Sarcoma_CellminerCDB: A tool to interrogate the genomic and functional characteristics of a comprehensive collection of sarcoma cell lines

Camille Tlemsani, Christine M. Heske, Fathi Elloumi, Lorinc Pongor, Prashant Khandagale, Sudhir Varma, Paul S. Meltzer, Javed Khan, William C. Reinhold, Yves Pommier

PII: S2589-0042(24)01003-4

DOI: https://doi.org/10.1016/j.isci.2024.109781

Reference: ISCI 109781

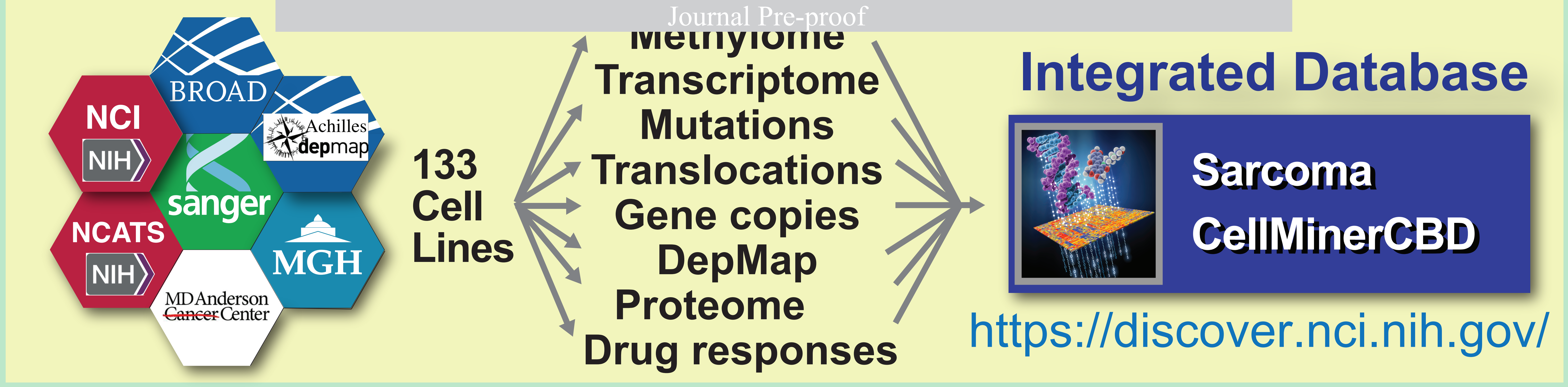
To appear in: *ISCIENCE*

Received Date: 26 October 2023

Revised Date: 28 December 2023

Accepted Date: 15 April 2024

Published by Elsevier Inc.



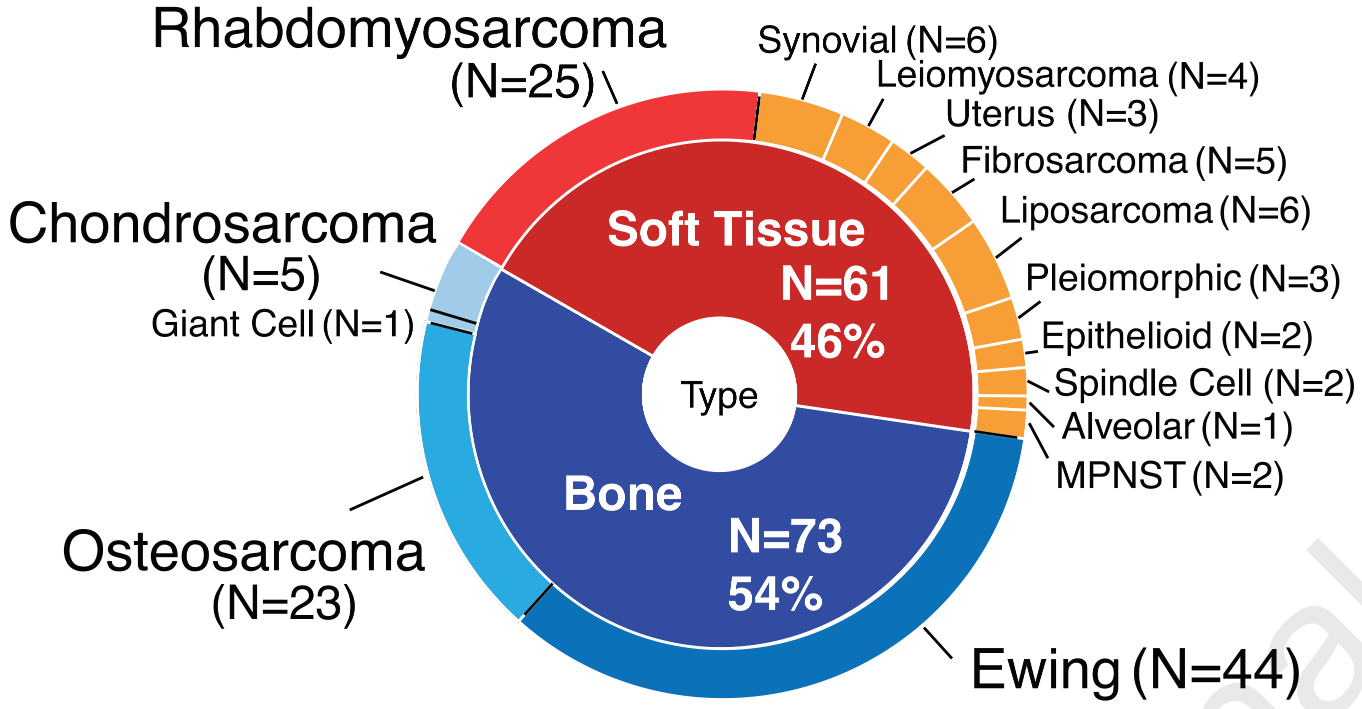
133 Cell Lines

Methylome
Transcriptome
Mutations
Translocations
Gene copies
DepMap
Proteome
Drug responses

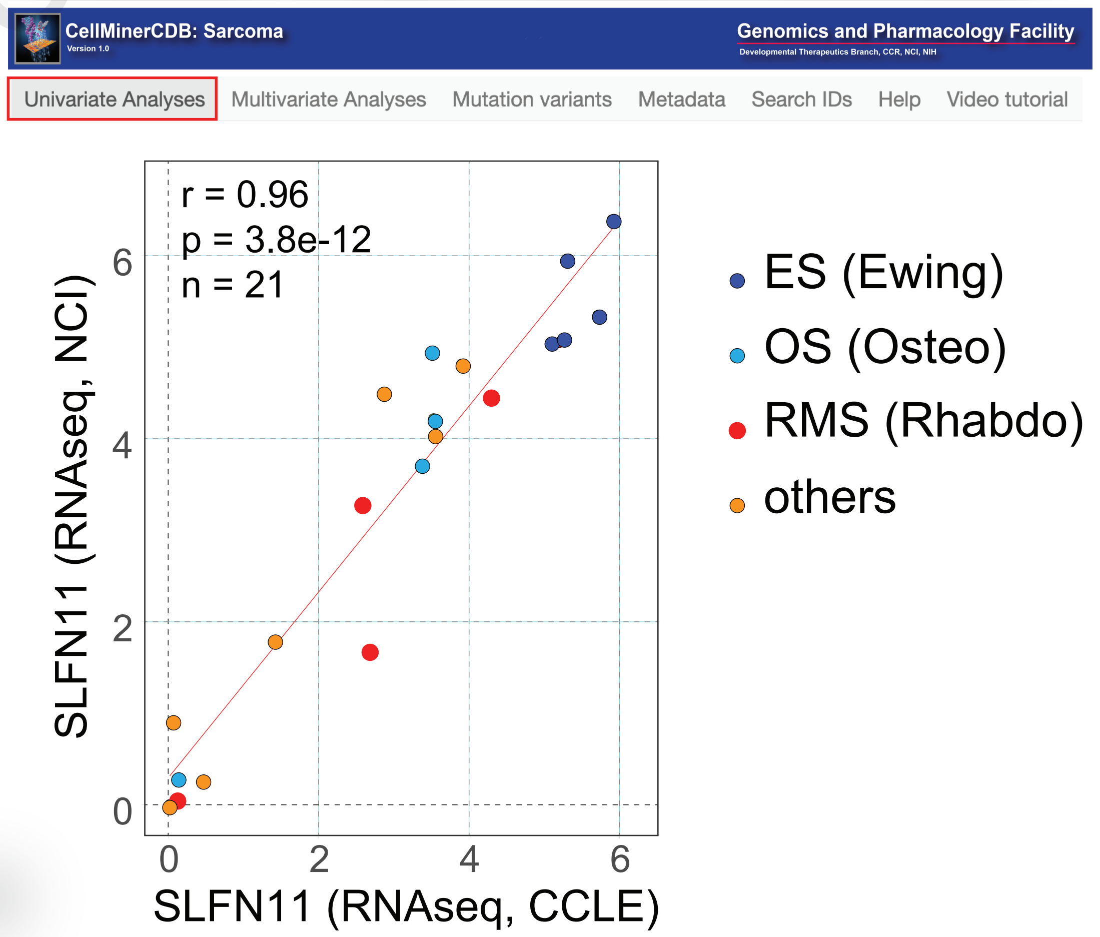
Integrated Database

Sarcoma CellMinerCBD

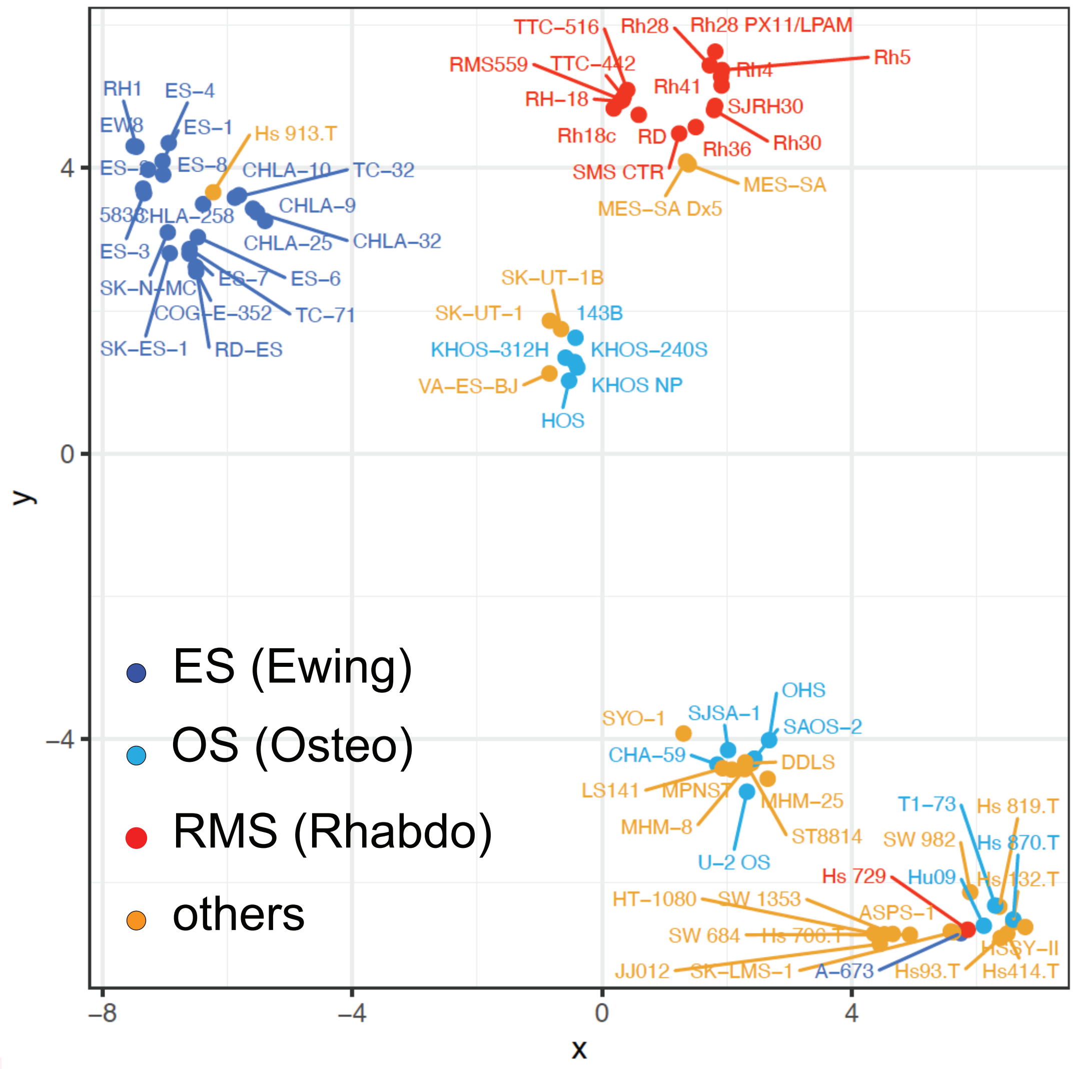
https://discover.nci.nih.gov/



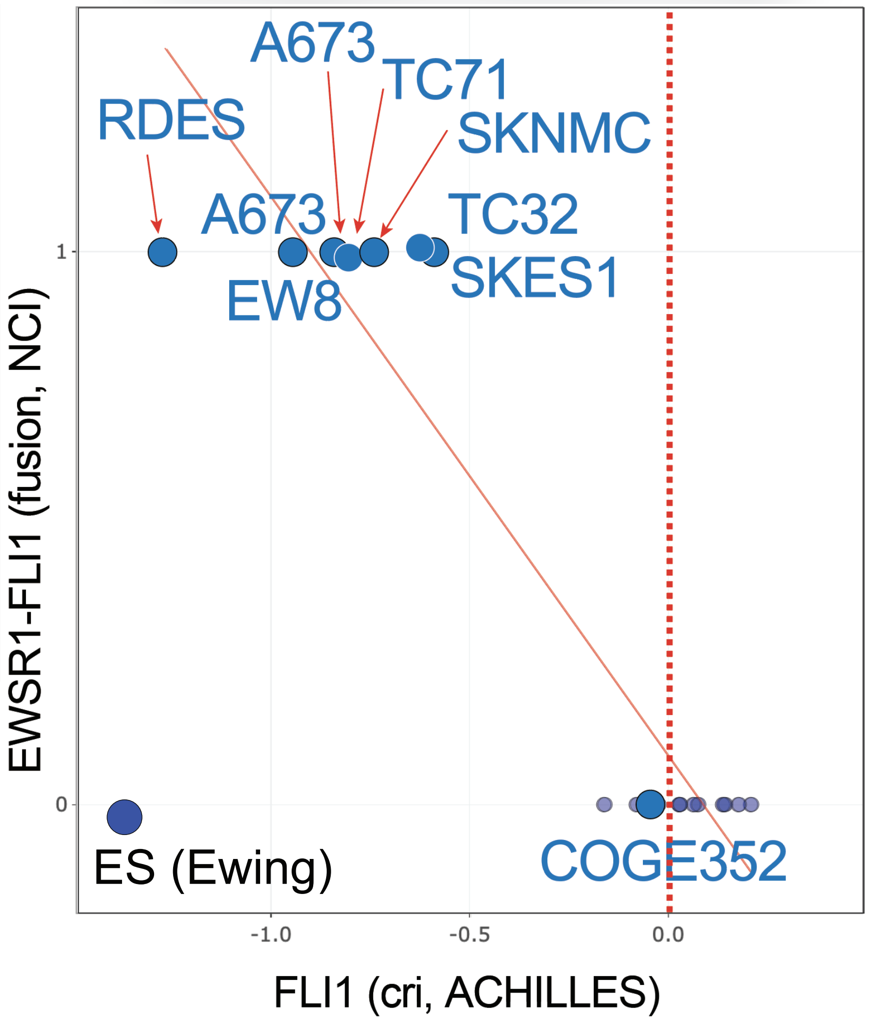
Data reproducibility



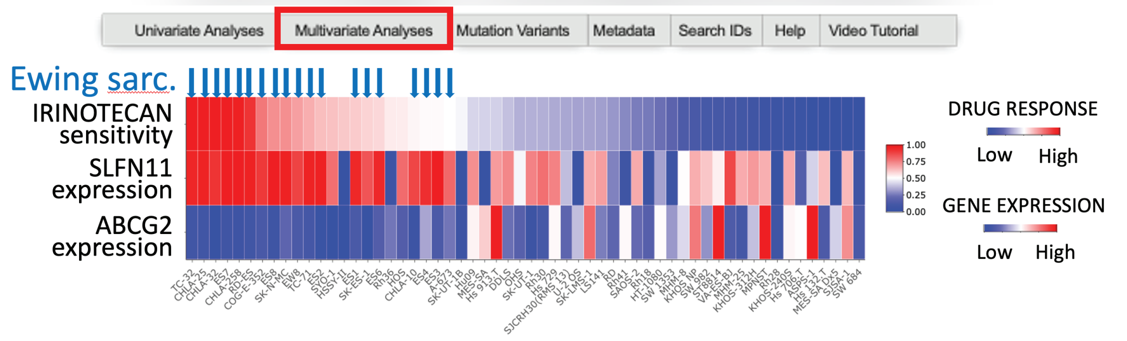
Methylome subtypes



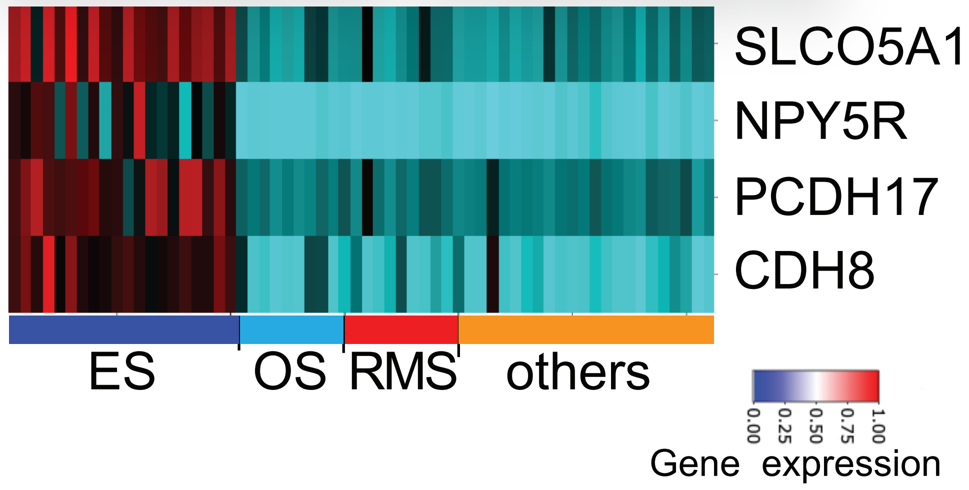
Fusions & Gene Dependency



Drug response



Surface markers



Sarcoma_CellminerCDB: A tool to interrogate the genomic and functional characteristics of a comprehensive collection of sarcoma cell lines

Camille Tlemsani[1,2,3^+], Christine M. Heske[4^+], Fathi Elloumi[1], Lorinc Pongor[1,5], Prashant Khandagale[1], Sudhir Varma[1], Paul S. Meltzer[6], Javed Khan[6], William C. Reinhold[1], Yves Pommier[1+*]

^ These authors contributed equally.
+ Co-corresponding authors
* Lead contact

Affiliations:

[1] Developmental Therapeutics Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland, 20892, USA.

[2] Department of Medical Oncology, Cochin Hospital, Paris Cancer Institute CARPEM, Université Paris Cité, APHP. Centre, Paris, France.

[3] Institut Cochin, INSERM U1016, CNRS UMR8104, Paris Cancer Institute CARPEM, Université Paris Cité, Paris, France.

[4] Pediatric Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, 20892, USA

[5] Hungarian Centre of Excellence for Molecular Medicine, Cancer Genomics and Epigenetics Core Group, Szeged, Hungary

[6] Genetics Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland, 20892, USA.

Declaration of Interests: The authors declare no competing interests.

Acknowledgments: This work was supported by Funding from the Intramural Research Program of the Center for Cancer Research, NCI, NIH. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Correspondence:

Camille Tlemsani, MD PhD
Department of Oncology
Cochin Hospital
123 Boulevard Port Royal
75014 Paris
Telephone: +33 1 58411926
Email: Camille.tlemsani@aphp.fr

Christine M. Heske, MD
Pediatric Oncology Branch
Center for Cancer Research
National Cancer Institute, NIH
10 Center Drive, 10-CRC, Room 1-3816
Bethesda, MD 20892
Telephone: 240-760-6197
Email: christine.heske@nih.gov

Yves Pommier, MD PhD
Developmental Therapeutics Branch
Center for Cancer Research
National Cancer Institute, NIH
Building 37, Room 5068
Bethesda, MD 20892
Telephone: 240-760-6142
Email: pommier@nih.gov

Summary

Sarcomas are a diverse group of rare malignancies comprised of multiple different clinical and molecular subtypes. Due to their rarity and heterogeneity, basic, translational, and clinical research in sarcoma has trailed behind that of other cancers. Outcomes for patients remain generally poor due to an incomplete understanding of disease biology and a lack of novel therapies. To address some of the limitations impeding preclinical sarcoma research, we have developed Sarcoma_CellMinerCDB, a publicly-available interactive tool that merges publicly-available sarcoma cell line data and newly generated omics data to create a comprehensive database of genomic, transcriptomic, methylomic, proteomic, metabolic, and pharmacologic data on 133 annotated sarcoma cell lines. The reproducibility, functionality, biological relevance, and therapeutic applications of Sarcoma_CellMinerCDB described herein are powerful tools to address and generate biological questions and test hypotheses for translational research. Sarcoma_CellMinerCDB (https://discover.nci.nih.gov/SarcomaCellMinerCDB) aims to contribute to advancing the preclinical study of sarcoma.

Introduction

Sarcomas represent a heterogeneous group of rare cancers that affect children, adolescents, and adults. This broad group of malignancies includes numerous distinct clinical and molecular subtypes. There are two main categories of sarcomas, those that arise from bone, such as osteosarcoma and Ewing sarcoma, and those that arise from soft tissue, such as rhabdomyosarcoma, liposarcoma, leiomyosarcoma, and others. Moreover, even within these general categories, sarcomas demonstrate strikingly heterogenous biology, which affects the pathological diagnosis, clinical presentation, progression, and outcome of these cancers.[1]

Biologically, sarcomas exhibit extreme molecular diversity. Some subtypes have recurrent oncogenic alterations, including oncogenic fusion proteins or recurrent amplifications and point mutations, whereas other subtypes demonstrate more complex genomic profiles.[2,3] Due to the relative rarity of sarcoma compared to other cancers, the collection of data, including genomic clinical data, and the compilation of large datasets lags behind other malignancies.[4] In addition, the field suffers from a dearth of publicly shared and well-annotated preclinical models, particularly for the rarer subtypes, which has hindered progress towards new therapeutic advances.[5] As a result, very few sarcomas are treated with a targeted approach, and therapy for most subtypes still consists of multiagent systemic chemotherapy and local control, a highly toxic approach that fails to cure all patients.[6-8] Consequently, outcomes for patients with most sarcoma subtypes have not improved in several decades due to an incomplete understanding of disease biology and a lack of novel therapies.[1,9-11]

To remedy some of the limitations hindering the preclinical study of sarcoma, we have developed Sarcoma_CellMinerCDB, a publicly-available interactive tool for the research community (https://discover.nci.nih.gov/rsconnect/SarcomaCellMinerCDB/). Sarcoma_CellMinerCDB merges both publicly-available sarcoma cell line data with additional novel omics data to create a comprehensive database containing genomic, transcriptomic, methylomic, proteomic, metabolic, and pharmacologic data on 133 annotated sarcoma cell lines.

Importantly, the data are displayed using an interface that allows for easy visualization, analysis of the information, and development of hypotheses by users. Herein we introduce these unique resources and data assembled in Sarcoma_CellMinerCDB, and describe the reproducibility, functionality, biological relevance, and therapeutic applications of this database.

Results

Summary of resources and data included in Sarcoma_CellMinerCDB

Sarcoma_CellMinerCDB assembles data from 133 sarcoma cell lines, representing 15 distinct clinical entities (4 bone and 11 soft tissue) as shown in Figure 1A. Osteosarcoma (n=23 cell lines), Ewing sarcoma (n=42 cell lines), and rhabdomyosarcoma (n=26 cell lines), which are the most common clinical entities, represent the largest groups of cell lines. However, rarer sarcomas, including alveolar soft part sarcoma and fibrosarcoma are also represented. Genomic and drug response data are compiled from six publicly-available sources (CTRP, CCLE, GDSC, MD Anderson, Achilles, NCATS)[12,13] with the addition of new previously unpublished data from NCI. Data types include mutation (exome), gene expression (Affymetrix and RNAseq), gene copy number, methylation, microRNA (Nanostring), pharmacologic, proteomic, metabolic and CRISPR Cas9 knockout screen results. As previously described for SCLC-CellMiner,[12,14] we also developed Global Z-score expression sets by regrouping all datasets for expression (Affymetrix and RNAseq) using Z-score normalization. This enables analyses of transcriptomic data across all datasets. In addition, we used Affymetrix SNP6.0 Array data to generate GDSC copy number. Figure 1B shows a summary of data sources and the number of cell lines per data source. Detailed information regarding the data source and name of each individual cell line, as well as available clinical data, is shown in Figure 1C and Supplemental Table S1. For many cell lines, similar analyses were derived from multiple sources (data overlap). The extent and type of data overlap are described in Figures 1C-E and Supplemental Figures S1A-B.

Reproducibility and functionality of Sarcoma_CellMinerCDB

Sarcoma_CellMinerCDB has multiple capabilities that are summarized in Table 1. To validate the data across individual datasets, we first compared their general reproducibility by performing Pearson and Spearman correlation analyses on methylation, expression, and copy number data for overlapping cell lines from the different genomic databases. We found that datasets are highly reproducible, with median Pearson correlations between 0.65 and 0.86 for methylation data, between 0.68 and 0.92 for expression data, and between 0.55 and 0.83 for copy number data across the data sources (Figure 2A and Supplemental Figure S2A). In addition, comparison of our Global Z-score for expression to the expression data from the other data sources revealed strong correlations (Pearson correlations between 0.92 and 0.97; Spearman correlations between 0.89 and 0.95) (Supplemental Figures S2B-C). Taken together, these analyses demonstrate that the expression, methylation, and copy number data are reproducible across the data sources, and that the cell lines grown and analyzed independently have overall conserved genotypes.

To illustrate the high reproducibility across datasets, we used the Sarcoma_CellMinerCDB Univariate Analysis tool and visualized the expression of an exemplary gene (*SLFN11; Schlafen 11*),[15] across the broad sarcoma cell line panel (Figure 2B). In this example, *SLFN11* expression from the NCI database is highly correlated with *SLFN11* expression from the CCLE database (r=0.96) and with the Global Z-score for *SLFN11* expression (r=0.99). Furthermore, when assessed by individual sarcoma subtype, *SLFN11* expression level was found to be highest in Ewing sarcoma and highly variable across cell lines in the other subtypes, across each of the datasets. This both demonstrates that there is consistency across the database and confirms what has been previously described regarding expression of *SLFN11* in Ewing sarcoma.[16,17]

In addition to exploring and validating gene expression across databases, the Univariate Analysis tool can be used to interrogate correlations between data on two genes, either in the full sarcoma cell line panel or within individual sarcoma subtypes. For example, when expression of the ETS family transcription factor *FLI1* is plotted against the expression of *SLFN11* or *CD99* (encoding a cell surface and T-cell adhesion glycoprotein) in the full sarcoma cell line panel, the scatter plots reveal highly significant correlations between *FLI1* and both *SLFN11* and *CD99* expression in the majority of Ewing sarcoma cells, which is consistent with known disease biology (Figure 2C, left and middle panels)[16,18]. Importantly, even in subtypes with fewer available cell lines, the Univariate Analysis tool can identify distinct biological features characteristic of those subtypes. For example, the expression of *MDM2* (a key p53 ubiquitin ligase that downregulates p53) is highest in liposarcoma cell lines LS141 and DDLS and correlates with high *MDM2* copy number (Figure 2C, right panel). This too is expected based on known disease biology.[19]

The Multivariate Analysis tool can be used to examine relationships between multiple parameters across the cell line panel. In the example shown, expression of two genes *SLFN11* and *ABCG2* (which encodes the drug efflux transporter BCRP) are plotted against sensitivity to irinotecan across the broad sarcoma cell line panel (Figure 1D). This analysis demonstrates a strong relationship between high *SLFN11* expression, low *ABCG2* expression, and high sensitivity to irinotecan, particularly among Ewing sarcoma cell lines. High SLFN11 expression has been described as a biomarker for topoisomerase I (TOP1) inhibitor sensitivity in cancers including Ewing sarcoma.[16,17] While *ABCG2* expression has been described as a biomarker for TOP1 inhibitor sensitivity in some tumor types,[20] it has yet to be established in Ewing sarcoma. Taken together, these examples highlight the reproducibility and functionality of Sarcoma_CellMinerCDB to confirm known biological features of sarcoma subtypes in the cell line models and potentially making new discoveries that can be translated in the clinic.

Biological relevance of sarcoma cell lines as preclinical models based on oncogenic fusions

Oncogenic fusions are a hallmark of several subtypes of sarcomas, including Ewing sarcoma,[21] fusion-positive rhabdomyosarcoma,[22] synovial sarcoma,[23] and alveolar soft part sarcoma.[24] Identified fusions in the Sarcoma_CellMinerCDB cell lines are listed in Supplemental Table S2. Fusion status can be accessed under the “NCI” dropdown in the x- or y-Axis Cell Line Set and the “gene fusions” dropdown in the x- or y-Axis Data Type. The desired fusion can be typed into the identifier field. Full fusion data from NCI cell lines can be downloaded from the Metadata tab by again selecting the “NCI” Cell Line Set and “gene fusions” Data Type.

In fusion-driven sarcoma subtypes, the fusions act as the main drivers of disease biology and impact numerous downstream processes. Examination of the relationship of fusion status to other available omics data can be used to mine the biology in cell lines in the context of the disease of origin, discover new correlations for future study, and define “outlier” cell lines that may be useful for interrogating particular experimental questions. For example, Univariate Analysis based on EWS-FLI1 fusion reveals the presence of the fusion in all Ewing sarcoma cell lines, with two exceptions COG-E-352 and CHLA-25 (Figure 3A), which are known to harbor the EWS-ERG fusion variant type.[25] In addition, Sarcoma_CellMinerCDB shows that none of the non-Ewing sarcoma cell lines express the EWS-FLI1 fusion. When the EWS-FLI1 fusion status is compared to expression of other known markers of Ewing sarcoma, notable data emerge. *FLI1* expression is tightly correlated with EWS-FLI1 fusion status, with the lowest *FLI1* expression seen in the EWS-ERG fusion cell lines. In contrast, *CD99*, a pathologic marker which is positive in nearly all Ewing sarcoma tumors,[18] is highly expressed in all but two Ewing sarcoma cell lines (TC32 and RD-ES) but is independent of the fusion type. High expression of *SLFN11* is universal in Ewing sarcoma lines and is also independent of fusion type. *NROB1* (encoding the Nuclear Receptor Subfamily O Group B Member 1 that acts as a dominant-negative transcription regulator), which has been described as a downstream target of EWS-FLI1,[26] is highly expressed only in EWS-FLI1 fusion cell lines, but not universally (RD-ES does not express it highly). As expected, *CD99, SLFN11*, and *NROB1* are all expressed more highly in Ewing cell lines than in any other sarcoma cell lines, supporting the biological relevance of sarcoma cell lines as models.

Analyzing fusion-positive rhabdomyosarcoma cell lines demonstrates the value of studying fusion in concert with other omics data (Figure 3B). The presence of a PAX3-FOXO1 fusion is identified in 2/5 rhabdomyosarcoma cell lines in the NCI dataset. Published data describing differential gene expression between fusion-positive and fusion-negative rhabdomyosarcoma patient tumors have identified a number of genes, including *MYOG*, *NOS1*, *OLIG2*, and *PIPOX,* that are highly expressed in tumors with PAX3-FOXO1 fusions, compared to those that lack the fusion (Figure 3B).[27-30] Interrogation of the relationship of these four genes with PAX3-FOXO1 using Sarcoma_CellMinerCDB reveals concordance between the cell line data and the tumor data for each of these genes. Furthermore, no other sarcoma cell lines highly express these genes, suggesting that across sarcomas, they may be specific to fusion-positive rhabdomyosarcoma and can be used as a classifier of this subgroup. Additionally, as the biological function of *NOS1*, *OLIG2*, and *PIPOX* remain largely unexplored in rhabdomyosarcoma, there may be opportunities to study them as prognostic factors and potential therapeutic targets.

Beyond the correlations between fusions and gene expression, the Univariate Analysis tool of Sarcoma_CellMinerCDB can be utilized to identify genetic dependencies. Using the oncogenic fusions to illustrate this, CRISPR data from the Achilles database, which are integrated in Sarcoma_CellMinerCDB (under the “Achilles” dropdown in the x- or y-Axis Cell Line Set), confirms that *FLI1* is an essential gene for EWS-FLI1 fusion positive Ewing sarcoma cell lines, but not for COG-E-352, the EWS-ERG fusion positive Ewing sarcoma cell line (Figure 3C). Similarly, *FOXO1* (encoding the Forkhead Box O1 transcription factor) is essential only in rhabdomyosarcoma cell lines harboring the PAX-FOXO1 fusion (Figure 3D). Taken together, the fusion data highlight the authenticity of cell lines as biologically relevant models, and demonstrate the diverse functionality of integrating transcriptomic, fusion, and CRISPR knockout data.

Mutation and mutational burden characteristics of sarcoma cell lines

The presence of somatic mutations and tumor mutational burden (TMB) are key features that can be used to confirm diagnoses and/or dictate therapeutic decisions in certain subtypes of sarcoma.[31-33] To first capture the overall mutational burden of each sarcoma cell line, we calculated TMB using exome data (Figure 4A). Ewing sarcoma cell lines exhibited the lowest TMB of all the sarcoma subtypes, which matches the known low mutation burden observed in patient tumors and is a well-known feature of the disease.[34,35] In contrast, soft-tissue sarcomas (excluding rhabdomyosarcoma) exhibited the largest TMB range, with some cell lines having greater than 35 mutations/megabase, and others fewer than 10. Although this is an expected consequence of analyzing such a heterogeneous group of tumors together, these data may be useful to identify models of specific subtypes with certain features to be used for preclinical study. For example, there is a clinical subset of leiomyosarcoma known to have *MSH2* mutations and high microsatellite instability (MSI).[36] Sarcoma_CellMinerCDB identifies leiomyosarcoma cell lines SK-UT-1 and SK-UT-1B as having the highest TMB of all the cell lines. Further analysis reveals that these cell lines have pathogenic mutations in the mismatch repair gene *MSH2* (Figure 4B). This is associated with low expression of *MSH2*, potentially conferring a mismatch repair deficiency phenotype and explaining the accumulation of mutations and high TMB (Figure 4B). These cell lines may represent valuable models to preclinically study this subgroup of leiomyosarcomas.

Sarcoma_CellMinerCDB can also be applied to explore mutations for a particular disease entity in specific cell lines. To test this feature for known disease-specific mutations, we looked for cell lines with isocitrate dehydrogenase (*IDH)* mutations. We identified just two cell lines in the whole panel with *IDH* mutations (one for *IDH1* and another for *IDH2*), both of which were in chondrosarcoma cell lines (Supplemental Figure S3A). This result is consistent with the clinical disease biology of chondrosarcoma, as *IDH* mutations are a known genetic feature of chondrosarcoma and are found in approximately half of patients.[37]

We also identified deleterious mutations in the *STAG2* gene (encoding Stromal Antigen 2, a component of the cohesin complex) in seven Ewing sarcoma cell lines. Comparison of the *STAG2* mutations present in the Sarcoma_CellMinerCDB Ewing sarcoma cell lines to those described in Ewing sarcoma patient tumors confirmed that in the cell lines, as in the patient tumors, the mutations are not hot-spot mutations (Figure 4C). Furthermore, based on the Achilles data integrated in Sarcoma_CellMinerCDB, we confirmed that presence of STAG2 is not a dependency in Ewing sarcoma cell lines (Supplemental Figure S3B), which is consistent with the known biological role of STAG2 mutation as a marker of poor prognosis in Ewing sarcoma tumors.[35,38] However, our analysis did reveal that the frequency of *STAG2* mutations in the Ewing sarcoma cell lines (35%) was higher than what would be expected in patient tumors (Figure 4D).[35,39] Although this is a limitation associated with the use of all cell line models, the tool has sufficient data to allow users to determine whether there are differences between the cell line models and the patients, reducing the chance of misinterpretation of the data.

Using the Compare Patterns tool of the Univariate Analysis page of Sarcoma_CellMinerCDB,[12] we identified a correlation between *STAG2* mutational status and sensitivity to the tyrosine kinase inhibitor cabozantinib in the Ewing sarcoma cell lines (Figure 4E). Cabozantinib was recently studied in a clinical trial for patients with bone sarcoma, and in patients with Ewing sarcoma, 25% of patients experienced a partial response.[40] Since no biomarkers of response were studied in this clinical trial, our findings suggest that it may be worth evaluating *STAG2* mutations as a predictive biomarker of response to this therapy. In summary, the mutation data readily accessible in Sarcoma_CellMinerCDB is a rich resource for examining both tumor mutational burden, as well as specific gene mutations in sarcoma cell lines.

Alternative Lengthening of Telomeres (ALT) in the sarcoma cell lines

We next sought to characterize the presence of *TERT* mutations, as *TERT* promoter mutations represent the most common non-coding mutations in cancer cells and have been described in a subset of soft-tissue sarcomas.[41-44] *TERT* encodes the catalytic subunit of telomerase which maintains telomere length, and mutations in *TERT* can reduce telomerase function.[42] *TERT* negative cancer cells use an alternative pathway called the ALT pathway which is active in about 10-15% of cancers, especially in osteosarcomas, and may have therapeutic implications.[45-48] Using Sarcoma_CellMinerCDB, we analyzed osteosarcoma cell lines for *TERT* mutations and expression. Among the 59 sarcoma cell lines sequenced in the NCI database, the 27 cell lines of the GDSC and the 30 cell lines of the CCLE, deleterious *TERT* mutations were only found in one cell line sequenced at the NCI: the CHLA-59 osteosarcoma cell line.

Notably, as expected, known ALT positive osteosarcoma cell lines (U2OS, SAOS2, CAL-72, Hu09 and NY),[49,50] lack *TERT* expression (Supplemental Figure S4), whereas known ALT negative cell lines express *TERT*.[49] Because for many sarcoma cell lines, the ALT status is unknown, examination of *TERT* expression may therefore provide clues to their ALT status. For example, the osteosarcoma cell line Hs 870.T and the spindle cell sarcoma Hs 321.T cells have no *TERT* expression and may represent additional ALT positive bone sarcoma cell lines (Supplemental Figure S4A). Additionally, based on lack of *TERT* expression, there may be a significant number of other types of sarcoma cell lines with ALT positivity, including 3 out of the 25 rhabdomyosarcoma cell lines of the Sarcoma_CellMinerCDB database (Rh30, Rh41, and SJCRH30) and 2 out of the 5 fibrosarcoma cell lines (Hs 414.T, Hs913.T and Hs 93.T). Notably, a significant number of cell lines lack DAXX expression which is a signature of ALT (Supplemental Figure S4B). Further experiments are warranted to expand these results in the cell lines and to determine whether the frequency of ALT and its therapeutic implication are underappreciated in soft tissue sarcomas.[51]

Methylome and methylation profiling

Given the heterogeneity of sarcomas, particularly non-rhabdomyosarcoma soft-tissue sarcomas, the diagnosis of certain subtypes remains a challenge. New approaches using promoter methylation data as an adjunct to traditional pathological and molecular techniques are increasingly being utilized.[52] In addition, a lack of knowledge regarding the cell(s) of origin giving rise to sarcomas remains a knowledge gap in the field, and access to methylation data may help to answer this question. Using promoter methylation data from each cell line, we were able to broadly classify sarcoma subtypes (Figure 5A).

All the Ewing sarcoma cell lines clustered tightly together, with the exception of A673, a widely used cell line bearing the EWS-FLI1 fusion, that has been shown to paradoxically maintain a normal growth rate in experiments silencing EWS-FLI1.[53] Hs 913.T, a fibrosarcoma cell line, clusters with the Ewing sarcoma cell lines, although the reason for this is unclear. Similarly, most of the rhabdomyosarcoma cell lines cluster together, with an apparent separation between alveolar/fusion-positive and embryonal/fusion-negative lines, as has been previously reported for patient samples.[54] The one exception to this is the Hs 729 cell line, which is a pleiomorphic rhabdomyosarcoma and biologically distinct from the embryonal and fusion-positive subtypes; it clusters with the non-rhabdomyosarcoma soft-tissue sarcomas. As expected, osteosarcomas and other soft-tissue sarcomas do not separate as clearly into distinct clusters, likely due their heterogeneity and complex genomic features.

A comparison of overall promoter methylation between the sarcoma subtypes demonstrates that soft-tissue sarcomas, including rhabdomyosarcoma, have higher levels of global promoter methylation than bone sarcomas (Figure 5B). Using hierarchical clustering, comparison of promoter methylation profiles between each of the sarcoma subtypes demonstrates the presence of six gene clusters (Figure 5C, Supplemental Tables S4, S5). Pathway enrichment analysis identifies three clusters (1,2, and 4) with significant pathway enrichment (Supplemental Figure S5A). As is the case with other methylation studies, this clustering may be more reflective of the cell(s) of origin than the oncogenic pathways themselves.[55]

Sarcoma_CellMinerCDB readily allows the visualization of the relationship between expression and promoter methylation status (Table 1). As a representative gene, we used *MYOD1*, a key gene in rhabdomyosarcoma, as an example. As expected, there is a negative correlation between *MYOD1* expression and *MYOD1* promoter methylation in rhabdomyosarcoma cell lines (Figure 5D). In contrast, there is no relationship between *MYOD1* expression and promoter methylation in other sarcoma cell lines, suggesting that the expression of *MYOD1* is regulated by promoter methylation status specifically in rhabdomyosarcoma.

Recent work has shown that in addition to promoter methylation, gene body methylation may be important for predicting gene expression.[56-58] In Sarcoma_CellMinerCDB, we integrated gene body methylation data to augment the predictive value of the promoter methylation data. A representative comparison of gene expression and methylation of *SLFN11* at the promoter versus at the gene body, shows the expected negative correlation with promoter methylation and a positive correlation with body methylation. Importantly, both correlations show highly significant p-values (Supplemental Figure S5B). Furthermore, the use of promoter and body methylation together improves the significance of the predicted gene expression (Supplemental Figure S5C). Taken together, the Sarcoma_CellMinerCDB methylation data provide comprehensive resource of methylation status for sarcoma cell lines and could serve as the foundation for further epigenetic studies, as we demonstrated for the small lung cancer cell lines.[58]

Predictive biomarkers of drug response

As an overarching feature of preclinical models is to uncover biological mechanisms that may translate into a clinical impact on patient outcomes, a unique feature of Sarcoma_CellMinerCDB is the inclusion of drug activity data in the cell lines. These data incorporate drug response data from the NCI, the Broad and Sanger Institutes, as well as the recent data from the National Center for Advancing Translational Science (NCATS) reporting on drug activity for >2500 compounds in 183 cancer cell lines (see Figure 1C and 1E).[13]

Figure 6A displays data generated from Sarcoma_CellMinerCDB showing seven conventionally used therapeutics for the treatment of sarcoma, and their relative activity across the sarcoma cell lines. These data are consistent with what is clinically known about the activity of these agents for specific subtypes of sarcomas; for example, the high sensitivity of rhabdomyosarcoma to dactinomycin, osteosarcoma to methotrexate, and some non-rhabdomyosarcoma soft-tissue sarcomas to pazopanib. In addition, they suggest some new insights for drugs that are not currently being implemented as upfront standard of care in the clinic, such as the exquisite sensitivity of Ewing sarcoma to irinotecan. While irinotecan is a common agent in relapse regimens for Ewing sarcoma, it is not presently part of first-line therapy, and in these results, its activity appears to exceed that of current front-line agents. To understand the genomic determinants of drug sensitivity, the Compare Patterns tool from the Univariate Analysis page can be used to identify biomarkers of response.[12] This unbiased approach reveals that *SLFN11* expression is highly correlated with irinotecan sensitivity in the Ewing sarcoma cell lines (Figure 6B). *SLFN11* expression was also correlated with response to the PARP inhibitor, talazoparib, in Ewing sarcoma cell lines (Figure 6B), which is consistent with independent publications.[16,17,59]

Presently, there are very few examples of targeted therapies that are effective in sarcomas, and even fewer examples of immunotherapeutic approaches that have shown efficacy in patients with sarcoma.[60-62] To interrogate the potential utility of repurposing approved targeted therapies for other cancers in sarcoma, we generated a heatmap showing RNA expression of surface targets with approved antibody-drug conjugates (ADCs) across the full cell line panel (Supplemental Figure S6A). For the three most common sarcoma subtypes, Ewing sarcoma, osteosarcoma, and rhabdomyosarcoma, none of the targets of approved therapies were highly expressed within or across the sarcoma subtypes, suggesting that approved targeted therapies are unlikely to be an effective approach for most patients with these malignancies. However, within the non-rhabdomyosarcoma soft-tissue sarcoma group, there were a small number of cell lines expressing potential targets. For example, two chondrosarcoma cell lines (SW 1353 and JJ012) exhibit high expression of CD33. This has not been reported in the literature but may suggest that there is a role for testing an anti-CD33 therapeutic in a subset of patients with chondrosarcoma, or at the very least examining chondrosarcoma tumors for CD33 positivity, given that gentuzumab ozagamicin, an anti-CD33 ADC conjugated to the antineoplastic antibiotic calicheamicin, is approved for acute myeloid leukemia and is available.

Sarcoma_CellMinerCDB can be used to potentially identify new targets of diagnostic and therapeutic interest for sarcoma. As an example, we generated a heatmap illustrating RNA expression of surface markers across the panel of sarcoma cell lines (Figure 6C, Supplemental Table S5). For each subtype, distinct expression patterns emerged. We focused further on the genes from the heatmap for Ewing sarcoma, selecting four genes with high transcript levels compared to other sarcoma types: *SLCO5A1, NPY5R, PCDH17,* and *CDH8*. Using *FLI1* expression as a comparator, we verified high gene expression for each of the four genes in all Ewing sarcoma cell lines (Figure 6D). Notably, the Ewing sarcoma cell line with an EWS-ERG fusion also expressed high levels of each gene, and no non-Ewing cell lines expressed high levels, confirming the high degree of specificity of these surface markers for Ewing sarcoma. Notably, *NPY5R* expression was recently described as correlating with high SUV measurements in FDG-PET scans of Ewing sarcoma tumors.[63] There are currently no reports describing an association with or function of *SLCO5A1, PCDH17, or CDH8* in Ewing sarcoma.

Given that gene expression levels in cancer cells can be indicative of an overlapping biology of normal tissues, we next examined the expression of the surface markers which are overexpressed in Ewing sarcoma in 7862 normal tissues from 32 tissues of origin (Supplemental Figure S6B). Several genes, including *ITM2A* and *FCGRT*, both of which had been previously identified as EWS-FLI1 target genes,[64,65] were widely expressed across nearly all normal tissue, limiting their potential as therapeutic targets. In contrast, *SLCO5A1, NPY5R, PCDH17,* and *CDH8* expression was low in most normal tissues, suggesting that these targets may be suitable for future development of diagnostics, ADCs, and cellular immune therapies in Ewing sarcoma. Overall, Sarcoma_CellMinerCDB is a powerful tool that can be used to elucidate patterns of drug response and resistance in cell lines according to their genomic characteristics, which is crucial given the overall heterogeneity of sarcoma subtypes. In addition, its multi-functionality can be used to discover new therapeutic targets for patients with sarcoma.

Discussion

Herein we have described Sarcoma_CellMinerCDB, a unique web-based and exploratory resource integrating comprehensive data from multiple sources together with novel data into a single multi-omic research tool allowing easy interrogation of specific genomics and pharmacological features of sarcoma cell line models. Building on new genomic data and the existing previously unlinked databases comprising RNA expression, mutation analyses, and promoter methylation, Sarcoma_CellMinerCDB allows the cross-comparison and full exploitation of those data including gene fusion status, mutations, gene expression, TMB, genome body methylation and large-scale drug screening from multiple sources comprising NCATS. In addition, we built an integrated function, the Global Z-score to facilitate comparisons between 110 cell lines and across the different but highly reproducible data sources. Given the heterogeneity of sarcoma, Sarcoma_CellMinerCDB enables users to select analyses that either incorporate the full group of sarcomas together, the major subtypes of sarcoma, or the rarest diagnoses represented. These functionalities make Sarcoma_CellMinerCDB a unique resource to deeply characterize preclinical sarcoma models, drive new biological questions, and generate hypotheses for translational research. To our knowledge, this is the first public multi-omic resource of its kind.

Although cell lines remain a mainstay for the progress of cancer research, including for sarcoma, concerns have increasingly been raised regarding the reliability of cell lines as models for disease biology. Potential limitations include the effects of immortalization and selection for growth on plastic, which might transform the features of cells or reflect inherently altered biology. However, cell lines have also been shown to be representative models, for example accurately predicting drug responses and gene expression.[66-68] Encouragingly, we were able to use Sarcoma_CellMinerCDB to confirm concordance between the genomic and drug response features of the sarcoma cell lines in this database and those of human tumors, providing evidence that these cell line models represent biologically relevant entities. Furthermore, Sarcoma_CellMinerCDB provides extensive characterization of each cell line and enables comparisons between cell lines, which is especially helpful given the heterogeneity of sarcomas overall. This information can be used to identify relevant differences between the models and inform their use in particular experimental settings. For example, we demonstrated that in the Ewing sarcoma cell lines TC32 and RD-ES, the surface marker *CD99* is not highly expressed. This could be an important factor to consider when selecting representative cell lines for experiments related to *CD99* expression and function, and this type of information can be rapidly located using Sarcoma_CellMinerCDB. This tool may also potentially identify outlier cell lines. For example, we showed through methylation analysis that the Ewing sarcoma cell line A673 did not cluster with the rest of the Ewing sarcoma cell lines, which may suggest there is something different about its origin or biology. This may be an important consideration before using certain cell lines for experimentation or interpreting data generated from their use.

Sarcoma_CellMinerCDB can also be used to identify new avenues for biological discovery. Herein, we present examples of several novel hypothesis-generating insights. Specifically, we describe an association between the PAX3-FOXO1 fusion in rhabdomyosarcoma and several highly expressed genes, namely *NOS1*, *OLIG2*, and *PIPOX.* We show that these genes are exclusively upregulated in rhabdomyosarcoma cell lines bearing the PAX3-FOXO1 fusion and not in any other subtypes of sarcoma. Presently, the function of these genes in fusion-positive rhabdomyosarcoma has not been described, nor have they been studied as potential therapeutic targets for this cancer. Thus, these preliminary findings suggest a new research direction for understanding biological mechanisms in this rare malignancy. An additional biological insight revealed by the Sarcoma_CellMinerCDB tool is the power of combining methylation data from the gene promoter and body regions to better predict gene expression and pathway analyses. We showed that the combined use of promoter and body methylation data increases the significance of predicted expression for a particular gene.[58] Since high quality DNA is easier to obtain than high quality RNA, particularly in clinical specimens of bone tumors,[69,70] assays relying exclusively on DNA may be more likely to provide insights on gene expression and pathway activity. Thus, a DNA-based assay that reports both promoter and body methylation and accurately reproduces RNA-based expression data may be an acceptable alternative when RNA-based sequencing fails. This is of particular importance to clinical translation, as methylation assays are increasingly being used as part of clinical specimen analysis. Currently, clinical methylation assays report mostly promoter methylation. However, the increased predictive value of adding body methylation data[58] may justify development of more comprehensive methylation assays for the future.

Finally, given the relative dearth of novel clinical interventions for sarcoma, perhaps the most impactful aspect of Sarcoma_CellMinerCDB is its ability to provide preliminary data on translationally relevant research questions. Here we report several examples of how the tool can identify novel translational insights for further study. First, a clinical subgroup of leiomyosarcoma with mismatch repair deficiency has been recently described.[36] In our cell line panel of leiomyosarcoma, we identified a subset with pathogenic *MSH2* mutations and MSI phenotype. These cell lines may represent robust preclinical models for this clinical subgroup. In addition, it may be informative to compare the behavior of these *MSH2*-mutated leiomyosarcoma cell lines to that of other cell lines representing MSI-high colon and non-leiomyosarcoma uterine cancers to better evaluate whether therapies targeting MSI could be of use in this newly described subset of leiomyosarcoma. A second illustrative example relates to the challenge of identifying specific markers for diagnosis, prognostication, and therapeutic targeting in sarcoma, due to subtype heterogeneity. Using an original approach based on RNA expression of cell surface markers, we identified *CD33* expression as a candidate marker for chondrosarcoma, a sarcoma subtype with limited therapeutic options.[71] Given that CD33-targeting therapies are currently approved and in use for other malignancies, such as CD33+ acute myeloid leukemia,[72] this finding offers a novel and potentially promising therapeutic strategy for patients with chondrosarcoma. Although further preclinical validation is required, there is reason to be enthusiastic about future clinical testing of the approved anti-CD33 drug conjugate gentuzumab ozagamicin in patients with chondrosarcoma. The mechanism of action of gentuzumab ozagamicin is based on the payload calicheamicin, which acts to induce double-stranded DNA breaks.[73] Some forms of chondrosarcoma are known to be sensitive to other systemic cytotoxic agents, such as doxorubicin, which shares this mechanism,[74,75] suggesting that a subset of patients with chondrosarcoma may benefit from this therapy. Further, ADC technology offers the promise of more targeted tumor delivery, which may enhance antitumor efficacy and decrease systemic toxicity for this subset of patients. Our approach further identified additional candidate surface markers for other sarcoma subtypes, most notably Ewing sarcoma, that may be tractable targets for the future development of diagnostics and therapies.

The sarcoma field has historically suffered from slow progress due to the rarity and heterogeneity of the disease and a lack of models and novel therapeutics. Sarcoma_CellMinerCDB, a publicly-available and interactive resource, is a unique and multifunctional tool that is designed to address some of these limitations. Overall, this resource represents a crucial novel contribution for sarcoma researchers that has the ability to substantially advance the preclinical study of multiple subtypes of sarcoma going forward.

Limitations of the study

A major limitation to this work is the reliance on cell lines as models of disease biology. As previously described, these limitations include the effects of immortalization and selection for growth on plastic, which might select for certain features that lack fidelity with human tumor biology. In addition, cell lines lack the heterogeneity present in human tumors and do not reflect microenvironmental conditions. An additional limitation is the small number of cell lines for some of the sarcoma subtypes in the panel, particularly some of the non-rhabdomyosarcoma soft tissue sarcomas. Small sample sizes make it more difficult to generate and test hypothesis and likely do not reflect the full spectrum of disease for these subtypes. Our goal is to update Sarcoma_CellMinerCDB with additional cell lines for these rarer subtypes as they become available. In addition, we anticipate the development of a similar tool that will include data from sarcoma patient samples.

Author Contributions

Conceptualization: W.C.R., Y.P.
Methodology: F.E., L.P., S.V., W.C.R., Y.P.
Software: F.E., L.P., S.V., W.C.R.
Validation: C.T., C.M.H., F.E., L.P., P.K., S.V., W.C.R., Y.P.
Formal Analysis: F.E., L.P., S.V., W.C.R.
Investigation: C.T., C.M.H., F.E., L.P., P.K., S.V., W.C.R., Y.P.
Resources: J.K., W.C.R., Y.P.
Data Curation: F.E., L.P., S.V.
Writing – Original Draft: C.T., C.M.H.
Writing – Review & Editing: C.T., C.M.H., F.E., L.P., P.K., S.V., P.S.M., J.K., W.C.R., Y.P.
Visualization: C.T., C.M.H., F.E., L.P., P.K., S.V., W.C.R., Y.P.
Supervision: W.C.R., Y.P.
Project Administration: Y.P.
Funding Acquisition: Y.P.

Declarations of Interests: The authors declare no competing interests.

Figure Titles and Legends

Figure 1: Summary of resources and data of Sarcoma_CellMinerCDB

(A) Donut plot summarizing the distribution of 133 sarcoma cell lines by subtype and tissue of origin.

(B) Summary of the genomic, proteomic, metabolic and drug response data in Sarcoma_CellMinerCDB. For mutations, expression (Affymetrix microarray), copy number, methylation and CRISPR data (Achilles), the numbers indicate the number of genes included. For the RNAseq data, the numbers indicate the number of transcripts. For the microRNA data, the numbers indicate the number of microRNA included. For the drug response data, including NCATS, the numbers indicate the number of drugs included. For the proteomic and metabolic data, the numbers indicate the number of proteins included. The bottom row shows the total number of cell lines with data in each category, which are part of Sarcoma_CellMinerCDB. Yellow highlighting indicates Sarcoma_CellMinerCDB data not previously publicly available.

(C) Data available from each data source (listed at the top) for each individual cell line (listed in the left column). Different sarcoma subtypes (Ewing sarcoma, osteosarcoma, other bone sarcoma, rhabdomyosarcoma, other soft tissue sarcoma) are identified by color. Black fill-in represents presence of data from indicated source for each cell line (NCI database (NCATS (n=14), Achilles (n = 54), MD Anderson (n = 32), GDSC (n = 52), CCLE (n = 42), CTRP (n = 30), and NCI (n = 78). Full cell line details are provided in Table S1.

(D) Cell line overlap between data sources.

(E) Drug response data overlap between data sources.

Figure 2: Reproducibility and functionalities of Sarcoma_CellMinerCDB

(A) Violin plots depicting reproducibility between the data sources for gene expression data using Pearson correlations. Correlations between the indicated individual data sources for matched cell lines were 0.92 for NCI_{exp}/NCI_{RNAseq}, 0.76 for NCI_{RNAseq} /$CCLE_{RNAseq}$, 0.76 for $GDSC_{exp}/CCLE_{exp}$, 0.77 for $NCI_{exp}/CCLE_{exp}$ and 0.68 for $NCI_{exp}/GDSC_{exp}$.

(B) Representative scatter plot from Sarcoma_CellMinerCDB showing reproducibility of SLFN11 gene expression in the NCI database (y-axis) vs. the CCLE database (x-axis, left panel) and the Global dataset (x-axis, right panel) across the common cell lines. Each dot represents an individual cell line; sarcoma subtype is indicated by color. Ewing sarcoma cell lines (dark blue dots) have the highest SLFN11 expression compared to all other sarcoma subtype cell lines.

(C) Representative scatter plots from Sarcoma_CellMinerCDB demonstrating the Univariate Analysis tool. In each plot, individual dots represent a cell line; sarcoma subtypes are indicated by color. Left panel shows a significant positive correlation between *FLI1* expression (x-axis) and *SLFN11* expression (y-axis) across all sarcoma cell lines. Middle panel shows *CD99* expression (x-axis) v. *FLI1* expression (y-axis) in all sarcoma cell lines, which Ewing sarcoma cell lines (dark blue) demonstrating the highest expression of *CD99*. Cell lines with the lowest *FLI1* expression represents COG-E-352, which harbors an EWS-ERG fusion. Right panel shows a significant positive correlation between *MDM2* expression (x-axis) and copy number (y-axis). The highest *MDM2* expression level is found in liposarcoma cell lines LS141 and DDLS, which are highlighted in orange and is expected based on known disease biology.

(D) Representative example from Sarcoma_CellMinerCDB demonstrating the Multivariate Analysis tool using the Linear Regression option. Irinotecan sensitivity is highly associated with high *SLFN11* expression (microarray) and low *ABCG2* expression (microarray). This is particularly the case for Ewing sarcoma cell lines, indicated with the red arrows.

Figure 3: Gene fusions and alternative lengthening of telomeres (ALT) in the Sarcoma_CellMinerCDB cell lines

(A) Representative scatter plots from Sarcoma_CellMinerCDB showing correlations between presence of the *EWSR1-FLI1* fusion in Ewing sarcoma cell lines (red) and other sarcomas (blue) (x-axis: 0 means EWSR1-FLI1 fusion is absent, 1 means *EWSR1-FLI1* fusion is present) and four key genes known to be upregulated in Ewing sarcoma. Each dot represents a cell line. The plots show the high correlation between the presence of *EWSR1-FLI1* fusion and *FLI1*, *CD99*, *SLFN11*, and *NROB1* gene expression (y-axes).

(B) Representative scatter plots from Sarcoma_CellMinerCDB showing correlations between the presence of the *PAX3-FOXO1* fusion in rhabdomyosarcoma cell lines (red) and other sarcomas (blue) (x-axis: 0 means *PAX3-FOXO1* fusion is absent, 1 means PAX3-FOXO1 fusion is present) and four key genes involved in *PAX3-FOXO1* fusion-positive rhabdomyosarcoma. Each dot represents a cell line. The plots show the high correlation between the presence of the *PAX3-FOXO1* fusion *MYOG*, *NOS1*, *OLIG2* and *PIPOX* gene expression (y-axes).

(C) Representative scatter plots from Sarcoma_CellMinerCDB demonstrating the essentiality of the *EWSR1-FLI1* fusion in Ewing sarcoma cell lines (red). Each dot represents a cell line. The *EWSR1-FLI1* fusions are shown on the y-axis: 0 and 1 mean absence or presence of *EWSR1-FLI1* fusion, respectively. The dependency score is shown in the x-axis, based on CRISPR knockout of *FLI1* in the CCLE Achilles project (see Fig. 1).

(D) Representative scatter plots from Sarcoma_CellMinerCDB showing the essentiality of the *PAX3-FOXO1* fusion in rhabdomyosarcoma cell lines (red). On the y-axis: 0 means *PAX3-FOXO1* fusion is absent, and 1 indicates *PAX3-FOXO1* fusion. The dependency score is shown in the x-axis, based on CRISPR knockout of *FOXO1* in the Achilles project (see Fig. 1).

Figure 4: Mutations in Sarcoma_CellMinerCDB cell lines

(A) Tumor mutational burden (TMB) (number of mutations per megabase, y-axis) for Ewing sarcoma cell lines (dark blue), osteosarcoma cell lines (light blue), rhabdomyosarcoma cell lines (red), and other soft tissue sarcoma cell lines (orange). Each circle represents a cell line. For each category, the median (dashed line) and standard deviations (dotted lines) are represented. The median number of mutations per megabase is statistically lower in Ewing sarcoma cell lines compared to osteosarcoma (p<0.0001), rhabdomyosarcoma (p=0.0015) and soft tissue sarcomas (STS) (p=0.0016) cell lines.

(B) Representative scatter plot from Sarcoma_CellMinerCDB showing the correlations between TMB (tumor mutational burden, the number of non-inherited mutations per megabase, x-axis) and *MSH2* mutation score (y-axis) (left plot). Each dot represents a cell line. The cell lines with a high mutational burden (SK-UT-1 and SK-UT-1B, both uterine leiomyosarcomas) have a high *MSH2* mutation score. The plot to the right shows the correlation between *MSH2* expression (x-axis) and the *MSH2* mutation score (y-axis). The cell lines with a high *MSH2* mutation score (SK-UT-1 and SK-UT-1B) also have low *MSH2* expression. The cell line with a moderate *MSH2* mutation score (SW-684, synovial sarcoma) has an intermediate *MSH2* expression.

(C) Lollipop plot representing published *STAG2* mutations in Ewing sarcoma tumor patients from the Institut Curie 2014 cohort. Black dots indicate truncating driver mutation, gray dots indicate truncating variant of unknown significance (VUS), and orange dots indicate splice driver mutations. The location of the seven *STAG2* mutations identified in Ewing sarcoma cell lines through Sarcoma_CellMinerCDB are designated in red.

(D) Proportion of STAG2 mutations in Ewing sarcoma cell lines.

(E) Representative scatter plot plots from Sarcoma_CellMinerCDB showing the correlation between *STAG2* mutations (x-axis) and cabozantinib sensitivity (y-axis) (left plot). Each dot represents a cell line. Sarcoma subtypes are defined by colors, with light blue for osteosarcoma, dark blue for Ewing sarcoma, red for rhabdomyosarcoma, orange for other soft tissue sarcoma (see legend). The majority of cell lines with *STAG2* mutations exhibit increased sensitivity to cabozantinib compared to those without STAG2 mutations.

Figure 5: Gene promoter methylation of the Sarcoma_CellMinerCDB cell lines

(A) t-Distributed stochastic neighbor embedding clustering plot using methylation data from the 79 sarcoma cell lines from the NCI data source. Each dot represents a cell line (Ewing sarcoma in dark blue, osteosarcoma in light blue, rhabdomyosarcoma in red, and other soft tissue sarcomas (STS) in orange).

(B) Violin plots showing the median levels of promoter methylation in the sarcoma cell lines. Each point represents the median methylation level of an individual cell line for the total set of 23,202 genes. Ewing sarcoma cell lines are in dark blue, osteosarcoma in light blue, rhabdomyosarcoma in red, and other soft tissue sarcomas in orange. **p=0.0075; ****p<0.0001.

(C) Comparison of promoter methylation profiles for 79 sarcoma cell lines from the NCI data source according to sarcoma subtype (rhabdomyosarcoma, Ewing sarcoma, osteosarcoma, and other soft tissue sarcomas (STS). The heatmap displays the levels of methylation of 744 genes with a high dynamic range. Six gene clusters are obtained using hierarchical clustering. Clusters 1, 2, 3, 4, 5 and 6 include 82, 172, 161, 45, 136, and 148 genes, respectively. The details of the cell lines and gene names by cluster are provided in Supplemental Tables 3 and 4.

(D) Representative scatter plots from Sarcoma_CellMinerCDB showing the negative correlation between *MYOD1* expression (x-axis) and *MYOD1* promoter methylation (y-axis) for the rhabdomyosarcoma cell lines. The left panel shows the correlation with rhabdomyosarcomas cell lines only and the right panel shows the correlation with all sarcoma cell lines.

Figure 6: Therapeutic implications of Sarcoma_CellMinerCDB

(A) Relative drug sensitivity of the Sarcoma_CellMinerCDB NCI cell lines to standard therapeutic agents. Each dot represents a cell line. Sarcoma subtype is indicated by the color of the dot (Ewing sarcoma in dark blue, osteosarcoma in light blue, rhabdomyosarcoma in red, and other soft tissue sarcomas (STS) in orange). Arrows represent agents that are part of the therapy for upfront or relapsed disease for each specific subtype. Drug activity is presented for each of the drugs across the (x-axis) and was calculated using -log10 IC_{50} molar measurements converted to z-scores across cell lines (y-axis).

(B) Representative scatter plots from Sarcoma_CellMinerCDB showing *SLFN11* expression (x-axis) versus irinotecan sensitivity (left panel) and talazoparib sensitivity (right panel). Each dot represents an individual cell line. Ewing sarcoma cell lines, shown in red, highly express *SLFN11* and demonstrate high sensitivity to both irinotecan and talazoparib, as compared to the other sarcoma cell types.

(C) Heat map showing RNA expression of genes coding for surface markers. Highly expressed genes specific to each sarcoma subtype are listed on the left. Sarcoma subtypes are indicated by the colored bar on the top. Individual cell line names are shown at the bottom. Additional information is included in Supplemental Figure S5.

(D) Representative scatter plots showing *FLI1* expression (x-axis) versus *SLCOSA1* expression (y-axis) (upper left panel), NPY5R expression (y-axis) (upper right panel), PCDH*17* expression (y-axis) (lower left panel), and *CDH8* expression (y-axis) (lower right panel) in all CCLE cell lines. Each dot represents a cell line. Ewing sarcoma cell lines are shown in red.

STAR METHODS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for reagents may be directed to and will be fulfilled by Lead Contact Yves Pommier (pommier@nih.gov).

Materials availability

The Sarcoma_CellMinerCDB software is the same as CellMinerCDB and is freely available, open source, and hosted in GitHub at github.com/CBIIT/cellminercdb.

Data and code availability

- **Data:** Data are from CellMinerCDB (https://discover.nci.nih.gov/rsconnect/cellminercdb) and the NCI database (https://sarcomacelllines.cancer.gov/sarcoma**).** The data sources and the method used to obtain the new generated data are detailed below in Method Details. The data are publicly available at https://discover.nci.nih.gov/SarcomaCellMinerCDB).

- **Code:** All codes used are publicly available in GitHub at github.com/CBIIT/cellminercdb.
- **Other:** Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The cell line sets included in Sarcoma_CellMinerCDB are from the National Cancer Institute (NCI) Sarcoma cell lines from the Developmental Therapeutics Program (DTP) and Center for Cancer Research (CCR), Cancer Cell Line Encyclopedia (CCLE), Genomics and Drug Sensitivity in Cancer (GDSC), Cancer Therapeutics Response Portal (CTRP), MD-Anderson, Achilles, National Center for Advancing Translational Sciences (NCATS) and a new resource, Global Z-score. The data source details are described in "Help" section of the Sarcoma_CellMinerCDB website.

METHOD DETAILS

Data Sources

Sarcoma_CellMinerCDB is a dedicated CellMinerCDB version for sarcoma cell lines [76-79] https://discover.nci.nih.gov/cellminercdb/).

Most of the data including drug activity and genomics experiments were processed at the institute of origin and were downloaded from their website or provided from their principal investigator. The genomic data from CTRP and CCLE are common for the overlapping cell lines. However, expression, methylation, mutation and copy number data were processed at Development Therapeutics Branch (DTB), CCR, NCI to generate a gene level summary as described previously.[80-84] The new Global Z-score was developed at DTB by merging the gene expression of all the data sources.[14]

Newly Generated Data

The following section will detail methods for generation and analysis of previously unpublished data present in Sarcoma_CellMinerCDB.

NCI Mutation data

We ran the CCBR exome sequencing pipeline as previously described (https://github.com/mtandon09/CCBR_GATK4_Exome_Seq_Pipeline). In summary, BWA MEM (version 0.7.17) was run to map reads to Hg38 reference genome. Then Mutect2 in GATK 4.2 was used to call the variants with a panel of normal (PON). We processed the variants in a similar fashion to that detailed for the previous dataset.[77] The variants were filtered for those with 6 or more reads. For insertions and deletions we had a further filter of Quality>60 while for other mutations we had a filter of Quality>30. The processing was the same, resulting in a gene mutation summary between 0 and 100 for each gene for each sample as previously described.[77]

Tumor Mutation Burden (TMB)

The TMB was computed using the R package MAFtools based on Mutect2 variants and the following filtering criteria:

1. Minimum read depth of 20
2. VAF>=10%
3. Population AF<0.5% (in the ExAC database)
4. Amino acid changing variant (any one of "frameshift", "missense", "nonframeshift", "nonsense", "read_through", "splicesense" or "initiation_loss")

Fusion Data

Fusion data were obtained for 40 cell lines using RNAseq from NCI. Expressed fusion transcript were detected using three different algorithms FusionCatcher,[85] Star Fusion,[86] and TophatFusion[87] and further filtered using the following criteria:

1. Remove fusions present in normal samples
2. Keep fusions classified as one of these "Tier 1.1", "Tier 1.2", "Tier 1.3", "Tier 2.1"

3. Remove fusions called by only "Star Fusion" or "Fusioncatcher", but to keep any called by two or more callers regardless of spanning reads
4. Keep fusions with spanning reads >= 5 for Tophat only.
5. For right gene intact look at the gene expression value of the right gene, if high then likely true positive
6. In frame fusions were classified as having more credence.

Promoter Gene Level Methylation Data

Promoter gene-level methylation using the 850K Illumina Infinium MethylationEPIC BeadChip array was summarized based on.[76] In short, methylation data were normalized using the minfi package using default parameters, where probe-level beta-values and detection p-values were calculated for each probe. This provided 866,091 methylation probe measurements. Methylation probe beta-values for individual cell lines with detection p-values >=10-3 were set to missing. Also probes with median p-value >=10-6 were set to missing for all cells and removed from the analysis. Probe locations on the human genome (hg19 version) defined by Illumina was used for the analysis, annotating proximal gene transcripts and CpG islands. Probes were designated as category “1” or “2”, with category “1” considered to be most informative. Category “1” probes overlapped CpG islands and they overlapped either the TSS region within a 1.5kb distance, the first exon or 5’-UTR region. Additionally, probes on the upstream shore of a CpG island with a maximal distance of 200bp from the TSS were also included as category “1” probes. Category “2” probes were positioned either in the upstream- or downstream shore of a CpG island and overlapping the first exon, or on the downstream shore of CpG islands overlapping a 200bp region from the TSS, or in 5’-UTR. In case of genes with multiple transcript start sites, the transcript methylation with the most negative correlation to the gene level expression was used. The analysis resulted in gene-level methylation values for 23,202 genes.

Gene Body Level Methylation

We used the gene body methylation quantification introduced in [58]. Briefly, raw methylation files (idat) format were processed in R using the minfi (v1.34.0) package,[88] and the gene body methylation was computed for each gene in each sample as the average methylation of the probes overlapping gene bodies (excluding CpG probes, and probes that overlap promoter areas). For genes with multiple transcripts, the transcript with the most positive correlation was selected.

Copy Number Analysis

Genome wide copy number for the cell lines was estimated from the methylation array data using the Chip Analysis Methylation Pipeline (*ChAMP*) package.[89] *ChAMP* returns lists of genomic segments with putative copy number estimates. However, the estimate is not valid for regions with high methylation detection p-values. For this reason, regions spanning more than 1kb with at least 5 probes with high detection p-values ($p>0.05$) were filtered out. The copy number estimates were set to missing for those areas. Gene level copy number (for n=25,568 genes) was calculated for each gene individually, by calculating the average estimate between the transcription start sites and transcription end sites.

Global Expression Data

We generated a new Global Z-score using all combined cell line resources: NCI, CCLE, CTRP and GDSC. The data sources have a mixture of microarray and RNA-seq gene expression. For each experiment, genes were scaled across all cell lines to create a z-score normalized dataset. The Global Z-score expression was calculated by averaging the z-scored gene expressions from all sources.

QUANTIFICATION AND STATISTICAL ANALYSES

t-SNE Clustering of NCI Sarcoma Cell Lines Using Promoter Methylation

Sarcoma cell line grouping was performed with the gene expression data from the NCI promoter methylation dataset using the t-SNE algorithm in R (v3.5.1). The random seed was set to 1, the Euclidean distance of genes was calculated with the *dist()* function with default settings. The t-SNE grouping was calculated using the *Rtsne()* function from the Rtsne [90] package (v0.15) using the calculated distance matrix, with perplexity set to 10, and 5k maximum iterations.

Methylome Cluster Analysis

The methylation cluster analysis was performed using the methylation data from the 79 NCI-sarcoma cell lines. Genes with high standard deviation (>0.25) in the NCI sarcoma cell lines were selected for the analysis. The number of reported clusters was selected based on the *cutreeDynamic()* function of the *dynamicTreeCut* R package (v1.63-1), which split genes into 6 main clusters and sarcoma cell line subtypes (rhabdomyosarcoma, Ewing sarcoma, osteosarcoma, non-rhabdomyosarcoma soft tissue sarcoma as reported in the figure). The methylation heatmap was created with the *ComplexHeatmap* [91] R package (version 1.20.0).

Drug Analysis

The scatter plot of the drug activities of 12 standard of care sarcoma drug activities was created using drug activity data downloaded from the Sarcoma_CellMinerCDB\ Metadata for NCI selections. The 61 NCI cell lines were analyzed. The data was z scored across cell lines and then plotted using R Computing.[92]

The cluster image map (CIM) in the drug analysis section was generated using the Genomics and Pharmacology Facilities CIMMiner tool (https://discover.nci.nih.gov/cimminer/) selecting the One Matrix CIM, with the Equal width Binning method. The input data is from Sarcoma_CellMinerCDB\Metadata\NCI cell line set\exp: mRNA Expression (log2) microarray data.

Statistical Methods

Correlations, heatmaps, and histograms were generated mostly using The R Project for Statistical Computing. we clustered the cell lines based on gene expression using the raw data and the normalized data in R using the *hclust()* for clustering, and the *ape* package (version 5.3) to create the clustering dendrograms.

Some plots and analysis (such as the Kruskal Willis test) were generated using Partek Genomics suite v7.17.1222 (https://www.partek.com/partek-genomics-suite/), The Xena Functional Genomics Explorer portal (https://xenabrowser.net/),[93] or using Sarcoma_CellMinerCDB and CellMinerCDB (http://discover.nci.nih.gov/cellminercdb).

Wilcoxon rank-sum tests were used to test the difference between continuous variables such as drug sensitivity or gene expression. We considered changes significant if p-values were below 0.05. In the figures, p-values below 0.00005 were summarized with four asterisks, p-values below 0.0005 were summarized with three asterisks, p-values below 0.005 were summarized with two asterisks and p-values below 0.05 were summarized with one asterisk.

KEY RESOURCES TABLES

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
CellMinerCDB cell line data	Rajapakse et al[94]	https://discover.nci.nih.gov/cellminercdb/
Sarcoma cell line data	NCI DTP	https://sarcomacelllines.cancer.gov/sarcoma/
Software and algorithms		
ChAMP	Tian et al[89]	https://bioconductor.org/packages/release/bioc/html/ChAMP.html
STAR aligner	Dobin et al[95]	https://github.com/alexdobin/STAR
Cufflinks	Trapnell et al[96]	http://cole-trapnell-lab.github.io/cufflinks/
Mutation pipeline	CCBR mutation pipeline	https://github.com/mtandon09/CCBR_GATK4_Exome_Seq_Pipeline
ape	Paradis et al[97]	https://cran.r-project.org/web/packages/ape/index.html
relaimpo	Gromping et al[98]	https://cran.r-project.org/web/packages/relaimpo/index.html
dynamicTreeCut	Langfelder et al[99]	https://cran.r-project.org/web/packages/dynamicTreeCut/index.html
ComplexHeatmap	Gu et al[91]	https://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html
Rtsne	Van der Maaten et al[90]	https://cran.r-project.org/web/packages/Rtsne/index.html
clusterProfiler	Yu et al[100]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
ReactomePA	Yu et al[101]	https://bioconductor.org/packages/release/bioc/html/ReactomePA.html
Partek Genomics Suite (software for analysis of microarray data)	Partek	http://www.partek.com/partek-genomics-suite/
Xena Functional Genomics Explorer portal	[93]Goldman et al	https://xenabrowser.net/
The cluster image map (CIM)	CIM	https://discover.nci.nih.gov/cimminer/
GraphPad Prism 10 (software for drawing graphs and statistics analysis)	GraphPad	N/A
Analysis scripts	This paper	The Sarcoma_CellMinerCDB software is the same as CellMinerCDB and is freely available, open source, and hosted in GitHub at github.com/CBIIT/cellminercdb

References

1. HaDuong, J.H., Martin, A.A., Skapek, S.X., and Mascarenhas, L. (2015). Sarcomas. Pediatric Clinics of North America *62*, 179-200. 10.1016/j.pcl.2014.09.012.
2. Choi, J.H., and Ro, J.Y. (2021). The 2020 WHO Classification of Tumors of Bone: An Updated Review. Adv Anat Pathol *28*, 119-138. 10.1097/PAP.0000000000000293.
3. Sbaraglia, M., Bellan, E., and Dei Tos, A.P. (2021). The 2020 WHO Classification of Soft Tissue Tumours: news and perspectives. Pathologica *113*, 70-84. 10.32074/1591-951X-213.
4. Lyu, H.G., Haider, A.H., Landman, A.B., and Raut, C.P. (2019). The opportunities and shortcomings of using big data and national databases for sarcoma research. Cancer *125*, 2926-2934. 10.1002/cncr.32118.
5. Imle, R., Kommoss, F.K.F., and Banito, A. (2021). Preclinical In Vivo Modeling of Pediatric Sarcoma-Promises and Limitations. J Clin Med *10*. 10.3390/jcm10081578.
6. Ferguson, J.L., and Turner, S.P. (2018). Bone Cancer: Diagnosis and Treatment Principles. Am Fam Physician *98*, 205-213.
7. Ray-Coquard, I., Serre, D., Reichardt, P., Martin-Broto, J., and Bauer, S. (2018). Options for treating different soft tissue sarcoma subtypes. Future Oncol *14*, 25-49. 10.2217/fon-2018-0076.
8. Yechieli, R.L., Mandeville, H.C., Hiniker, S.M., Bernier-Chastagner, V., McGovern, S., Scarzello, G., Wolden, S., Cameron, A., Breneman, J., Fajardo, R.D., and Donaldson, S.S. (2021). Rhabdomyosarcoma. Pediatr Blood Cancer *68 Suppl 2*, e28254. 10.1002/pbc.28254.
9. Grunewald, T.G.P., Cidre-Aranaz, F., Surdez, D., Tomazou, E.M., de Alava, E., Kovar, H., Sorensen, P.H., Delattre, O., and Dirksen, U. (2018). Ewing sarcoma. Nat Rev Dis Primers *4*, 5. 10.1038/s41572-018-0003-x.
10. Skapek, S.X., Ferrari, A., Gupta, A.A., Lupo, P.J., Butler, E., Shipley, J., Barr, F.G., and Hawkins, D.S. (2019). Rhabdomyosarcoma. Nat Rev Dis Primers *5*, 1. 10.1038/s41572-018-0051-2.
11. Beird, H.C., Bielack, S.S., Flanagan, A.M., Gill, J., Heymann, D., Janeway, K.A., Livingston, J.A., Roberts, R.D., Strauss, S.J., and Gorlick, R. (2022). Osteosarcoma. Nat Rev Dis Primers *8*, 77. 10.1038/s41572-022-00409-y.
12. Luna, A., Elloumi, F., Varma, S., Wang, Y., Rajapakse, V.N., Aladjem, M.I., Robert, J., Sander, C., Pommier, Y., and Reinhold, W.C. (2021). CellMiner Cross-Database (CellMinerCDB) version 1.2: Exploration of patient-derived cancer cell line pharmacogenomics. Nucleic Acids Res *49*, D1083-D1093. 10.1093/nar/gkaa968.
13. Reinhold, W.C., Wilson, K., Elloumi, F., Bradwell, K.R., Ceribelli, M., Varma, S., Wang, Y., Duveau, D., Menon, N., Trepel, J., et al. (2023). CellMinerCDB: NCATS Is a Web-Based Portal Integrating Public Cancer Cell Line Databases for Pharmacogenomic Explorations. Cancer Res *83*, 1941-1952. 10.1158/0008-5472.CAN-22-2996.
14. Tlemsani, C., Pongor, L., Elloumi, F., Girard, L., Huffman, K.E., Roper, N., Varma, S., Luna, A., Rajapakse, V.N., Sebastian, R., et al. (2020). SCLC-CellMiner: A Resource for Small Cell Lung Cancer Cell Line Genomics and Pharmacology Based on Genomic Signatures. Cell Rep *33*, 108296. 10.1016/j.celrep.2020.108296.
15. Jo, U., and Pommier, Y. (2022). Structural, molecular, and functional insights into Schlafen proteins. Exp Mol Med *54*, 730-738. 10.1038/s12276-022-00794-0.
16. Tang, S.W., Bilke, S., Cao, L., Murai, J., Sousa, F.G., Yamade, M., Rajapakse, V., Varma, S., Helman, L.J., Khan, J., et al. (2015). SLFN11 Is a Transcriptional Target of EWS-FLI1 and a Determinant of Drug Response in Ewing Sarcoma. Clin Cancer Res *21*, 4184-4193. 10.1158/1078-0432.CCR-14-2112.

17. Gartrell, J., Mellado-Largarde, M., Clay, M.R., Bahrami, A., Sahr, N.A., Sykes, A., Blankenship, K., Hoffmann, L., Xie, J., Cho, H.P., et al. (2021). SLFN11 is Widely Expressed in Pediatric Sarcoma and Induces Variable Sensitization to Replicative Stress Caused By DNA-Damaging Agents. Mol Cancer Ther *20*, 2151-2165. 10.1158/1535-7163.MCT-21-0089.

18. Ambros, I.M., Ambros, P.F., Strehl, S., Kovar, H., Gadner, H., and Salzer-Kuntschik, M. (1991). MIC2 is a specific marker for Ewing's sarcoma and peripheral primitive neuroectodermal tumors. Evidence for a common histogenesis of Ewing's sarcoma and peripheral primitive neuroectodermal tumors from MIC2 expression and specific chromosome aberration. Cancer *67*, 1886-1893. 10.1002/1097-0142(19910401)67:7<1886::aid-cncr2820670712>3.0.co;2-u.

19. Thway, K. (2019). Well-differentiated liposarcoma and dedifferentiated liposarcoma: An updated review. Semin Diagn Pathol *36*, 112-121. 10.1053/j.semdp.2019.02.006.

20. Polgar, O., Robey, R.W., and Bates, S.E. (2008). ABCG2: structure, function and role in drug response. Expert Opin Drug Metab Toxicol *4*, 1-15. 10.1517/17425255.4.1.1.

21. Delattre, O., Zucman, J., Plougastel, B., Desmaze, C., Melot, T., Peter, M., Kovar, H., Joubert, I., de Jong, P., Rouleau, G., and et al. (1992). Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. Nature *359*, 162-165. 10.1038/359162a0.

22. Galili, N., Davis, R.J., Fredericks, W.J., Mukhopadhyay, S., Rauscher, F.J., 3rd, Emanuel, B.S., Rovera, G., and Barr, F.G. (1993). Fusion of a fork head domain gene to PAX3 in the solid tumour alveolar rhabdomyosarcoma. Nat Genet *5*, 230-235. 10.1038/ng1193-230.

23. Storlazzi, C.T., Mertens, F., Mandahl, N., Gisselsson, D., Isaksson, M., Gustafson, P., Domanski, H.A., and Panagopoulos, I. (2003). A novel fusion gene, SS18L1/SSX1, in synovial sarcoma. Genes Chromosomes Cancer *37*, 195-200. 10.1002/gcc.10210.

24. Ladanyi, M., Lui, M.Y., Antonescu, C.R., Krause-Boehm, A., Meindl, A., Argani, P., Healey, J.H., Ueda, T., Yoshikawa, H., Meloni-Ehrig, A., et al. (2001). The der(17)t(X;17)(p11;q25) of human alveolar soft part sarcoma fuses the TFE3 transcription factor gene to ASPL, a novel gene at 17q25. Oncogene *20*, 48-57. 10.1038/sj.onc.1204074.

25. May, W.A., Grigoryan, R.S., Keshelava, N., Cabral, D.J., Christensen, L.L., Jenabi, J., Ji, L., Triche, T.J., Lawlor, E.R., and Reynolds, C.P. (2013). Characterization and drug resistance patterns of Ewing's sarcoma family tumor cell lines. PLoS One *8*, e80060. 10.1371/journal.pone.0080060.

26. Boro, A., Pretre, K., Rechfeld, F., Thalhammer, V., Oesch, S., Wachtel, M., Schafer, B.W., and Niggli, F.K. (2012). Small-molecule screen identifies modulators of EWS/FLI1 target gene expression and cell survival in Ewing's sarcoma. Int J Cancer *131*, 2153-2164. 10.1002/ijc.27472.

27. Lae, M., Ahn, E.H., Mercado, G.E., Chuai, S., Edgar, M., Pawel, B.R., Olshen, A., Barr, F.G., and Ladanyi, M. (2007). Global gene expression profiling of PAX-FKHR fusion-positive alveolar and PAX-FKHR fusion-negative embryonal rhabdomyosarcomas. J Pathol *212*, 143-151. 10.1002/path.2170.

28. Rudzinski, E.R., Anderson, J.R., Lyden, E.R., Bridge, J.A., Barr, F.G., Gastier-Foster, J.M., Bachmeyer, K., Skapek, S.X., Hawkins, D.S., Teot, L.A., and Parham, D.M. (2014). Myogenin, AP2beta, NOS-1, and HMGA2 are surrogate markers of fusion status in rhabdomyosarcoma: a report from the soft tissue sarcoma committee of the children's oncology group. Am J Surg Pathol *38*, 654-659. 10.1097/PAS.0000000000000195.

29. Kaleta, M., Wakulinska, A., Karkucinska-Wieckowska, A., Dembowska-Baginska, B., Grajkowska, W., Pronicki, M., and Lastowska, M. (2019). OLIG2 is a novel immunohistochemical marker associated with the presence of PAX3/7-FOXO1

translocation in rhabdomyosarcomas. Diagn Pathol *14*, 103. 10.1186/s13000-019-0883-4.

30. Raghavan, S.S., Mooney, K.L., Folpe, A.L., and Charville, G.W. (2019). OLIG2 is a marker of the fusion protein-driven neurodevelopmental transcriptional signature in alveolar rhabdomyosarcoma. Hum Pathol *91*, 77-85. 10.1016/j.humpath.2019.07.003.
31. Fletcher, J.A., and Rubin, B.P. (2007). KIT mutations in GIST. Curr Opin Genet Dev *17*, 3-7. 10.1016/j.gde.2006.12.010.
32. Parsons, D.W., Janeway, K.A., Patton, D.R., Winter, C.L., Coffey, B., Williams, P.M., Roy-Chowdhuri, S., Tsongalis, G.J., Routbort, M., Ramirez, N.C., et al. (2022). Actionable Tumor Alterations and Treatment Protocol Enrollment of Pediatric and Young Adult Patients With Refractory Cancers in the National Cancer Institute-Children's Oncology Group Pediatric MATCH Trial. J Clin Oncol *40*, 2224-2234. 10.1200/JCO.21.02838.
33. Pestana, R.C., Beal, J.R., Parkes, A., Hamerschlak, N., and Subbiah, V. (2022). Impact of tissue-agnostic approvals for patients with sarcoma. Trends Cancer *8*, 135-144. 10.1016/j.trecan.2021.11.007.
34. Brohl, A.S., Solomon, D.A., Chang, W., Wang, J., Song, Y., Sindiri, S., Patidar, R., Hurd, L., Chen, L., Shern, J.F., et al. (2014). The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutation. PLoS Genet *10*, e1004475. 10.1371/journal.pgen.1004475.
35. Tirode, F., Surdez, D., Ma, X., Parker, M., Le Deley, M.C., Bahrami, A., Zhang, Z., Lapouble, E., Grossetete-Lalami, S., Rusch, M., et al. (2014). Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. Cancer Discov *4*, 1342-1353. 10.1158/2159-8290.CD-14-0622.
36. Choi, J., Manzano, A., Dong, W., Bellone, S., Bonazzoli, E., Zammataro, L., Yao, X., Deshpande, A., Zaidi, S., Guglielmi, A., et al. (2021). Integrated mutational landscape analysis of uterine leiomyosarcomas. Proc Natl Acad Sci U S A *118*. 10.1073/pnas.2025182118.
37. Vuong, H.G., Ngo, T.N.M., and Dunn, I.F. (2021). Prognostic importance of IDH mutations in chondrosarcoma: An individual patient data meta-analysis. Cancer Med *10*, 4415-4423. 10.1002/cam4.4019.
38. Shulman, D.S., Chen, S., Hall, D., Nag, A., Thorner, A.R., Lessnick, S.L., Stegmaier, K., Janeway, K.A., DuBois, S.G., Krailo, M.D., et al. (2022). Adverse prognostic impact of the loss of STAG2 protein expression in patients with newly diagnosed localised Ewing sarcoma: A report from the Children's Oncology Group. Br J Cancer *127*, 2220-2226. 10.1038/s41416-022-01977-2.
39. Shern, J.F., Chen, L., Chmielecki, J., Wei, J.S., Patidar, R., Rosenberg, M., Ambrogio, L., Auclair, D., Wang, J., Song, Y.K., et al. (2014). Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors. Cancer Discov *4*, 216-231. 10.1158/2159-8290.CD-13-0639.
40. Italiano, A., Mir, O., Mathoulin-Pelissier, S., Penel, N., Piperno-Neumann, S., Bompas, E., Chevreau, C., Duffaud, F., Entz-Werle, N., Saada, E., et al. (2020). Cabozantinib in patients with advanced Ewing sarcoma or osteosarcoma (CABONE): a multicentre, single-arm, phase 2 trial. Lancet Oncol *21*, 446-455. 10.1016/S1470-2045(19)30825-3.
41. Campanella, N.C., Penna, V., Abrahao-Machado, L.F., Cruvinel-Carloni, A., Ribeiro, G., Soares, P., Scapulatempo-Neto, C., and Reis, R.M. (2016). TERT promoter mutations in soft tissue sarcomas. Int J Biol Markers *31*, e62-67. 10.5301/jbm.5000168.
42. Dratwa, M., Wysoczanska, B., Lacina, P., Kubik, T., and Bogunia-Kubik, K. (2020). TERT-Regulation and Roles in Cancer Formation. Front Immunol *11*, 589929. 10.3389/fimmu.2020.589929.

43. Zhang, Y., Chen, Y., Yang, C., Seger, N., Hesla, A.C., Tsagkozis, P., Larsson, O., Lin, Y., and Haglund, F. (2021). TERT promoter mutation is an objective clinical marker for disease progression in chondrosarcoma. Mod Pathol *34*, 2020-2027. 10.1038/s41379-021-00848-0.

44. Nacev, B.A., Sanchez-Vega, F., Smith, S.A., Antonescu, C.R., Rosenbaum, E., Shi, H., Tang, C., Socci, N.D., Rana, S., Gularte-Merida, R., et al. (2022). Clinical sequencing of soft tissue and bone sarcomas delineates diverse genomic landscapes and potential therapeutic targets. Nat Commun *13*, 3405. 10.1038/s41467-022-30453-x.

45. Bryan, T.M., Englezou, A., Dalla-Pozza, L., Dunham, M.A., and Reddel, R.R. (1997). Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. Nature medicine *3*, 1271-1274.

46. Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.d.L., Coviello, G.M., Wright, W.E., Weinrich, S.L., and Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. Science *266*, 2011-2015.

47. Sohn, E.J., Goralsky, J.A., Shay, J.W., and Min, J. (2023). The Molecular Mechanisms and Therapeutic Prospects of Alternative Lengthening of Telomeres (ALT). Cancers (Basel) *15*. 10.3390/cancers15071945.

48. Heaphy, C.M., Subhawong, A.P., Hong, S.M., Goggins, M.G., Montgomery, E.A., Gabrielson, E., Netto, G.J., Epstein, J.I., Lotan, T.L., Westra, W.H., et al. (2011). Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes. Am J Pathol *179*, 1608-1615. 10.1016/j.ajpath.2011.06.018.

49. Mason-Osann, E., Dai, A., Floro, J., Lock, Y.J., Reiss, M., Gali, H., Matschulat, A., Labadorf, A., and Flynn, R.L. (2018). Identification of a novel gene fusion in ALT positive osteosarcoma. Oncotarget *9*, 32868-32880. 10.18632/oncotarget.26029.

50. Yost, K.E., Clatterbuck Soper, S.F., Walker, R.L., Pineda, M.A., Zhu, Y.J., Ester, C.D., Showman, S., Roschke, A.V., Waterfall, J.J., and Meltzer, P.S. (2019). Rapid and reversible suppression of ALT by DAXX in osteosarcoma cells. Scientific reports *9*, 4544. 10.1038/s41598-019-41058-8.

51. Flynn, R.L., Cox, K.E., Jeitany, M., Wakimoto, H., Bryll, A.R., Ganem, N.J., Bersani, F., Pineda, J.R., Suva, M.L., Benes, C.H., et al. (2015). Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors. Science *347*, 273-277. 10.1126/science.1257216.

52. Koelsche, C., Schrimpf, D., Stichel, D., Sill, M., Sahm, F., Reuss, D.E., Blattner, M., Worst, B., Heilig, C.E., Beck, K., et al. (2021). Sarcoma classification by DNA methylation profiling. Nat Commun *12*, 498. 10.1038/s41467-020-20603-4.

53. Smith, R., Owen, L.A., Trem, D.J., Wong, J.S., Whangbo, J.S., Golub, T.R., and Lessnick, S.L. (2006). Expression profiling of EWS/FLI identifies NKX2.2 as a critical target gene in Ewing's sarcoma. Cancer Cell *9*, 405-416. 10.1016/j.ccr.2006.04.004.

54. Sun, W., Chatterjee, B., Wang, Y., Stevenson, H.S., Edelman, D.C., Meltzer, P.S., and Barr, F.G. (2015). Distinct methylation profiles characterize fusion-positive and fusion-negative rhabdomyosarcoma. Mod Pathol *28*, 1214-1224. 10.1038/modpathol.2015.82.

55. Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M., Cherniack, A.D., Thorsson, V., et al. (2018). Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell *173*, 291-304 e296. 10.1016/j.cell.2018.03.022.

56. Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., and Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer Cell *26*, 577-590. 10.1016/j.ccr.2014.07.028.

57. Bacolod, M.D., and Barany, F. (2021). MGMT Epigenetics: The Influence of Gene Body Methylation and Other Insights Derived from Integrated Methylomic, Transcriptomic, and

Chromatin Analyses in Various Cancer Types. Curr Cancer Drug Targets *21*, 360-374. 10.2174/1568009621666210203111620.

58. Pongor, L.S., Tlemsani, C., Elloumi, F., Arakawa, Y., Jo, U., Gross, J.M., Mosavarpour, S., Varma, S., Kollipara, R.K., Roper, N., et al. (2022). Integrative epigenomic analyses of small cell lung cancer cells demonstrates the clinical translational relevance of gene body methylation. iScience *25*, 105338. 10.1016/j.isci.2022.105338.

59. Federico, S.M., Pappo, A.S., Sahr, N., Sykes, A., Campagne, O., Stewart, C.F., Clay, M.R., Bahrami, A., McCarville, M.B., Kaste, S.C., et al. (2020). A phase I trial of talazoparib and irinotecan with and without temozolomide in children and young adults with recurrent or refractory solid malignancies. Eur J Cancer *137*, 204-213. 10.1016/j.ejca.2020.06.014.

60. Lim, J., Poulin, N.M., and Nielsen, T.O. (2015). New Strategies in Sarcoma: Linking Genomic and Immunotherapy Approaches to Molecular Subtype. Clin Cancer Res *21*, 4753-4759. 10.1158/1078-0432.CCR-15-0831.

61. Koumarianou, A., and Duran-Moreno, J. (2021). The Sarcoma Immune Landscape: Emerging Challenges, Prognostic Significance and Prospective Impact for Immunotherapy Approaches. Cancers (Basel) *13*. 10.3390/cancers13030363.

62. Moreno Tellez, C., Leyfman, Y., D'Angelo, S.P., Wilky, B.A., and Dufresne, A. (2022). Immunotherapy in Sarcoma: Where Do Things Stand? Surg Oncol Clin N Am *31*, 381-397. 10.1016/j.soc.2022.03.004.

63. Prexler, C., Knape, M.S., Erlewein-Schweizer, J., Roll, W., Specht, K., Woertler, K., Weichert, W., von Luettichau, I., Rossig, C., Hauer, J., et al. (2022). Correlation of Transcriptomics and FDG-PET SUVmax Indicates Reciprocal Expression of Stemness-Related Transcription Factor and Neuropeptide Signaling Pathways in Glucose Metabolism of Ewing Sarcoma. Cancers (Basel) *14*. 10.3390/cancers14235999.

64. Riggi, N., Suva, M.L., Suva, D., Cironi, L., Provero, P., Tercier, S., Joseph, J.M., Stehle, J.C., Baumer, K., Kindler, V., and Stamenkovic, I. (2008). EWS-FLI-1 expression triggers a Ewing's sarcoma initiation program in primary human mesenchymal stem cells. Cancer Res *68*, 2176-2185. 10.1158/0008-5472.CAN-07-1761.

65. Cidre-Aranaz, F., and Alonso, J. (2015). EWS/FLI1 Target Genes and Therapeutic Opportunities in Ewing Sarcoma. Front Oncol *5*, 162. 10.3389/fonc.2015.00162.

66. Wang, H., Huang, S., Shou, J., Su, E.W., Onyia, J.E., Liao, B., and Li, S. (2006). Comparative analysis and integrative classification of NCI60 cell lines and primary tumors using gene expression profiling data. BMC Genomics *7*, 166. 10.1186/1471-2164-7-166.

67. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603-607. 10.1038/nature11003.

68. Weinstein, J.N. (2012). Drug discovery: Cell lines battle cancer. Nature *483*, 544-545. 10.1038/483544a.

69. Carter, L.E., Kilroy, G., Gimble, J.M., and Floyd, Z.E. (2012). An improved method for isolation of RNA from bone. BMC Biotechnol *12*, 5. 10.1186/1472-6750-12-5.

70. Cepollaro, S., Della Bella, E., de Biase, D., Visani, M., and Fini, M. (2018). Evaluation of RNA from human trabecular bone and identification of stable reference genes. J Cell Physiol *233*, 4401-4407. 10.1002/jcp.26319.

71. Rock, A., Ali, S., and Chow, W.A. (2022). Systemic Therapy for Chondrosarcoma. Curr Treat Options Oncol *23*, 199-209. 10.1007/s11864-022-00951-7.

72. Swaminathan, M., and Cortes, J.E. (2023). Update on the role of gemtuzumab-ozogamicin in the treatment of acute myeloid leukemia. Ther Adv Hematol *14*, 20406207231154708. 10.1177/20406207231154708.

73. Zein, N., Sinha, A.M., McGahren, W.J., and Ellestad, G.A. (1988). Calicheamicin gamma 1I: an antitumor antibiotic that cleaves double-stranded DNA site specifically. Science *240*, 1198-1201. 10.1126/science.3240341.

74. van Maldegem, A.M., Gelderblom, H., Palmerini, E., Dijkstra, S.D., Gambarotti, M., Ruggieri, P., Nout, R.A., van de Sande, M.A., Ferrari, C., Ferrari, S., et al. (2014). Outcome of advanced, unresectable conventional central chondrosarcoma. Cancer *120*, 3159-3164. 10.1002/cncr.28845.

75. Italiano, A., Mir, O., Cioffi, A., Palmerini, E., Piperno-Neumann, S., Perrin, C., Chaigneau, L., Penel, N., Duffaud, F., Kurtz, J.E., et al. (2013). Advanced chondrosarcomas: role of chemotherapy and survival. Ann Oncol *24*, 2916-2922. 10.1093/annonc/mdt374.

76. Reinhold, W.C., Varma, S., Sunshine, M., Rajapakse, V., Luna, A., Kohn, K.W., Stevenson, H., Wang, Y., Heyn, H., Nogales, V., et al. (2017). The NCI-60 Methylome and Its Integration into CellMiner. Cancer Res *77*, 601-612. 10.1158/0008-5472.CAN-16-0655.

77. Reinhold, W.C., Varma, S., Sousa, F., Sunshine, M., Abaan, O.D., Davis, S.R., Reinhold, S.W., Kohn, K.W., Morris, J., Meltzer, P.S., et al. (2014). NCI-60 whole exome sequencing and pharmacological CellMiner analyses. PLoS One *9*, e101670. 10.1371/journal.pone.0101670.

78. Reinhold, W.C., Sunshine, M., Liu, H., Varma, S., Kohn, K.W., Morris, J., Doroshow, J., and Pommier, Y. (2012). CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. Cancer Res *72*, 3499-3511. 10.1158/0008-5472.CAN-12-1370.

79. Reinhold, W.C., Varma, S., Sunshine, M., Elloumi, F., Ofori-Atta, K., Lee, S., Trepel, J.B., Meltzer, P.S., Doroshow, J.H., and Pommier, Y. (2019). RNA sequencing of the NCI-60: Integration into CellMiner and CellMiner CDB. Cancer Res *79*, 3514-3524. 10.1158/0008-5472.CAN-18-2047.

80. Polley, E., Kunkel, M., Evans, D., Silvers, T., Delosh, R., Laudeman, J., Ogle, C., Reinhart, R., Selby, M., Connelly, J., et al. (2016). Small Cell Lung Cancer Screen of Oncology Drugs, Investigational Agents, and Gene and microRNA Expression. J Natl Cancer Inst *108*. 10.1093/jnci/djw122.

81. Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature *483*, 570-575. 10.1038/nature11005.

82. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603-307. http://www.nature.com/nature/journal/v483/n7391/abs/nature11003.html#supplementary-information.

83. McMillan, E.A., Ryu, M.J., Diep, C.H., Mendiratta, S., Clemenceau, J.R., Vaden, R.M., Kim, J.H., Motoyaji, T., Covington, K.R., Peyton, M., et al. (2018). Chemistry-First Approach for Nomination of Personalized Treatment in Lung Cancer. Cell *173*, 864-878 e829. 10.1016/j.cell.2018.03.028.

84. Krushkal, J., Silvers, T., Reinhold, W.C., Sonkin, D., Vural, S., Connelly, J., Varma, S., Meltzer, P.S., Kunkel, M., Rapisarda, A., et al. (2020). Epigenome-wide DNA methylation analysis of small cell lung cancer cell lines suggests potential chemotherapy targets. Clin Epigenetics *12*, 93. 10.1186/s13148-020-00876-8.

85. Nicorici, D., Satalan, M., Edgren, H., Kangaspeska, S., Murumagi, A., Kallioniemi, O., Virtanen, S., and Kilkku, O. (2014). FusionCatcher – a tool for finding somatic fusion genes in paired-end RNA-sequencing data. bioRxiv 011650; doi: https://doi.org/10.1101/011650.

86. Haas, B.J., Dobin, A., Li, B., Stransky, N., Pochet, N., and Regev, A. (2019). Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. Genome Biol *20*, 213. 10.1186/s13059-019-1842-9.

87. Kim, D., and Salzberg, S.L. (2011). TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol *12*, R72. 10.1186/gb-2011-12-8-r72.

88. Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363-1369. 10.1093/bioinformatics/btu049.

89. Tian, Y., Morris, T.J., Webster, A.P., Yang, Z., Beck, S., Feber, A., and Teschendorff, A.E. (2017). ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics *33*, 3982-3984. 10.1093/bioinformatics/btx513.

90. van der Maaten, L. (2014). Accelerating t-SNE using Tree-Based Algorithms. J Mach Learn Res *15*, 3221-3245.

91. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847-2849. 10.1093/bioinformatics/btw313.

92. The R Project for Statistical Computing. *https://www.r-project.org*.

93. Goldman, M.J., Craft, B., Hastie, M., Repecka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol *38*, 675-678. 10.1038/s41587-020-0546-8.

94. Rajapakse, V.N., Luna, A., Yamade, M., Loman, L., Varma, S., Sunshine, M., Iorio, F., Sousa, F.G., Elloumi, F., Aladjem, M.I., et al. (2018). CellMinerCDB for Integrative Cross-Database Genomics and Pharmacogenomics Analyses of Cancer Cell Lines. iScience *10*, 247-264. 10.1016/j.isci.2018.11.029.

95. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21. 10.1093/bioinformatics/bts635.

96. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578. 10.1038/nprot.2012.016.

97. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics *20*, 289-290. 10.1093/bioinformatics/btg412.

98. Gromping, U. (2006). Relative Importance for Linear Regression in R: The package relaimpo. J Stat Softw *17*, 1-27.

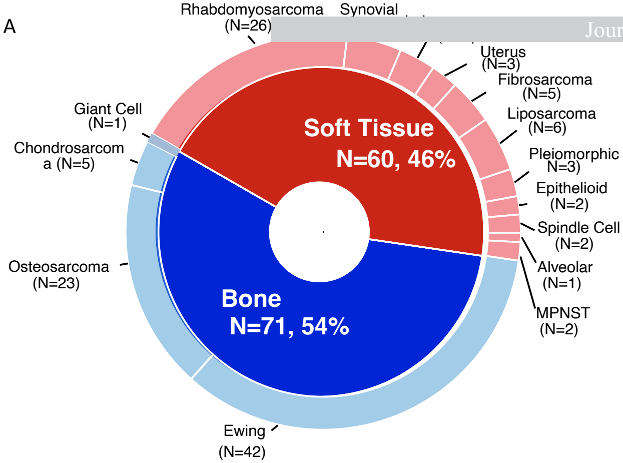
99. Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics *24*, 719-720. 10.1093/bioinformatics/btm563.

100. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284-287. 10.1089/omi.2011.0118.

101. Yu, G., and He, Q.Y. (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst *12*, 477-479. 10.1039/c5mb00663e.

Table 1: Examples of Sarcoma_CellMinerCDB capabilities.

	Sarcoma_CellMinerCDB Explores & Validates	Method	Examples	Examples of Findings
1	Cell line reproducibility, robustness & consistency	Univariate Analyses: Plot Data: Expression of the same gene across different datasets (X & Y)	Fig. 2B	Cell lines are highly reproducible across datasets.
3	Integrates all the Sarcoma cell line genomic datasets under Global Z score (NCI, GDSC, CCLE, CTRP)	Use the pull-down tabs for Cell Line Sets and choose "Global"	Figs. 2B ; S2C-D	110 sarcoma cell lines can be compared using gene expression status. Among them, highest SLFN11 expression are observed in Ewing sarcoma cell lines.
5	Select and compare subsets of cell lines based on sarcoma subtypes	Univariate Analyses: select Y axis: Select Tissue or Show color	Figs. 2B-C; 3C-D-F ; 4E ; 5D ; 6B-D	SLCOSA1, NPYSR, PCDH17 and CDH8 surface cell markers are selectively expressed in Ewing sarcoma cell lines.
6	Select and compare subsets of sarcoma cell lines based on recurrent fusion data	Univariate Analyses: select NCI : gene fusions and write the gene of interest	Figs. 3A-B-C-D	Fusion positive rhabdomyosarcoma cell lines selectively over expressed specific genes such as MYOG, NOS1, OLIG2, PIPOX.
	Identify essential genes (Achilles) in a subset of sarcoma cell lines	Univariate Analyses: Plot Data select Achilles and write the gene of interest (X & Y)	Figs. 3C-D	FOXO1 is essential only in fusion positive rhabdomyosarcoma cell lines.
	Select and compare subsets of sarcoma cell lines based on the Tumor Mutation Burden (TMB)	Univariate Analyses: select NCI : mda: Signatures, Miscellaneous data and TMB	Fig. 4B	Identification of a subset of leiomyosarcoma cell lines with a high TMB.
	Mutation visualization for each gene	Mutation variants: select NCI : and write the gene name	Figs. 4C-D ; S3	50% of the chondrosarcoma cell lines of our collection have a IDH1 or IDH2 mutation
8	Epigenetics: promoter and body methylation for any given gene	Univariate analyses: Plot Data: Expression of a given gene vs its body or promoter methylation (X & Y Data Type) within a given Cell Line Set or across datasets (independent datasets can be tested for missing Data Type and confirmation)	Fig. S4B	Both promoter and body methylation of SLFN11 are highly negatively and positively respectively correlated with SLFN11 expression.
9	Gene amplification and deletions for any given gene	Univariate analyses: Plot Data: Expression of a given gene vs copy number (X & Y Data Type) within a given Cell Line Set or across datasets (independent datasets can be tested for validation and missing Data Type)	Fig. 2C	MDM2 expression is as expected driven by copy number variation in dedifferentiated liposarcoma cell lines.
10	Integrate and complement different datasets for common cell lines	Univariate Analyses: Plot Data: Plot different parameters (Data Type for genomic or drug response) across Cell Line Sets (X & Y) to counter missing data in one dataset	Figs. 3C-D ; S3B	FLI1 is essential (CRISPR/Achilles data) only in ES sarcoma cell lines with a recurrent EWSR1-FLI1 fusion (NCI data).
12	Discover determinants of drug response and targeted drug delivery	Univariate Analyses: Plot Data: Compare Patterns: Coregulated genes for a given gene (X or Y) within a given dataset (independent datasets can be tested for confirmation)	Figs. 4E ; 6B	SLFN11 expression is associated with response to talozoparib in Ewing sarcoma cell lines.
14	Examine drug correlations: COMPARE analyses	Univariate Analyses: Plot Data: Data Type: drug vs drug (X or Y); also select Compare patterns to identify drug-drug ou drug/molecular data correlations	Fig. 4E	STAG2 mutation in Ewing sarcoma cell lines is associated with response to cabozantinib.
15	Multivariate models of drug response & genomic features	Multivariate Analyses: Cell Line Set; Response Data Type; Predictor Data Type/s; Predictor Identifier: enter drug and genomic parameters to be tested as indentifier or use LASSO to discover additional non-redundant determinants of response or compare response according to sarcoma subtypes	Figs 2D ; 6A	ABCG2 transporter expression is negatively correlated with SLFN11 expression and response to irinotecan in sarcoma cell lines.
16	Data download	Univariate Analyses: View Data: Download tabs or Multivariate Analyses: Download tab	-	Allow further in depth analyses and data download in Excel
17	Drug identifier conversion	Not applicable	-	Allow drug identification across different sources
	Integration with CellMinerCDB	Open in parallel: http://discover.nci.nih.gov/cellminercdb	-	Identify genes that are selective for sarcomas comparing with the entire NCI cell line collection including several tissues.



B

	Mutations (exome)	Expression (Affymetrix)	RNAseq	Copy number	Methylation	microRNA (Nanostring)	Drugs	Proteomic	Metabolic	CRISPR Cas9 KO screen
NCI 78 cell lines	7711 33 cell lines	18432 67 cell lines	54283 66 cell lines	25476 77 cell lines	23202 77 cell lines (450k) Body methylation: 23808 71 cell lines (850k)	800 55 cell lines	291 63 cell lines	-	-	-
CTRP 30 cell lines	1667 28 cell lines	19851 30 cell lines	-	23316 30 cell lines	-	-	481 30 cell lines	-	-	-
CCLE 42 cell lines	1667 35 cell lines	19851 42 cell lines	52604 40 cell lines	23316 42 cell lines	19880 30 cell lines (RRBS)	734 39 cell lines	24 21 cell lines	214 33 cell lines	225 32 cell lines	-
GDSC 52 cell lines	18099 51 cell lines	19562 49 cell lines	37263 27 cell lines	24502 50 cell lines	19864 50 cell lines (450k)	-	297 52 cell lines	-		-
MD Anderson 32 cell lines	-	-	-	-	-	-	-	452 32 cell lines		-
NCATS 14 cell lines	-	-	-	-	-	-	2675 14 cell lines			
Achilles 53 cell lines	-	-	-	-	-	-	-	-		18119 53 cell lines
Total 133 cell lines	63 cell lines	97 cell lines	80 cell lines	108 cell lines	104 cell lines	69 cell lines	92 cell lines	45 cell lines	32 cell lines	53 cell lines

Global Z score
61639 genes
110 cell lines

D

Cell lines overlap between datasets

	NCI	CCLE	GDSC	CTRP	ACHILLES	MDA	GLOBAL	NCATS
NCI	78	27	28	19	26	27	78	11
CCLE		42	26	30	22	18	42	10
GDSC			54	24	18	17	54	8
CTRP				30	18	16	30	8
ACHILLES					54	20	32	9
MDA						32	30	6
GLOBAL							110	13
NCATS								14

E

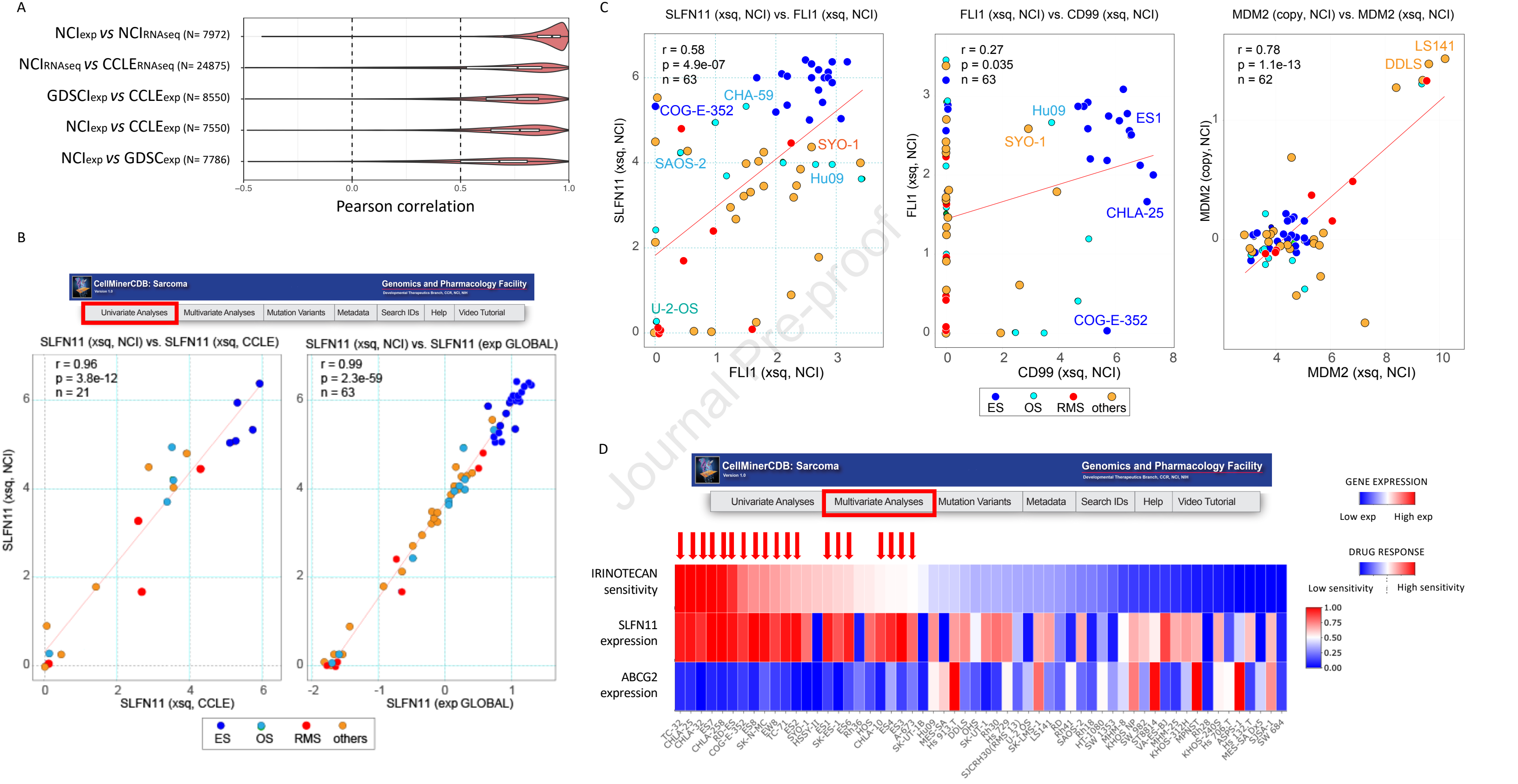
Drug overlap between datasets

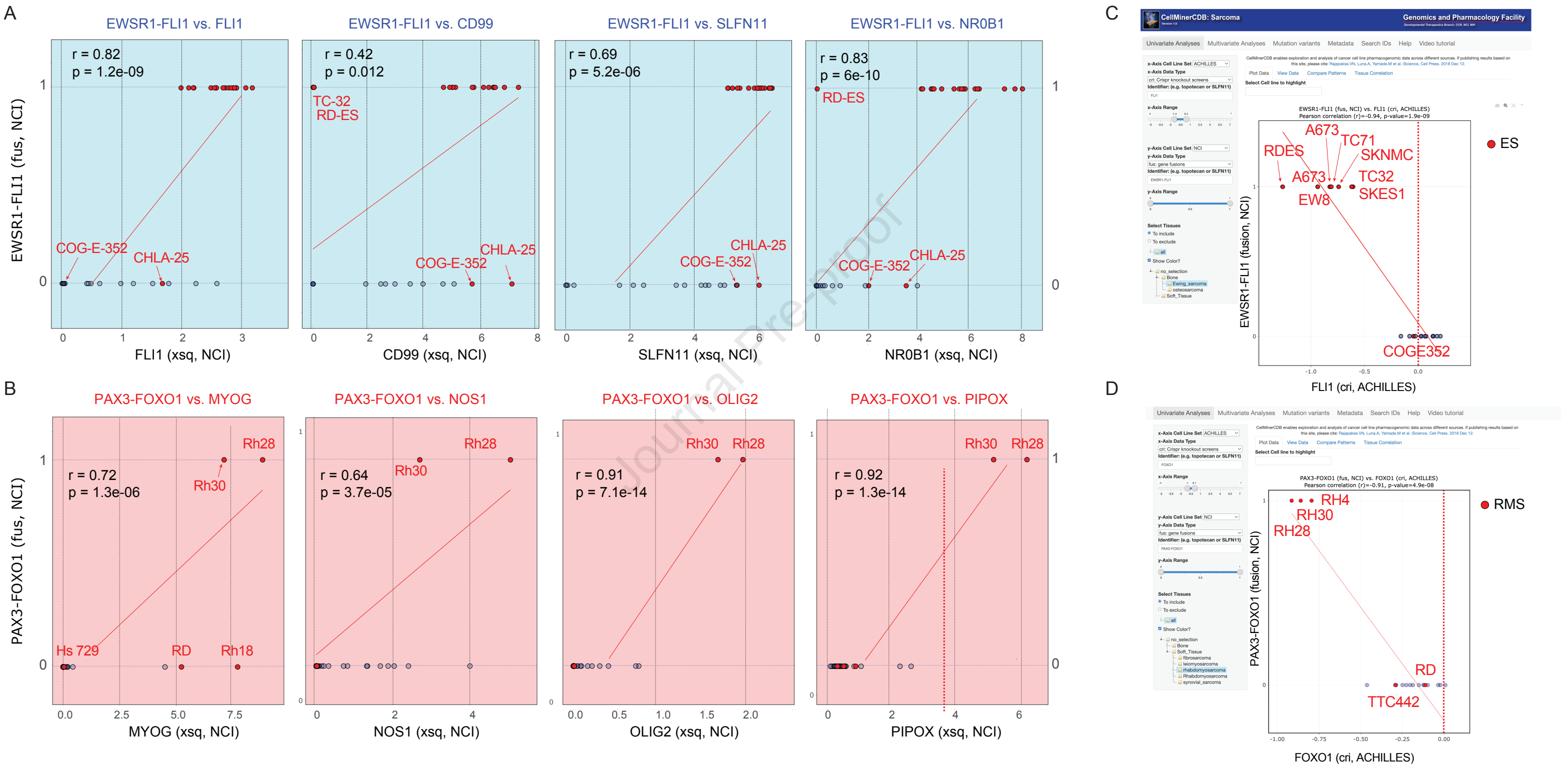
	NCI	CCLE	GDSC	CTRP	GDSC1	GDSC2	NCATS
NCI	291	18	91	90	97	79	232
CCLE		24	16	14	16	14	22
GDSC			297	77	235	114	201
CTRP				481	83	66	165
GDSC1					402	120	245
GDSC2						295	181
NCATS							2675

🟦	Ewing Sarcoma
🟦	Osteosarcoma
🟦	Chondrosarcoma
🟥	Rhabdomyosarcoma
🟨	Other STS
⬜	Not specified

□	not in the database
■	in the database

Cell Line	Subtype	NCATS	Achilles	MD	GDSC	CCLE	CTRP	NCI
A-673	Ewing Sarcoma	□	■	■	■	■	■	■
SK-ES-1	Ewing Sarcoma	□	■	■	■	■	■	■
TC-71	Ewing Sarcoma	■	■	■	■	■	■	■
RD-ES	Ewing Sarcoma	■	■	■	□	■	■	■
SK-N-MC	Ewing Sarcoma	□	■	□	□	■	■	■
TC-32	Ewing Sarcoma	■	■	■	□	■	□	■
EW8	Ewing Sarcoma	■	■	□	□	■	□	■
CADO-ES1	Ewing Sarcoma	□	□	□	■	■	■	□
MHH-ES-1	Ewing Sarcoma	□	□	□	■	■	■	□
ES1	Ewing Sarcoma	□	□	□	■	□	□	■
ES3	Ewing Sarcoma	□	□	□	■	□	□	■
ES4	Ewing Sarcoma	□	□	□	■	□	□	■
ES6	Ewing Sarcoma	□	□	□	■	□	□	■
ES7	Ewing Sarcoma	□	□	□	■	□	□	■
ES8	Ewing Sarcoma	□	□	□	■	□	□	■
Rh1	Ewing Sarcoma	□	□	□	■	□	□	■
CHLA-10	Ewing Sarcoma	□	□	■	□	□	□	■
CHLA-258	Ewing Sarcoma	□	□	■	□	□	□	■
CHLA-9	Ewing Sarcoma	□	□	■	□	□	□	■
COG-E-352	Ewing Sarcoma	□	■	□	□	□	□	■
5838	Ewing Sarcoma	□	□	□	□	□	□	■
CHLA-25	Ewing Sarcoma	□	□	□	□	□	□	■
CHLA-32	Ewing Sarcoma	□	□	□	□	□	□	■
ES2	Ewing Sarcoma	□	□	□	□	□	□	■
EWS502	Ewing Sarcoma	□	■	□	□	■	□	□
ES5	Ewing Sarcoma	□	□	□	■	□	□	□
EW-1	Ewing Sarcoma	□	□	□	■	□	□	□
EW-11	Ewing Sarcoma	□	□	□	■	□	□	□
EW-12	Ewing Sarcoma	□	□	□	■	□	□	□
EW-13	Ewing Sarcoma	□	□	□	■	□	□	□
EW-16	Ewing Sarcoma	□	□	□	■	□	□	□
EW-18	Ewing Sarcoma	□	□	□	■	□	□	□
EW-22	Ewing Sarcoma	□	□	□	■	□	□	□
EW-24	Ewing Sarcoma	□	□	□	■	□	□	□
EW-3	Ewing Sarcoma	□	□	□	■	□	□	□
EW-7	Ewing Sarcoma	□	□	□	■	□	□	□
SKNEP1	Ewing Sarcoma	□	□	□	■	□	□	□
SKPNDW	Ewing Sarcoma	□	□	□	■	□	□	□
TC106	Ewing Sarcoma	□	■	□	□	□	□	□
TC138	Ewing Sarcoma	□	■	□	□	□	□	□
TC205	Ewing Sarcoma	□	■	□	□	□	□	□
CHLA-57	Ewing Sarcoma	□	■	□	□	□	□	□
SJSA-1	Osteosarcoma	□	■	■	■	■	■	■
U-2 OS	Osteosarcoma	□	■	■	■	■	■	■
HOS	Osteosarcoma	□	□	■	■	■	■	■
SAOS-2	Osteosarcoma	■	■	□	■	■	■	■
G-292, clone A141B1	Osteosarcoma	■	■	□	■	■	■	□
MG-63	Osteosarcoma	□	□	■	■	■	■	□
143B	Osteosarcoma	□	■	□	□	■	□	■
Hu09	Osteosarcoma	□	■	□	■	□	□	■
Hs 870.T	Osteosarcoma	□	□	□	□	■	□	■
T1-73	Osteosarcoma	□	□	□	□	■	□	■
KHOS-240S	Osteosarcoma	□	□	■	□	□	□	■
KHOS-312H	Osteosarcoma	□	□	■	□	□	□	■
Hs 888.T	Osteosarcoma	□	□	□	□	■	■	□
CHA-59	Osteosarcoma	□	□	□	□	□	□	■
KHOS NP	Osteosarcoma	□	□	□	□	□	□	■
OHS	Osteosarcoma	□	□	□	□	□	□	■
CAL-72	Osteosarcoma	□	■	□	■	□	□	□
EW7476	Osteosarcoma	□	□	□	■	□	□	□
HuO-3N1	Osteosarcoma	■	□	□	■	□	□	□
NOS-1	Osteosarcoma	□	□	□	■	□	□	□
NY	Osteosarcoma	□	□	□	■	□	□	□
SARC9371	Osteosarcoma	□	□	□	■	□	□	□
OS252	Osteosarcoma	□	■	□	□	□	□	□
SW 1353	Chondrosarcoma	□	□	■	□	■	■	■
Hs 819.T	Chondrosarcoma	□	□	□	□	■	□	■
OUMS-27	Chondrosarcoma	□	□	□	□	■	□	□
JJ012	Chondrosarcoma	□	□	□	□	□	□	■
CAL78	Chondrosarcoma	□	■	□	□	□	□	□
Hs 706.T	Giant Cell Sarcoma	□	□	□	□	■	□	■
RD (Rh18dm)	Rhabdomyosarcoma	■	■	■	■	■	■	■
Rh18	Rhabdomyosarcoma	□	■	□	■	■	■	■
Rh41	Rhabdomyosarcoma	■	□	□	■	■	■	■
Rh30	Rhabdomyosarcoma	■	■	□	□	■	■	■
Hs 729	Rhabdomyosarcoma	□	□	□	□	■	■	■
SJCRH30(RMS 13)	Rhabdomyosarcoma	□	□	□	■	■	□	■
KYM-1	Rhabdomyosarcoma	□	□	□	■	■	■	□
TE 441.T	Rhabdomyosarcoma	□	□	□	■	■	■	□
Rh28	Rhabdomyosarcoma	□	■	□	□	□	□	■
Rh4	Rhabdomyosarcoma	□	■	□	□	□	□	■
TTC-442	Rhabdomyosarcoma	□	■	□	□	□	□	■
TE 617.T	Rhabdomyosarcoma	□	□	□	□	■	■	□
Rh18c	Rhabdomyosarcoma	□	□	□	□	□	□	■
Rh28 PX11/LPAM	Rhabdomyosarcoma	□	□	□	□	□	□	■
Rh36	Rhabdomyosarcoma	■	□	□	□	□	□	■
Rh5	Rhabdomyosarcoma	□	□	□	□	□	□	■
RMS559	Rhabdomyosarcoma	■	□	□	□	□	□	■
SMSCTR	Rhabdomyosarcoma	□	■	□	□	□	□	■
TTC-516	Rhabdomyosarcoma	□	□	□	□	□	□	■
TE 125.T	Rhabdomyosarcoma	□	□	□	□	■	□	□
TE 159.T	Rhabdomyosarcoma	□	□	□	□	■	□	□
RHJT	Rhabdomyosarcoma	□	■	□	□	□	□	□
CW9019	Rhabdomyosarcoma	□	■	□	□	□	□	□
JR	Rhabdomyosarcoma	□	■	□	□	□	□	□
SCMCRM2	Rhabdomyosarcoma	□	■	□	□	□	□	□
CCA	Rhabdomyosarcoma	■	□	□	□	□	□	□
SW982	Synovial Sarcoma	□	■	■	■	□	□	■
SYO-1	Synovial Sarcoma	□	■	□	□	□	□	■
HSSY-II	Synovial Sarcoma	□	□	□	□	□	□	■
CME1	Synovial Sarcoma	□	■	□	□	□	□	□
SCS214	Synovial Sarcoma	□	■	□	□	□	□	□
YAMATO	Synovial Sarcoma	□	■	□	□	□	□	□
HT-1080	Fibrosarcoma	■	■	■	■	■	■	■
SW684	Fibrosarcoma	□	□	□	■	□	□	■
Hs 913.T	Fibrosarcoma	□	□	□	□	□	□	■
Hs 414.T	Fibrosarcoma	□	□	□	□	□	□	■
Hs 93.T	Fibrosarcoma	□	□	□	□	□	□	■
SK-LMS-1	Leiomyosarcoma	□	□	■	■	■	■	■
SK-UT-1	Leiomyosarcoma	□	■	□	■	■	■	■
RKN	Leiomyosarcoma	□	■	□	■	■	■	□
SK-UT-1B	Leiomyosarcoma	□	□	□	□	□	□	■
MES-SA	Uterine Sarcoma	□	□	■	■	■	■	■
ESS-1	Uterine Sarcoma	□	■	■	■	■	■	□
MES-SA Dx5	Uterine Sarcoma	□	□	□	□	□	□	■
LS141 (LPS141)	Dedifferentiated Liposarcoma	□	■	□	□	□	□	■
DDLS	Dedifferentiated Liposarcoma	□	□	□	□	□	□	■
LPS853	Dedifferentiated Liposarcoma	□	■	□	□	□	□	□
LPS27	Liposarcoma	□	■	□	□	□	□	□
LPS6	Liposarcoma	□	■	□	□	□	□	□
95T1000	Well differentiated liposarcoma	□	■	□	□	□	□	□
GCT	Pleomorphic sarcoma	□	■	□	■	■	■	□
CCLFPEDS0003T	Pleomorphic sarcoma	□	■	□	□	□	□	□
S117	Pleomorphic sarcoma	□	■	□	□	□	□	□
VA-ES-BJ	Epithelioid Sarcoma	□	□	■	■	□	□	■
CCLFPEDS0008T	Epithelioid Sarcoma	□	■	□	□	□	□	□
Hs 132.T	Spindle Cell Sarcoma	□	□	□	□	□	□	□
Hs 312.T	Spindle Cell Sarcoma	□	□	□	□	□	□	■
MPNST	MPNST	□	□	□	□	□	□	■
ST8814	MPNST	□	□	□	□	□	□	■
ASPS-1	Alveolar Soft Part Sarcoma	□	□	□	□	□	□	■
MHM-25	Not specified	□	□	□	□	□	□	■
MHM-8	Not specified	□	□	□	□	□	□	■


A
NCIexp vs NCIRNAseq (N= 7972)
NCIRNAseq vs CCLERNAseq (N= 24875)
GDSCIexp vs CCLEexp (N= 8550)
NCIexp vs CCLEexp (N= 7550)
NCIexp vs GDSCexp (N= 7786)
−0.5
0.0
0.5
1.0
Pearson correlation
B
CellMinerCDB: Sarcoma
Version 1.0
Genomics and Pharmacology Facility
Developmental Therapeutics Branch, CCR, NCI, NIH
Univariate Analyses
Multivariate Analyses
Mutation Variants
Metadata
Search IDs
Help
Video Tutorial
SLFN11 (xsq, NCI) vs. SLFN11 (xsq, CCLE)
r = 0.96
p = 3.8e-12
n = 21
SLFN11 (xsq, NCI)
SLFN11 (xsq, CCLE)
SLFN11 (xsq, NCI) vs. SLFN11 (exp GLOBAL)
r = 0.99
p = 2.3e-59
n = 63
SLFN11 (exp GLOBAL)
ES
OS
RMS
others
C
SLFN11 (xsq, NCI) vs. FLI1 (xsq, NCI)
r = 0.58
p = 4.9e-07
n = 63
COG-E-352
CHA-59
SAOS-2
SYO-1
Hu09
U-2-OS
SLFN11 (xsq, NCI)
FLI1 (xsq, NCI)
FLI1 (xsq, NCI) vs. CD99 (xsq, NCI)
r = 0.27
p = 0.035
n = 63
Hu09
SYO-1
ES1
CHLA-25
COG-E-352
FLI1 (xsq, NCI)
CD99 (xsq, NCI)
MDM2 (copy, NCI) vs. MDM2 (xsq, NCI)
r = 0.78
p = 1.1e-13
n = 62
LS141
DDLS
MDM2 (copy, NCI)
MDM2 (xsq, NCI)
ES
OS
RMS
others
D
CellMinerCDB: Sarcoma
Version 1.0
Genomics and Pharmacology Facility
Developmental Therapeutics Branch, CCR, NCI, NIH
Univariate Analyses
Multivariate Analyses
Mutation Variants
Metadata
Search IDs
Help
Video Tutorial
GENE EXPRESSION
Low exp
High exp
DRUG RESPONSE
Low sensitivity
High sensitivity
1.00
0.75
0.50
0.25
0.00
IRINOTECAN sensitivity
SLFN11 expression
ABCG2 expression
TC-32
CHLA-25
CHLA-32
ES7
CHLA-258
RD-ES
COG-E-352
ES8
SK-N-MC
EW8
TC-71
ES2
SYO-1
HSSY-II
ES1
SK-ES-1
ES6
Rh36
HOS
CHLA-10
ES4
ES3
A-673
SK-UT-1B
Hu09
MES-SA
Hs 913.T
DDLS
OHS
SK-UT-1
Rh30
Hs 729
SJCRH30(RMS 13)
U-2 OS
SK-LMS-1
LS141
RD
Rh41
SAOS-2
Rh18
HT-1080
SW 1353
MHM-8
KHOS NP
SW 982
ST8814
VA-ES-BJ
MHM-25
KHOS-312H
MPNST
Rh28
KHOS-240S
Hs 706.T
ASPS-1
Hs 132.T
MES-SA Dx5
SJSA-1
SW 684



C

CellMinerCDB enables exploration and analysis of cancer cell line pharmacogenomic data across different sources. If publishing results based on this site, please cite: Rajapakse.VN, Luna.A, Yamade.M et al. iScience, Cell Press. 2018 Dec 12.

x-Axis Cell Line Set: ACHILLES
x-Axis Data Type: cri: Crispr knockout screens
Identifier: (e.g. topotecan or SLFN11): FLI1
x-Axis Range: -1.4 – 0.4
y-Axis Cell Line Set: NCI
y-Axis Data Type: fus: gene fusions
Identifier: (e.g. topotecan or SLFN11): EWSR1-FLI1
y-Axis Range: 0 – 1
Select Tissues: ● To include ○ To exclude
all
☑ Show Color?
no_selection
Bone
Ewing_sarcoma
osteosarcoma
Soft_Tissue

Select Cell line to highlight

EWSR1-FLI1 (fus, NCI) vs. FLI1 (cri, ACHILLES)
Pearson correlation (r)=-0.94, p-value=1.9e-09

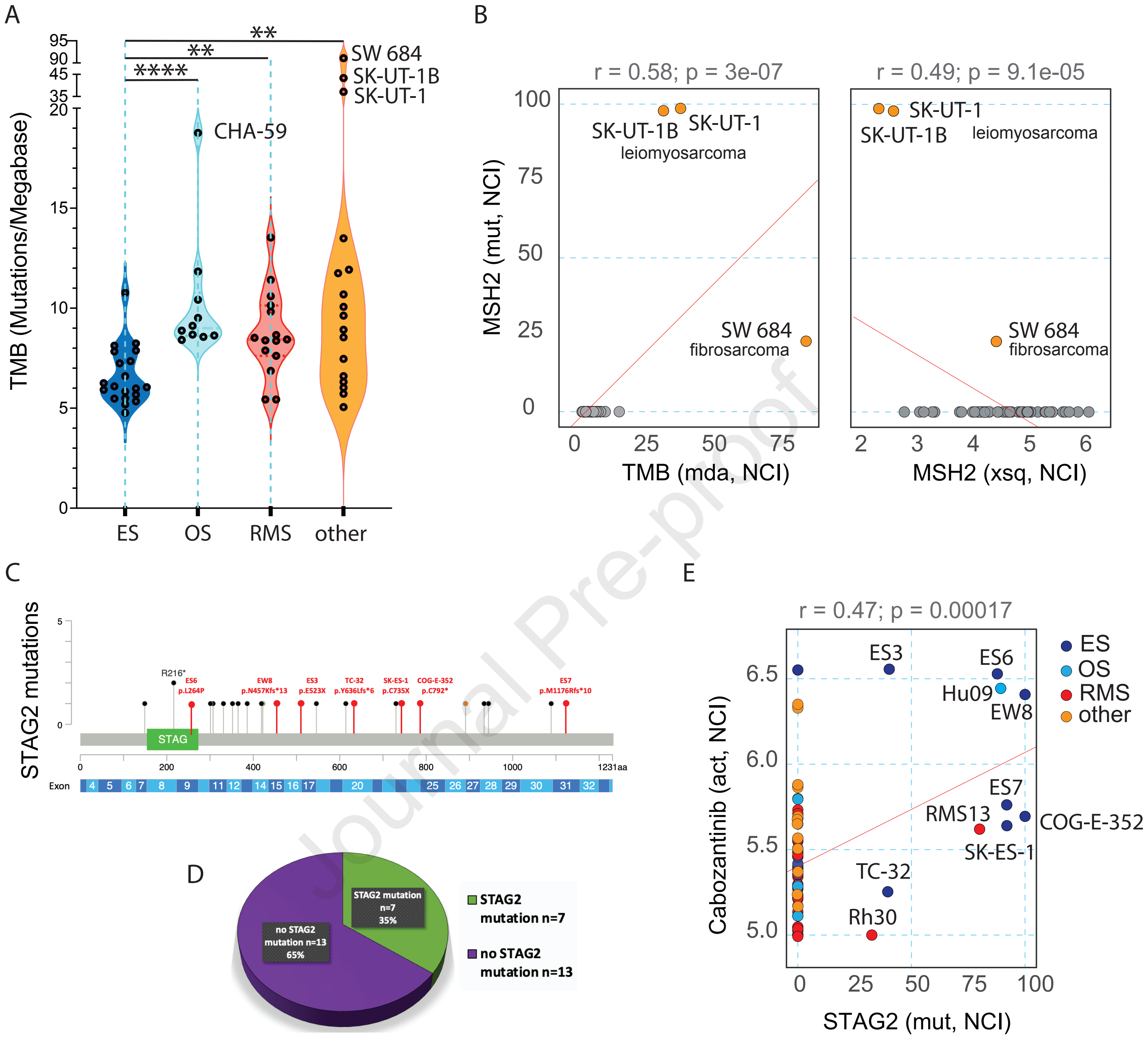
D

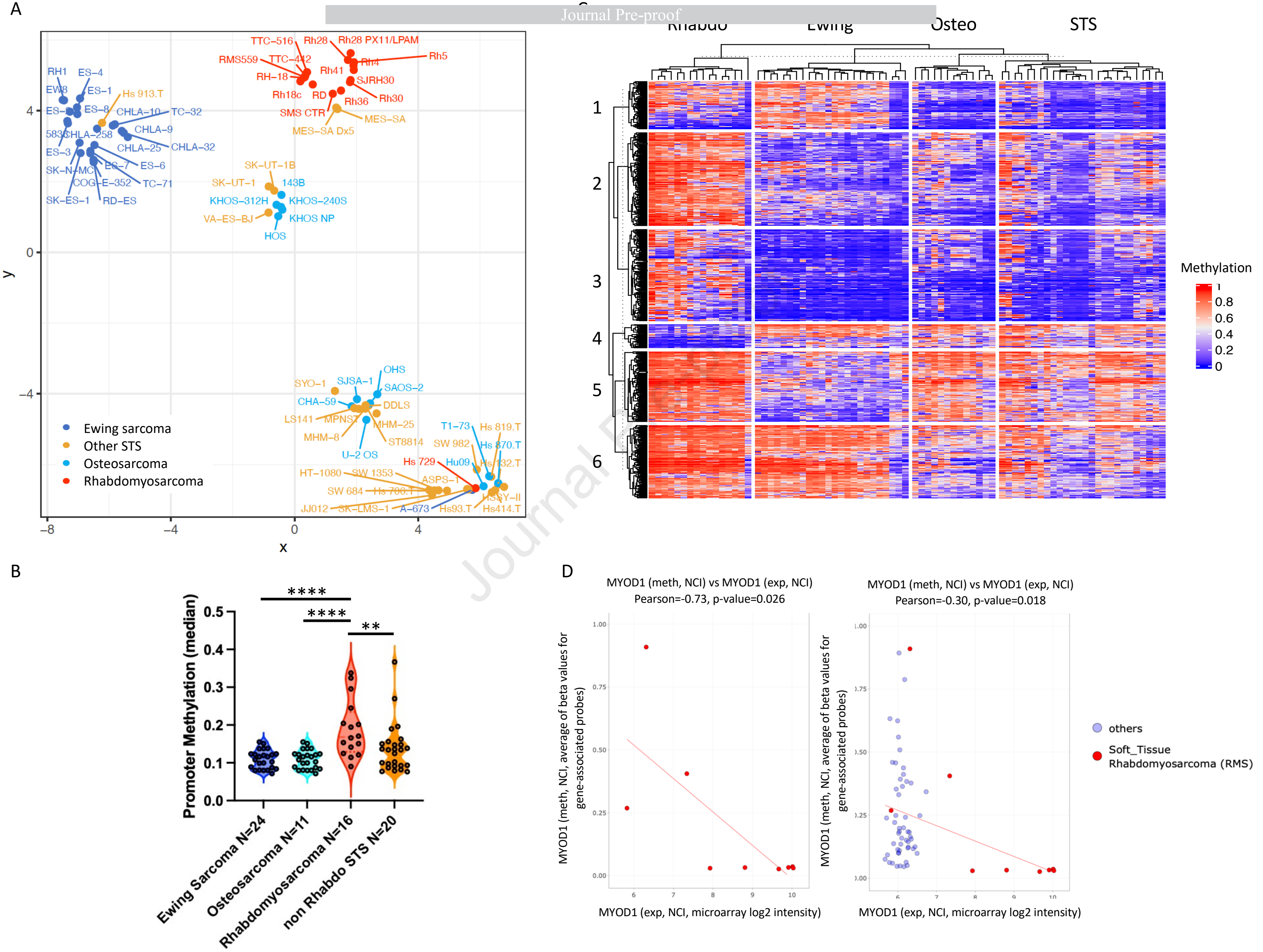
CellMinerCDB enables exploration and analysis of cancer cell line pharmacogenomic data across different sources. If publishing results based on this site, please cite: Rajapakse.VN, Luna.A, Yamade.M et al. iScience, Cell Press. 2018 Dec 12.

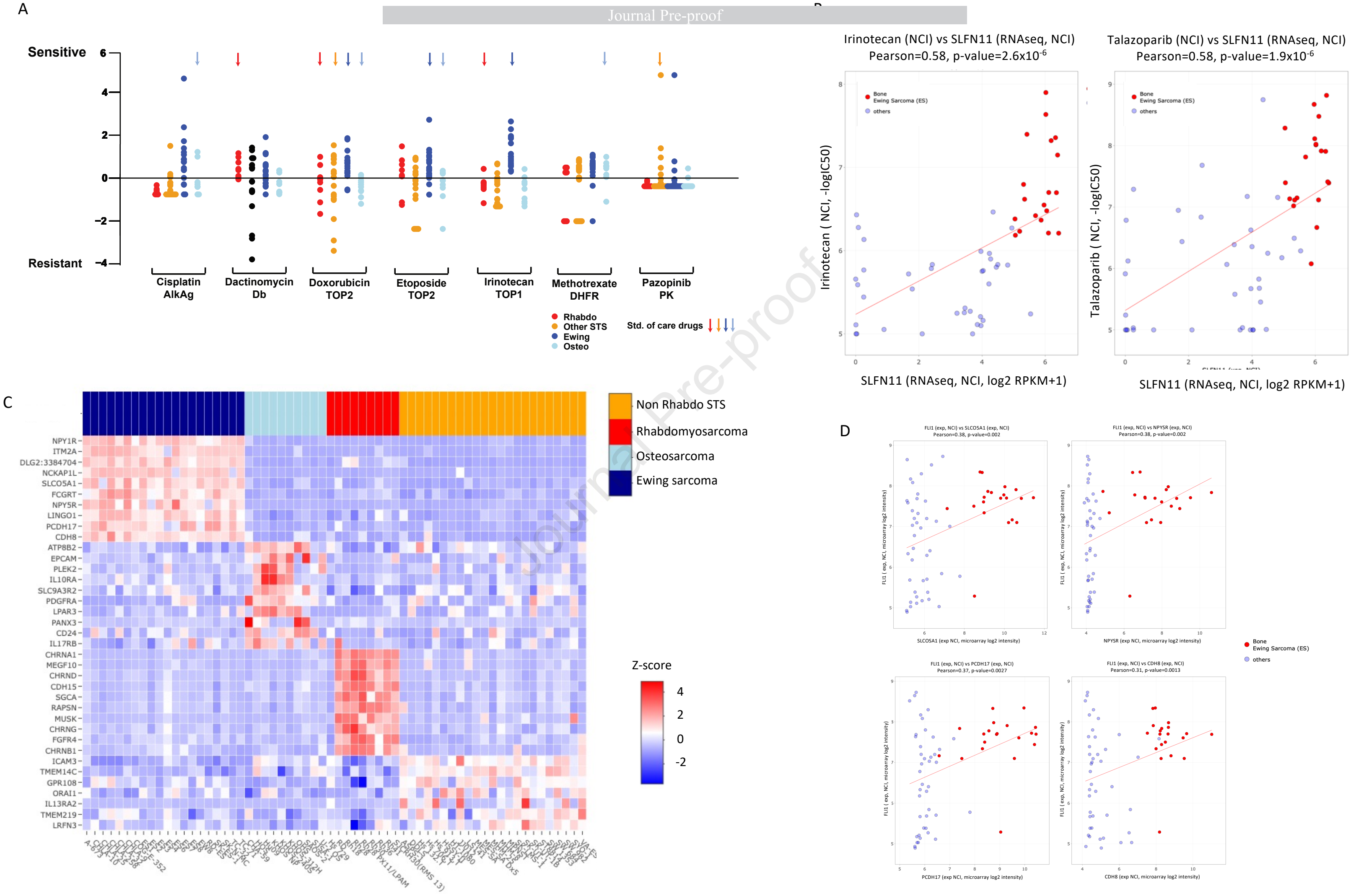
x-Axis Cell Line Set: ACHILLES
x-Axis Data Type: cri: Crispr knockout screens
Identifier: (e.g. topotecan or SLFN11): FOXO1
x-Axis Range: -1 – 0.1
y-Axis Cell Line Set: NCI
y-Axis Data Type: fus: gene fusions
Identifier: (e.g. topotecan or SLFN11): PAX3-FOXO1
y-Axis Range: 0 – 1
Select Tissues: ● To include ○ To exclude
all
☑ Show Color?
no_selection
Bone
Soft_Tissue
fibrosarcoma
leiomyosarcoma
rhabdomyosarcoma
Rhabdomyosarcoma
synovial_sarcoma

Select Cell line to highlight

PAX3-FOXO1 (fus, NCI) vs. FOXO1 (cri, ACHILLES)
Pearson correlation (r)=-0.91, p-value=4.9e-08


A
TMB (Mutations/Megabase)
95
90
45
35
20
15
10
5
0
**
**
****
SW 684
SK-UT-1B
SK-UT-1
CHA-59
ES
OS
RMS
other
B
r = 0.58; p = 3e-07
r = 0.49; p = 9.1e-05
MSH2 (mut, NCI)
100
75
50
25
0
SK-UT-1B
SK-UT-1
leiomyosarcoma
SW 684
fibrosarcoma
0
25
50
75
TMB (mda, NCI)
SK-UT-1
SK-UT-1B
leiomyosarcoma
SW 684
fibrosarcoma
2
3
4
5
6
MSH2 (xsq, NCI)
C
STAG2 mutations
5
0
R216*
ES6
p.L264P
EW8
p.N457Kfs*13
ES3
p.E523X
TC-32
p.Y636Lfs*6
SK-ES-1
p.C735X
COG-E-352
p.C792*
ES7
p.M1176Rfs*10
STAG
0
200
400
600
800
1000
1231aa
Exon
4
5
6
7
8
9
11
12
14
15
16
17
20
25
26
27
28
29
30
31
32
D
STAG2 mutation
n=7
35%
no STAG2
mutation n=13
65%
🟩 STAG2 mutation n=7
🟪 no STAG2 mutation n=13
E
r = 0.47; p = 0.00017
Cabozantinib (act, NCI)
6.5
6.0
5.5
5.0
ES3
ES6
Hu09
EW8
ES7
RMS13
COG-E-352
SK-ES-1
TC-32
Rh30
0
25
50
75
100
STAG2 (mut, NCI)
ES
OS
RMS
other


A
RH1
ES–4
EW8
ES–1
Hs 913.T
ES–5
ES–8
CHLA–10
TC–32
5838
CHLA–258
CHLA–9
CHLA–25
CHLA–32
ES–3
ES–7
ES–6
SK–N–MC
COG–E–352
TC–71
SK–ES–1
RD–ES
TTC–516
Rh28
Rh28 PX11/LPAM
Rh5
RMS559
TTC–442
Rh4
Rh41
RH–18
SJRH30
Rh18c
RD
Rh30
Rh36
SMS CTR
MES–SA
MES–SA Dx5
SK–UT–1B
SK–UT–1
143B
KHOS–312H
KHOS–240S
VA–ES–BJ
KHOS NP
HOS
OHS
SYO–1
SJSA–1
SAOS–2
CHA–59
DDLS
LS141
MPNST
MHM–25
MHM–8
ST8814
U–2 OS
T1–73
Hs 819.T
SW 982
Hs 870.T
Hs 729
Hu09
Hs 132.T
HT–1080
SW 1353
ASPS–1
SW 684
Hs 706.T
HSSY–II
JJ012
SK–LMS–1
A–673
Hs93.T
Hs414.T
Ewing sarcoma
Other STS
Osteosarcoma
Rhabdomyosarcoma
y
x
4
0
–4
–8
B
Promoter Methylation (median)
0.0
0.1
0.2
0.3
0.4
0.5
****
****
**
Ewing Sarcoma N=24
Osteosarcoma N=11
Rhabdomyosarcoma N=16
non Rhabdo STS N=20
Rhabdo
Ewing
Osteo
STS
1
2
3
4
5
6
Methylation
1
0.8
0.6
0.4
0.2
0
D
MYOD1 (meth, NCI) vs MYOD1 (exp, NCI)
Pearson=-0.73, p-value=0.026
MYOD1 (meth, NCI, average of beta values for gene-associated probes)
MYOD1 (exp, NCI, microarray log2 intensity)
1.00
0.75
0.50
0.25
0.00
6
7
8
9
10
MYOD1 (meth, NCI) vs MYOD1 (exp, NCI)
Pearson=-0.30, p-value=0.018
MYOD1 (meth, NCI, average of beta values for gene-associated probes)
MYOD1 (exp, NCI, microarray log2 intensity)
others
Soft_Tissue Rhabdomyosarcoma (RMS)


A
Sensitive
Resistant
6
4
2
0
−2
−4
Cisplatin
AlkAg
Dactinomycin
Db
Doxorubicin
TOP2
Etoposide
TOP2
Irinotecan
TOP1
Methotrexate
DHFR
Pazopinib
PK
Rhabdo
Other STS
Ewing
Osteo
Std. of care drugs
B
Irinotecan (NCI) vs SLFN11 (RNAseq, NCI)
Pearson=0.58, p-value=2.6x10^-6
Talazoparib (NCI) vs SLFN11 (RNAseq, NCI)
Pearson=0.58, p-value=1.9x10^-6
Bone
Ewing Sarcoma (ES)
others
Irinotecan (NCI, -logIC50)
Talazoparib (NCI, -logIC50)
SLFN11 (RNAseq, NCI, log2 RPKM+1)
SLFN11 (RNAseq, NCI, log2 RPKM+1)
C
Non Rhabdo STS
Rhabdomyosarcoma
Osteosarcoma
Ewing sarcoma
NPY1R
ITM2A
DLG2:3384704
NCKAP1L
SLCO5A1
FCGRT
NPY5R
LINGO1
PCDH17
CDH8
ATP8B2
EPCAM
PLEK2
IL10RA
SLC9A3R2
PDGFRA
LPAR3
PANX3
CD24
IL17RB
CHRNA1
MEGF10
CHRND
CDH15
SGCA
RAPSN
MUSK
CHRNG
FGFR4
CHRNB1
ICAM3
TMEM14C
GPR108
ORAI1
IL13RA2
TMEM219
LRFN3
Z-score
4
2
0
-2
D
FLI1 (exp, NCI) vs SLCO5A1 (exp, NCI)
Pearson=0.38, p-value=0.002
FLI1 (exp, NCI) vs NPY5R (exp, NCI)
Pearson=0.38, p-value=0.002
FLI1 (exp, NCI) vs PCDH17 (exp, NCI)
Pearson=0.37, p-value=0.0027
FLI1 (exp, NCI) vs CDH8 (exp, NCI)
Pearson=0.31, p-value=0.0013
FLI1 (exp, NCI, microarray log2 intensity)
SLCO5A1 (exp NCI, microarray log2 intensity)
NPY5R (exp NCI, microarray log2 intensity)
PCDH17 (exp NCI, microarray log2 intensity)
CDH8 (exp NCI, microarray log2 intensity)
Bone
Ewing Sarcoma (ES)
others

Highlights

- Sarcoma_CellMinerCDB merges publicly-available and new sarcoma cell line data
- It includes reproducible genomic and pharmacologic data for 133 sarcoma cell lines
- It is a novel comprehensive resource including the methylome of sarcoma cell lines
- Its multi-functionality can be used to identify new therapeutic targets for sarcoma