**ORIGINAL ARTICLE**

*DCX and CRABP2 are candidate genes for differential diagnosis between pre-chemotherapy embryonic and alveolar rhabdomyosarcoma in pediatric patients*

Nian Sun1* | Yeran Yang2* | Shengcai Wang1 | Jie Zhang1 | Jingang Gui3 | Jun Tai4 | Lejian He5 | Jiatong Xu5 | Yanzhen Li1 | Xuexi Zhang1 | Qiaoyin Liu1 | Zhiyong Liu1 | Yongli Guo2 | Xin Ni1

1Department of Otolaryngology, Head and Neck Surgery, Beijing Children’s Hospital, Capital Medical University, National Center for Children’s Health, Beijing, China
2Beijing Key Laboratory for Pediatric Diseases of Otolaryngology, Head and Neck Surgery, MOE Key Laboratory of Major Diseases in Children, Beijing Pediatric Research Institute, Beijing Children’s Hospital, Capital Medical University, National Center for Children’s Health, Beijing, China
3Laboratory of Tumor Immunology, Beijing Pediatric Research Institute, Beijing Children’s Hospital, Capital Medical University, National Center for Children’s Health, Beijing, China
4Laboratory of Otorhinolaryngology, Children’s Hospital, Capital Institute of Pediatrics, Beijing, China
5Department of Otorhinolaryngology, Children’s Hospital, Capital Medical University, National Center for Children’s Health, Beijing, China

**Correspondence**

Xin Ni, Beijing Children’s Hospital, Capital Medical University, National Center for Children’s Health, Beijing 100045, China
Email: nixin@bch.com.cn
Yongli Guo, Beijing Pediatric Research Institute, Beijing Children’s Hospital, Capital Medical University, National Center of Children’s Health, Beijing 100045, China
Email: guoyongli@bch.com.cn

*These authors contribute equally to this study.

**Funding source**

‘Beijing Hospitals Authority’ Ascent Plan (20191201), Beijing Municipal Science and Technology Commission Fund (Z201100005520077), and National Natural Science Foundation of China (81702787)

Received: 16 March, 2021
Accepted: 26 May, 2021

**ABSTRACT**

**Importance:** Rhabdomyosarcoma (RMS) is the most common soft tissue sarcoma in children. More than 90% of cases are classified as embryonic RMS (ERMS) or alveolar RMS (ARMS). ERMS has a worse prognosis than ARMS. Early differential diagnosis is of paramount importance for optimization of treatment.

**Objective:** To identify genes that are differentially expressed between ARMS and ERMS, which can be used for accurate rhabdomyosarcoma classification.

**Methods:** Three Gene Expression Omnibus datasets composed of ARMS and ERMS samples were screened and 35 differentially expressed genes (DEGs) were identified. Receiver operating characteristic curve analysis and area under the curve analysis was performed for these 35 DEGs and seven candidate genes with the best differential expression scores between ARMS and ERMS were determined. The expression of these seven candidate genes was validated by immunohistochemical analysis of pre-chemotherapy ARMS and ERMS specimens.

**Results:** The levels of DCX and CRABP2 were confirmed to be remarkably different between paraffin-embedded ARMS and ERMS tissues, while EGFR abundance was only marginally different between these two RMS subtypes.

**Interpretation:** DCX and CRABP2 are potential biomarkers for distinguishing ARMS from ERMS in pre-chemotherapy pediatric patients.

**KEYWORDS**

Alveolar rhabdomyosarcoma, Embryonic rhabdomyosarcoma, DCX, CRABP2, Immunohistochemistry
INTRODUCTION

Rhabdomyosarcoma (RMS), a high-grade malignant neoplasm of mesenchymal origin and is the most common soft tissue sarcoma in children. RMS can be classified into four histological subtypes based on different locations or origins according to the World Health Organization: embryonal (ERMS), alveolar (ARMS), pleomorphic, and spindle cell/sclerosing. ERMS and ARMS are the two major subtypes accounting for about 90% of RMS and they have distinct molecular and clinical profiles. Accurate classification of these two subtypes is essential for optimization of treatment protocols. Currently, the classification of ARMS and ERMS is mainly based on histological and cytological features, as well as molecular characteristics. Approximately 60% of ARMS tumors are accompanied by a gene fusion between PAX3 or PAX7 and FOXO1, which leads to a higher probability of bone marrow metastasis. Genetic mutations have been found in some ARMS cases, such as in MYCN and CDK4 genes. Other genetic variants including in TP53, NRAS, Kras, HRAS, PIK3CA, CTNNB1 and FGFR4 have been documented in ERMS cases. However, PAX3/7-FOXO1 fusion is only a partial indicator of ARMS because around 40% of ARMS cases lack this fusion and share some clinical features with ERMS. Therefore, additional genetic markers are needed as supplementary factors to help distinguish ARMS from ERMS.

Microarray analysis is an efficient tool in oncology studies. Although many previous studies have used microarray technology to acquire gene expression profiles of sarcoma, investigation of RMS is very limited. In this regard, identifying differentially expressed genes (DEGs) would help discriminate ARMS from other subtypes of RMS, especially ERMS.

In this study, we applied bioinformatics to interrogate previously published data from ARMS and ERMS samples. We identified DEGs between ARMS and ERMS and validated the expression of critical genes by immunohistochemical analysis of tumor specimens.

METHODS

Ethical approval

This research was approved by the Ethics Committees Board of Beijing Children’s Hospital (2018-k-148), and followed the principles of the Helsinki Declaration II. The informed consent was obtained from guardians of patients.

Dataset selection

To identify the DEGs between ARMS and ERMS, the datasets of gene expression profiles with sequence numbers of GSE114621, GSE52252 and GSE967 were retrieved from the Gene Expression Omnibus (GEO) database. GSE114621 (Dataset 1) contained 43 ARMS and 35 ERMS tissue samples; GSE52252 (Dataset 2) included three ARMS and three ERMS samples; GSE967 (Dataset 3) included nine ARMS and three ERMS samples. The selection criteria for these three datasets were as follows: (1) including both ARMS and ERMS tissue samples; (2) accessible to all the raw data.

Preprocess of data and screen of DEGs

We used an online tool, GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r), for each dataset separately. Preprocessing and standardization of the original dataset were the basis for obtaining accurate data. The matrix was constructed by the GEO2R tool with Limma R packages and the Affymetrix gene expression microarray data were read. We considered DEGs as fold-change > 1.2 or log-fold-change < −1.2 and P < 0.05. Online software Venn diagram analysis (http://bilinfogp.cnb.csic.es/tools/venny/index.html, VENNY 2.1) was used to intersect DEGs between ARMS and ERMS after obtaining three profiles of differentially expressed Affymetrix gene IDs. Datasets were represented by three different colored areas, and cross areas represented the common DEGs. We used Database for Annotation, Visualization and Integration Discovery (DAVID) to convert Gene IDs into symbols. After removing uncharacterized and duplicated genes, 35 common DEGs from three datasets were generated. Since dataset 1 (GSE114621) was identified to have a larger sample size of ARMS and ERMS tissues, heatmaps and volcano plots were generated of this dataset in order to visualize DEGs using Morpheus (https://software.broadinstitute.org/morpheus, version 1.0) and Main software (BOA Bioinformatics, version 8.1).

Immunohistochemistry (IHC)

A total of 10 pre-chemotherapeutic RMS tumor specimens, including 5 ARMS and 5 ERMS specimens were collected from Beijing Children’s Hospital, Capital Medical University.

The pre-chemotherapeutic tumor specimens were fixed in 10% formalin, embedded in paraffin and sectioned. After dewaxing, unstained 4-μm FFPE sections were heated for 15 min at 95°C with 10 mM sodium citrate buffer (pH 6.0) to retrieve tissue antigen. In order to block endogenous peroxidase, the sections were incubated with 3% hydrogen peroxide for 10 min at room temperature. After phosphate-buffered saline (PBS) rinsing, the appropriate diluted primary antibody (RPSAP58: dilution 1:500, H00388524-B01P, Novus Biologicals; DCX: dilution 1:100, sc-271390, Santa Cruz; CRABP2: dilution 1:100, sc-166897, Santa Cruz; GAS2: dilution 1:100, ab109762, Abcam; LPAR1: dilution 1:100, sc-515665, Santa Cruz; EGFR: dilution 1:100, sc-373746, Santa Cruz; SLC1A3: dilution 1:100, ab181036, Abcam) were added to each section and
incubated overnight at 4°C. After washing the primary antibody with PBS, the sections were incubated with secondary antibody (rabbit HRP Polymer PV-9001 or mouse HRP Polymer PV-9002, Beijing Zhongshan Golden Bridge Biotechnology Co., Ltd., China) and developed with DAB (3-3’ diaminobenzidine-4HCl, Sigma). Finally, the staining signal were evaluated under a light microscope under 200× bright-field and 400× bright-field, respectively.

The staining signal of RPSAP58, DCX, CRABP2, GAS2, LPAR1, EGFR, and SLC1A3 were scored by two senior pathologists independently. The intensity of immunostaining was estimated as follows: 0, negative; 1, weak positive; 2, positive; or 3, strong positive. We observe the percentage of cell staining intensity in the field of view of 5 high power lenses (200×) to the same type of cell in the field of view. No coloring was regarded as negative, and 0 points were recorded. Less than 10% of cells showing brown coloring were regarded as weak positive, and scored 1 point. Ten to thirty percent of cells with strong tan staining or less than 70% of cells with weak or medium strength brown staining were considered positive, scored 2 points. More than 30% of cells with strong or more than 70% medium strength tan staining were regarded as strong positive, scored 3 points.

Statistical analysis

Univariate and multivariable logistic regression analysis were performed to identify the DEGs that may help in separating ARMS from ERMS using SPSS software (version 16.0; SPSS, Inc). Age, gender and stage were included in multivariable logistic regression analysis. The area under the receiver operating characteristic (ROC) curve was carried out to evaluate the DEGs as potentially predictive molecules for differential diagnosis.

RESULTS

Identification of DEGs between ARMS and ERMS

To search for differential gene expression between ARMS and ERMS, data from three published datasets that include 43 ARMS and 35 ERMS tissue samples were processed. Data were normalized and processed according to the flowchart shown in Figure 1. In short, differential gene expression in these three datasets was extracted using GEO2R. Duplicated and uncharacterized data were removed and genes differentially expressed in all three datasets were selected. Thirty-five DEGs with log2(fold-change) ≥1.2 between ARMS and ERMS samples were found, including 19 that were up-regulated and 16 that were down-regulated in ARMS samples compared with ERMS samples (Figure 2).

Screening of critical DEGs between ARMS and ERMS

To find the most relevant DEGs between ARMS and ERMS, we first used univariate and multivariable logistic regression analysis to confirm the accuracy of DEG selection. All of the 35 DEGs were significantly different ($P < 0.05$) regardless of which mode was chosen for analysis (data not shown). We then processed the data by ROC curve analysis. The ROC curves and area under the curve (AUC) values of DEGs are presented in Table 1 and Figure S1. All AUC values exceeded 0.5, a standard criterion for data separation. Then DEGs possessing AUC values greater than 0.8 were selected as critical candidate genes for further validation.

Validation of critical DEGs by IHC

We chose seven genes (RPSAP58, DCX, CRABP2, GAS2, LPAR1, EGFR and SLC1A3) for validation whose AUC values were > 0.8. Their expression levels were quantified

![Figure 1](https://example.com/figure1.png)
and compared in pre-chemotherapy RMS tumor specimens by IHC. We collected tumor specimens from ten pre-chemotherapy RMS pediatric patients (seven males and three females) with ages ranging from 20 to 93 months. Five patients were diagnosed as ARMS and five as ERMS. Three of the ARMS patients were FOXO1 gene positive. Eight patients were in the stage III and medium-risk group, while two were in the stage IV and high-risk group.

The demographic and clinical characteristics of these 10 RMS patients are detailed in Table S1. The intensity of immunohistochemical staining was scored by two senior pathologists, and the results are shown in Figures 3 and S2. Among the seven DEGs, the abundance of DCX was significantly higher in ARMS, CRABP2 abundance was significantly lower in ARMS, while EGFR was only marginally different between ARMS and ERMS. The other four DEGs (RPSAP58, GAS2, LPAR1 and SLC1A3), despite having high AUC scores, showed no differences in protein abundance between the two RMS subtypes.

### DISCUSSION

Cancer is essentially a genetic disease in which many genetic variations accumulate during the multistep process of carcinogenesis, leading to abnormal unrestrained cell growth and malignancy. RMS is a malignant solid tumor occurring most commonly in children. If diagnosis is made after metastasis has occurred, the 5-year survival rate of RMS, even with the best treatment, is below 20%. Considering the poor prognosis of ARMS, it is essential to distinguish it from other subtypes of RMS, especially ERMS. We hypothesized that this could be achieved by examining their differential gene expression profiles. In the present study, seven candidate DEGs were identified by comparing microarray gene expression data between ERMS and ARMS in three published databases. We then validated the expression levels of these seven genes in paraffin-embedded tumor sections and found that two genes (DCX and CRABP2) were strongly associated with the ERMS and ARMS subtypes. DCX is an essential factor in neurogenesis and plays a role in central nervous system tumors and prostate cancer. The expression of DCX was enhanced in ARMS, which was consistent with a previous report showing that DCX was up-regulated in PAX-FoxO1 gene-fusion-positive ARMS. CRABP2 is associated with increased circulating levels of low-

### TABLE 1 AUC values of DEGs

| Gene ID | AUC  | Gene ID | AUC  |
|---------|------|---------|------|
| Up-regulated |      | Down-regulated |      |
| RPSAP58 | 0.8133 | CRABP2 | 0.8492 |
| DCX    | 0.8073 | GAS2    | 0.8385 |
| GCA    | 0.7980 | LPAR1   | 0.8179 |
| CNR1   | 0.7934 | EGFR    | 0.8153 |
| EDN3   | 0.7880 | SLC1A3  | 0.8153 |
| SEL1L3 | 0.7794 | PLAG1   | 0.7914 |
| CEL2B  | 0.7588 | NELL2   | 0.7781 |
| CEL2A  | 0.7588 | NAV3    | 0.7615 |
| NELL1  | 0.7522 | NNAT    | 0.7608 |
| CHRN1B | 0.7515 | WIFI1   | 0.7508 |
| ALK    | 0.7508 | CCN1    | 0.7342 |
| ELOVL2 | 0.7309 | NFM     | 0.7150 |
| TFAP2B | 0.7150 | PRAME   | 0.7096 |
| FGFR2  | 0.7143 | ITM2A   | 0.6957 |
| JAKMIP2| 0.7130 | MMP16   | 0.6525 |
| ERVW-1 | 0.7050 | MYF5    | 0.5972 |
| GRED1  | 0.6990 |        |      |
| TTN    | 0.6844 |        |      |
| BMP5   | 0.6638 |        |      |

**AUC**, area under the curve; **DEGs**, differentially expressed genes.
density lipoprotein cholesterol and its expression level is connected to lung, breast and cervical cancer.\textsuperscript{23-25} In our study, the expression of CRABP2 was upregulated in ERMS, which has not been reported previously. The EGFR family, which showed minimal expression difference in this study, consists of four genes encoding receptors that play important roles in cell proliferation, division and survival.\textsuperscript{26} Several studies have demonstrated the expression of EGFR in RMS cell lines, indicating that the proliferative ability of RMS cells was impaired by reduced EGFR expression.\textsuperscript{27} In our study, albeit with a limited sample size, we found upregulation of EGFR in two out of five ERMS samples. This was in agreement with a previous demonstration of EGFR downregulation in ARMS.\textsuperscript{22} Surprisingly, the other four genes (RPASPS8, GAS2, LPAR1, and SLC1A3) that we chose with great confidence based on DEG and AUC analysis showed no differential expression in ARMS and ERMS tumor sections.

There are several limitations in the present study. First, the three datasets we analyzed were all microarray data that did not include information of fusion status. Three of the ARMS tumor samples chosen for IHC validation were positive for PAX-FOXO1 gene fusion. However, because of the limited sample size we could not identify an obvious relationship between different fusion types and DCX/CRABP2 expression. In the future, comparison of DCX and CRABP2 expression in ARMS tumors with and without gene fusion will be valuable for understanding the molecular mechanism of tumor progression. We suggest that expression of these two genes might constitute a supplementary index for ARMS. In many difficult cases, for example, PAX3/7-FOXO1-negative ARMS cases, it could be an alternative diagnostic parameter. Second, the data size was small and was mainly from the Affymetrix Human Exon Array. In the future, we will use more data, including RNA-Seq data, to validate this finding and to discover more markers. Third, because of the limited sample size tested, scores overlapped between ERMS and ARMS. ERMS and ARMS do share many clinical features, which is why it is difficult to distinguish these two tumor types at early stages, particularly before chemotherapy. Because of the rarity of rhabdomyosarcoma tissue samples, especially pre-chemotherapy samples, our results require further validation using larger cohorts.

In summary, we performed a comprehensive bioinformatics analysis of DEGs between ARMS and ERMS based on three public datasets and identified 35 DEGs. The results of the analysis and IHC validation indicated that DCX and CRABP2 are candidate diagnostic molecules for distinguishing ARMS from ERMS.

CONFLICT OF INTEREST
None.

REFERENCES
1. Ognjanovic S, Linabery AM, Charbonneau B, Ross JA. Trends in childhood rhabdomyosarcoma incidence and survival in the United States, 1975-2005. Cancer. 2009;115:4218-4226.
2. Rudzinski ER, Anderson JR, Hawkins DS, Skapek SX, Parham DM, Teot LA. The World Health Organization Classification of skeletal muscle tumors in pediatric rhabdomyosarcoma: A report from the Children’s Oncology Group. Arch Pathol Lab Med. 2015;139:1281-1287.
3. Parham DM, Barr FG. Classification of rhabdomyosarcoma and its molecular basis. Adv Anat Pathol. 2013;20:387-397.
4. Sorensen PH, Lynch JC, Qualman SJ, Tirabosco R, Lim JF, Maurer HM, et al. PAX3-FKHR and PAX7-FKHR gene fusions are prognostic indicators in alveolar rhabdomyosarcoma: A report from the children’s oncology group. J Clin Oncol. 2002;20:2672-2679.
5. Scrable H, Cavenee W, Ghavimi F, Lovell M, Morgan K, Sapienza C. A model for embryonal rhabdomyosarcoma tumorigenesis that involves genome imprinting. Proc Natl Acad Sci U S A. 1989;86:7480-7484.
6. Martinek S, McDowell HP, Vigne SD, Kokai G, Uccini S, Tartaglia M, et al. RAS signaling dysregulation in human
embryonal rhabdomyosarcoma. Genes Chromosomes Cancer. 2009;48:975-982.
7. Taylor AC, Shu L, Danks MK, Poquette CA, Shetty S, Thayer MJ, et al. P53 mutation and MDM2 amplification frequency in pediatric rhabdomyosarcoma tumors and cell lines. Med Pediatr Oncol. 2000;35:96-103.
8. Stratton MR, Fisher C, Gusterson BA, Cooper CS. Detection of point mutations in N-ras and K-ras genes of human embryonal rhabdomyosarcomas using oligonucleotide probes and the polymerase chain reaction. Cancer Res. 1989;49:6324-6327.
9. Shukla N, Ameur N, Yilmaz I, Nafa K, Lau CY, Marchetti A, et al. Oncogene mutation profiling of pediatric solid tumors reveals significant subsets of embryonal rhabdomyosarcoma and neuroblastoma with mutated genes in growth signaling pathways. Clin Cancer Res. 2012;18:748-757.
10. Taylor JG 6th, Cheuk AT, Tsang PS, Chung JY, Desai K, et al. Identification of FGFR4-activating mutations in human rhabdomyosarcomas that promote metastasis in xenotransplanted models. J Clin Invest. 2009;119:3395-3407.
11. Oberlin O, Rey A, Lyden E, Bisogno G, Stevens MC, Meyer WH, et al. Prognostic factors in metastatic rhabdomyosarcomas: results of a pooled analysis from United States and European cooperative groups. J Clin Oncol. 2008;26:2384-2389.
12. Li Q, Zhang L, Jiang J, Zhang Y, Wang X, Zhang Q, et al. CDK1 and CCNB1 as potential diagnostic markers of rhabdomyosarcoma: validation following bioinformatics analysis. BMC Med Genomics. 2019;12:198.
13. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44-57.
14. Xu L, Zheng Y, Liu J, Rakheja D, Singleterry S, Laetsch TW, et al. Integrative Bayesian analysis identifies rhabdomyosarcoma disease genes. Cell Rep. 2018;24:238-251.
15. Wang X, Bledsoe KL, Graham RP, Asmann YW, Viswanatha DS, Lewis JE, et al. Recurrent PAX3-MAML3 fusion in biphenotypic sinonasal sarcoma. Nat Genet. 2014;46:666-668.
16. Baer C, Nees M, Breit S, Selle B, Kulozik AE, Schaefer KL, et al. Profiling and functional annotation of mRNA gene expression in pediatric rhabdomyosarcoma and Ewing’s sarcoma. Int J Cancer. 2004;110:687-694.
17. Kaemmerer D, Peter L, Lupp A, Schulz S, Sänger J, Baum RP, et al. Comparing of IRS and Her2 as immunohistochemical scoring schemes in gastroenteropancreatic neuroendocrine tumors. Int J Clin Exp Pathol. 2012;5:187-194.
18. 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;38:1043-1048.
19. Wang X, Feng J, Li Z, Zhang X, Chen J, Feng G. Characteristics and prognosis of embryonal rhabdomyosarcoma in children and adolescents: An analysis of 464 cases from the SEER database. Pediatr Investig. 2020;4:242-249.
20. Mauffrey P, Tchitchev N, Barroca V, Bemelmans AP, Firlej V, Allory Y, et al. Progenitors from the central nervous system drive neurogenesis in cancer. Nature. 2019;569:672-678.
21. Luty M, Piwowarczyk K, Łąbędź-Masłowska A, Wróbel T, Szczypiel M, Catapano J, et al. FenoFibrate augments the sensitivity of drug-resistant prostate cancer cells to docetaxel. Cancers (Basel). 2019;11:77.
22. Davicioni E, Finckenstein FG, Shahbazzian V, Buckley JD, Triche TJ, Anderson MJ. Identification of a PAX-FKHR gene expression signature that defines molecular classes and determines the prognosis of alveolar rhabdomyosarcomas. Cancer Res. 2006;66:6936-6946.
23. Kim DJ, Kim WJ, Lim M, Hong Y, Lee SJ, Hong SH, et al. Plasma CRABP2 as a novel biomarker in patients with non-small cell lung cancer. J Korean Med Sci. 2018;33:e178.
24. Feng X, Zhang M, Wang B, Zhou C, Mu Y, Li J, et al. CRABP2 regulates invasion and metastasis of breast cancer through hippo pathway dependent on ER status. J Exp Clin Cancer Res. 2019;38:361.
25. Arellano-Ortiz AL, Salcedo-Vargas M, Vargas-Requena CL, López-Díaz JA, De la Mora-Covarrubias A, Silva-Espinoza JC, et al. DNA methylation of cellular retinoic acid-binding proteins in cervical cancer. Genet Epigenet. 2016;8:53-57.
26. Yarden Y, Sliekowski MX. Untangling the ErbB signalling network. Nat Rev Mol Cell Biol. 2001;2:127-137.
27. De Giovanni C, Landuzzi L, Frabetti F, Nicoletti G, Griffoni C, Rossi I, et al. Antisense epidermal growth factor receptor transfection impairs the proliferative ability of human rhabdomyosarcoma cells. Cancer Res. 1996;56:3898-3901.

SUPPORTING INFORMATION
Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Sun N, Yang Y, Wang S, Zhang J, Gui J, Tai J, et al. DCX and CRABP2 are candidate genes for differential diagnosis between pre-chemotherapy embryonic and alveolar rhabdomyosarcoma in pediatric patients. Pediatr Investig. 2021;5:106-111. https://doi.org/10.1002/ped4.12278