BMC Cancer

RESEARCH

Open Access

Classification of pediatric soft and bone sarcomas using DNA methylation-based profiling

Felipe Luz Torres Silva[1,2], Mayara Ferreira Euzébio[1,2], Juliana Silveira Ruas[1], Mayra Troiani Franco[3], Alejandro Enzo Cassone[3], Thais Junqueira[3], Danielle Ribeiro Lucon[1], Izilda Aparecida Cardinalli[3], Luis Henrique Pereira[3], Priscila Pini Zenatti[1,2], Patricia Yoshioka Jotta[1,2] and Mariana Maschietto[1,2*]

Abstract

Pediatric sarcomas present heterogeneous morphology, genetics and clinical behavior posing a challenge for an accurate diagnosis. DNA methylation is an epigenetic modification that coordinates chromatin structure and regulates gene expression, determining cell type and function. DNA methylation-based tumor profiling classifier for sarcomas (known as sarcoma classifier) from the German Cancer Research Center (Deutsches Krebsforschungszentrum) was applied to 122 pediatric sarcomas referred to a reference pediatric oncology hospital. The classifiers reported 88.5% of agreement between histopathological and molecular classification confirming the initial diagnosis of all osteosarcomas and Ewing sarcomas. The Ewing-like sarcomas were reclassified into sarcomas with *BCOR* or *CIC* alterations, later confirmed by orthogonal diagnostic techniques. Regarding the CNAs profile, osteosarcomas had several chromosomal gains and losses as well as chromothripsis, whereas Ewing sarcomas had few large events, such as amplifications of chromosomes 8 and 12. The molecular classification together with clinical and histopathological assessment could improve the diagnosis of pediatric sarcomas although there are limitations to deal with more rare classes. This study provides an increase in the number of sarcomas evaluated for DNA methylation profiling in the pediatric population.

Keywords Pediatric sarcomas, Soft and bone sarcomas, Molecular classification, DNA methylation, Copy number alteration

Background

Sarcomas comprise a rare and diverse group of soft tissue and bone mesenchymal tumors with varied clinical manifestations and complex morphological structures [1]. The clinical characteristics together with the radiological image and morphological analysis are integrated for diagnosis [2]. The most common entities are rhabdomyosarcomas, osteosarcomas, Ewing sarcomas, synovial sarcomas and other rarer histologies. Several of these entities can occur in similar anatomic locations adding a layer of difficulty for diagnosis [3]. Though the heterogeneity can be challenging for the diagnosis, even when immunohistochemical markers are used, especially for rarer entities, it reflects the genetics/epigenetics variation [4, 5], which include a combination of mutations, translocations (gene fusions) [6], CNAs [7] and DNA methylation variation [8]. The presence of a gene fusion

*Correspondence:
Mariana Maschietto
marianamasc@gmail.com
[1] Research Center, Boldrini Children's Hospital, Rua Marcia Mendes, 619, Cidade Universitaria, CEP 13083-884, Campinas, São Paulo, Brazil
[2] Postgraduate program in Genetics and Molecular Biology, Institute of Biology, State University of Campinas (UNICAMP), Campinas, São Paulo, Brazil
[3] Boldrini Children's Hospital, Campinas, São Paulo, Brazil

is regarded as the gold standard marker for the diagnosis of a mesenchymal tumor [3, 9].

The implementation of molecular markers improved the classification of sarcomas, such as the *EWSR1-FLI1* fusion present in 85% of Ewing sarcomas [10]. Following this discovery, other molecular markers were found allowing the differential diagnosis of several sarcomas: rearrangements involving *CIC* (*CIC*-rearranged sarcoma), *BCOR* (*BCOR*-rearranged sarcoma), *CREB3L2* (Low-grade fibromyxoid sarcoma), *ALK* (Inflammatory myofibroblastic tumor), *PAX3*/*PAX7*/*FOXO1* (alveolar rhabdomyosarcomas) and *ETV6-NTRK3* fusion (infantile fibrosarcoma), among others [11–13]. While several sarcomas present a gene fusion that undisputedly leads to a diagnosis [14], these molecular markers were not yet characterized for all sarcomas.

The genome-wide DNA methylation pattern in cancer likely reflects both the cell type of origin and the modifications acquired over the course of tumor development [15, 16]. This concept resulted in the development of a classifier based on the methylation profiles of 1,077 bone and soft tissue sarcomas as well as their normal counterparts [17], further validated by other studies [17–19]. In this first version of the sarcoma classifier, 75% of the samples received a molecular class, with an agreement of 91% with the original histological diagnosis, while 9% of the classifier predictions resulted in a revised diagnosis in favor of the predicted methylation class after histological review and confirmation by other molecular techniques [17]. Under- or no-represented entities decrease the power of the classification but with the increased number of cases being profiled, the classifier's precision is likely to improve [18].

Besides the methylation pattern, several sarcomas present copy number alterations (CNAs) that may impact the accuracy of the diagnosis, prognosis and therapy for some pediatric tumors [14, 20]. For instance, *MDM2* copy number levels are used for the differential diagnosis between well- and dedifferentiated liposarcomas [21]. Therefore, although rare, CNAs could be used as additional information to characterize the molecular classes identified by methylation profiling.

Here, we profiled the DNA methylation of 122 tumors to characterize the molecular classes from a single institutional cohort of pediatric patients diagnosed with bone and soft-tissue sarcomas. Despite some limitations, the methylation-based classification confirmed or improved the diagnosis of most cases.

Materials and methods

Patient eligibility, sample collection and clinical data

This study was approved by the Boldrini's Ethical Committee with informed consent from patients and/or their legal guardians. The study included 107 consecutive patients diagnosed with pediatric sarcomas and referred to one specialized oncopediatric hospital. Only cases that reached the required ethical criteria were retrieved for this study.

A total of 122 samples (out of 107 patients) were selected according to the original diagnosis, comprising 44 osteosarcomas, 26 Ewing sarcomas, 9 synovial sarcomas, 6 embryonal rhabdomyosarcomas, among others (Supplemental Table 1). We also included 2 chondroblastomas and 2 osteoblastomas because they are represented in the DKFZ Sarcoma classifier to test the agreement between methylation-based classification and histopathology of all entities covered by the classifier with an available sample. All samples were stored in RNAlater at the Institutional Biobank. Patients' clinical information and fluorescence in situ hybridization (FISH) results were obtained from institutional medical records using *CIC* break apart (Zytovision) or *BCOR/CCNB3* dual fusion (Cytocell).

DNA isolation and hybridization on the EPIC beadchip arrays

DNA was extracted with GenELUTE Mammalian Genomic DNA Miniprep (SIGMA-Aldrich, St. Louis, MO, USA; cat. ID: G1N70-1KT) and quantified with Qubit™ dsDNA BR Assay Kits (Invitrogen™, Waltham, MA, USA; cat. ID: Q32853). Around 500 ng of DNA was bisulfite converted using EZ DNA Methylation Direct (Zymo Research, Irvine, CA, USA; cat. ID: D5002)), hybridized in the EPIC BeadChip Methylation arrays (Illumina, San Diego, CA, USA) and scanned on the NextSeq550 (Illumina Inc, San Diego, CA, USA) using system version 4.0.2.

RNA isolation and sequencing

RNA was extracted with RNeasy Kit (QIAGEN, Hilden, Germany; cat. ID: 74106) and quantified with High Sensitivity Qubit™ RNA Assay Kits (Invitrogen™, Waltham, MA, USA; cat. ID: Q32855). Transcriptome assay was performed using Illumina Stranded Total RNA Prep, Ligation with Ribo-Zero Plus (Illumina, San Diego, CA, USA) and sequenced on the NextSeq550 or NOVAseqX (Illumina Inc, San Diego, CA, USA). Paired-end runs were sequenced in 74 cycles per read (2×74), with an average of 41.5 million reads per sample.

Bioinformatic analysis of methylation data

Samples that passed the quality checks, as given by both BeadArray Controls Reporter and minfi QCreport function [22] were normalized using FunNorm [23].

Raw IDAT files for all 122 samples were uploaded to the DKFZ Sarcoma Classifier (v12.2) classifier. We used

the DKFZ Brain tumor (mnp v12.5) classifier for 7 cases with low score by the DKFZ Sarcoma classifier (3 CIC-rearranged sarcomas, 1 embryonal rhabdomyosarcoma, 2 high-grade neuroepithelial tumor with MN1:CXXC5 fusion and 1 atypical teratoid rhabdoid tumor MYC activated) and if they pointed to a molecular class with high score, the typical alterations were validated with orthogonal techniques [15, 17]. Therefore, the higher scores between the two classifiers were used as final classification for each sample. A calibrated score higher than 0.9 was considered for molecular classification. Samples with calibrated score ranging between 0.26 and 0.9 were evaluated as follows: if methylation-based classification was concordant with histopathological diagnosis, molecular classification was maintained and considered concordant; if histopathological diagnosis and methylation-based classification were discrepant, molecular classification was discarded.

Next, IDAT files were analyzed using the minfi R package [22] and normalized using functional normalization (FunNorm) method [23]. XYS and low-quality probes were removed, remaining 758,591 probes for further analysis. The 10,000 most variable probes were used for unsupervised non-linear dimension reduction based on standard deviation. Heatmaps and hierarchical clustering were performed using pheatmap (version 1.0.12) with Ward method and euclidean for distance matrix. MethylResolver R package 0.1.0 was applied to normalized B-values and used to estimate tumor immune cell infiltration and tumor cell purity. It uses a deconvolution method on B-values of bulk tissues by comparing to the methylation signature extracted from 11 leukocyte cell types [24].

Conumee2 R package [25] was used to estimate CNA from methylation arrays. Both methylated and unmethylated signal intensities are added together and compared to a healthy pool of normal tissues (flat genome). This ratio is plotted for each genome segment [26]. Conumee2 CNV.summaryplot() function was used to plot the CNA percentages across samples. CNV.write(what = "overview") function was used to generate a data frame with chromosomal segments, genomic positions and CNA type, further used to calculate the total number of gains and losses for each sample.

Transcriptome fusion calling

Transcriptome raw data quality was verified with FASTQC [27]. We used STAR-fusion [28] to identify and annotate fusion transcripts based on discordant read alignments with default configurations. ChimeraViz was used to plot fusion genes [29].

Statistical analysis

Survival analysis was performed using the Kaplan–Meier method and survival curves were plotted using the 'survival' (v 3.6-4) and 'survminer' (v 0.4.9) R packages. The minimum purity required for a classification score greater than 0.9 was calculated using a linear regression model, dividing (0.9 - intercept) by the purity slope coefficient (beta_1).

Results

Methylation profile predicted the classification of 88.5% of pediatric sarcomas

The DKFZ classifier was applied to 122 samples from 107 patients (Supplemental Table 1). None of the samples failed the QC parameters from the classifier. The average age was 11.9 years (1 year – 22.5 years), most patients were male (64.8%) and 30.3% of patients died of the disease.

Analysis of the classifier reported that 39 (31.96%) samples were predicted with low calibrate score (< 0.90) (12 samples in agreement with histopathological diagnosis and 27 samples had discrepancies). The remaining 83 (68.04%) samples were predicted to belong to one of the defined methylation classes (calibrate score > 0.90), from which 65 agreed with the histopathological diagnosis, 10 samples did not and 8 samples were classified as inflammatory myofibroblastic tumors ($n=2$) or Langerhans cell histiocytosis ($n=6$) probably due to an incorrect or biased sampling of tumor. The diagnosis of Langerhans cell histiocytosis and inflammatory myofibroblastic tumor can be associated purely with high content of inflammatory cells or low tumor cell purity [30, 31].

Eighty-seven samples were considered for comparison between histopathological diagnosis and molecular classification (75 samples with high calibrated score and 12 samples with subthreshold score, but in agreement with the histopathological diagnosis). Of the 87 samples, 77 (88.5%) agreed and maintained the initial diagnosis and the remaining 10 (11.5%) samples were considered discrepant and were further investigated by orthogonal techniques to substantiate or reject the predicted diagnosis or molecular class. The discrepant cases were molecular classified as CIC-rearranged sarcoma ($n=1$), BCOR-rearranged sarcoma ($n=3$), NET_CXXC5 sarcoma ($n=2$), Ewing sarcoma ($n=2$), ATRT_MYC tumor ($n=1$) and embryonal rhabdomyosarcoma ($n=1$), all with high calibrated score (Fig. 1A; Table 1).

Twelve cases had a biopsy and a consecutive sample collected (biopsy and either relapse or metastasis). Molecular classification for both samples agreed in 6 (50%) of the cases (Supplemental Table 2).

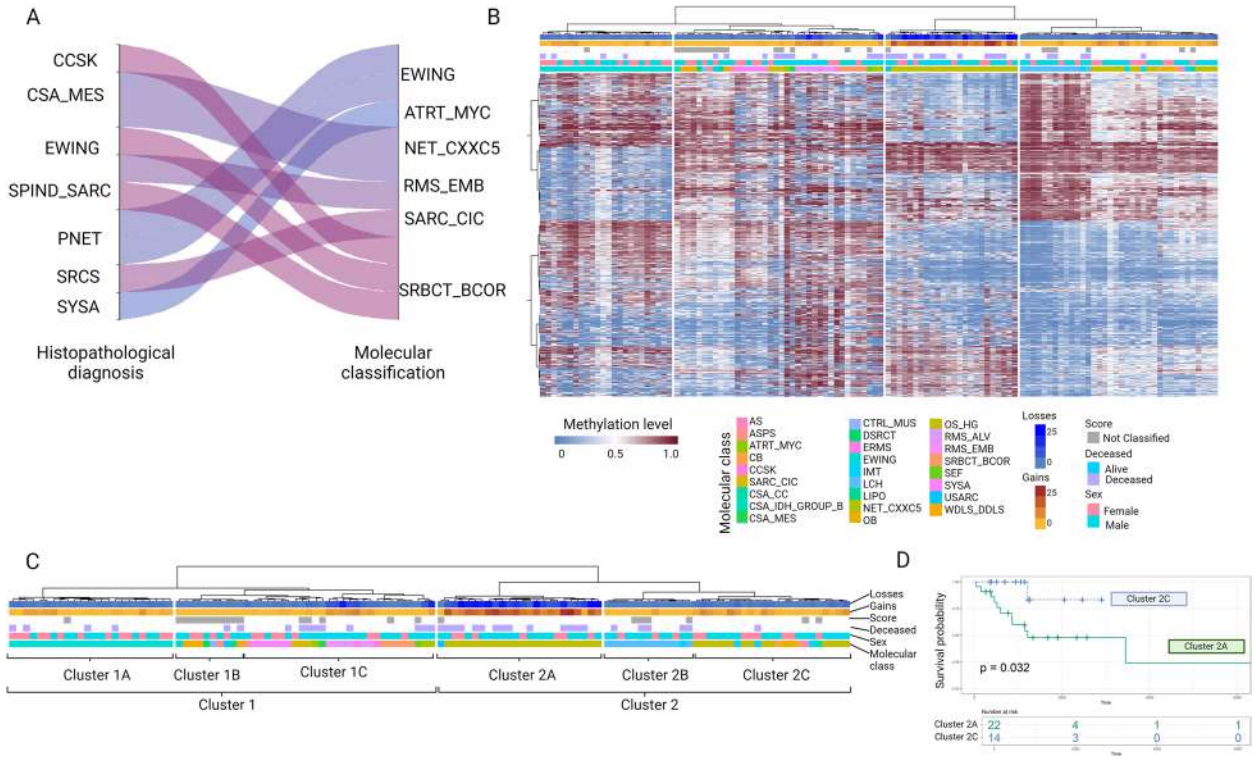

Fig. 1 Molecular classification of pediatric sarcomas. **A** Sankey plot showing molecular classification (right) compared to histopathological diagnosis (left) Altered. **B** Unsupervised hierarchical clustering baked on the 10 thousand most variable probes (Red: methylated, blue: unmethylated CpG site) Individual samples are color-coded labeled in the respective molecular class. **C** Unsupervised hierarchical clustering of samples with cluster names. **D** Survival analysis showing the impact of osteosarcoma methylation subgroups (Overall survival by Kaplan–Meier analysis). SARC_CIC, CIC rearranged sarcoma (Brain classifier); *OB* Osteoblastoma, *EWING* Ewing sarcoma, *OS_HG* High-grade osteosarcoma, *SBRCT_BCOR* BCOR rearranged sarcoma, *CSA_MES* Mesenchymal chondrosarcoma, *ERMS* Embryonal rhabdomyosarcoma, *RMS_EMB* Embryonal rhabdomyosarcoma, *ARMS* Alveolar rhabdomyosarcoma, *CSA_CC* Clear cell chondrosarcoma, *SPIND_SARC* spindle cell sarcoma, *USARC* Undifferentiated sarcoma, *IMT* Inflammatory myofibroblastic tumor, *LCH* Langerhans cell histiocytosis, *CSA_IDH_GROUP_B* Chondrosarcoma IDH group B, *ATRT_MYC* Atypical teratoid rhabdoid tumor, *MYC activated* SYSA, *Synovial sarcoma CB* Chondroblastoma, *AS* Alveolar sarcoma, *ASPS* Alveolar soft part sarcoma, *CCSK* Clear cell sarcoma of the kidney, *CTRL_MUS* Control tissue: muscle, *DSRCT* Desmoplastic small round cell tumors, *LIPO* Lipoma, *SEF* Sclerosing epithelioid fibrosarcoma. Created with BioRender.com (o99u352)

All samples were submitted to hierarchical clustering based on the methylation levels of the 10 thousand most variable CpG sites (Fig. 1B and C). Samples were discriminated into two clusters. Cluster 1 contains the majority of Ewing sarcomas (1 A), not classified tissues (1B), and Ewing-like sarcoma, embryonal rhabdomyosarcomas and high-grade neuroepithelial tumor with *MN1:CXXC5* fusion sarcomas (1 C). Cluster 2 was composed of osteosarcomas showing a high CNA burden (2 A), as well as Langerhans cell histiocytosis, chondroblastoma, osteoblastoma and a subset of osteosarcomas (2B).

Identification of chromosomal structures to confirm the diagnosis of the classifier

To verify the classification given by DNA methylation, we chose entities in which the detection of a translocation is used for diagnosis. These included 16 cases of Ewing sarcoma, *NET_CXXC5* sarcoma, and *BCOR*- and *CIC*-rearranged sarcomas that were submitted to FISH as a routine tool (performed at the diagnosis of the case −2 cases with no RNA available) or RNAseq (for 14 samples which RNA was available in the institutional biobank).

We confirmed that all Ewing sarcoma samples presented *EWSR1-FLI1* or *EWSR1-ERG* translocations, as well as the previously diagnosed as Ewing-like sarcomas that showed *BCOR*- or *CIC*- rearrangement and *NET_CXXC5* (Fig. 2, Supplementary Fig. 1).

Characterization of the recurrent copy number alterations in bone sarcomas

CNA analysis provides a good overview of pediatric sarcomas regarding structural variants. At least one CNA was found in 103 out of 122 sarcoma samples and 19 had a normal profile. Overall, 8q (30%), 14q (20%), 5p (15%) and chr21 (20%) gains and 6q (20%), chr10 (18%), 13q (18%) and 8p (15%) losses were the most recurrent

Table 1 Revised classification based on the classifier and orthogonal techniques results

Case	Histological diagnosis	Classifier prediction	Calibrated score	Identification of the translocation	CNAs
SARC_11	Small round cell sarcoma	*CIC*-rearranged sarcoma	0.93	FISH *CIC*-BREAKAPART	chr6, chr7 losses
SARC_05[a]	Mesenchymal chondrosarcoma	high-grade neuroepithelial tumor with *MN1:CXXC5* fusion	0.98	RNAseq *MN1-CXXC5*	*chr10, chr16, chr18, chr8 and chr9 losses chr5 and chr13 focal losses*
SARC_38	Synovial sarcoma	Atypical teratoid rhabdoid tumor, *MYC* activated	0.99	-	chr6 and chr7 losses, chr12 focal loss
SARC_77[a]	Mesenchymal chondrosarcoma	high-grade neuroepithelial tumor with *MN1:CXXC5* fusion	1.00	RNAseq *MN1-CXXC5*	*chr21q gain chr12, chr3 and chr5 focal gains chr1, chr10q, chr12, chr13, chr16, chr18, chr21, chr3, chr5, chr6, chr8p and chr9 focal losses*
SARC_44	Spindle cell sarcoma	Embryonal Rhabdomyosarcoma	1.00	-	*chr9 (CDKN2A/B) loss chr8 gain*[b]
SARC_111	Primitive neuroectodermal tumor	Ewing sarcoma	1.00	RNAseq *EWSR1-FLI1*	*chr6 and chr8 loss chr7 gain*
SARC_110	Primitive neuroectodermal tumor	Ewing sarcoma	1.00	RNAseq *EWSR1-FLI1*	*Flat genome*
SARC_09	Clear cells sarcoma of the kidney	*BCOR*-rearranged sarcoma	1.00	RNAseq *BCOR-CCNB3*	*chr6 loss*
SARC_29	Ewing sarcoma	*BCOR*-rearranged sarcoma	1.00	RNAseq *BCOR-CCNB3*	chr6 loss
SARC_43	Spindle cell sarcoma	*BCOR*-rearranged sarcoma	1.00	FISH BCOR/CCNB3 inv (X) (p11.4p11.22)	chr10p loss

[a] SARC_77 and SARC_05 are samples from the same individual (biopsy and relapse, respectively). *MN1-CXXC5* was confirmed in the biopsy sample

[b] Typical alteration found in 40% of Embryonal rhabdomyosarcomas [15]

alterations (Fig. 3A). The distribution of the total number of copy number shifts per sample and the number of chromosomes displaying CNAs were significantly higher in malignant tumors (osteosarcoma and embryonal rhabdomyosarcomas) than in benign lesions (Wilcoxon test, $p < 0.05$).

Osteosarcoma was the tumor class with more CNAs (Wilcoxon test, $p < 0.05$, Fig. 3B) and presented the most complex profiles, with many CNAs distributed throughout the genome, in addition to chromothripsis events. This high number of CNAs is more evident in a subset of osteosarcomas (Fig. 1B, cluster 2 A). Although the number of patients is small, these patients had a poorer overall survival compared to those cases that grouped in cluster 2 C (Fig. 1D).

In embryonal rhabdomyosarcomas, 8q (55%), focal 7 (35%), focal 12 (35%) and focal gains in chr22 (35%) as well as 6q (80%), chr9 (45%), chr3 (45%) and chr4 (45%) losses were the most recurrent alterations. While 35% of Ewing sarcomas had 8 gain and 20% had chr12 gain, the Ewing-like sarcomas showed a flat genome with few recurrent focal CNAs (Fig. 3C). Ewing-like sarcomas included BCOR- (10q loss in 20% samples) and *CIC*-rearranged sarcoma (3p gain in 45%, 3q loss in 45% and 21p gain in 45% samples). The synovial sarcomas presented 13p and chr21 gain in less than 20% of the cases.

Nonmalignant molecular classes were attributed to 17 cases: 4 inflammatory myofibroblastic tumors, 1 control muscle and 12 Langerhans cell histiocytosis. Control muscle showed a flat genome, corresponding to normal tissue. However, inflammatory myofibroblastic tumors and Langerhans cell histiocytosis (histopathological diagnosis of osteosarcoma and Ewing sarcoma) presented CNAs, with 6q loss (20%) and 8q gain (15%) being the most frequent. One sample, SARC_28, with the histopathological diagnosis of osteosarcoma and methylation-based classification of Langerhans cell histiocytosis (calibrated score 0.9) showed *MYC* amplification. These findings suggest that non-cancer cell infiltration may contribute to the misassignment of a methylation class, however canonical DNA alterations (here exemplified by CNA) can still be identified, showing that sample cellular content included sarcoma cells, albeit to a lesser extent.

Using a deconvolution analysis, we found that samples classified as inflammatory environment had a higher content of neutrophils. Tumor purity was also investigated and, as expected, tumor cells content was correlated with methylation-based calibrated score (Fig. 4).

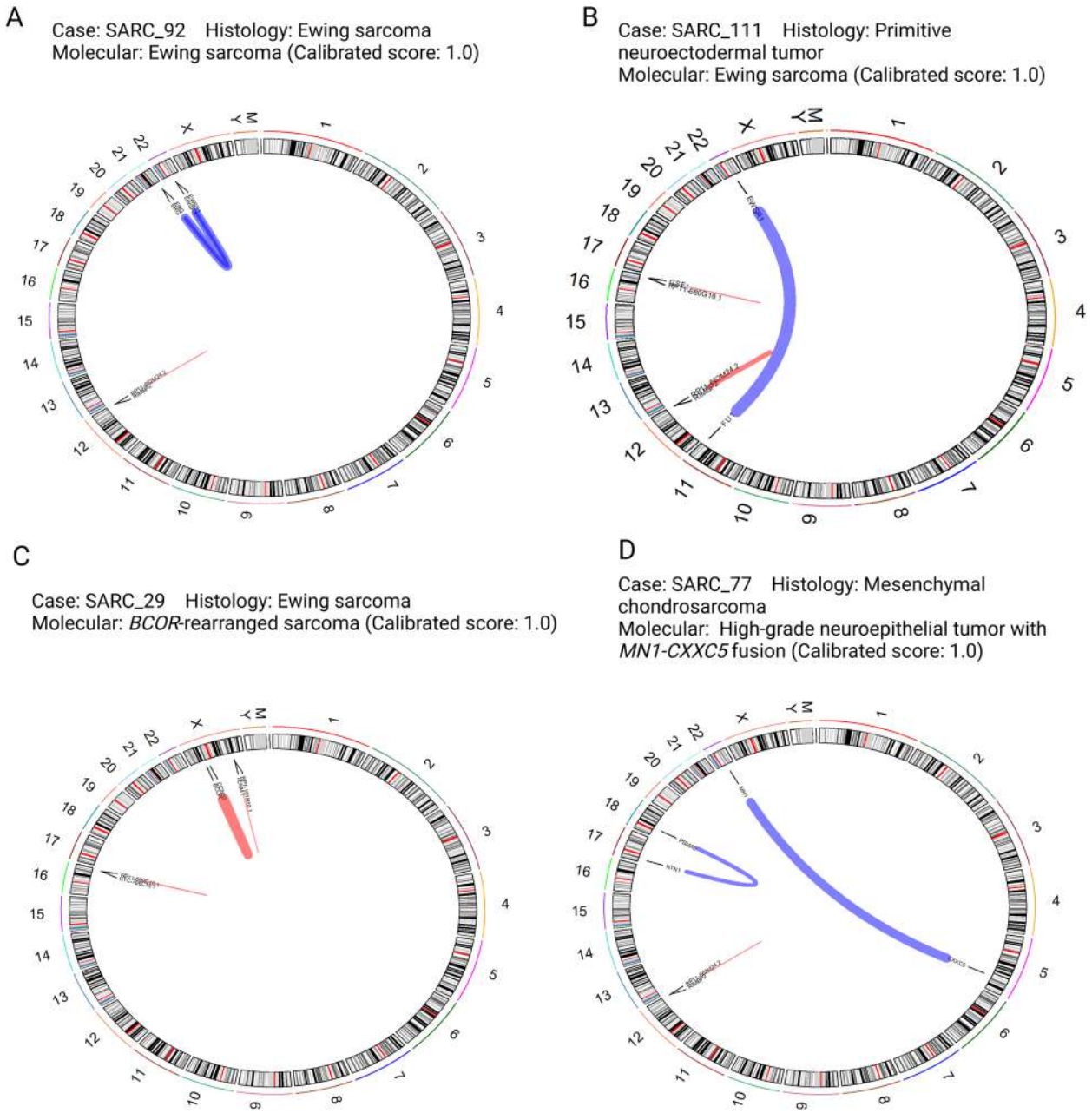

Fig. 2 Identification of the translocation pointed by the methylation-based classification. Example of 4 cases with diagnostic fusion genes, representing **A** Ewing sarcoma, **B** PNET, **C** and **D** as original diagnostic that changed to *BCOR*-rearranged sarcoma and high-grade neuroepithelial tumor with *MN1:CXXC5* fusion, respectively. Circus plot showing 22 autosomes and 2 sexual chromosomes as well as the mitochondrial genome, all chromosomes with cytoband information. Fusion events as shown as links between locations in chromosomes (with gene names), red and blue lines indicate intrachromosomal and interchromosomal fusions, respectively. The width of each line varies according to how many reads support the fusion event. Created with BioRender.com (z83l932)

Discussion

Current WHO classification of mesenchymal tumors distinguishes embryonal, alveolar, pleomorphic and spindle cell/sclerosing rhabdomyosarcoma based on histologic, genetic and clinical features. Some additional entities are recognized based on the specific molecular alteration [32]. Several of the new entities were firstly recognized by genome-wide DNA methylation pattern, with further characterization of molecular alterations by orthogonal techniques, characterization of the molecular

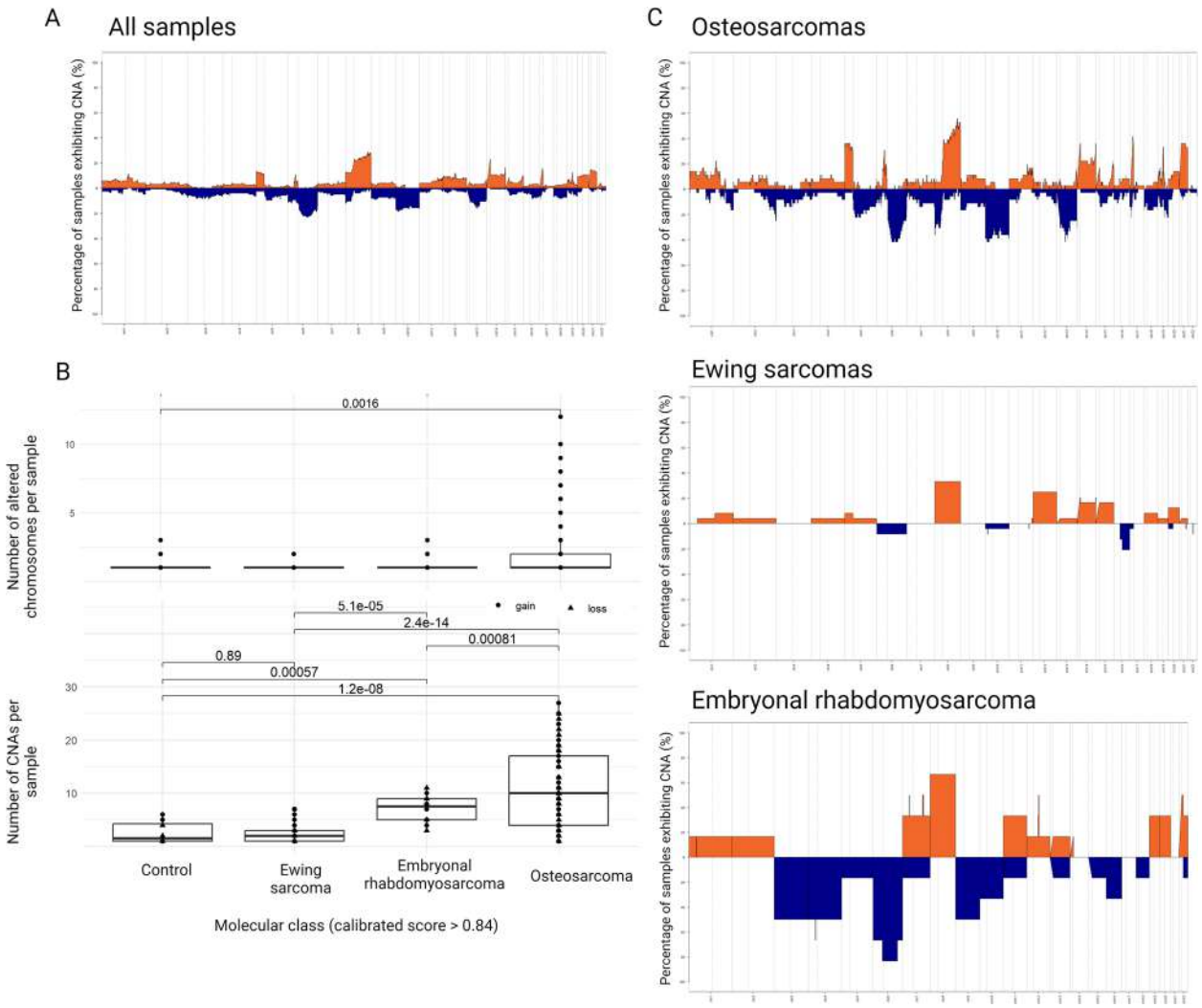

Fig. 3 Copy number profiles of pediatric sarcomas. **A** Recurrent copy number alterations (%) in all pediatric sarcomas and among the more common molecular classes. Gray line means a chromosome and dashed gray lines centromeric regions dividing p and q arms. Zero means two copies, above zero gain and below zero loss. **B** Number of altered chromosomes and number of copy number alterations among the more common molecular classes, including benign samples. **C** Percentage of copy number alterations among the more common molecular classes. Orange: gain, Blue: loss. Created with BioRender.com (x24n652)

pathogenesis, clinical behavior and actionable vulnerabilities, such as sarcomas with *FUS-TFCP2* or *EWSR1-TFCP2* fusions [33]. The recognition of the new entities enables future researchers and requires the characterization of a high number of samples. An accurate diagnosis directly impacts the prognosis of patients, modifying stratification and management. The agreement between both histopathology and molecular classification is similar to other studies [17, 19]. All osteosarcomas and Ewing sarcomas were confirmed by the molecular classification. Ewing-like sarcomas were reclassified as sarcomas with *BCOR* or *CIC* rearrangements, which were later verified using orthogonal techniques (RNAseq or FISH). Even though these tumors are distinct from Ewing sarcoma, the management of the patients is similar [34], however *CIC*-rearranged sarcomas are extremely aggressive and have a far lower prognosis than Ewing sarcoma [35]. One case (with two samples, SARC-05 and SARC_77) was diagnosed as mesenchymal chondrosarcoma and showed low calibrated score (0.2) using the Sarcoma (v12.2) classifier but was classified as high-grade neuroepithelial tumor with MN1:CXXC5 fusion (calibrated score 1.0) using Brain tumor (mnp v12.5) classifier. This fusion was further confirmed by RNAseq. We believe that this case was a sarcoma with *MN1-CXXC5* fusion, but this class is not yet represented in the Sarcoma classifier (v12.2).

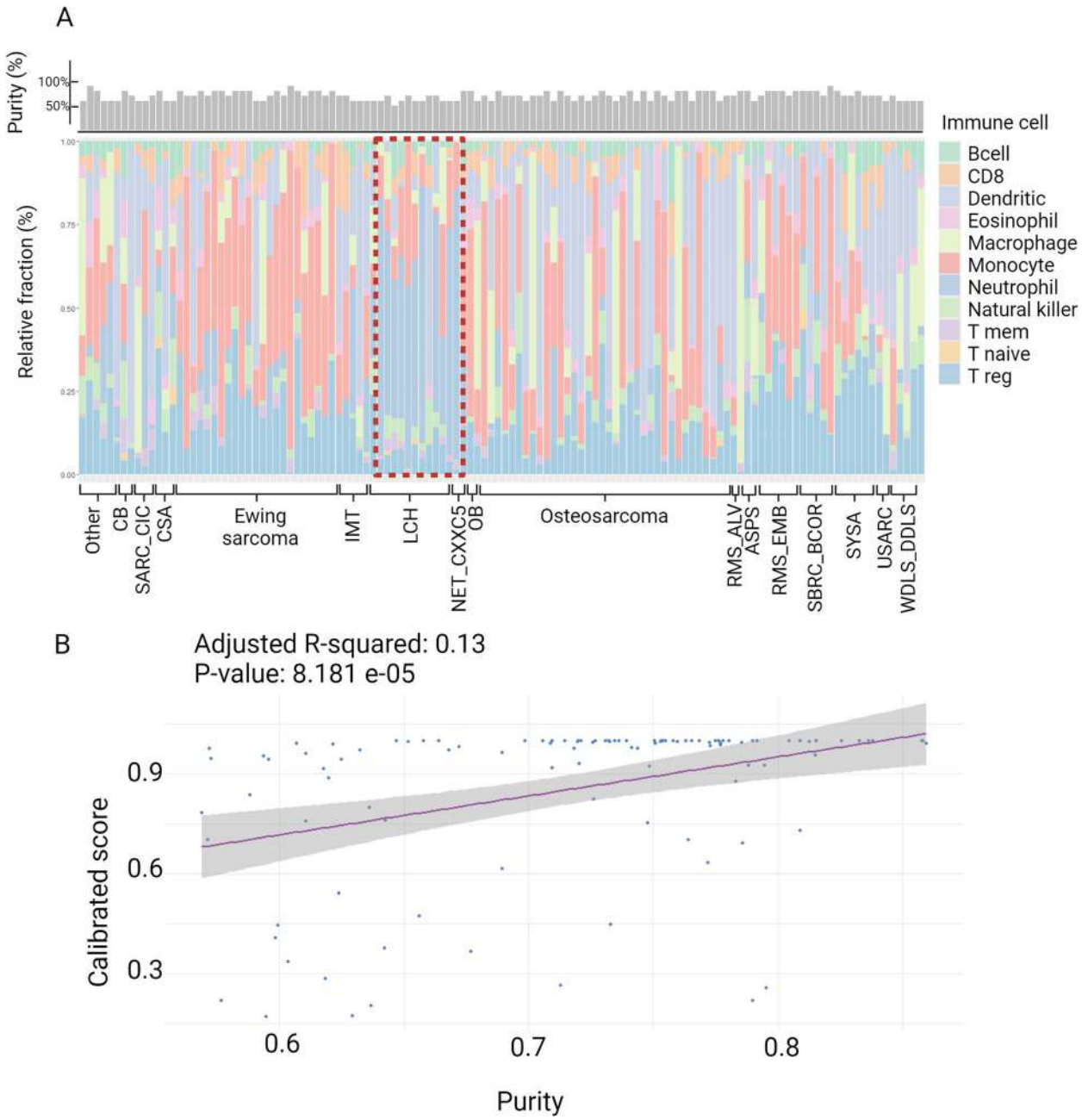

Fig. 4 **A** Tumor infiltrating leukocytes, red line highlighting samples with higher content of neutrophils, classified as inflammatory environment, top panel showing percentage of tumor cell purity in the assayed samples and **B** Correlation between tumor purity and calibrated score. NET_CXXC5, high-grade neuroepithelial tumor with MN1:CXXC5 fusion; SARC_CIC, CIC rearranged sarcoma (Brain classifier); OB, Osteoblastoma; EWING, Ewing sarcoma; OS_HG, High grade osteosarcoma; SBRCT_BCOR, BCOR rearranged sarcoma; CSA_MES, Mesenchymal chondrosarcoma; ERMS, Embryonal rhabdomyosarcoma; RMS_EMB, Embryonal rhabdomyosarcoma; ARMS, Alveolar rhabdomyosarcoma; CSA_CC, Clear cell chondrosarcoma; USARC, Undifferentiated sarcoma; IMT, Inflammatory myofibroblastic tumor; LCH, Langerhans cell histiocytosis; CSA_IDH_GROUP_B, Chondrosarcoma IDH group B; ATRT_MYC, Atypical teratoid rhabdoid tumor, MYC activated; SYSA, Synovial sarcoma; CB, Chondroblastoma. Other, classes represented by one sample (AS, Alveolar sarcoma; ASPS, Alveolar soft part sarcoma; CCSK, Clear cell sarcoma of the kidney; CTRL_MUS, Control tissue: muscle; DSRCT, Desmoplastic small round cell tumors; LIPO, Lipoma; SEF, Sclerosing epithelioid fibrosarcoma; respectively). Created with BioRender.com (v43j268)

Samples that were not classified may reflect the low number of classes represented in the classifier which will evolve with the inclusion of additional studies. A practical example was the cases with molecular classification of NET_CXXC5 or CNS_SARC_CIC. They did not have a high calibrated score using the Sarcoma (v12.2) classifier, but by the Brain tumor (mnp v12.5) classifier reached a high calibrated score, which demonstrates a limitation of Sarcoma (v12.2) to deal with more rare entities [19]. In our cohort, four cases of clear cell sarcoma of the kidney were evaluated, from which two did not reach the cut-off of 0.9 calibrated-score, one showed agreement between methylation class and histopathological diagnosis, and one was classified as *BCOR*-rearranged sarcoma. The majority of CSSK have either *YWHAE-NUTM2B/E* fusion or BCOR-internal tandem duplication (ITD), while *BCOR-CCNB3* fusion is extremely rare in this context [36], but can be predicted by the Sarcoma classifier (v12.2).

We identified eight (6.56%) tumors classified as Langerhans cell histiocytosis or Inflammatory myofibroblastic tumor, which can be a consequence of the inflammatory infiltrated cells interfering with the molecular classification [17], suggesting that these classes may be related to the tissue sampling for the biobank. Lower tumor purity can lead to incorrect molecular classification as control tissues since the addition of non-cancer cells biases the methylation pattern [37]. This is a limitation of the study as the frozen samples were not reviewed by a pathologist before DNA extraction. Thus, in the 122 samples analyzed, there was a correlation between tumor cell proportion and molecular classification calibrated score, confirming that tissue sampling is important for molecular classification. An independent study showed that the prediction is related to the tumor purity content but not with the accuracy [19]. For practical clinical settings, it is of utmost importance to establish the minimum of neoplastic cells required by the methylation classifiers to avoid performing expensive methylation testing when the expectation for false positive prediction is high. Also, it was previously reported that Sarcoma classifier shows low specificity for both Langerhans histiocytosis and inflammatory myofibroblastic tumor, contributing to false positive assignments [38]. Accordingly, the analysis of tumor-infiltrating immune cells reported a group of samples with higher content of neutrophils, classified as Langerhans cell histiocytosis. Considering that the histological-based diagnosis is correct, these cells could be tumor-associated neutrophils (TANS) that are able to polarize to pro- or anti-tumor phenotype. TANS can secrete elastase to tumor cells, which activate *PI3K* pathway and contribute to cell proliferation. They also contribute to genetic instability and immunosuppression as well as mediate tumor cell lysis and CD8 cell stimulation [39].

The hierarchical clustering based on the methylation levels of the most variable CpGs grouped all Ewing sarcomas into the same cluster (1 A) and discriminated the osteosarcomas into two distinct clusters (2 A and 2 C) with an impact in overall survival. The small number of cases is a limitation for this finding, and thus requiring further studies to explore both groups of osteosarcomas. Cluster 2 A was composed of 23 high-grade osteosarcomas, with higher CNA burden when compared to other 14 samples in cluster 2 C. Although the variance analysis identified two groups, all osteosarcomas share some common alterations compared to bone tissues. These alterations were enriched for biological processes related to skeletal development and cell differentiation and together with *DNMT3B* overexpression suggest that DNA methylation is disrupted in osteosarcomas [40]. Future studies are needed to investigate molecular and clinical differences between osteosarcoma subgroups.

Cluster 2 C was also composed of benign tumors (chondroblastoma and osteoblastoma). Cluster 1B was formed by samples that were not classified by DNA methylation, which may indicate novel molecular classes not yet represented in the Sarcoma Classifier [18, 19]. All samples classified as Langerhans cell histiocytosis or inflammatory myofibroblastic tumors fell in the same cluster (2B), in between two osteosarcoma subclusters and the sample classified as control muscle tissue felt in cluster 1B, close to Ewing sarcomas (1 A). *MYC* amplification was observed in one sample classified as Langerhans cell histiocytosis (sarc_28). Alterations in *MYC* and *MYC*-associated genes are enriched in the LCH cells [41, 42].

Similar to other pediatric solid tumors, sarcomas show only few recurrent mutations [43, 44] meaning that other alteration's types are associated with cell transformation. Our finding shows that malignant molecular classes had a much larger distribution of CNAs than benign neoplasms and control tissues, similar to other studies [45]. The recurrent CNAs from this study are similar to other studies in sarcomas: 8q (30%) and 14q (20%), 5p (15%) and 21 (20%) gains and 6q (20%), 10 (18%), 13q (18%) and 8p (15%) losses are the most common events in pediatric sarcomas, in agreement to other studies [7, 43]. CNA profile of sarcomas can be helpful to stratify diagnosis, but the lack of information combining multiple CNAs is a disadvantage to obtain diagnosis suggestions.

In osteosarcomas, single cell analysis found small subclonal diversification based on conserved CNAs over tumor evolutionary process or acquired CNAs during chemotherapy, suggesting that early catastrophic events, such as chromothripsis and chromoanasynthesis, may

be the primary mechanism of osteosarcoma development [46, 47]. Chromothripsis were first described in osteosarcoma [48] and defined as grouped chromosomal shattering followed by random chromosomal restitching, resulting in tens to hundreds of inter- and intra-chromosomal rearrangements. Chromothripsis is linked to aggressive malignancies and poor patient survival and are related to failure of DNA repair pathways [49]. The high frequency of structural variants were hypothesized to be related to loss-translocation-amplification (LTA) chromothripsis, which is preceded by a single double-strand break, *TP53* inactivation and segmental amplifications, often amplifying oncogenes to high copy numbers and mediates a rapid malignant evolution [50]. In agreement with this hypothesis, 15% of high grade osteosarcomas present *MYC* amplification that is generated by a single chromothripsis-driven amplification event on chromosome 8q with multiple translocations [51].

Ewing sarcomas are more genomic stable tumors when compared to osteosarcomas [45] with few recurrent CNAs, such as trisomy of chromosomes 8 (50%) and 12 (25%), along with 1q amplification (19%) and 16q deletions (17%) [52]. We found gain of chromosomes 8 and 12 in 35% and 20%, respectively, and loss of chromosome 16q in 19% of the cases. The *EWS-FLI1* translocation is necessary but not sufficient to Ewing sarcoma onset [53]. Therefore, searching for other somatic mechanisms, such as DNA methylation and CNAs are important to better understand the genomic landscape of those tumors.

Embryonal rhabdomyosarcomas also present some few recurrent copy number gains involving chromosomes 2, 8, 12, and 20 as well as recurrent loss of heterozygosity (LOH) of chromosomes 1, 7, 14, and X [7]. Gain of chromosome 8 with *MYC* amplification, which seems to be an early molecular event on rhabdomyosarcomatogenesis and involved in tumor initiation. It seems to be a good prognosis marker for these cases while chr3q loss is observed in a few cases (approximately 2%) and is related to poor prognosis [54].

The molecular classification based on DNA methylation and CNAs could benefit the patients by providing a more accurate and agnostic diagnosis. Ideally, the analysis of the sample should have the histology evaluated by the pathologist to select the best area for the assay. The DNA methylation-based diagnosis is not sufficient by itself to render a confident diagnosis in all cases, requiring an integrated diagnostic approach, similar to brain tumors. This is more relevant for cases when the DNA methylation classification results are discrepant from the histopathologic diagnosis. Nevertheless, the molecular analysis disclosed entities not considered by histopathology alone (CIC- or BCOR-altered tumors and *MN1-CXXC5* fusion tumors) and suggested the existence of two groups of high-grade osteosarcomas. Pediatric cancer is rare and often less represented in studies and databases, which gives additional importance for this study.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-024-13159-9.

Supplementary Material 1.

Acknowledgements
The authors acknowledge the Biobank and Hospital Medical Registry Center (SAME), as well as the nurses from the Boldrini Children's Hospital.

Authors' contributions
F.L.T.S.: data curation, conceptualization, methodology, validation, formal analysis, original draft preparation. M.F.E.: methodology, data curation. J.S.R: methodology. D.R.L: patients' information. M.T.: patients' information. T.J.: patients' information. I.A.C.: sample collection, patients' information. A.E.C.: sample collection, patients' information. L.H.P.: patients' information. P.P.Z : conceptualization, supervision, final revision, project administration, funding acquisition. P.Y.J.: conceptualization, formal analysis and final revision, project administration, funding acquisition. M.M.: conceptualization, formal analysis, supervision, original draft preparation and final revision, project administration, funding acquisition.

Funding
This research was funded by PRONON (SIPAR 25000.012259/2019-42 and 25000.211368/2019-41) and FAPESP 2021/06782-7. Fellowship was provided for F.L.T.S (FAPESP 2022/04781-6) and MM (CNPq 311141/2021-8).

Data availability
Relevant data is provided within the manuscript, supplementary information files or made available online, as described in the manuscript.

Declarations

Ethics approval and consent to participate
The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of Boldrini Children's Hospital (CAAE: 28386820.7.0000.5376, CAAE: 22737219.1.0000.5376, CAAE: 44219021.6.0000.5376). Informed consent was obtained from all subjects involved in the study.

Consent for publication
Not Applicable. shared information does not compromise individuals' anonymity.

Competing interests
The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Received: 26 July 2024 Accepted: 7 November 2024
Published online: 20 November 2024

References

1. Hingorani P, Janeway K, Crompton BD, Kadoch C, Mackall CL, Khan J, et al. Current state of pediatric sarcoma biology and opportunities for future discovery: a report from the sarcoma translational research workshop. Cancer Genet. 2016;209:182–94.
2. Weaver R, O'Connor M, Carey Smith R, Halkett GK. The complexity of diagnosing sarcoma in a timely manner: perspectives of health professionals, patients, and carers in Australia. BMC Health Serv Res. 2020;20:711.

3. Grünewald TG, Alonso M, Avnet S, Banito A, Burdach S, Cidre-Aranaz F, et al. Sarcoma treatment in the era of molecular medicine. EMBO Mol Med. 2020;12:e11131.
4. Riggi N, Cironi L, Suvà M-L, Stamenkovic I. Sarcomas: genetics, signalling, and cellular origins. Part 1: the fellowship of TET. J Pathol. 2007;213:4–20.
5. Gutiérrez-Jimeno M, Alba-Pavón P, Astigarraga I, Imízcoz T, Panizo-Morgado E, García-Obregón S, et al. Clinical value of NGS genomic studies for Clinical Management of Pediatric and Young Adult Bone Sarcomas. Cancers. 2021;13(21):5436.
6. Toguchida J, Nakayama T. Molecular genetics of sarcomas: applications to diagnoses and therapy. Cancer Sci. 2009;100:1573–80.
7. Cheng L, Pandya PH, Liu E, Chandra P, Wang L, Murray ME, et al. Integration of genomic copy number variations and chemotherapy-response biomarkers in pediatric sarcoma. BMC Med Genomics. 2019;12(Suppl 1):23.
8. Nacev BA, Jones KB, Intlekofer AM, Yu JSE, Allis CD, Tap WD, et al. The epigenomics of sarcoma. Nat Rev Cancer. 2020;20:608–23.
9. Nacev BA, Sanchez-Vega F, Smith SA, Antonescu CR, Rosenbaum E, Shi H, et al. Clinical sequencing of soft tissue and bone sarcomas delineates diverse genomic landscapes and potential therapeutic targets. Nat Commun. 2022;13:3405.
10. Delattre O, Zucman J, Plougastel B, Desmaze C, Melot T, Peter M, et al. Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. Nature. 1992;359:162–5.
11. Hettmer S, Linardic CM, Kelsey A, Rudzinski ER, Vokuhl C, Selfe J, Ruhen O, Shern JF, Khan J, Kovach AR, Lupo PJ, Gatz SA, Schäfer BW, Volchenboum S, Minard-Colin V, Koscielniak E, Hawkins DS, Bisogno G, Sparber-Sauer M, Venkatramani R, Merks JHM, Shipley J. Molecular testing of rhabdomyosarcoma in clinical trials to improve risk stratification and outcome: a consensus view from European paediatric soft tissue sarcoma Study Group, Children's Oncology Group and Cooperative Weichteilsarkom-Studiengruppe. Eur J Cancer. 2022;172:367–86.
12. Huang X, Li G, Li L, Wang J, Shen J, Chen Y, et al. Establishing an RNA fusions panel in soft tissue sarcoma with clinical validation. Sci Rep. 2023;13:4403.
13. Miettinen M, Felisiak-Golabek A, Luiña Contreras A, Glod J, Kaplan RN, Killian JK, et al. New fusion sarcomas: histopathology and clinical significance of selected entities. Hum Pathol. 2019;86:57–65.
14. Choi JH, Ro JY. The 2020 WHO classification of tumors of bone: an updated review. Adv Anat Pathol. 2021;28:119–38.
15. Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. Nature. 2018;555:469–74.
16. Capper D, Stichel D, Sahm F, Jones DTW, Schrimpf D, Sill M, et al. Practical implementation of DNA methylation and copy-number-based CNS tumor diagnostics: the Heidelberg experience. Acta Neuropathol. 2018;136:181–210.
17. Koelsche C, Schrimpf D, Stichel D, Sill M, Sahm F, Reuss DE, et al. Sarcoma classification by DNA methylation profiling. Nat Commun. 2021;12:498.
18. Roohani S, Ehret F, Perez E, Capper D, Jarosch A, Flörcken A, et al. Sarcoma classification by DNA methylation profiling in clinical everyday life: the Charité experience. Clin Epigenetics. 2022;14:149.
19. Lyskjaer I, De Noon S, Tirabosco R, Rocha AM, Lindsay D, Amary F, et al. DNA methylation-based profiling of bone and soft tissue tumours: a validation study of the DKFZ Sarcoma Classifier. J Pathol Clin Res. 2021;7:350–60.
20. Desjardins P, Conklin D. NanoDrop microvolume quantitation of nucleic acids. J Vis Exp. 2010. https://doi.org/10.3791/2565.
21. Sciot R. MDM2 amplified sarcomas: a Literature Review. Diagnostics (Basel). 2021;11(3):496.
22. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30:1363–9.
23. Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. 2014;15:503.
24. Arneson D, Yang X, Wang K. MethylResolver—a method for deconvoluting bulk DNA methylation profiles into known and unknown cell contents. Commun Biol. 2020;3:1–13.
25. Daenekas B, Pérez E, Boniolo F, Stefan S, Benfatto S, Sill M, et al. Conumee 2.0: enhanced copy-number variation analysis from DNA methylation arrays for humans and mice. Bioinformatics. 2024;40(2):btae029.
26. Hovestadt VZM, Zapatka M. Conumee: enhanced copy-number variation analysis using Illumina DNA methylation arrays R Package version 190. 2020.
27. Andrews S. FastQC A quality control tool for high throughput sequence data. Babraham bioinformatics. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Accessed 18 Apr 2024.
28. Haas BJ, Dobin A, Li B, et al. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. Genome Biol. 2021;20:213.
29. Lågstad S, Zhao S, Hoff AM, Johannessen B, Lingjærde OC, Skotheim RI. Chimeraviz: a tool for visualizing chimeric RNA. Bioinformatics. 2017;33:2954–6.
30. Tomassen T, Koelsche C, de Leng WWJ, Kommoss FKF, Voijs CMA, Peeters T, et al. Calcifying fibrous tumor and inflammatory myofibroblastic tumor are epigenetically related: a comparative genome-wide methylation study. Ann Diagn Pathol. 2019;41:102–5.
31. Nakamine H, Yamakawa M, Yoshino T, Fukumoto T, Enomoto Y, Matsumura I. Langerhans Cell histiocytosis and Langerhans Cell Sarcoma: current understanding and Differential diagnosis. J Clin Exp Hematop. 2016;56:109–18.
32. Anderson WJ, Doyle LA. Updates from the 2020 World Health Organization Classification of Soft Tissue and Bone Tumours. Histopathology. 2021;78:644–57.
33. Schöpf J, Uhrig S, Heilig CE, Lee K-S, Walther T, Carazzato A, et al. Multi-omic and functional analysis for classification and treatment of sarcomas with FUS-TFCP2 or EWSR1-TFCP2 fusions. Nat Commun. 2024;15:51.
34. Ewing Sarcoma Treatment (PDQ®): Health Professional Version. PDQ Pediatric Treatment Editorial Board. 2002. https://www.ncbi.nlm.nih.gov/books/NBK66045.3/. Accessed 18 Apr 2024.
35. Yoshida A. Ewing and ewing-like sarcomas: a morphological guide through genetically-defined entities. Pathol Int. 2023;73:12–26.
36. Dorwal P, Abou-Seif C, Ng J, Super L, Chan Y, Rathi V. Clear cell sarcoma of the kidney (CCSK) with BCOR-CCNB3 Fusion: a rare case Report with a brief review of the literature. Pediatr Dev Pathol. 2023;26:149–52.
37. Wenger A, Carén H. Methylation profiling in diffuse gliomas: diagnostic value and considerations. Cancers. 2022;14:5679.
38. Miettinen M, Abdullaev Z, Turakulov R, Quezado M, Luiña Contreras A, Curcio CA, et al. Assessment of the utility of the sarcoma DNA methylation Classifier in Surgical Pathology. Am J Surg Pathol. 2024;48:112–22.
39. Masucci MT, Minopoli M, Carriero MV. Tumor Associated neutrophils. Their role in Tumorigenesis, Metastasis, Prognosis and Therapy. Front Oncol. 2019;9:1146.
40. Pires SF, de Barros JS, da Costa SS, de Oliveira Scliar M, Van Helvoort Lengert A, Boldrini É, et al. DNA methylation patterns suggest the involvement of DNMT3B and TET1 in osteosarcoma development. Mol Genet Genomics. 2023;298:721–33.
41. Halbritter F, Farlik M, Schwentner R, Jug G, Fortelny N, Schnöller T, et al. Epigenomics and single-cell sequencing define a developmental hierarchy in Langerhans Cell histiocytosis. Cancer Discov. 2019;9:1406–21.
42. Jean Gogusev Louise Telvi Ichiro Murakami Alexandre Stojkoski Christophe Glorion Francis Jaubert. Molecular cytogenetic alterations of C-MYC, CCND1 and HER-2/neu oncogenes in Langherhans cell histiocytosis. Proc Am Assoc Cancer Res. 2006;66:992.
43. Gröbner S, Jones BC, Hettmer S, Niemeyer C, Chavez L, Zapatka M, et al. The Landscape of Genomic Alterations Across Childhood Cancers. Nature. 2018;555(7696):321–7.
44. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. Nature. 2020;578:82–93.
45. Köster J, Piccinelli P, Arvidsson L, Vult von Steyern F, Bedeschi Rego, De Mattos C, Almquist M, et al. The diagnostic utility of DNA copy number analysis of core needle biopsies from soft tissue and bone tumors. Lab Invest. 2022;102:838–45.
46. Rajan S, Zaccaria S, Cannon MV, Cam M, Gross AC, Raphael BJ, et al. Structurally complex osteosarcoma genomes exhibit limited heterogeneity within individual tumors and across Evolutionary Time. Cancer Res Commun. 2023;3:564–75.
47. Pires SF, de Barros JS, da Costa SS, do Carmo GB, de Oliveira Scliar M, Lengert HA, et al. Analysis of the Mutational Landscape of Osteosarcomas

Identifies Genes Related to Metastasis and Prognosis and Disrupted Biological Pathways of Immune Response and Bone Development. Int J Mol Sci. 2023;24(13):10463.
48. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell. 2011;144:27–40.
49. Pellestor F, Gaillard JB, Schneider A, Puechberty J, Gatinois V. Chromoanagenesis, the mechanisms of a genomic chaos. Semin Cell Dev Biol. 2022;123:90–9.
50. Valle-Inclan JE, De Noon S, Trevers K, Elrick H, Tanguy M, Butters T et al. Mechanisms underpinning osteosarcoma genome complexity and evolution. bioRxiv. 2023;2023.12.29.573403.
51. De Noon S, Ijaz J, Coorens TH, Amary F, Ye H, Strobl A, et al. MYC amplifications are common events in childhood osteosarcoma. J Pathol Clin Res. 2021;7:425–31.
52. Jahromi MS, Jones KB, Schiffman JD. Copy number alterations and methylation in Ewing's Sarcoma. Sarcoma. 2011;2011:362173.
53. Toomey EC, Schiffman JD, Lessnick SL. Recent advances in the molecular pathogenesis of Ewing's sarcoma. Oncogene. 2010;29:4504–16.
54. Dehner CA, Bell RC, Cao Y, He K, Chrisinger JSA, Armstrong AE, et al. Loss of chromosome 3q is a prognostic marker in Fusion-Negative Rhabdomyosarcoma. JCO Precis Oncol. 2023;7:e2300037.

Publisher's Note