



OPEN ACCESS

EDITED BY

Nicoletta Coccaro,
University of Bari Aldo Moro, Italy

REVIEWED BY

Bruce Frederic Smith,
Auburn University, United States
Michael Kinnaman,
Regeneron Pharmaceuticals, Inc., United States

*CORRESPONDENCE

James E. Jacobs,
✉ jacojam@ohsu.edu

RECEIVED 27 September 2023

ACCEPTED 01 March 2024

PUBLISHED 18 March 2024

CITATION

Jacobs JE, Davis L and McWeeney S (2024),
Single nucleotide variants in nuclear pore
complex disassembly pathway associated with
poor survival in osteosarcoma.
Front. Genet. 15:1303404.
doi: 10.3389/fgene.2024.1303404

COPYRIGHT

© 2024 Jacobs, Davis and McWeeney. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in
other forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Single nucleotide variants in nuclear pore complex disassembly pathway associated with poor survival in osteosarcoma

James E. Jacobs*, Lara Davis and Shannon McWeeney

Oregon Health & Science University, Portland, OR, United States

Introduction: The bone tumor, osteosarcoma, remains challenging to treat in children and young adults, especially when patients present with metastatic disease. Developing new therapies based on genomic data from sequencing projects has proven difficult given the lack of recurrent genetic lesions across tumors. MYC overexpression has been associated with poor outcomes in osteosarcoma. However, other genomic markers of disease severity are lacking.

Materials and Methods: We utilized whole genome sequencing of 106 tumors and matched normal controls in order to define genomic characteristics that correlate with overall survival. Single nucleotide variants were overlaid onto annotated molecular pathways in order to define aberrant pathway signatures specific to aggressive osteosarcoma. Additionally, we calculated differential gene expression in a subsample of 71 tumors. Differentially expressed genes were then queried for known MYC-responsive genes.

Results: Molecular pathways specific to nuclear pore complex disassembly (NPCD) show significant correlation with poor overall survival in osteosarcoma when mutations were present. Genes involved in immune response and immune regulation are enriched in the differential expression analysis of samples with and without NPCD pathway aberrations. Furthermore, neither MYC nor MYC-responsive genes show differential expression between NPCD-aberrant and non-aberrant groups. The NPCD pathway mutations are dominated by regulatory region variants rather than protein-altering mutations, suggesting that dysregulation of genetic regulatory networks may be the underlying mechanism for their relation to osteosarcoma phenotype.

Discussion: Overall survival is significantly worse in patients whose tumors show aberrations in the NPCD pathway. Moreover, this difference in survival is not driven by MYC-overexpression, suggesting a novel mechanism for some

Abbreviations: OS, overall survival; SVs, structural variants; CNAs, copy number alterations; SNVs, single nucleotide variants; WGS, whole genome sequencing; RNA-Seq, RNA sequencing; PCGP, Pediatric Cancer Genome Project; G4K, St Jude Clinical Genomics and Genomes for Kids; VEP, Variant Effect Predictor; NE, nuclear envelope; NPCs, nuclear pore complexes; TPR, translocated promoter region protein.

aggressive osteosarcomas. These findings add light to the evolving understanding of the drivers of osteosarcoma and may aid in the search for new treatments based on patient-specific genetic data.

KEYWORDS

osteosarcoma, single nucleotide variants, genomic instability, nuclear pore complex, differential gene expression, MYC-overexpression

Introduction

The bone tumor, osteosarcoma, continues to be a significant burden to the health and wellbeing of children and young adults, with about 500 new cases per year in the United States. Overall survival (OS) at 10 years is approximately 65% for those with localized disease *versus* 25% for the one in five patients that present with metastatic disease (Voute et al., 1999; Bielack et al., 2002; Kager et al., 2003; Ferrari et al., 2005). Survival rates for osteosarcoma—utilizing multi-agent chemotherapy and surgical local control—have not improved significantly over the last 2 decades (Mirabello et al., 2009). Moreover, osteosarcoma survivors experience late effects related to therapy that can include cardiomyopathy, hearing loss, kidney dysfunction, second malignancies, infertility as well as significant physical limitations secondary to surgery (Ward et al., 2014). Thus, research leading to an increased biological understanding of this devastating tumor and development of new treatment options represents a vital need in the pediatric population and a potentially high-impact addition to our knowledge of sarcoma biology.

Studying the genomics of osteosarcoma has proven challenging, as individual tumors show tremendous heterogeneity characterized by structural variants (SVs) and copy number alterations (CNAs) (Chen et al., 2014; Perry et al., 2014). This complex genomic landscape has made it difficult to identify common mutations across patients. However, some consistent findings have emerged. Large scale sequencing studies have revealed recurrent alterations in the tumor suppressor genes, TP53 and RB1, as well as in known cancer-associated genes such as ATRX and DLG2 (Chen et al., 2014). Furthermore, common cancer-associated pathways such as the phosphatidylinositol 3-kinase/mammalian target of rapamycin (PI3K/mTOR) pathway and insulin-like growth factor (IGF) signaling have also been found to be abnormal in osteosarcoma (Perry et al., 2014; Behjati et al., 2017). While these alterations are dominated by SVs, it should be noted that germline and somatic single nucleotide variants (SNVs) have also been identified in these genes (Martin et al., 2012; Chen et al., 2014). Another consistent finding in osteosarcoma is the relationship between MYC overexpression and tumor progression leading to poor overall survival (Gamberi et al., 1998; Scionti et al., 2008; Shimizu et al., 2010; Chen et al., 2018).

Genomic instability, leading to SVs and CNAs, has been implicated as the driving force in the development of osteosarcoma (Chen et al., 2014) and specific types of genomic instability have been described in many osteosarcomas. These include the localized hypermutable regions associated with SVs, known as kataegis (Nik-Zainal et al., 2012; Chen et al., 2014), as well as massive genomic rearrangement events called chromothripsis (Behjati et al., 2017; Cortes-Ciriano et al., 2020). While these SVs and CNAs are not consistent or universal across all osteosarcomas, targeting of copy number variable genes, matched to a patient's particular genomic profile has shown to be effective in patient-

derived xenograft models (Sayles et al., 2019). Thus, SVs and CNAs appear to be the key factors in the oncogenesis of osteosarcoma. However, what leads to this underlying genomic instability and how this relates to other consistent genomic and clinical characteristics of osteosarcoma remains to be discovered. SNVs are often thought of as byproducts of the genomic instability of osteosarcoma and are believed to be passenger events (Chen et al., 2014; Bousquet et al., 2016). However, it is possible that specific patterns of mutation exist that may contribute to the severity of these tumors. Here we show that by analyzing SNVs overlaid onto annotated molecular pathways across multiple patient samples, we can identify patterns in the underlying genomics of osteosarcoma that illuminate mutations in genes that correlate with previously described genomic and clinical characteristics of osteosarcoma.

Materials and methods

Patient enrollment and sample collection

Whole genome sequencing (WGS) and RNA sequencing (RNA-Seq) datasets were downloaded from the St Jude Cloud (McLeod et al., 2021). All sequencing data were previously obtained as part of the Pediatric Cancer Genome Project (PCGP) and the St Jude Clinical Genomics and Genomes for Kids (G4K) Clinical Pilot Project (Rusch et al., 2018). Patient consent and sample collection were conducted as part of the PCGP and G4K Clinical Pilot Project (Rusch et al., 2018). Please refer to original study for ethics committee info. We restricted our WGS analysis to human subjects in which both the tumor and matched control sequences were available. We restricted our RNA-Seq analysis to human subjects in which a matching WGS dataset was available.

Sequence analysis and molecular pathway annotation

WGS data were analyzed using the GATK Best Practices (Broad Institute, 2020). The downloaded files were first reverted back to an unaligned format (Broad Institute, 2019). Sequencing reads were then aligned to the most recent version of the human genome (GRCh38/hg38) which incorporates population-based contigs (Li and Durbin, 2010; Broad Institute, 2017a). The aligned reads were then separated by read group and duplicate reads were marked (Cibulskis et al., 2013; Broad Institute, 2015b). To decrease erroneous mutation calling, systemically inaccurate base quality scores were recalibrated (Broad Institute, 2015a; Broad Institute, 2017b). MuTect2 was then used to identify mutations in the tumor samples relative to the matched normal samples (Cibulskis et al., 2013; Broad Institute, 2016). The Ensemble

Variant Effect Predictor (VEP) (McLaren et al., 2016) was then used to classify the called somatic mutations based on their predicted effects. The final mutation set was limited to those predicted to cause 3' or 5' untranslated region variants, splice site variants, transcription-factor binding site variants, protein altering variants, stop loss or gain variants, coding sequence variants, start loss variants, frameshift variants or regulatory region variants.

Following VEP filtering, mutations were mapped to curated molecular pathways using Reactome. Covering over 10,000 human proteins, 13,000 biochemical reactions and 2,400 pathways, Reactome is the most comprehensive open source, manually curated biological pathway database (Milacic et al., 2012; Fabregat et al., 2016). After selecting for pathways in which the aberrant group and the non-aberrant group were represented by at least 10 samples each, we searched for pathways that distinguished patients based on overall survival.

Survival analysis

Of the 82 patients, 81 had available clinical data for time-to-event analyses. Of those, 79 had corresponding WGS data. Overall survival was evaluated in these 79 patients. Kaplan-Meier curves were generated with R (R.C. Team, 2020; Kassambara A and Biecek, 2021; Therneau, 2020) between samples with and without aberrant NPCD. Other cancer types were also analyzed for differences in overall survival due to aberrant NPCD. Mutation and survival data for individual cancer types were obtained from the cBioPortal for Cancer Genomics (Cerami et al., 2012; Gao et al., 2013). These data were then mapped to molecular pathways using Reactome (Milacic et al., 2012; Fabregat et al., 2016). Kaplan-Meier curves were then generated between samples with and without aberrant pathways.

Gene differential expression analysis

Expression data (RNA-Seq) were available for a subset of 71 osteosarcoma samples. These samples were analyzed as follows: First, the downloaded RNA-Seq files were reverted back to an unaligned format (Broad Institute, 2019) and aligned to the most recent version of the human genome (GRCh38/hg38) using the STAR Aligner with the “quantMode TranscriptomeSAM GeneCounts” option (Dobin et al., 2013). Gene expression values for each sample were then estimated using RSEM (Li and Dewey, 2011). Finally differential gene expression between samples with and without mutations in NPCD was calculated in R using the DESeq2 package (Love et al., 2014). To account for multiple hypotheses testing, only genes with a false discovery rate of ≤ 0.05 were considered to be significantly differentially expressed.

Results

Mutation calling from whole genome sequencing reads

We analyzed WGS datasets from 81 unique patients, representing 106 total sequenced biopsies. Contamination of

tumor samples was calculated at 0.0024 ± 0.00026 (Ave \pm std). Somatic mutations in the tumor samples were identified relative to a matched normal sample from the same patient. The called mutations were then limited to those predicted to cause coding sequence variants (including 5' and 3' untranslated regions), and variants in splice sites or regulatory regions. The full list of called mutations can be found in [Supplementary Material S1](#). The per-sample mutational burden is shown in [Figure 1](#). The number of mutations per sample showed a wide variation with a mean and standard deviation of 605 ± 278 for untreated samples and $1,540 \pm 987$ for recurrent samples. The vast majority of mutations were single nucleotide variants (SNVs). However, small insertions (INS), deletions (DEL) and substitutions (SUB) were also noted. This analysis did not evaluate larger structural variants (SVs) such as copy number alterations (CNAs).

Molecular pathway analysis

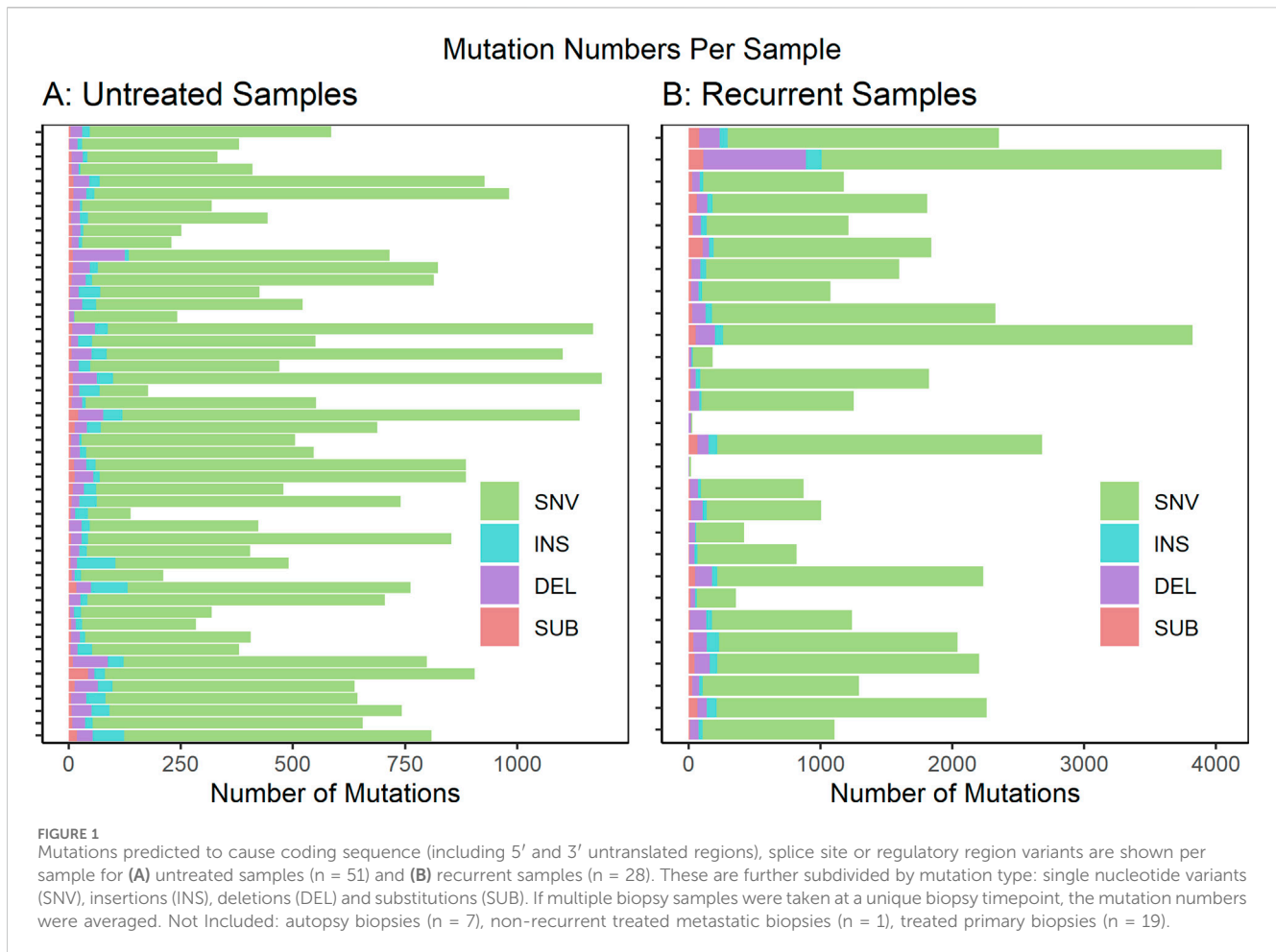
Mutations were mapped to annotated molecular pathways and aberrant pathways were then further analyzed in the context of clinical outcomes. Seventy-nine of the patients with mutation data also had available clinical data for time-to-event analyses (clinical data shown in [Table 1](#)). We found that patients with mutations in the nuclear pore complex disassembly (NPCD) pathway (Reactome ID: R-HSA-3301854) had significantly poorer overall survival than those without mutations (FDR = 0.03). The survival difference between groups with and without aberrant NPCD is shown in [Figure 2](#).

A full list of called mutations associated with the NPCD pathway can be found in [Supplementary Material S2](#). No differences were noted in the variant allele frequency (VAF) between NPCD pathway mutations in untreated primary biopsy samples (0.242 ± 0.160) vs samples from recurrent lesions (0.251 ± 0.171).

As aberrations in NPCD could represent a general signature for aggressive malignancies rather than an osteosarcoma-specific signature, we analyzed overall survival in various cancer types with regards to the presence of mutations in this pathway. No differences were noted in overall survival in breast invasive carcinoma, lung squamous cell carcinoma, prostate adenocarcinoma or glioblastoma multiforme with respect to aberrations in NPCD ([Supplementary Figures S1–S4](#)) (Hoadley et al., 2018). Adult soft tissue sarcomas and Ewing sarcoma were also analyzed and again, no difference in overall survival was noted ([Supplementary Figures S5, S6](#)) (Edsc Cancer Genome Atlas Research Network Cancer Genome Atlas Research Network, 2017). Furthermore, the sarcoma samples had very few aberrations in the NPCD pathway (Tirode et al., 2014). These findings suggest that the survival difference noted in osteosarcoma patients with aberrations in NPCD is specific to the disease process of osteosarcoma.

Gene expression correlation

Gene expression data were available for 72 of the osteosarcoma samples. Of these, 37 showed aberrations in the NPCD pathway genes. Differential expression analysis of the aberrant vs non-



aberrant groups was performed using DESeq2 (Love et al., 2014). Interestingly, neither MYC nor known MYC-responsive genes were found to be differentially expressed secondary to NPCD pathway aberrations (Figure 3). This is in contrast to MYC overexpression which was seen to be associated with poor survival (Supplementary Figure S7) as has been demonstrated in previous studies (Chen et al., 2018). These findings suggest a possible alternate mechanism for the development of aggressive osteosarcomas. Similarly, none of the NPCD pathway genes were differentially expressed between NPCD-aberrant and non-aberrant samples. A full list of differentially expressed genes can be found in Supplementary Material S3. GO-terms analysis (Eden et al., 2007; Eden et al., 2009) showed that of the 10 most significantly enriched GO-terms processes, eight relate to immune function. Notably, these genes largely showed decreased expression in the aberrant vs. the non-aberrant samples (Figure 3). A full list of GO-terms processes enriched in the differentially expressed genes can be found in Supplementary Material S4.

Patterns in osteosarcoma nuclear pore complex mutations

Many of the genes responsible for the regulation of nuclear pore complex disassembly were found to be mutated in this

cohort of osteosarcoma samples (Laurell et al., 2011). These mutations across samples showed a largely non-overlapping pattern as shown in Figure 4. Perhaps surprisingly, the majority of the NPCD-related mutations occurred in genetic regulatory regions rather than within the coding region of the gene itself.

Availability of data and materials

The WGS and RNA-Seq datasets analyzed for this study are available from the St Jude Cloud: <https://www.stjude.cloud/>

All code used in this project can be found in the following repositories:

- WGS Workflow: <https://github.com/jejacobs23/osteo>
- RNA-Seq Workflow: <https://github.com/jejacobs23/osteo-RNA-Seq-Workflow>

Discussion

Here we show that across a cohort of osteosarcoma patients, mutations in genes responsible for the structure and function of nuclear pore complex disassembly (NPCD) portend a poorer

TABLE 1 Clinical characteristics of the 79 patients with available clinical data. No statistically significant differences were detected between groups with and without NPCD aberrations with regards to patient age at diagnosis, sex, or race. A statistically significant difference was noted in duration of follow up between aberrant and non-aberrant groups as would be expected. The various histological subtypes lacked adequate numbers of samples for statistically meaningful comparison. A statistically significant difference was noted in the incidence of metastatic disease at the time of diagnosis between aberrant and non-aberrant groups and the aberrant group had more post-treatment samples available for analysis.

		Total (n = 79)	NPCD aberrant (n = 42)	NPCD non-aberrant (n = 37)	p-value
Age at diagnosis (years; Ave ± std)		13.6 ± 3.9	13.5 ± 4.5	13.0 ± 3.9	0.54 ^a
Sex	Female	36	19	17	1**
	Male	43	23	20	
Race	Black	17	11	6	0.60***
	White	54	27	27	
	Other	8	4	4	
Tumor Histology	Conventional: NOS	16	7	9	NA
	Conventional: Osteoblastic	15	9	6	
	Conventional: Chondroblastic	4	2	2	
	Conventional: Fibroblastic	0	0	0	
	Telangiectatic	3	3	0	
	Small Cell	0	0	0	
	Mixed	3	1	2	
	Not Available	38	20	18	
Sample Type	Primary (untreated)	42	20	22	0.024***
	Post-Treatment	55	41	14	
	Metastatic (untreated)	7	4	3	
Metastatic Disease Present at Diagnosis		35	24	11 [†]	0.043**
Duration of Follow Up (Months; Ave ± std)		81.0 ± 64.0	64.4 ± 58.6	103.0 ± 65.4	0.008 ^a

^aTwo-Sample *t*-test, **Chi-squared test with Yates' continuity correction, ***Fisher's Exact Test, [†]Two non-aberrant subjects had unknown metastatic disease status at diagnosis, NOS: Not Otherwise Specified—Conventional osteosarcoma without subclassification into osteoblastic, chondroblastic or fibroblastic subtypes.

prognosis in terms of overall survival when compared to patients without such mutations. The nuclear envelope is a crucial cellular structure in eukaryotes and is involved in a multitude of cellular processes such as the coordinated transport between the nucleus and cytoplasm, chromatin organization and regulation of gene expression (Guttinger et al., 2009). Moreover, the disassembly and reassembly of the nuclear envelope—including the nuclear pore complexes (NPCs)—is a key process in the transition between G2 and M phases of the cell cycle (Elder et al., 2000; Li et al., 2005).

Abnormalities in the NPCs have been associated with a variety of diseases including multiple types of cancer. NUP98—and to a lesser extent, NUP214—have been found to be part of various fusion events involved in the development of several different myeloid and lymphoid leukemias (Cronshaw and Matunis, 2004). Additionally, fusions involving the translocated promoter region protein (TPR) gene have been implicated in various sarcomas including osteosarcoma (Park et al., 1986; Vriens et al., 2009). Exactly how these abnormal NPC gene products are associated with malignancy is not fully understood. However, defects in the NPCs' normal role in mitotic checkpoint regulation is an intriguing possibility

(Cronshaw and Matunis, 2004). Our results highlight a downregulation of the expression of many immune-related genes in patients with NPCD aberrations. This would suggest that immune dysregulation may be a feature of more aggressive osteosarcomas, as supported by recent publications (Tang et al., 2023; Wu et al., 2023). Furthermore, the lack of protein altering mutations in the coding regions of the NPCD genes in our study suggests that dysregulation of these gene networks—rather than direct gene product dysfunction—may be responsible for the oncogenic phenotype associated with NPCD mutations in osteosarcoma.

Overexpression of the proto-oncogene, MYC, is a well-described characteristic of aggressive osteosarcomas (Chen et al., 2018). Normally, MYC is involved in the transcriptional regulation of many downstream genes. When overexpressed, this intricate cascading network becomes disrupted which can result in oncogenesis. Here we show that the differential expression of MYC and MYC-responsive genes is not present when comparing NPCD-aberrant and non-aberrant groups. These findings may point to an alternate mechanism for the development of aggressive osteosarcomas and NPCD aberrations could prove to be a valuable prognostic marker for osteosarcoma. Further studies will

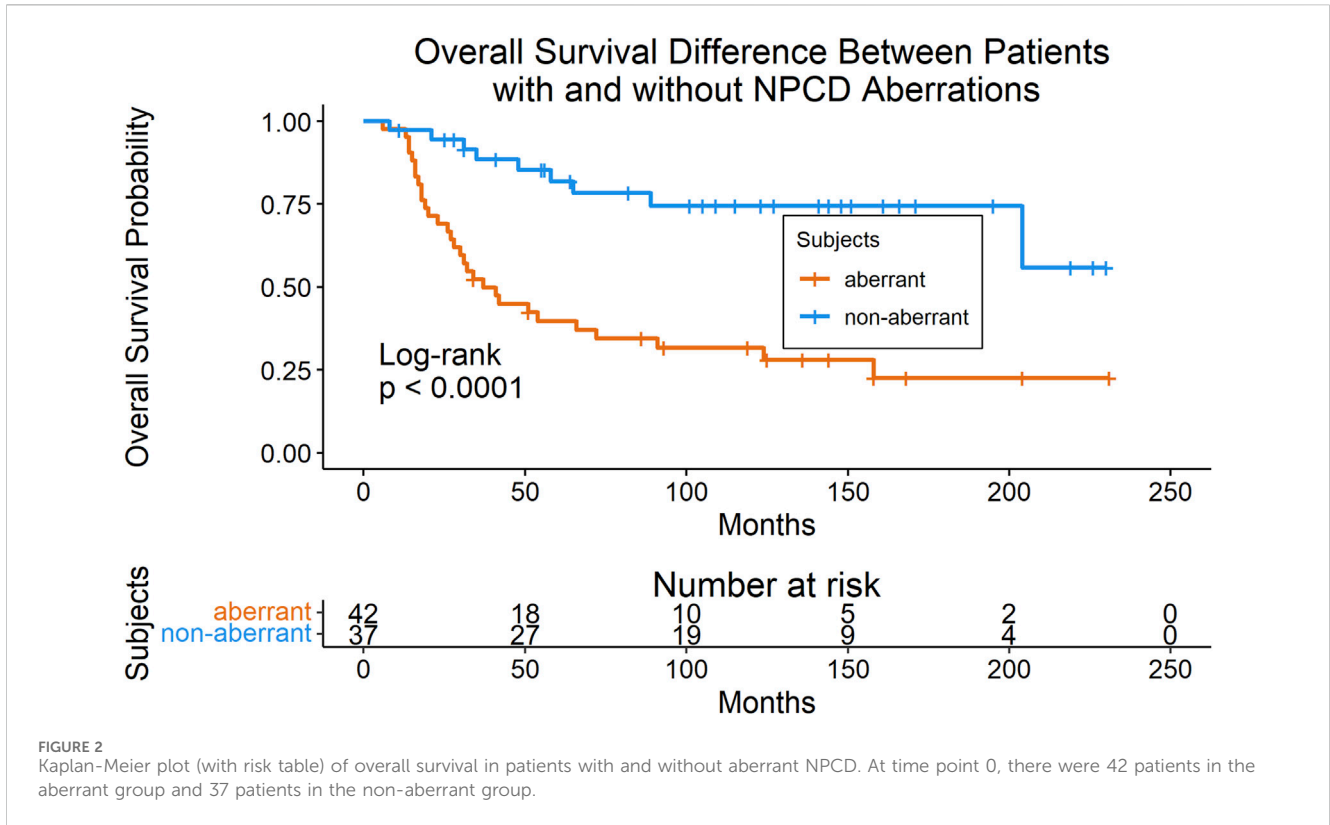


FIGURE 2 Kaplan-Meier plot (with risk table) of overall survival in patients with and without aberrant NPCD. At time point 0, there were 42 patients in the aberrant group and 37 patients in the non-aberrant group.

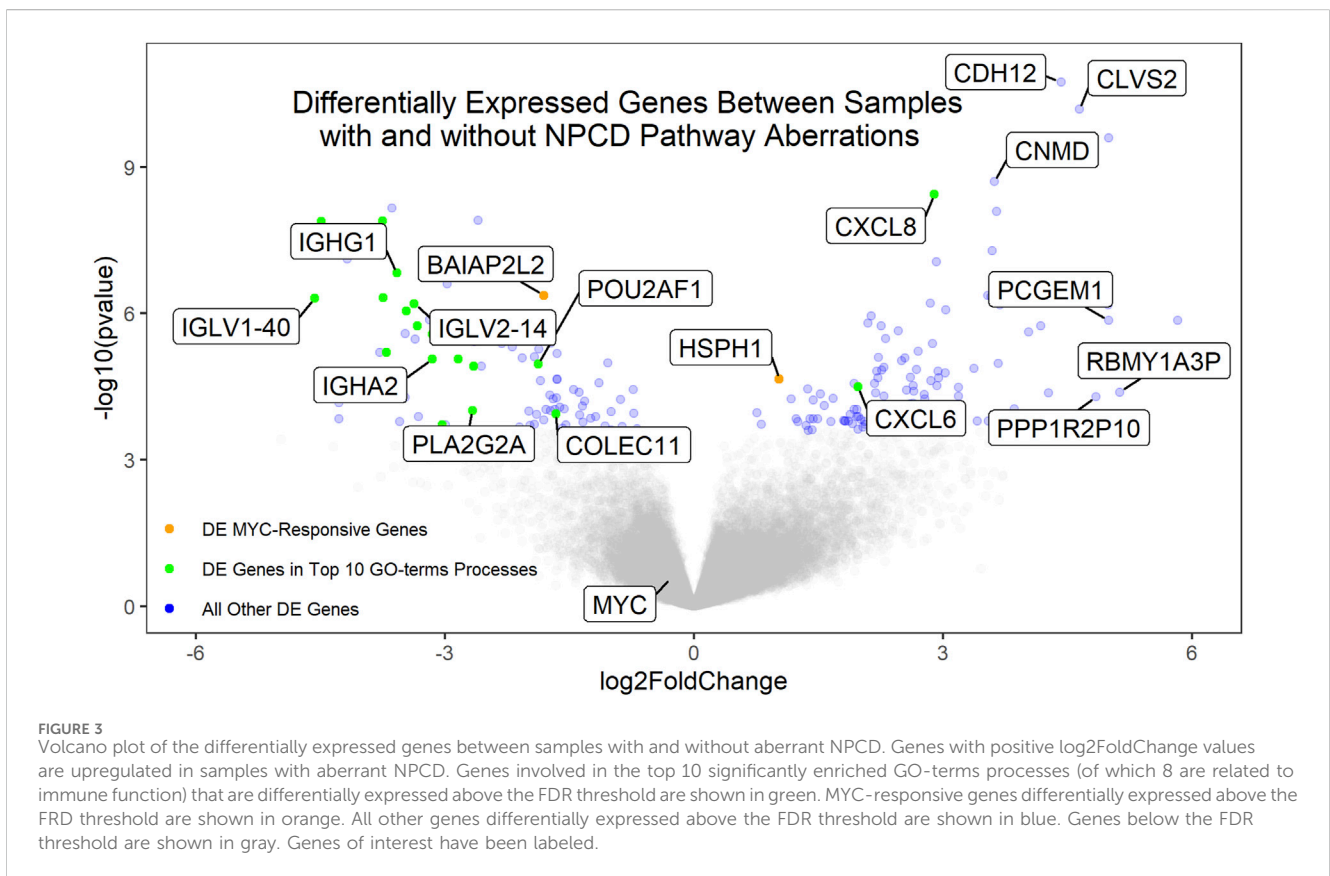
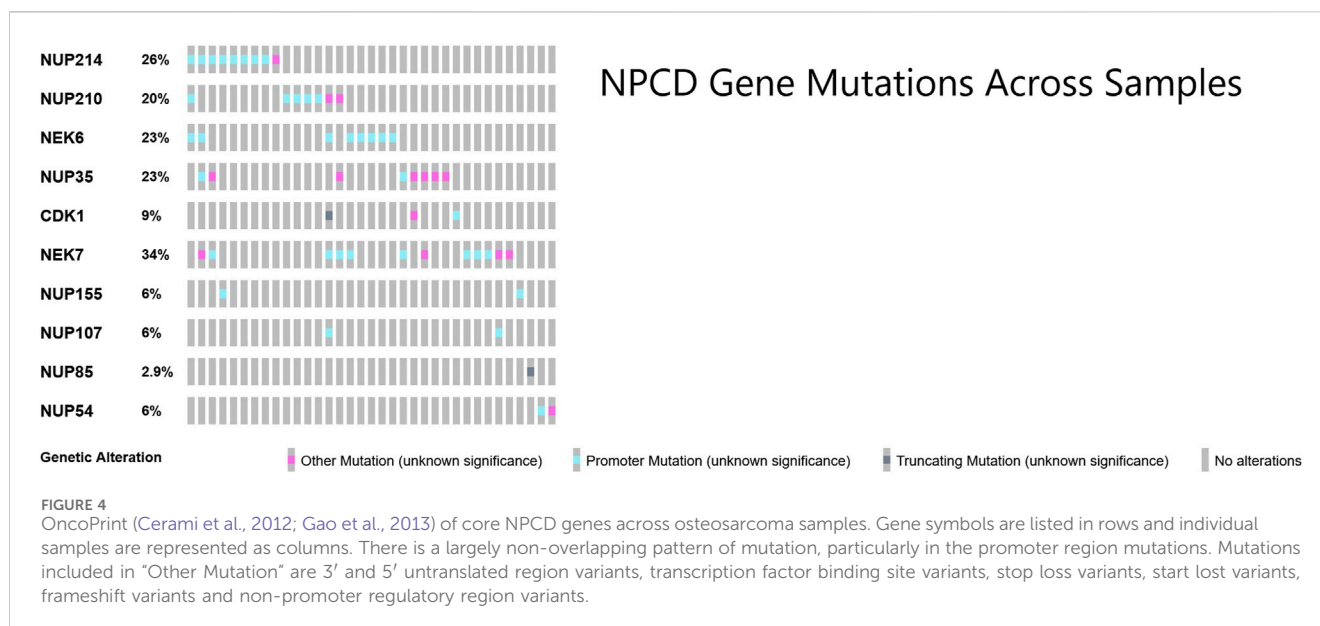


FIGURE 3 Volcano plot of the differentially expressed genes between samples with and without aberrant NPCD. Genes with positive log₂FoldChange values are upregulated in samples with aberrant NPCD. Genes involved in the top 10 significantly enriched GO-terms processes (of which 8 are related to immune function) that are differentially expressed above the FDR threshold are shown in green. MYC-responsive genes differentially expressed above the FRD threshold are shown in orange. All other genes differentially expressed above the FDR threshold are shown in blue. Genes below the FDR threshold are shown in grey. Genes of interest have been labeled.



be required to validate these findings and determine if NPCD pathway aberrations are a suitable target for therapy.

While the difference in survival between groups of patients with and without mutations related to the NPCD genes is statistically significant, the inference of why these mutations affect prognosis is entirely speculative and cannot be resolved through purely computational analyses. Our study highlights the benefits of hypothesis-driven computational analyses guiding further bench experimentation as a way to streamline the process of discovery in the highly complex area of cancer genomics.

Acknowledgement of genetic data sources: This study makes use of data generated by the St Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project and/or Childhood Solid Tumor Network as well as the Clinical Genomics and Genomes for Kids (G4K) Clinical Pilot.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for this study because all consent for use of sequencing data was obtained at the time of tissue collection by the original study (see text for details). The studies were conducted in accordance with the local legislation and institutional requirements. We did not use any human samples in this study, but rather utilized sequencing data that had previously been obtained. Written informed consent to participate in this study was not required from the

participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

JJ: Writing–original draft, Formal Analysis, Conceptualization. LD: Conceptualization, Writing–review and editing. SMcW: Conceptualization, Writing–review and editing, Supervision.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We would like to thank the patients and their families for entrusting the PCGP and G4K to use their genetic information for the betterment of other children and young adults. We also would like to thank Kirby Birch and Clay McLeod at the St Jude Cloud as well as Colin Bailey of the St Jude Biorepository. Finally, we would like to thank Guanming Wu, PhD for his guidance and mentorship regarding the Reactome Pathway Database.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the

reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2024.1303404/full#supplementary-material>

References

- Behjati, S., Tarpey, P. S., Haase, K., Ye, H., Young, M. D., Alexandrov, L. B., et al. (2017). Recurrent mutation of IGF signalling genes and distinct patterns of genomic rearrangement in osteosarcoma. *Nat. Commun.* 8, 15936. doi:10.1038/ncomms15936
- Bielack, S. S., Kempf-Bielack, B., Delling, G., Exner, G. U., Flege, S., Helmke, K., et al. (2002). Prognostic factors in high-grade osteosarcoma of the extremities or trunk: an analysis of 1,702 patients treated on neoadjuvant cooperative osteosarcoma study group protocols. *J. Clin. Oncol. official J. Am. Soc. Clin. Oncol.* 20, 776–790. doi:10.1200/JCO.2002.20.3.776
- Bousquet, M., Noirot, C., Accabbled, F., Sales de Gauzy, J., Castex, M. P., Brousset, P., et al. (2016). Whole-exome sequencing in osteosarcoma reveals important heterogeneity of genetic alterations. *Ann. Oncol. official J. Eur. Soc. Med. Oncol.* 27, 738–744. doi:10.1093/annonc/mdw009
- Broad Institute (2015a). *Base quality score recalibration, GATK talks, Broad Institute.*
- Broad Institute (2015b). *Mapping, processing, and duplicate marking with Picard tools, GATK Talks.* Cambridge, MA: Broad Institute.
- Broad Institute (2016). *MuTect2: call somatic SNPs and indels via local re-assembly of haplotypes, GATK Documentation.* Cambridge, MA: Broad Institute.
- Broad Institute (2017a). *(How to) Map reads to a reference with alternate contigs like GRCh38, GATK Forum.* Cambridge, MA: Broad Institute.
- Broad Institute (2017b). *(howto) Recalibrate base quality scores = run BQSR, GATK Tutorials.* Cambridge, MA: Broad Institute.
- Broad Institute (2019). *How to Map and clean up short read sequence data efficiently GATK Documentation.* Cambridge, MA: Broad Institute.
- Broad Institute (2020). *Genome analysis toolkit.* Cambridge, MA: Broad Institute. 20201.
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2, 401–404. doi:10.1158/2159-8290.CD-12-0095
- Chen, D., Zhao, Z., Huang, Z., Chen, D. C., Zhu, X. X., Wang, Y. Z., et al. (2018). Super enhancer inhibitors suppress MYC driven transcriptional amplification and tumor progression in osteosarcoma. *Bone Res.* 6, 11. doi:10.1038/s41413-018-0009-8
- Chen, X., Bahrami, A., Pappo, A., Easton, J., Dalton, J., Hedlund, E., et al. (2014). Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell Rep.* 7, 104–112. doi:10.1016/j.celrep.2014.03.003
- Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., et al. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 213–219. doi:10.1038/nbt.2514
- Cortes-Ciriano, I., Lee, J. J., Xi, R., Jain, D., Jung, Y. L., Yang, L., et al. (2020). Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat. Genet.* 52, 331–341. doi:10.1038/s41588-019-0576-7
- Cronshaw, J. M., and Matunis, M. J. (2004). The nuclear pore complex: disease associations and functional correlations. *Trends Endocrinol. Metab.* 15, 34–39. doi:10.1016/j.tem.2003.11.005
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi:10.1093/bioinformatics/bts635
- Eden, E., Lipson, D., Yogev, S., and Yakhini, Z. (2007). Discovering motifs in ranked lists of DNA sequences. *PLoS Comput. Biol.* 3, e39. doi:10.1371/journal.pcbi.0030039
- Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinforma.* 10, 48. doi:10.1186/1471-2105-10-48
- Edsc Cancer Genome Atlas Research NetworkCancer Genome Atlas Research Network (2017). Comprehensive and integrated genomic characterization of adult soft tissue sarcomas. *Cell* 171, 950–965.e28. e28. doi:10.1016/j.cell.2017.10.014
- Elder, R. T., Yu, M., Chen, M., Edelson, S., and Zhao, Y. (2000). Cell cycle G2 arrest induced by HIV-1 Vpr in fission yeast (*Schizosaccharomyces pombe*) is independent of cell death and early genes in the DNA damage checkpoint. *Virus Res.* 68, 161–173. doi:10.1016/s0168-1702(00)00167-2
- Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., et al. (2016). The reactome pathway knowledgebase. *Nucleic acids Res.* 44, D481–D487. doi:10.1093/nar/gkv1351
- Ferrari, S., Smeland, S., Mercuri, M., Bertoni, F., Longhi, A., Ruggieri, P., et al. (2005). Neoadjuvant chemotherapy with high-dose Ifosfamide, high-dose methotrexate, cisplatin, and doxorubicin for patients with localized osteosarcoma of the extremity: a joint study by the Italian and Scandinavian Sarcoma Groups. *J. Clin. Oncol. official J. Am. Soc. Clin. Oncol.* 23, 8845–8852. doi:10.1200/JCO.2004.00.5785
- Gamberi, G., Benassi, M. S., Bohling, T., Ragazzini, P., Molendini, L., Sollazzo, M. R., et al. (1998). C-myc and c-fos in human osteosarcoma: prognostic value of mRNA and protein expression. *Oncology* 55, 556–563. doi:10.1159/000011912
- Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., et al. (2013). Integrative cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal* 6, 11. doi:10.1126/scisignal.2004088
- Guttinger, S., Laurell, E., and Kutay, U. (2009). Orchestrating nuclear envelope disassembly and reassembly during mitosis. *Nat. Rev. Mol. Cell Biol.* 10, 178–191. doi:10.1038/nrm2641
- Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., et al. (2018). Cell-of-Origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* 173, 291–304 e6. doi:10.1016/j.cell.2018.03.022
- Kager, L., Zoubek, A., Potschger, U., Kastner, U., Flege, S., Kempf-Bielack, B., et al. (2003). Primary metastatic osteosarcoma: presentation and outcome of patients treated on neoadjuvant Cooperative Osteosarcoma Study Group protocols. *J. Clin. Oncol. official J. Am. Soc. Clin. Oncol.* 21, 2011–2018. doi:10.1200/JCO.2003.08.132
- Kassambara A, K. M., and Biecek, P. (2021). *survminer: drawing Survival Curves using "ggplot2".*
- Laurell, E., Beck, K., Krupina, K., Theerthagiri, G., Bodenmiller, B., Horvath, P., et al. (2011). Phosphorylation of Nup98 by multiple kinases is crucial for NPC disassembly during mitotic entry. *Cell* 144, 539–550. doi:10.1016/j.cell.2011.01.012
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinforma.* 12, 323. doi:10.1186/1471-2105-12-323
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi:10.1093/bioinformatics/btp698
- Li, L., Li, H. S., Pauza, C. D., Bukrinsky, M., and Zhao, R. Y. (2005). Roles of HIV-1 auxiliary proteins in viral pathogenesis and host-pathogen interactions. *Cell Res.* 15, 923–934. doi:10.1038/sj.cr.7290370
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8
- Martin, J. W., Squire, J. A., and Zielenska, M. (2012). The genetics of osteosarcoma. *Sarcoma* 2012, 627254. doi:10.1155/2012/627254
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., et al. (2016). The ensembl variant effect predictor. *Genome Biol.* 17, 122. doi:10.1186/s13059-016-0974-4
- McLeod, C., Gout, A. M., Zhou, X., Thrasher, A., Rahbarinia, D., Brady, S. W., et al. (2021). St. Jude Cloud: a pediatric cancer genomic data-sharing ecosystem. *Cancer Discov.* 11, 1082–1099. doi:10.1158/2159-8290.CD-20-1230
- Milacic, M., Haw, R., Rothfels, K., Wu, G., Croft, D., Hermjakob, H., et al. (2012). Annotating cancer variants and anti-cancer therapeutics in reactome. *Cancers* 4, 1180–1211. doi:10.3390/cancers4041180

- Mirabello, L., Troisi, R. J., and Savage, S. A. (2009). Osteosarcoma incidence and survival rates from 1973 to 2004: data from the surveillance, Epidemiology, and end results Program. *Cancer* 115, 1531–1543. doi:10.1002/cncr.24121
- Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., et al. (2012). Mutational processes molding the genomes of 21 breast cancers. *Cell* 149, 979–993. doi:10.1016/j.cell.2012.04.024
- Park, M., Dean, M., Cooper, C. S., Schmidt, M., O'Brien, S. J., Blair, D. G., et al. (1986). Mechanism of met oncogene activation. *Cell* 45, 895–904. doi:10.1016/0092-8674(86)90564-7
- Perry, J. A., Kiezun, A., Tonzi, P., Van Allen, E. M., Carter, S. L., Baca, S. C., et al. (2014). Complementary genomic approaches highlight the PI3K/mTOR pathway as a common vulnerability in osteosarcoma. *Proc. Natl. Acad. Sci. U. S. A.* 111, E5564–E5573. doi:10.1073/pnas.1419260111
- R.C. Team (2020). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rusch, M., Nakitandwe, J., Shurtleff, S., Newman, S., Zhang, Z., Edmonson, M. N., et al. (2018). Clinical cancer genomic profiling by three-platform sequencing of whole genome, whole exome and transcriptome. *Nat. Commun.* 9, 3962. doi:10.1038/s41467-018-06485-7
- Sayles, L. C., Breese, M. R., Koehne, A. L., Leung, S. G., Lee, A. G., Liu, H. Y., et al. (2019). Genome-informed targeted therapy for osteosarcoma. *Cancer Discov.* 9, 46–63. doi:10.1158/2159-8290.CD-17-1152
- Scionti, I., Michelacci, F., Pasello, M., Hattinger, C. M., Alberghini, M., Manara, M. C., et al. (2008). Clinical impact of the methotrexate resistance-associated genes C-MYC and dihydrofolate reductase (DHFR) in high-grade osteosarcoma. *Ann. Oncol. official J. Eur. Soc. Med. Oncol.* 19, 1500–1508. doi:10.1093/annonc/mdn148
- Shimizu, T., Ishikawa, T., Sugihara, E., Kuninaka, S., Miyamoto, T., Mabuchi, Y., et al. (2010). c-MYC overexpression with loss of Ink4a/Arf transforms bone marrow stromal cells into osteosarcoma accompanied by loss of adipogenesis. *Oncogene* 29, 5687–5699. doi:10.1038/onc.2010.312
- Tang, H., Liu, S., Luo, X., Sun, Y., Li, X., Luo, K., et al. (2023). A novel molecular signature for predicting prognosis and immunotherapy response in osteosarcoma based on tumor-infiltrating cell marker genes. *Front. Immunol.* 14, 1150588. doi:10.3389/fimmu.2023.1150588
- Therneau, T. (2020). *A package for survival analysis in R*.
- Tirode, F., Surdez, D., Ma, X., Parker, M., Le Deley, M. C., Bahrami, A., 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. doi:10.1158/2159-8290.cd-14-0622
- Voute, P. A., Souhami, R. L., Nooij, M., Somers, R., Cortes-Funes, H., van der Eijken, J. W., et al. (1999). A phase II study of cisplatin, ifosfamide and doxorubicin in operable primary, axial skeletal and metastatic osteosarcoma. European Osteosarcoma Intergroup (EOI). *Ann. Oncol. official J. Eur. Soc. Med. Oncol.* 10, 1211–1218. doi:10.1023/a:1008361612767
- Vriens, M. R., Suh, I., Moses, W., and Kebebew, E. (2009). Clinical features and genetic predisposition to hereditary nonmedullary thyroid cancer. *Thyroid* 19, 1343–1349. doi:10.1089/thy.2009.1607
- Ward, E., DeSantis, C., Robbins, A., Kohler, B., and Jemal, A. (2014). Childhood and adolescent cancer statistics. *CA a cancer J. Clin.* 64, 83–103. doi:10.3322/caac.21219
- Wu, C., Gong, S., Duan, Y., Deng, C., Kallendrusch, S., Berninghausen, L., et al. (2023). A tumor microenvironment-based prognostic index for osteosarcoma. *J. Biomed. Sci.* 30, 23. doi:10.1186/s12929-023-00917-3