were at high risk when diagnosed. Advanced NB is highly invasive and rapidly progressing, leading to difficulty in treatment and poor prognosis [5]. Staging and risk classification are two important prognostic factors of NB. The International Neuroblastoma staging system (INSS), which is the most widely used staging standard, classifies NB tumors into stage I, II, III, IV and IV-S according to the tumor resectability [6]. The biological and
clinical characteristics of NB are complex, which would be difficult to evaluate the development and risk before treatment only by relying on INSS. Therefore, the International Neuroblastoma Risk Group (INRG) proposed a new stage and risk classification system for pretreatment. According to INSS stage, age of diagnosis, MYCN amplification status, DNA ploidy, and histopathology, patients with NB are classified into low-risk group, intermediate-risk group, and high-risk group [7]. Children's Oncology Group (COG) suggests that the therapy strategy should be made according to the risk classification of NB [8,9].

Assessment of NB COG risk classification requires a lot of comprehensive information, and subjective factors of medical staff are involved. To reduce the uncertain subjectivity and improve the accuracy of risk classification, we performed epigenome-wide association study (EWAS) on methylation data of patients with NB to identify methylation variations that are related to NB COG risk classification. In addition, a logistic regression model was constructed using these CpG sites for accurate prediction of patients’ COG.

### Materials and methods

#### Methylation data of NB

The methylation data of NB were obtained from Therapeutically Applicable Research to Generate Effective Treatments (TARGET) database (https://ocg.cancer.gov/programs/target), which contains information of common tumors in children. We obtained the methylation data and corresponding clinical data of 236 NB patients from TARGET database. All 236 samples were tumor samples, of which 169 samples contained complete survival information. Detailed epidemiological features of those 169 samples are provided in Table 1. The methylation data were detected by Illumina Human Methylation 450 (HM450) arrays platform.

#### EWAS

Similar to genome-wide association study (GWAS), which finds genetic mutations in the whole-genome by comparing case group and control group, EWAS is to identify methylation variations that are related to disease by comparing different sample groups at epigenetic level. We used CpGassoc package (version 2.60, https://CRAN.R-project.org/package=CpGassoc) of R language to study the relationship between methylation level and phenotype. In this package, the matrix of methylation $\beta$ value and the sample phenotype information, here refers to COG risk group, were used as input, and the $P$-value that represented the relationship between each CpG site and phenotype was calculated by chi-square test. The smaller the $P$-value is, the closer the relationship between the CpG site and phenotype is. $P$-value $< 1e^{-8}$ was used as the threshold for the significant association between CpG site and COG risk group.

#### Construction of logistic regression model

As a common approach in classification, logistic regression is based on a set of variables to predict the classification results. In the present study, the $\beta$ value of methylation site is used to predict samples’ COG. The 236 samples were
randomly divided into training set and testing set with the same sample size. Based on the β value of CpG site that screened in the previous analysis in the training set, the logistic regression model was constructed with R language. Testing set was used to evaluate the accuracy of the regression model.

Statistical analysis
For the 169 samples with survival information, we used the survival package (version 2.42, https://www.rdocumentation.org/packages/survival) of R language (version 3.5, https://www.r-project.org/) to perform Kaplan–Meier survival analysis. Cox proportional hazards model, which introduced mitosis-karyorrhexis index (MKI), risk groups (COG risk group), and tumor percentage (percent_tumor) as variables, was used to calculate the hazard rate (HR) between the sample groups based on each variable. Fisher’s exact test was used to determine the significance of the overall survival between different sample groups.

Results
MKI and COG risk group significantly affect the prognosis and survival
Hazard ratio was calculated based on the Cox proportional hazards model. All 169 samples were divided into groups according to MKI (high, intermediate, low, and unknown), COG risk group (high risk, intermediate risk, and low risk), and tumor percentage (H: > 50%, L: < 50%), respectively. HR > 1 indicated that samples with the factor had a higher risk of death than reference (high MKI, low COG risk group, or low tumor percentage), and HR < 1 indicated samples with the factor had a lower risk of death than reference. Lower.95 and upper.95 are 95% confidence intervals, and P-value < 0.05 was considered as threshold. Figure 1A showed that NB patients with low and intermediate MKI had lower risk of death compared with high MKI. Compared with high COG, low and intermediate COG samples had lower risk of death. Tumor percentage showed no significant correlation with prognosis. Kaplan–Meier survival analysis results also supported that low MKI (Figure 1B) and low COG (Figure 1C) correlated with superior NB prognosis.

EWAS analysis
Samples within different groups based on COG showed significantly different survival time according to previous analysis. To further study the relationship between DNA methylation and COG, we used CpGassoc package to perform EWAS. P-value was used to measure the relationship between the CpG sites methylation levels and tumor COG. The smaller the P-value, the closer the relationship between them. We screened 81 CpG sites with P < 1e-8 as the threshold, and 13 CpG sites with P < 1e-10 as the threshold. As shown in Figure 2A, the 13 CpG sites were distributed on chromosomes 1, 2, 5, 10, 15, and 22. For details, Figure 2B–G were enlarged Manhattan plots of CpG positions on chromosomes 1, 2, 5, 10, 15, and 22, respectively.

© 2020 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).
Construction and evaluation of the logistic regression model in predicting COG risk group

Six of the 13 CpG sites were not annotated to any gene, and seven of them were found with corresponding genes, which were cg25241559 (SNED1), cg20989926 (RLBP1), cg22124648 (FGR), cg23109891 (SOX10), cg23049458 (L1TD1), cg04044188 (PDSS1), and cg16157016 (PICK1). As shown in Figure 3A, we analyzed the methylation levels of these seven sites between high COG and low COG sample group. P-values calculated by Wilcoxon method were all less than 0.05, indicating that the methylation levels of these seven CpG sites were significantly different between the high and low COG groups, and the methylation level in the high COG group was significantly higher than that in the low COG group. A logistic regression model based on the methylation levels of these seven CpG sites in a random divided training set with half of all samples was constructed to determine the COG risk groups. To verify the accuracy of this model, we used the other half of the samples for prediction. As shown in Figure 3B, the inflection points of the curve were all close to the upper left and the area under curve (AUC) was greater than 0.8 (AUC = 0.8663), indicating the high accuracy of the model. Using information-gain-based approach to sort the reliability of the seven CpG sites in predicting COG risk group from high to low, we found that the prediction accuracy of cg25241559 + cg20989926 reached 0.84, while the accuracy of the next points did not significantly increase, indicating that the model constructed by these two CpG sites was accurate enough for COG group prediction (Figure 3C).

Discussion

The risk stratification based on clinical and biologic factors of patients with tumor has been used to determine the appropriate treatment and is criteria for predicting prognosis for decades. Since 2006, the COG has been collecting INRG data and revising the COG risk classification system to achieve better predictive performance [10]. However, patients with same risk stratification may have different clinical courses after receiving same therapy [11]. On the basis of the assumption that the occurrence and development of tumor are driven by genetic and biological characteristics, researchers have identified a large number of potential prognostic markers and established classification for improving treatment strategies [12–14]. Although several well-characterized genetic abnormalities, including DNA content [15], gain of chromosome arm 17q, and deletion of chromosome arm 1p and 11q have been discovered to be associated with outcome prediction in NB [16], features of these molecules are inadequate to explain the clinical heterogeneity [17].

There is increasing evidence that abnormal DNA methylation is highly related to the development and progression of many cancers including NB. Hypermethylation was found in aggressive tumors compared with low-grade ones [18],
and the tumorigenic properties of NB was inhibited by reversing epigenetic changes in CpG island with demethylating agent [19]. Capper et al. reported that DNA methylation patterns might be useful for the accurate classification of olfactory NB [20]. In this study, we analyzed the methylation data of NB in TARGET database, and found that the DNA methylation in the whole-genome that was associated with MKI, an indicator of NB histological classification [7], was significantly related to prognosis and survival. Meanwhile, COG risk group significantly affected the prognosis and survival, also suggesting that the methylation characteristics showed good potential in NB prognosis prediction.

Furthermore, seven CpG sites with significantly different methylation levels in the high- and low-COG risk groups were identified, including cg25241559, cg20989926, cg22124648, cg23109891, cg23049458, cg04044188 and cg16157016, and logistic regression model with these CpG sites was constructed to accurately determine the COG
risk of the samples. In addition, by analyzing the prediction reliability of these CpG sites, we found that the model constructed using cg25241559 and cg20989926 was accurate enough for COG risk group prediction. cg25241559 is a probe representing Sushi Nidogen and EGF-like domains 1 (SNED1), which encodes a secreted protein composed of characteristic domains commonly found in ECM proteins [21]. The crucial role of SNED1 in the development of several tumors has been discovered. Longati et al. identified SNED1 as a stromal marker that induces cisplatin resistance in head and neck squamous carcinoma [22]. Naba et al. reported SNED1 as characteristic protein of highly metastatic in mammary carcinoma, and the overexpression of SNED1 was associated with poor outcome for patients with ER−/PR− breast cancer [23]. The homozygous deletions of SNED1 might be involved in tumor metastasis of NB [24], while DNA methylation change of SNED1 was associated with the clinical course of NB [25]. However, the mechanisms of SNED1 involved in invasive phenotype remains unclear. Cg20989926 is a probe representing RLBP1, a retinaldehyde-binding protein gene that encodes human cellular retinaldehyde-binding protein [26]. Mutation of RLBP1 is associated with several inherited retinal disorders, such as Bothnia dystrophy [27], newfoundland rod-cone dystrophy [28], retinitis pigmentosa [29], and retinitis punctata albscens [30]. In oral cancer, the overexpression of RLBP1 is associated with increased glucose uptake and aerobic glycolysis-mediated ATP synthesis [31]. The hypermethylation of RLBP1 is also recognized as one of the features of high-risk NB [25]. On the other hand, 5-hydroxymethylcytosine (5-hmC), a key component of DNA methylation, was proved to be related to hypoxic response of NB cell lines [32]. In another study on prognostic factors of NB using COG cohorts, 5-hmC profiles could be used as a biomarker for NB in children [33]. Perhaps the transcriptional network essential for promoting different NB phenotypes is regulated by distinct pattern of DNA modification, which requires further research in the future.

**Conclusion**

With the discovery of key somatic and germline genomic alterations, new tumor classifications relying on molecular profiles of tumor and host are emerging to provide evidences for treatment strategy of NB. In the present study, the methylation data of NB from TARGET database were analyzed to investigate the relationship between DNA methylation and COG risk groups. Seven CpG sites that significantly associated with risk classification were identified, and the combination of SNED1 and RLBP1 were accurate enough for risk stratification. The present study provided a basis for the refinements in risk classification, although additional research will be needed to clarify the biological functions of these methylation variations to achieve higher accuracy.

**Competing Interests**

The authors declare that there are no competing interests associated with the manuscript.

**Funding**

The authors declare that there are no sources of funding to be acknowledged.

**Author Contribution**

Hao Zhao, Xiaojun Zhou and Hu Sun put forward the ideas of this article, wrote this article and analyzed the data. Dongyun Zhao, Hongfei Liu and Bin Huang helped in acquisition of data and analysis and interpretation of data. Xingang Li and Yinghao Gu helped in revising the manuscript. All authors read and approved the final manuscript.

**Abbreviations**

AUC, area under curve; COG, Children’s Oncology Group; CpG, CpG island; ECM, extracellular matrix; EGF, Epidermal Growth Factor; EWAS, epigenome-wide association study; HR, hazard rate; INRG, International Neuroblastoma Risk Group; INSS, International Neuroblastoma staging system; MKI, mitosis-karyorrhexis index; MYCN, MYCN proto-oncogene, bHLH transcription factor; NB, neuroblastoma; SNED1, Sushi Nidogen and EGF-like domains 1; TARGET, Therapeutically Applicable Research to Generate Effective Treatments; 5-hmC, 5-hydroxymethylcytosine.

**References**

1. Ward, E., DeSantis, C., Robbins, A., Kohler, B. and Jemal, A. (2014) Childhood and adolescent cancer statistics, 2014. *CA Cancer J. Clin.* **64**, 83–103, [https://doi.org/10.3322/caac.21219](https://doi.org/10.3322/caac.21219)

2. Newman, E.A., Abdessalam, S., Afdink, J.H. et al. (2019) Update on neuroblastoma. *J. Pediatr. Surg.* **54**, 383–389, [https://doi.org/10.1016/j.jpedsurg.2018.09.004](https://doi.org/10.1016/j.jpedsurg.2018.09.004)

3. Davidoff, A.M. (2012) Neuroblastoma. *Semin. Pediatr. Surg.* **21**, 2–14, [https://doi.org/10.1053/j.sempedsurg.2011.10.009](https://doi.org/10.1053/j.sempedsurg.2011.10.009)

4. Maris, J.M., Hogarty, M.D., Bagatell, R. and Cohn, S.L. (2007) Neuroblastoma. *Lancet* **369**, 2106–2120, [https://doi.org/10.1016/S0140-6736(07)60983-0](https://doi.org/10.1016/S0140-6736(07)60983-0)
5 Maris, J.M. (2010) Recent advances in neuroblastoma. N. Engl. J. Med. 362, 2202–2211, https://doi.org/10.1056/NEJMra0804577

6 Smith, E.I., Haase, G.M., Seeger, R.C. and Brodeur, G.M. (1989) A surgical perspective on the current staging in neuroblastoma—the International Neuroblastoma Staging System proposal. J. Pediatr. Surg. 24, 386–390, https://doi.org/10.1016/0022-3468(89)80277-5

7 Monclair, T., Brodeur, G.M., Ambros, P.F. et al. (2009) The International Neuroblastoma Risk Group (INRG) staging system: an INRG task force report. J. Clin. Oncol. 27, 298–303, https://doi.org/10.1200/JCO.2008.16.6876

8 Strother, D.R., London, W.B., Schmidt, M.L. et al. (2012) Outcome after surgery alone or with restricted use of chemotherapy for patients with low-risk neuroblastoma: results of Children’s Oncology Group study P9641. J. Clin. Oncol. 30, 1842–1848, https://doi.org/10.1200/JCO.2011.37.9990

9 Pinto, N., Naranjo, A., Hibbits, E. et al. (2019) Predictors of differential response to induction therapy in high-risk neuroblastoma: a report from the Children's Oncology Group. Eur. J. Cancer 112, 66–79, https://doi.org/10.1016/j.ejca.2019.02.003

10 Naranjo, A., Irwin, M.S., Hogarty, M.D., Cohn, S.L., Park, J.R. and London, W.B. (2018) Statistical framework in support of a revised Children’s Oncology Group Neuroblastoma Risk Classification System. JCO Clin. Cancer Inform. 2, 1–15, https://doi.org/10.1200/CC17.00140

11 De Preter, K., Mestdagh, P. and Vermeulen, J. et al. (2011) miRNA expression profiling enables risk stratification in archived and fresh neuroblastoma tumor samples. Clin. Cancer Res. 17, 7684–7692, https://doi.org/10.1158/1078-0432.CCR-11-0610

12 Kaz, A.M. and Grady, W.M. (2014) Epigenetic biomarkers in esophageal cancer. Cancer Lett. 342, 193–199, https://doi.org/10.1016/j.canlet.2013.11.041

13 Van Neste, L., Groskopf, J., Grizzle, W.E. et al. (2017) Epigenetic risk score improves prostate cancer risk assessment. Prostate 77, 1259–1264, https://doi.org/10.1002/pros.23385

14 Pashayan, N., Reisel, D. and Widschwendter, M. (2016) Integration of genetic and epigenetic markers for risk stratification: opportunities and challenges. Per Med. 13, 93–95, https://doi.org/10.2217/pme.15.53

15Colechia, V., Petroni, M., Guarguaglini, G. et al. (2017) PARP inhibitors enhance replication stress and cause mitotic catastrophe in MYCN-dependent neuroblastoma. Oncogene 36, 4682–4691, https://doi.org/10.1038/onc.2017.40

16 Costa, R.A. and Seuanez, H.N. (2018) Investigation of major genetic alterations in neuroblastoma. Mol. Biol. Rep. 45, 287–295, https://doi.org/10.1007/s11033-018-4161-4

17 Yanez, Y., Grau, E., Rodriguez-Cortez, V.C. et al. (2015) Two independent epigenetic biomarkers predict survival in neuroblastoma. Clin Epigenetics 7, 16, https://doi.org/10.1186/s13480-015-0054-8

18 Olsson, M., Beck, S., Kogner, P., Martinsson, T. and Caren, H. (2016) Genome-wide methylation profiling identifies novel methylated genes in neuroblastoma tumors. Epigenetics 11, 74–84, https://doi.org/10.1002/epi.21348

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

20 Capper, D., Engel, N.W., Stichel, D. et al. (2018) DNA methylation-based reclassification of olfactory neuroblastoma. Acta Neuropathol. 136, 255–271, https://doi.org/10.1007/s00401-018-1854-7

21 Leimeister, C., Schumacher, N., Diez, H. and Gessler, M. (2004) Cloning and expression analysis of the mouse stroma marker Snep encoding a novel nidogen domain protein. Dev. Dyn. 230, 371–377, https://doi.org/10.1002/dvdy.20056

22 Longati, P., Jia, X., Eimer, J. et al. (2013) 3D pancreatic carcinoma spheroids induce a matrix-rich, chemoresistant phenotype offering a better model for drug testing. BMC Cancer 13, 95, https://doi.org/10.1186/1471-2407-13-95

23 Naba, A., Clauer, K.R., Lamar, J.M., Carr, S.A. and Hynes, R.O. (2014) Extracellular matrix signatures of human mammary carcinoma identify novel metastasis promoters. Elife 3, e01308, https://doi.org/10.7554/eLife.01308

24 Eterovic, A.K., Maher, O.M., Chandra, J., Chen, K., Huse, J. and Zaky, W. (2018) Metachronous medulloblastoma in a child with successfully treated neuroblastoma: case report and novel findings of DNA sequencing. J. Natl. Compr. Canc. Netw. 16, 683–691, https://doi.org/10.6004/jncn.2018.7009

25 Gomez, S., Castellano, G., Mayol, G. et al. (2015) DNA methylation fingerprint of neuroblastoma reveals new biological and clinical insights. Epigenetics 7, 1137–1153, https://doi.org/10.2217/epi.15.49

26 Burstedt, M.S. and Golovleva, I. (2010) Central retinal findings in Bothnia dysphoria caused by mutations in the cellular retinaldehyde-binding protein gene (RLBP1) on chromosome 15q26. Invest. Ophthalmol. Vis. Sci. 51, 995–1000

27 Burstedt, M.S. and Golovleva, I. (2010) Central retinal findings in Bothnia dysphoria caused by RLBP1 sequence variation. Arch. Ophthalmol. 128, 989–995, https://doi.org/10.1001/archophthalmol.2010.167

28 Eichers, E.R., Green, J.S., Stockton, D.W. et al. (2002) Newfoundland rod-cone dystrophy, an early-onset retinal dystrophy, is caused by splice-junction mutations in RLBP1. Am. J. Hum. Genet. 70, 955–964, https://doi.org/10.1086/339688

29 Burstedt, M.S., Forsman-Semb, K., Golovleva, I., Janunger, T., Wachtmeister, L. and Sandgren, O. (2001) Ocular phenotype of bothnia dysphoria, an autosomal recessive retinitis pigmentosa associated with an R234W mutation in the RLBP1 gene. Arch. Ophthalmol. 119, 260–267

30 Fishman, G.A., Roberts, M.F., Derlacki, D.J. et al. (2004) Novel mutations in the cellular retinaldehyde-binding protein gene (RLBP1) associated with retinitis punctata albescens: evidence of interfamilial genetic heterogeneity and fundus changes in heterozygotes. Arch. Ophthalmol. 122, 70–75, https://doi.org/10.1001/archophthalmol.122.1.70

31 Leong, C., Ma, J. and Lai, W. (2019) RLBP1 regulates oral cancer cells via Akt and is a novel target of miR-148a-3p and miR-148b-3p. J. Oral Pathol. Med. 48, 919–926, https://doi.org/10.1111/jop.12936

32 Mariani, C.J., Vasanthanakumar, A., Madzio, J. et al. (2014) TET1-mediated hydroxymethylation facilitates hypoxic gene induction in neuroblastoma. Cell Rep. 7, 1343–1352, https://doi.org/10.1016/j.celrep.2014.04.040

33 Mark, A.A., Erin, K.B., Jason, K. et al. (2019) 5-Hydroxymethylcytosine profiles are prognostic of outcome in neuroblastoma and reveal transcriptional networks that correlate with tumor phenotype. JCO Precis. Oncol. 3, 1–12