HHS Public Access

Author manuscript

JCO Precis Oncol. Author manuscript; available in PMC 2018 October 06.

Published in final edited form as:

JCO Precis Oncol. 2017 ; 2017: .

DNA Methylation–Based Classifier for Accurate Molecular Diagnosis of Bone Sarcomas

S. Peter Wu, MD[1,*], **Benjamin T. Cooper, MD**[1,*], **Fang Bu, MD**[2], **Christopher J. Bowman, MD**[2], **J. Keith Killian, MD, PhD**[3], **Jonathan Serrano**[2], **Shiyang Wang, MD, PhD**[4], **Twana M. Jackson, MD**[4], **Daniel Gorovets, MD**[1], **Neerav Shukla, MD**[5], **Paul A. Meyers, MD**[5], **David J. Pisapia, MD**[6], **Richard Gorlick, MD**[7], **Marc Ladanyi, MD**[3], **Kristen Thomas, MD**[2], **Matija Snuderl, MD**[2,†], and **Matthias A. Karajannis, MD, MS**[5,†]

[1]Department of Radiation Oncology, NYU Langone Medical Center, New York, NY

[2]Department of Pathology, NYU Langone Medical Center, New York, NY

[3]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY

[4]Department of Pediatrics, NYU Langone Medical Center, New York, NY

[5]Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY

[6]Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, NY

[7]Department of Pediatrics, Children's Hospital at Montefiore, Albert Einstein College of Medicine, Bronx, NY

Abstract

Purpose—Pediatric sarcomas provide a unique diagnostic challenge. There is considerable morphologic overlap between entities, increasing the importance of molecular studies in the diagnosis, treatment, and identification of therapeutic targets. We developed and validated a genome-wide DNA methylation based classifier to differentiate between osteosarcoma, Ewing's sarcoma, and synovial sarcoma.

Materials and Methods—DNA methylation status of 482,421 CpG sites in 10 Ewing's sarcoma, 11 synovial sarcoma, and 15 osteosarcoma samples were determined using the Illumina Infinium HumanMethylation450 array. We developed a random forest classifier trained from the 400 most differentially methylated CpG sites within the training set of 36 sarcoma samples. This classifier was validated on data drawn from The Cancer Genome Atlas (TCGA) synovial sarcoma, TARGET Osteosarcoma, and a recently published series of Ewing's sarcoma.

Results—Methylation profiling revealed three distinct patterns, each enriched with a single sarcoma subtype. Within the validation cohorts, all samples from TCGA were accurately classified

Corresponding Authors: Matthias A. Karajannis, MD, MS, Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, Phone: (212) 639-3171 | Fax: (212) 717-3239 | karajanm@mskcc.org, Matija Snuderl, MD, Departments of Pathology and Neurology, Division of Molecular Pathology and Diagnostics, 240 E 38th Street, 22nd Floor, New York, NY 10016, Phone: (646) 501-5281 | Fax: (212) 263-7916 | matija.snuderl@nyumc.org.

[*]S. Peter Wu and Benjamin T. Cooper contributed equally to this work.

[†]Matija Snuderl and Matthias A. Karajannis are senior co-authors.

Conflicts of Interest: None

as synovial sarcoma (10/10, sensitivity and specificity 100%), all but one sample from TARGET-OS were classified as osteosarcoma (85/86, sensitivity 98%, specificity 100%) and 14/15 Ewing's sarcoma samples classified correctly (sensitivity 93%, specificity 100%). The single misclassified osteosarcoma sample demonstrated high EWSR1 and ETV1 expression on RNA-seq although no fusion was found on manual curation of the transcript sequence. Two additional clinical samples, that were difficult to classify by morphology and molecular methods, were classified as osteosarcoma when previously suspected to be a synovial sarcoma and Ewing's sarcoma on initial diagnosis, respectively.

Conclusion—Osteosarcoma, synovial sarcoma, and Ewing's sarcoma have distinct epigenetic profiles. Our validated methylation-based classifier can be used to provide diagnostic assistance when histological and standard techniques are inconclusive.

Introduction

Pediatric sarcomas constitute a rare and diverse group of mesenchymal malignancies of soft tissue and bone. Ewing's Sarcoma (EWS), synovial sarcoma (SS), and osteosarcoma (OS) are among the most common malignant solid tumors in children [1]. While these tumors can occur in similar anatomical locations, optimal management and treatment strategies differ substantially depending on the tumor type [2, 3]. Accurate diagnosis is thus paramount for clinical management, but can be challenging.

Histologically, EWS is mainly composed of small round blue cells [4], however EWS occasionally consists of larger, more pleomorphic cells, making the diagnosis of EWS based solely on histopathological analysis unreliable [5]. The discovery of the EWSR1-FLI1 fusion detected by fluorescent *in situ* hybridization (FISH) has significantly improved diagnostic accuracy, but is only present in approximately 85% of samples that are histologically consistent with EWS [6]. In the remainder of cases, most tumors harbor a fusion of the *EWSR1* gene with a different member of the E-26 Transformation Specific (ETS) family of transcription factors [7][8][9][10][11].

Synovial sarcoma (SS) typically has a biphasic appearance consisting of epithelioid and fibroblast-like spindle cell components, however, a monophasic spindle cell variant is also commonly seen. Furthermore, a poorly differentiated (round cell) variant exists that is histologically indistinct from other poorly differentiated tumors, complicating diagnosis [12]. Analogous to EWS, SS are characterized by a pathognomonic t(X;18)(p11.2;q11.2) translocation that can be detected by cytogenetics and aid in the diagnosis [13]. Nonetheless, the clinical behavior of SS is varied, indicating biological heterogeneity [14].

Similar to SS and EWS, OS also has a variety of histologic appearances, with subtypes including fibroblastic, osteoblastic, chondroblastic, giant cell, telangiectatic, and small cell. Morphologic variants include spindle-cell OS (resembling fibrosarcoma or monophasic synovial sarcoma), high-grade pleomorphic OS (resembling undifferentiated pleomorphic sarcoma), or small round blue cell OS (resembling classic EWS) [12]. The presence of osteoid deposition is helpful in the histologic diagnosis, but may not be present in very poorly differentiated specimens or small biopsies. In contrast to the characteristic gene fusions found in EWS and SS, there is no pathognomonic molecular aberration that has been

recognized in OS. A wide range of copy number changes, most frequently including chromosomes 6p, 8q, 13q and 17p, have been observed [15–18]. Additionally, the presence of the *EWSR1-FLI1* translocation in a rare subset of small-cell OS further complicates diagnostics [19–22].

Previous work has demonstrated the feasibility to classify small round cell tumors using machine learning techniques and artificial neural networks trained on gene expression data [23], or support vector machine based classifiers to distinguish between sarcoma subtypes with variable sensitivity and specificity (range 50 – 100%) [24]. A new approach to aid in the diagnosis of solid tumors is based on the molecular signatures of genome-wide DNA methylation profiles. This technique has been pioneered as a powerful diagnostic tool in pediatric brain tumors, and has been shown to be superior for risk stratification compared to standard histopathology [25–27]. For the purposes of classification alone, the advantage of methylation is resolution. Older studies relied on cDNA microarrays with far fewer probes compared to methylation arrays exceeding 480,000 probes. There is also an advantage of decreased noise, as methylation is more invariant to formalin fixation, time to fixation, cold ischemia time, temperature out of the body, immune status of host, and several other factors that gene expression. Integrative DNA-methylation analysis has been previously examined as a tool to classify high grade soft-tissue sarcoma including SS, but not OS or EWS [28]. The aim of the current study was to determine if methylation profiling can be used to accurately distinguish between SS, OS, and EWS.

Methods

Tissue Collection and DNA Extraction

Eighty tissue samples from newly diagnosed OS, EWS, and SS patients were obtained from the archives of the Departments of Pathology at the NYU Langone Medical Center (NYULMC), Memorial Sloan Kettering Cancer Center (MSKCC), and Montefiore Medical Center. The study was conducted at NYULMC and approved by the Institutional Review Board in accordance with all local and federal regulations. EWS were screened by histology and the diagnosis confirmed with the presence of an *EWSR1* rearrangement by FISH. SS were diagnosed based on histologic features and, when available, by the presence of t(X;18) on FISH. OS were diagnosed based on histologic features and, where available, absence of fusions characteristics of EWS and SS. To develop the classifier, we selected histopathologically classic (OS, SS, EWS) and molecularly confirmed (SS, EWS) reference samples. The final cohort of samples included 36 total samples, including one secondary malignancy (OS). DNA was extracted from non-decalcified formalin-fixed paraffin-embedded (FFPE) specimens with the Promega Maxwell® 16 FFPE Plus LEV DNA Purification Kit (NYULMC and Montefiore samples), or from tumor lysate (MSKCC samples) using a Promega Maxwell® 16 instrument following manufacturer's instructions. The DNA was bisulfite-converted using the EZ-96 DNA Methylation Kit from Zymo Research. DNA from FFPE samples subsequently underwent restoration using the Illumina Infinium HD FFPE DNA Restore Kit.

Methylation Profiling, Pre-processing of Methylation Data, and Unsupervised Hierarchical Clustering

The HumanMethylation450 array (Illumina, San Diego, USA) was used to determine the DNA methylation status of 482,421 CpG sites, following manufacturer's instructions as previously reported [29]. Standard beta-mixture quantile normalization (BMIQ), background correction, quality control, and rule-based filtering of samples of probes were implemented using the RnBeads pipeline in order to calculate final beta values. [30]. Beta value is defined as the ratio of fluorescence intensity of the methylated probe over the overall intensity and was used in all visualization. Unsupervised hierarchical clustering was done with Euclidean measure for distance matrix and complete agglomeration method for clustering.

Statistical analysis

All statistical tests and modeling was completed in the open-source software R. M values were calculated from beta values and used for statistical tests. For the *i*th probe, the M-value is calculated as a $\log_2$ transform of a ratio of the beta value.

$$\mathrm{M_i}=\log_2(\mathrm{Beta_i}/(1-\mathrm{Beta_i}))$$

A one-way ANOVA via empirical bayes method (R package *limma*) was conducted to select for the most differentially methylated probes amongst the different sarcoma subtypes. An alpha cutoff of 0.01 after Bonferronni correction was applied to the F-test p-values to correct for multiple testing of 482,421 CpG sites. Over 8,556 probes were found to be significantly differentially methylated, however many were very highly correlated (Supplement S2); thus we selected the top 400 probes with the greatest median absolute deviation in beta values (Supplement S1).

Random Forest Classifier

The random forest (RF) algorithm was used for classification. We used the randomForest package in R for the implementation of the classifier (version 4.6–7) [31]. Classifier results were relatively insensitive to parameter choice as long as the number of trees was sufficiently large (>200). The number of trees was set at 400 with a number of variables tried at each node of 305 (tuned to reduce out-of-bag error), and minimum node size of 1.

Validation samples: TCGA SS, TARGET OS, and Huertas-Martinez *et al.* EWS

We validated our classifier using The Cancer Genome Atlas (TCGA), TARGET-OS (Children's Oncology Group and The Hospital for Sick Children in Toronto, Canada), and EWS samples from a recently published series. Raw signal intensity data (IDAT) files from TCGA SS (N = 10) were downloaded from the legacy archives of the Genomic Data Commons [32]. Similarly, IDAT files and normalized gene quantification from mRNA-Seq belonging to the discovery cohort of TARGET-OS (N = 86) were accessed through the TARGET data matrix (https://ocg.cancer.gov/programs/target/data-matrix). IDAT files from Huertas-Martinez *et al.* were obtained from the corresponding author [33].

Genomic Pathway Analysis

Probes in the classifier were annotated using the HumanMethylation450 manifest (v1.2, Illumina). Genomic information including DNA sequence and coordinates of gene coding regions were obtained from the University of California Santa Cruz Genome Browser database [34]. One-hundred thirteen probes corresponding to enhancers and gene coding regions were assessed using the Molecular Signatures Database (MSigDB), a gene-set based pathway analysis. In total, we interrogated the overlap of classifier genes with 4,729 curated gene sets, which include known chemical and genetic perturbations (3500), gene sets derived from the KEGG (186), BIOCARTA (217), REACTOME (674) pathways, and canonical pathways curated by domain experts (1329) [35].

Data access

IDAT files of the training set have been deposited in GEO, accession number GSE97529. R-script for classifier is available on Github <https://github.com/spw08536/Methylation>.

Results

Patient Characteristics

Of the 80 initial patients, 36 representative samples were chosen to construct the training set; EWS (n=10), SS (n=11), and OS (n=15) (Table 1). The median patient age in the training cohort was 23 years (range: 3–80). Gender was well balanced, with 53% of the patients being male. 78% of the samples were obtained from the primary tumor site. Sample location included the appendicular and axial skeleton, in 11/36 (31%) and 23/36 (64%), respectively; biopsy sites of 2/36 (5%) were unknown.

Sarcomas Show Distinct Patterns of Methylation

Unsupervised hierarchical clustering using the 400 most differentially methylated probes amongst the training set demonstrated three distinct molecular phenotypes corresponding to the pathologic diagnoses (Figure 1A). We observed that probes that were methylated in OS were generally unmethylated in SS. The methylation signature of EWS tumors showed a third, distinct pattern of methylation. None of the 400 probes are co-localized with known genetic probes of interest. For EWS, two probes (cg21242508 and cg06516502) are located on chromosome 22q, approximately 1Mb away from *EWSR1* within a tetratricopeptide repeat domain TTC28. The closest probe to *FLI1* is 5Mb away (cg13153466), corresponding to the promoter region of *ASAM*, a Coxsackie-And Adenovirus Receptor-Like Membrane protein. No probes lie near the cytogenetic band 18q11.2 belonging to the *SYT* translocation partner of synovial sarcoma – the closest four probes in chromosome 18 lie in a 0.1Mb strip at 18q23, notably hypomethylated in SS. Pathway analysis identified targets of polycomb group proteins SUZ12 and EED, possessing the H3K27 trimethylated mark, as highly enriched within classifier genes (Table 2).

An Accurate Methylation Based Classification of Bone Sarcomas

When the classifier was applied to the training cohort, all samples were accurately classified with a minimum margin score of 0.2, thus confirming internal validity (Table 3). Margin is

defined as the proportion of votes given to the correct class minus maximum proportion of votes for the other classes — thus, positive margin means correct classification. Within the validation cohorts, all samples (10/10) from TCGA were accurately classified as SS, 14 of 15 EWS samples were accurately classified as EWS, and all tumors from TARGET-OS (85/86) were classified as OS with the exception of one sample being classified as EWS (Table 4). A single case from the EWS validation set that classified as a synovial sarcoma did not have RNA data available for confirmatory studies.

The discrepant OS sample, TARGET-40-PASEFS (highlighted by the arrow in Figure 1A, displayed in Figure 1B), was submitted from an 18 year old Caucasian male who was enrolled and treated on Children's Oncology Group clinical trial AOST0331. The tumor was located in the proximal tibia, a common location for both EWS and OS. To determine whether the tumor might have been misclassified as EWS at the time of original diagnosis, we investigated *EWSR1, FLI1, ERG,* and *ETV1* mRNA transcript counts, available in the TARGET-OS dataset, to examine the common gene products of oncogenic EWS fusions. While the expression of *FLI1* was similar to other samples in the cohort, we noted a striking overexpression of *ETV1* (70 Transcripts / kb * million) and *EWSR1* (200 Transcripts / kb * million) compared to the other OS samples in the dataset (Figure 1C). Amongst all the TARGET-OS samples, this sample has the highest expression of *ETV1*, and third highest expression of *EWSR1*. We hypothesized that the tumor harbored a *EWSR1/ETV1* rearrangement consistent with EWS, however after manually curating the transcript sequence reads, there was no evidence of an ETV1 fusion; suggesting an alternative mechanism leading to ETV1 upregulation.

Clinical Application

To illustrate the clinical utility of our classifier, we present two challenging clinical cases. The first is a case of a 16 year old female with a past medical history of rhabdomyosarcoma of the right orbit initially diagnosed in 2007 and treated with chemotherapy and radiation therapy. She had recurrence of the tumor in 2009 and was treated with orbital exenteration and maxillectomy, chemotherapy, and re-irradiation. In mid-2016, recurrence was noted in the diploic cavity (Figure 2A). Biopsy demonstrated a predominantly spindled neoplasm with fascicular growth, necrosis, frequent mitotic activity, and nuclear pleomorphism (Figure 2B). Immunohistochemical stains were positive for vimentin, CD56 and desmin, but tumor cells were negative for myogenin, Myo-D1, S-100, CAM 5.2, AE 1/3, BCL-2, CD99, CD34 and EMA. INI-1/BAF47 expression was preserved. FISH testing using dual-color break-apart probe showed a rearrangement involving *SS18* (*SYT*) (88.0% of cells, data not shown), raising the concern that the tumor may represent SS. However, SatB2 was positive by immunohistochemistry, suggesting osteoblastic differentiation [36]. SSX1 and SSX2, the usual fusion partners for SYT in synovial sarcoma, had no abnormalities.

We performed methylation profiling on the 2016 recurrence sample. We analyzed the tumor using our Sarcoma Classifier, revealing a match with OS (Figure 2C). This case highlights the diagnostic value of our sarcoma classifier in samples that are difficult to diagnose using standard of care molecular methods. Moreover, it indicates robustness of methylation based sarcoma classifier not only in *de novo*, but also in radiation-induced OS.

The second difficult case was a 2 year old female who was found to have a mass involving T7–T9 of the thoracic spine with radiographic evidence of pulmonary metastases. Given the histologic appearance and diffuse CD99 membranous staining, a diagnosis of EWS was made, although molecular confirmation of an *EWSR1* rearrangement was never established. She received intensive multi-agent chemotherapy and involved field radiation therapy to the spine.

Three years later at the age of 5, she presented with recurrent disease, including a left sided pleural mass and hilar adenopathy. She received salvage chemotherapy including irinotecan and temozolomide, followed by a left thoracotomy and radical wedge resection. Pathology revealed viable disease with similar morphologic findings as her initial disease.

At the age of 15, she presented with headache, pupillary asymmetry, and slurred speech. A brain MR revealed a fronto-parietal mass, which was subsequently resected (Figure 2D). Pathology revealed a high grade pleomorphic and spindle cell sarcoma (Figure 2E). CD99 staining was focally membranous. Targeted DNA (MSK-IMPACT [37]) and RNA (Archer FusionPlex, ArcherDX, Inc.) was non-diagnostic, but revealed a genomically unstable tumor with numerous mutations, which would be unusual for translocation-driven sarcomas such as EWS. Quantitative predictive probabilities derived from the random forest model were used to classify this sample, with a predicted probability of 48% OS, 32% SS, 20% EWS. In the scenario in which OS and ES comprise the differential diagnosis, OS was the most likely and EWS the least likely grouping. (Figure 2F).

Discussion

In sarcomas lacking pathognomonic gene fusions, diagnostic differentiation can be extremely challenging, even with the help of modern diagnostic tools including immunohistochemistry and cytogenetics [38–41]. Additionally, tumor heterogeneity and sampling errors can significantly confound the diagnosis. Since therapeutic decisions depend on sarcoma subtype, novel methods are needed to improve diagnostic accuracy.

The least invasive histologic sampling method is fine needle aspiration biopsy [42], and more invasive techniques to obtain larger amounts of tumor tissue include open or core biopsy [43, 44].

However, even with an adequate specimen, a diagnosis based solely on histopathology and detailed clinical information is often inconclusive. Ancillary cytogenetic studies, such as FISH can assist in differentiating between morphologically similar tumor specimens [45], however, these tests frequently fail to identify a pathognomonic genetic abnormality and can occasionally produce false positive results.

The HumanMethylation450 array is a rapid and cost effective method for genome-wide quantitative profiling of the methylation of CpG loci [29]. To explore the utility of DNA-based methylation profiling in patients with sarcoma, we used a similar approach that has been shown to be highly accurate and reproducible in subclassifying other histologically similar tumors, i.e. pediatric brain tumors [46]. Using random forest modeling, we developed a classifier that we successfully validated using SS and OS samples from two

publically available datasets, and an EWS validation set obtained from a recently published series [33].

In our training set, our classifier correctly identified all 36 samples as the tumor type determined by a combination of clinical, histologic, and molecular genetic factors. When applied to our validation cohorts (TCGA and TARGET-OS), all 10 SS samples and 85/86 (99%) of the OS samples were correctly classified. Of particular interest is the sample that was classified incorrectly from the TARGET OS cohort as EWS. Our analysis shows that the case TARGET-40-PASEFS may represent a tumor that is biologically related to EWS [47], as indicated by overexpression of *EWSR1* and *ETV1*, illustrating the power and usefulness of methylation based profiling for discovery.

Our study highlights the epigenetic heterogeneity present in EWS despite a single recurrent oncogenic fusion driver. Some samples within the EWS cohort appear to be truly misclassified, and qualitatively the methylation pattern of the training and validation sets of EWS appear to be the most heterogeneous even after hierarchical clustering. As demonstrated recently, a larger number of EWS samples are needed to fully characterize the epigenetic heterogeneity [48].

A key advantage of methylation based analysis is the interrogation of multiple diagnoses using a single test, thereby expediting the diagnosis. Methylation also provides novel insights into sarcomabiology, for example OS appears to be characterized by a consistently hypomethylated strip along chromosome 1q43, among other features. Furthermore, with development and integration of additional sarcoma classifiers, such as rhabdomyosarcoma [49], molecular diagnostics of all major sarcoma subgroups could be cost efficiently performed using a single, methylation based platform. A prospective study comparing the performance and cost-effectiveness of our methylation classifier and traditional immunohistochemical and FISH testing could represent a fruitful avenue of future research.

Conclusion

In summary, we developed and validated a DNA methylation-based classifier that accurately differentiated three of the most common subtypes of bone sarcomas. Given their clinically and histologically overlapping features and markedly different clinical management, this novel methylation based classifier may provide a useful tool in the differential diagnosis of bone sarcomas.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The study was supported in part by KiDS of NYU Langone and The Friedberg Charitable Foundation. The authors would like to thank Dr. Oscar M. Tirado for providing Ewing sarcoma data from their study.

References

1. Yaw KM. Pediatric bone tumors. Semin Surg Oncol. 1999; 16(2):173–83. [PubMed: 9988871]
2. Werier J, et al. A systematic review of optimal treatment strategies for localized Ewing's sarcoma of bone after neo-adjuvant chemotherapy. Surg Oncol. 2016; 25(1):16–23. [PubMed: 26979636]
3. Isakoff MS, et al. Osteosarcoma: Current Treatment and a Collaborative Pathway to Success. J Clin Oncol. 2015; 33(27):3029–35. [PubMed: 26304877]
4. Ewing J. Diffuse endothelioma of bone. Proc NY Pathol Soc. 1921; (21):17–24.
5. Nascimento AG, et al. A clinicopathologic study of 20 cases of large-cell (atypical) Ewing's sarcoma of bone. Am J Surg Pathol. 1980; 4(1):29–36. [PubMed: 7361993]
6. Delattre O, et al. Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. Nature. 1992; 359(6391):162–5. [PubMed: 1522903]
7. Sankar S, Lessnick SL. Promiscuous partnerships in Ewing's sarcoma. Cancer Genet. 2011; 204(7): 351–65. [PubMed: 21872822]
8. Sorensen PH, et al. A second Ewing's sarcoma translocation, t(21;22), fuses the EWS gene to another ETS-family transcription factor ERG. Nat Genet. 1994; 6(2):146–51. [PubMed: 8162068]
9. Jeon IS, et al. A variant Ewing's sarcoma translocation (7;22) fuses the EWS gene to the ETS gene ETV1. Oncogene. 1995; 10(6):1229–34. [PubMed: 7700648]
10. Kaneko Y, et al. Fusion of an ETS-family gene, EIAF, to EWS by t(17;22)(q12;q12) chromosome translocation in an undifferentiated sarcoma of infancy. Genes Chromosomes Cancer. 1996; 15(2): 115–21. [PubMed: 8834175]
11. Peter M, et al. A new member of the ETS family fused to EWS in Ewing tumors. Oncogene. 1997; 14(10):1159–64. [PubMed: 9121764]
12. Goldblum JR, et al. Enzinger and Weiss's soft tissue tumors. :1. online resource.
13. Clark J, et al. Identification of novel genes, SYT and SSX, involved in the t(X;18)(p11.2;q11.2) translocation found in human synovial sarcoma. Nat Genet. 1994; 7(4):502–8. [PubMed: 7951320]
14. Vlenterie M, et al. Next generation sequencing in synovial sarcoma reveals novel gene mutations. Oncotarget. 2015; 6(33):34680–90. [PubMed: 26415226]
15. Smida J, et al. Genomic alterations and allelic imbalances are strong prognostic predictors in osteosarcoma. Clin Cancer Res. 2010; 16(16):4256–67. [PubMed: 20610556]
16. Yen CC, et al. Identification of chromosomal aberrations associated with disease progression and a novel 3q13.31 deletion involving LSAMP gene in osteosarcoma. Int J Oncol. 2009; 35(4):775–88. [PubMed: 19724913]
17. Kresse SH, et al. LSAMP, a novel candidate tumor suppressor gene in human osteosarcomas, identified by array comparative genomic hybridization. Genes Chromosomes Cancer. 2009; 48(8): 679–93. [PubMed: 19441093]
18. Man TK, et al. Genome-wide array comparative genomic hybridization analysis reveals distinct amplifications in osteosarcoma. BMC Cancer. 2004; 4:45. [PubMed: 15298715]
19. Dragoescu E, et al. Small cell osteosarcoma with Ewing sarcoma breakpoint region 1 gene rearrangement detected by interphase fluorescence in situ hybridization. Ann Diagn Pathol. 2013; 17(4):377–82. [PubMed: 22971270]
20. Hill DA, et al. Practical application of molecular genetic testing as an aid to the surgical pathologic diagnosis of sarcomas: a prospective study. Am J Surg Pathol. 2002; 26(8):965–77. [PubMed: 12170083]
21. Noguera RS, Navarro TJ. Triche, Translocation (11;22) in small cell osteosarcoma. Cancer Genet Cytogenet. 1990; 45(1):121–4. [PubMed: 2302678]
22. Oshima Y, et al. Abdominal small round cell tumor with osteoid and EWS/FLI1. Hum Pathol. 2004; 35(6):773–5. [PubMed: 15188147]
23. Khan J, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat Med. 2001; 7(6):673–9. [PubMed: 11385503]
24. Segal NH, et al. Classification and subtype prediction of adult soft tissue sarcoma by functional genomics. Am J Pathol. 2003; 163(2):691–700. [PubMed: 12875988]

25. Sturm D, et al. New Brain Tumor Entities Emerge from Molecular Classification of CNS-PNETs. Cell. 2016; 164(5):1060–72. [PubMed: 26919435]
26. Hovestadt V, et al. Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. Acta Neuropathol. 2013; 125(6):913–6. [PubMed: 23670100]
27. Pajtler KW, et al. Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups. Cancer Cell. 2015; 27(5):728–43. [PubMed: 25965575]
28. Renner M, et al. Integrative DNA methylation and gene expression analysis in high-grade soft tissue sarcomas. Genome Biol. 2013; 14(12):r137. [PubMed: 24345474]
29. Rohrich M, et al. Methylation-based classification of benign and malignant peripheral nerve sheath tumors. Acta Neuropathol. 2016; 131(6):877–87. [PubMed: 26857854]
30. Assenov Y, et al. Comprehensive analysis of DNA methylation data with RnBeads. 2014; 11(11): 1138–40.
31. Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002; 2(3):18–22.
32. Weinstein JN, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013; 45(10):1113–20. [PubMed: 24071849]
33. Huertas-Martinez J, et al. DNA methylation profiling identifies PTRF/Cavin-1 as a novel tumor suppressor in Ewing sarcoma when co-expressed with caveolin-1. Cancer Lett. 2017; 386:196–207. [PubMed: 27894957]
34. Raney BJ, et al. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. Bioinformatics. 2014; 30(7):1003–5. [PubMed: 24227676]
35. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102(43):15545–50. [PubMed: 16199517]
36. Conner JR, Hornick JL. SATB2 is a novel marker of osteoblastic differentiation in bone and soft tissue tumours. Histopathology. 2013; 63(1):36–49. [PubMed: 23701429]
37. Cheng DT, et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. J Mol Diagn. 2015; 17(3):251–64. [PubMed: 25801821]
38. Wang CK, et al. Characterization of bone and soft-tissue tumors with in vivo 1H MR spectroscopy: initial results. Radiology. 2004; 232(2):599–605. [PubMed: 15286325]
39. Schulte M, et al. Grading of tumors and tumorlike lesions of bone: evaluation by FDG PET. J Nucl Med. 2000; 41(10):1695–701. [PubMed: 11038000]
40. Costelloe CM, et al. Bone Windows for Distinguishing Malignant from Benign Primary Bone Tumors on FDG PET/CT. J Cancer. 2013; 4(7):524–30. [PubMed: 23983816]
41. Klaeser B, et al. PET/CT-guided biopsies of metabolically active bone lesions: applications and clinical impact. Eur J Nucl Med Mol Imaging. 2010; 37(11):2027–36. [PubMed: 20680270]
42. Kilpatrick SE, et al. The role of fine-needle aspiration biopsy in the initial diagnosis of pediatric bone and soft tissue tumors: an institutional experience. Mod Pathol. 1998; 11(10):923–8. [PubMed: 9796716]
43. Saghieh S, et al. The risk of local recurrence along the core-needle biopsy tract in patients with bone sarcomas. Iowa Orthop J. 2010; 30:80–3. [PubMed: 21045976]
44. Mankin HJ, Lange TA, Spanier SS. The hazards of biopsy in patients with malignant primary bone and soft-tissue tumors. J Bone Joint Surg Am. 1982; 64(8):1121–7. [PubMed: 7130225]
45. Bridge RS, et al. Molecular diagnosis of Ewing sarcoma/primitive neuroectodermal tumor in routinely processed tissue: a comparison of two FISH strategies and RT-PCR in malignant round cell tumors. Mod Pathol. 2006; 19(1):1–8. [PubMed: 16258512]
46. Northcott PA, et al. Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. Acta Neuropathol. 2012; 123(4):615–26. [PubMed: 22057785]
47. Smith SC, et al. CIC-DUX sarcomas demonstrate frequent MYC amplification and ETS-family transcription factor expression. Mod Pathol. 2015; 28(1):57–68. [PubMed: 24947144]

48. Sheffield NC, et al. DNA methylation heterogeneity defines a disease spectrum in Ewing sarcoma. 2017; 23(3):386–395.

49. Seki M, et al. Integrated genetic and epigenetic analysis defines novel molecular subgroups in rhabdomyosarcoma. 2015; 6:7557.

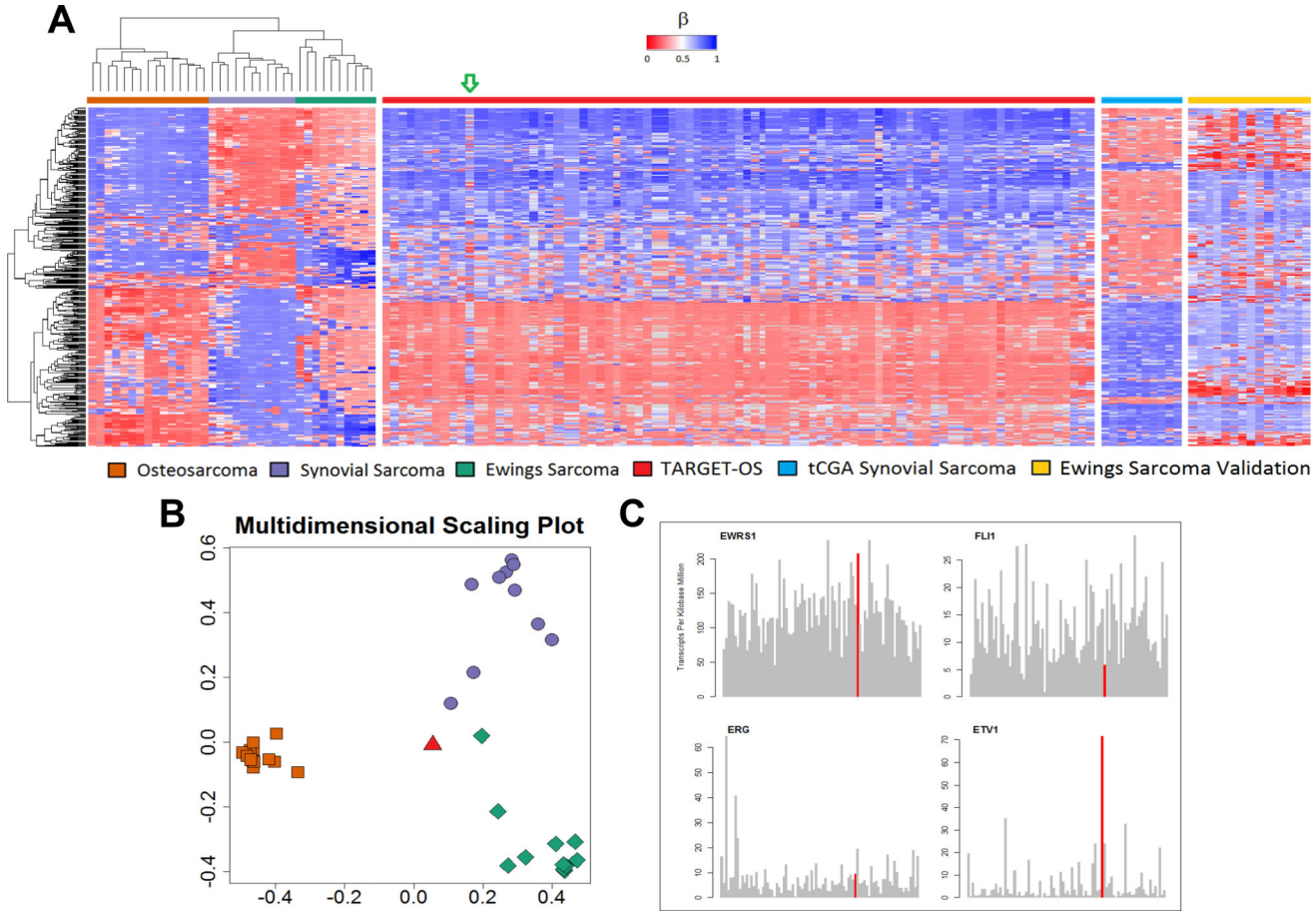

Figure 1.

(A) Unsupervised hierarchical clustering of validation and training sets shows three distinct clusters for osteosarcoma, synovial sarcoma, and Ewing's sarcoma. The green arrow indicates the sample from the TARGET-OS dataset (TARGET-40-PASEFS) that demonstrated hypermethylation in several CpG islands uncharacteristic of other osteosarcomas. **(B)** Multidimensional scaling plot of random forest classifier samples demonstrates this TARGET-OS sample (red triangle) was classified as Ewing's sarcoma. **(C)** RNAseq analysis of the sample TARGET-OS sample identified as EWS by methylation classifier. The TARGET-40-PASEFS sample shows overexpression of EWSR1 and ETV1. Other common fusion partners of EWSR1 including FLI1 and ERG, do not show increased transcription. Transcripts of interest are highlighted in red. RNA expression data are consistent with methylation sarcoma classifier diagnosis of EWS.

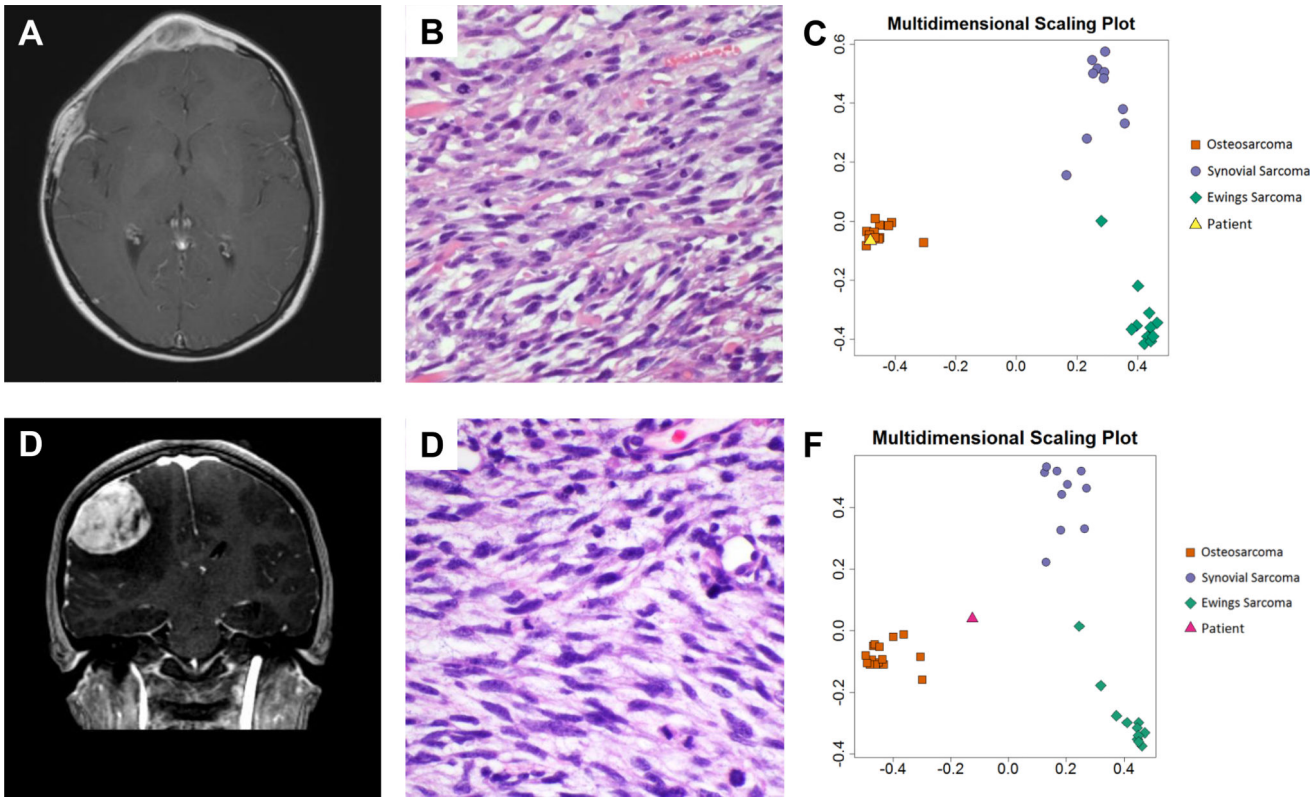

Figure 2. Case examples

(A) T1-weighted post-contrast MRI demonstrating sharply circumscribed homogeneously enhancing soft tissue mass. **(B)** The tumor consists of alternating epithelioid and plump spindled cells with interspersed coarse collagen fibers. Based on positive FISH analysis, the tumor was initially diagnosed as a synovial sarcoma, however, **(C)** multidimensional scaling plot of the classifier with the patient's tumor in yellow shows that the tumor classifies with osteosarcoma. **(D)** T1-weighted post-contrast MRI of the second clinical case demonstrating sharply circumscribed homogeneously enhancing soft tissue mass. **(E)** Histology shows sheets of poorly differentiated spindled tumor cells notably lacking distinctive morphological features of synovial sarcoma, Ewing Sarcoma or osteosarcoma. **(F)** The primary tumor was originally diagnosed as Ewing sarcoma; however *EWSR1* rearrnagement was never identified and the multidimensional scaling plot of the metastasis shows the patient sample in pink grouping most closely with osteosarcoma.

Table 1

Patient characteristics for the training set (n = 36).

Sample	Histology	Age	Sex	Primary Tumor location	Translocation or Fusion
1	EWS	14	M	Left Talus	N/A
2	EWS	3	F	Lateral side of the right knee	EWSR1-FLI1
3	EWS	10	F	Tibia	EWSR1-FLI1
4	EWS	18	M	Rib	EWSR1-FLI1
5	EWS	16	M	Back	EWSR1-FLI1
6	EWS	11	F	Mediastinal Lymph Node	EWSR1-FLI1
7	EWS	45	M	Rib	EWSR1-FLI1
8	EWS	46	F	Femur	EWSR1-FLI1
9	EWS	41	M	Epidural	EWSR1-FLI1
10	EWS	33	M	Tibia	EWSR1-FLI1
11	OS	14	F	Orbit (secondary OS)	N/A
12	OS	40	M	Lymph Node	N/A
13	OS	19	M	Femur	N/A
14	OS	38	F	Mandible	N/A
15	OS	15	F	Femur	N/A
16	OS	27	M	Femur	N/A
17	OS	14	F	Right Tibia	N/A
18	OS	80	F	Right Knee	N/A
19	OS	12	M	Right Distal Femur	N/A
20	OS	8	M	Left Tibia	N/A
21	OS	6	F	Right Femur	N/A
22	OS	10	F	Right Shoulder	N/A
23	OS	14	F	Left Pelvis	N/A
24	OS	38	M	Left Humerus	N/A
25	OS	65	M	Left Shoulder Wall	N/A
26	SS	39	M	Neck	SYT rearrangement
27	SS	17	M	Left knee	N/A
28	SS	13	F	Right Neck	X:18

Sample	Histology	Age	Sex	Primary Tumor location	Translocation or Fusion
29	SS	15	F	Left foot	X:18
30	SS	53	F	Thigh	SYT rearrangement
31	SS	35	M	Chest wall	N/A
32	SS	44	M	Foot	N/A
33	SS	48	F	Shoulder	X:18
34	SS	33	M	Abdomen	N/A
35	SS	48	F	Thigh	N/A
36	SS	43	M	Foot	SYT rearrangement

Table 2

Pathway analysis of 113 gene coding regions corresponding to probes within the classifier are highly enriched in genes bound by Polycomb proteins EED and SUZ12, as well as genes possessing the trimethylated H3K27 mark in their promoter.

Gene Set (# Genes)	Description	# Genes in Overlap [k]	p-value	FDR q-value
BENPORATH ES WITH H3K27ME3 (1118)	Genes identified by ChIP on chip.as possessing the trimethylated H3K27 mark in their promoters in human embryonic stem cells	**27**	$2.25\ e^{-19}$	$1.15\ e^{-15}$
BENPORATH EED TARGETS (1062)	Genes identified by ChIP on chip as targets of the Polycomb protein EED in human embryonic stem cells.	**24**	$1.29\ e^{-16}$	$3.29\ e^{-13}$
BENPORATH PRC2 TARGETS (652)	Polycomb Repression Complex 2 (PRC2) targets; identified by ChIP on chip on human embryonic stem cells as genes that: possess the trimethylated H3K27 mark in their promoters and are bound by SUZ12 and EED Polycomb proteins.	**17**	$5.57\ e^{-13}$	$7.11\ e^{-10}$
BENPORATH SUZ12 TARGETS (1038)	Genes identified by ChIP on chip as targets of the Polycomb protein SUZ12 in human embryonic stem cells.	**20**	$1.09\ e^{-12}$	$1.11\ e^{-9}$

Table 3

Contingency table demonstrating the random forest model accurately classifies all patients within the training set.

Training Data (N=36)		Random Forest Classification		
		Osteosarcoma	Synovial Sarcoma	Ewing's Sarcoma
Pathology Diagnosis	Osteosarcoma	15	0	0
	Synovial Sarcoma	0	11	0
	Ewing's Sarcoma	0	0	10

Table 4

Contingency table demonstrating that within the validation set, all patients were correctly classified within the TCGA (SS), Huertas-Martinez *et al.* (EWS), and TARGET-OS cohorts, with the exception of one sample from TARGET-OS (TARGET-40-PASEFS) that is classified as a Ewing's sarcoma and one Ewing's sarcoma sample from the Huertas-Martinez *et al.* cohort that classified as a synovial sarcoma.

Validation Data (N=110)		Random Forest Classification		
		Osteosarcoma	Synovial Sarcoma	Ewing's Sarcoma
Pathology Diagnosis	Osteosarcoma	85	0	1
	Synovial Sarcoma	0	10	0
	Ewing's Sarcoma	0	1	14