



OPEN

Genomic and transcriptomic characterization of pre-operative chemotherapy response in patients with osteosarcoma

Yongkun Yang^{1,2,4}, Zhen Huang^{1,2,4}, Mingming Yuan^{3,4}, Jinqiu Rui³, Rongrong Chen³, Tao Jin^{1,2}, Yang Sun^{1,2}, Zhiping Deng^{1,2}, Huachao Shan^{1,2}, Xiaohui Niu^{1,2} & Weifeng Liu^{1,2}✉

Osteosarcoma is a heterogeneous disease with regard to its chemotherapy response and clinical outcomes. This study aims to investigate the genomic and transcriptomic characteristics related to pre-operative chemotherapy response. Samples from 25 osteosarcoma patients were collected to perform both whole exome and transcriptome sequencing. Osteosarcoma had significant amount of chromosomal copy number variants (CNVs). Chemotherapy responders showed the higher chromosomal CNV burden than non-responders ($p = 0.0775$), but the difference was not significant. The percentage of COSMIC signature 3, associated with homologous recombination repair deficiency, was higher in responders (56%) than in non-responders (45%). Transcriptomic analysis suggested that 11 genes were significantly up-regulated in responders and 18 genes were up-regulated in non-responders. Both GSEA and KEGG enrichment analysis indicated that four pathways related to cardiomyopathy were up-regulated in responders, while neuroactive ligand – receptor interaction was up-regulated in non-responders. Finally, a previously published chemoresistant model was validated using our dataset, with the area under the curve of 0.796 (95% CI, 0.583–1.000). Osteosarcoma had the heterogeneous mutational profile with frequent occurrence of CNVs. Transcriptomic analysis identified several signaling pathways associated with chemotherapy responsiveness to osteosarcoma. Transcriptomic signatures provides a potential research direction for predicting the chemotherapy response.

Osteosarcoma is the most common primary malignant tumor of bone. Although it is a rare disease in all ages, primary osteosarcoma commonly occurs in children and adolescents aged 10–24 years. According to the statistical analyses of 5 016 osteosarcoma patients from the Surveillance, Epidemiology, and End Results program from 1975 to 2017¹, the age-adjusted incidence of osteosarcoma was 3.3/1,000,000 in all age groups and 7.2/1,000,000 in population aged 10–24 years. Pre-operative (neoadjuvant) chemotherapy followed by definite surgery and postoperative (adjuvant) chemotherapy is the mainstay of therapy for osteosarcoma, significantly improving the survival and prognosis in approximately two-thirds of patients with localized disease².

Despite the remarkable efficacy of chemotherapy, there is a lack of uniform and objective criteria for evaluating the chemotherapy response in osteosarcoma patients. How to evaluate the effectiveness of chemotherapy, especially pre-operative chemotherapy, is crucial for the adjustment of treatment plans and patient prognosis assessment. Currently, the commonly applied evaluation method to determine the chemotherapy response and tumor prognosis is based on the tumor cell necrosis rate^{3,4}. However, the evaluation and application of tumor cell necrosis rate have certain limitations: the procedure of tumor necrosis rate analysis is complex; heterogeneous tumors are difficult to carry out; the evaluation time required is long, resulting in a late outcome which is not conducive to subsequent treatment. Huvo⁵ first proposed a histological evaluation method for osteosarcoma necrosis rate in the 1970s, which has been used for more than 40 years. However, this method cannot be applied to evaluate the chemotherapy response before surgery, and it does not comprehensively consider clinical and

¹Department of Orthopaedic Oncology Surgery, Beijing Jishuitan Hospital, Capital Medical University, No. 31, Xijiekou East Street, Xicheng District, Beijing 100035, China. ²Fourth Medical College of Peking University, Beijing 100035, China. ³Geneplus-Beijing, Beijing 102206, China. ⁴These authors contributed equally: Yongkun Yang, Zhen Huang and Mingming Yuan. ✉email: liuweifengjst@126.com

imaging aspects. Therefore, it is necessary to understand the molecular characteristics related to chemotherapy response and find a new effective evaluation method for chemotherapy response in osteosarcoma.

The rapid development and widespread application of high throughput sequencing techniques promote the understanding of tumor development and progression from the molecular level, and thus change the clinical treatment modes and survival outcomes of patients in various of cancers, especially in non-small cell lung cancer and colorectal cancer^{6,7}. In addition to supporting treatment decisions for cancer patients, next-generation sequencing has also been utilized in multiple clinical settings, including detecting tumor susceptibility genes, early diagnosis and screening, evaluating patient prognosis and monitoring minimal residual diseases and tumor progression^{8–11}. Zeng et al.¹² developed a chemoresistant risk model using markers obtained from bulk RNA and single-cell RNA sequencing with the area under the curve of 0.82 in the TARGET-OS training cohort and 0.84 in the GSE33382 validation cohort. This study aims to investigate the genomic and transcriptomic characteristics of osteosarcoma using DNA- and RNA-based next-generation sequencing techniques, find clinical and molecular factors related to chemotherapy response and validate the performance of the previous model to predict chemotherapy response.

Results

Clinicopathologic characteristics

The clinicopathologic features of 25 patients with pathologic diagnoses of osteosarcoma were summarized in Supplementary Table S1. The median age at diagnosis was 15 years (range 8–32 years) and 75% patients were male. The most common site of osteosarcoma in this study was femur (60%), followed by tibia (20%), humerus (12%), fibula (4%) and multifocal (4%). Majority of the patients were conventional osteosarcoma (84%) and Surgical Staging System (SSS) stage IIB (92%). The tumor volume ranged from 24.633 to 14,705.856 cm³ with the median of 371.57 cm³. Pre-operative chemotherapy and surgery were conducted in 84% and 96% of patients, respectively. Pathological evaluation suggested that the tumor necrosis rate was higher than 90% in 40% of patients.

Genomic landscape of osteosarcoma

A total of 1570 somatic mutations were detected in 25 patients with osteosarcoma, including 1488 single nucleotide variants and small insertions and deletions, 72 CNVs and 7 structural variants. The median number of mutations was 49 (range 2–215). *MYC* (8q24.21), *NCOR1* (7p12–11.2), *PHOX2B* (4p13) and *TP53* (17p13.1) were the top 4 commonly mutated genes, and mutated in 32%, 28%, 28% and 24% of osteosarcomas, respectively (Fig. 1A). Osteosarcoma had the extremely low level of TMB (median 0.47 muts/Mb, range 0.04–1.62 muts/Mb) compared with other tumors¹³, and significant amount of chromosomal CNVs (median: 5, range 0–15). A total of 25 significantly amplified and 5 deleted regions were identified with the *q* value < 0.1 (Fig. 1B). COSMIC mutational signature analysis showed that osteosarcoma was mainly associated with signature 3 (55%, failure of DNA double-strand break-repair by homologous recombination), 22 (14%, aristolochic acid exposure) and 20 (13%, defective DNA mismatch repair) (Fig. 1C). To examine the relationship between copy number and expression level of CNVs, Spearman correlation analysis was performed and the result indicated that copy number for amplifications was positively correlated with Z-score, with the Spearman coefficient of 0.4168 (*p* = 0.0014) (Fig. 1D). The Spearman coefficient for deletions was -0.1071 (*p* = 0.8397) for (Fig. 1E), respectively.

Due to the association between signature 3 and 20 and osteosarcoma, we further analyzed the mutations on genes involved in DNA damage repair (DDR) and cell cycle pathways (Fig. 2). DDR and cell cycle pathway mutations were detected in 52% and 84% of osteosarcoma patients, and the functional mutations (including activating missense mutations, truncated mutations, amplifications and deletions) were detected in 40% and 56% of patients, respectively. The common functional mutations in these two pathways included *MYC*, *CCND3*, *CCND1* amplifications and *TP53* deletions.

Relationship between genomic features and chemotherapy response

To explore the relationship between genomic characteristics and chemotherapy response, fisher's exact test was used to find the differentially mutated genes between responders and non-responders with significance, and the result was negative, which may be explained by the low mutation rate in osteosarcoma (data not shown). Additionally, no difference in TMB between two groups was observed (median TMB for responders vs. non-responders: 0.5111 vs. 0.3778 muts/Mb, *p* = 0.3013) (Fig. 3A), and the relationship between TMB and survival time was not investigated in this study. Notably, the chromosomal CNV burden of chemotherapy responders appeared to be higher than that of non-responders with the median of 7 and 3.5, respectively, but the difference was not statistically significant (*p* = 0.0775) (Fig. 3B). The composition of COSMIC mutational signatures was also compared between chemotherapy responders and non-responders using chi-square test and the *p* value was smaller than 0.0001 (Fig. 3C). Mutational signatures 1, 3, 6, 17, 20, 22 and 28 were present in both groups. Apparently, the percentage of signature 3, which was associated with failure of DNA double-strand break-repair by homologous recombination, was higher in chemotherapy responders (56%) than in non-responders (45%), which suggested that signature 3 may be related to chemotherapy response. By contrast, the percentage of two signatures of unknown etiology (signature 17 and 28) was higher in non-responders (Non-responders vs. Responders: 12% vs. 4% for signature 17; 6% vs. 1% for signature 28).

Differentially expressed genes (DEGs) and pathway enrichment analysis

The raw expression counts for all genes were executed for the differentially expressed gene analysis between chemotherapy responders and non-responders. The thresholds for adjusted *p* (*padj*) and $|\log_2FC|$ were set as < 0.05 and ≥ 1 , respectively. As a result, the volcano plot showed that 11 genes were significantly up-regulated in

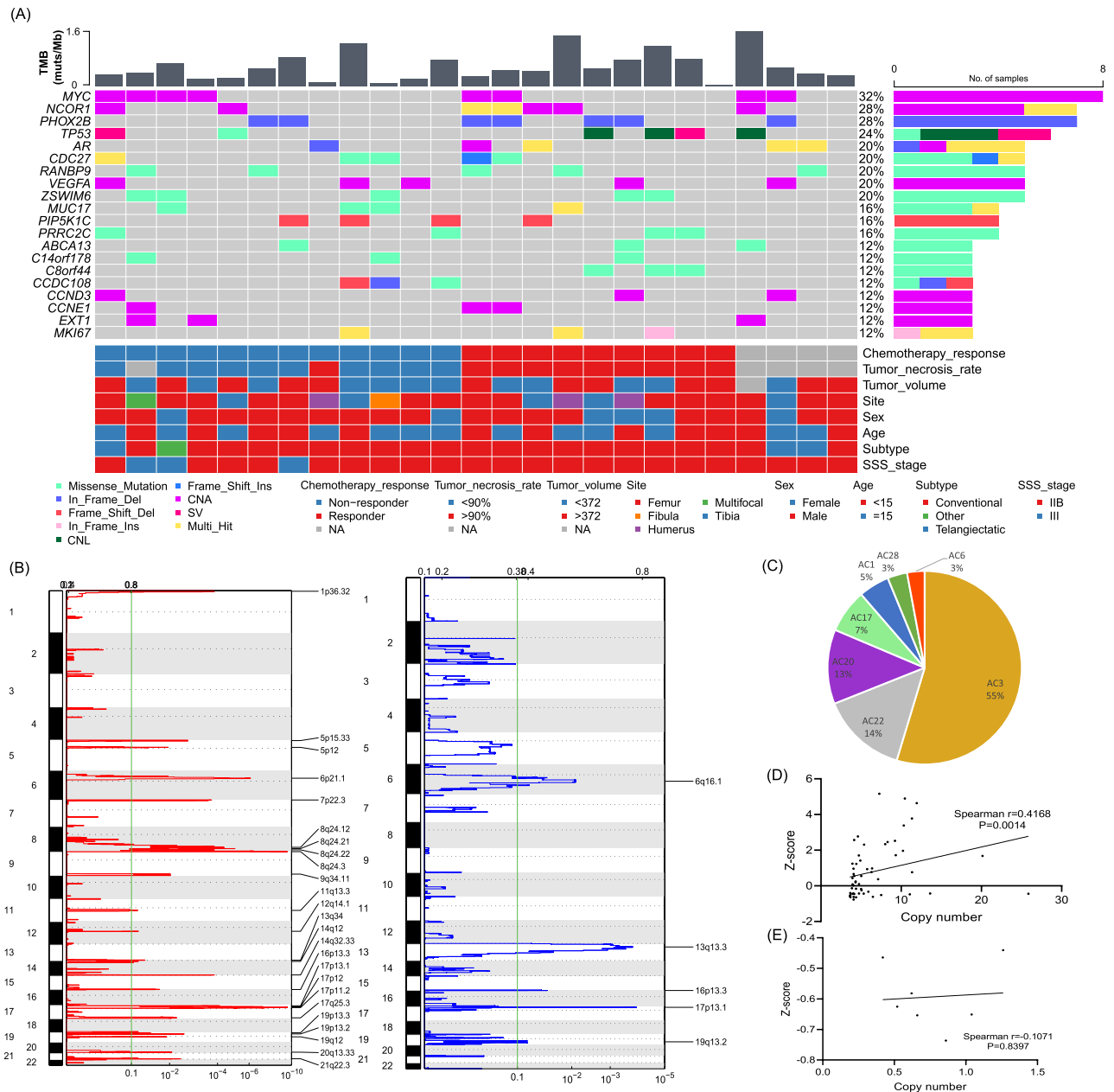


Figure 1. The genomic characteristics of osteosarcoma. **(A)** Clinical information and mutational landscape of 25 patients with osteosarcoma; **(B)** recurrent focal amplifications (left panel) and deletions (right panel) with the G-score and q values on the above and bottom of the figures, respectively; **(C)** inference of the Catalogue of Somatic Mutations in Cancer (COSMIC) mutational signatures composition in osteosarcoma; **(D)** the correlation between copy number of copy number amplifications on the genomic level and corresponding RNA expression on the transcriptomic level; **(E)** the correlation between copy number of copy number deletions on the genomic level and corresponding RNA expression on the transcriptomic level. CNA, copy number amplification; CNL, copy number loss; HRR, homologous recombination repair; MMR, mismatch repair; NA, not available; SSS, Surgical Staging System; SV, structural variant; TMB, tumor mutation burden.

responders and 18 genes were significantly up-regulated in non-responders with the adjusted p value smaller than 0.05 (Fig. 4A). GSEA enrichment analysis was performed to identify the up-regulated pathways in chemotherapy responders and non-responders. Four pathways related to cardiomyopathy were up-regulated in responders (Fig. 4B), while neuroactive ligand – receptor interaction and olfactory transduction pathways were up-regulated in non-responders (Fig. 4C).

Due to the relative small number of DEGs, genes with $p < 0.05$ and $|\log_2FC| \geq 1$ were included to perform conventional KEGG pathway enrichment analysis^{14–16} and PPI analysis. The significant KEGG pathways (adjusted $p < 0.05$) of 309 significantly expressed genes in responders included calcium signaling pathway, motor proteins, cGMP – PKG signaling pathway, circadian entrainment, arrhythmogenic right ventricular cardiomyopathy, as well as 4 pathways identified by GSEA enrichment analysis (Fig. 5A). In non-responders, neuroactive

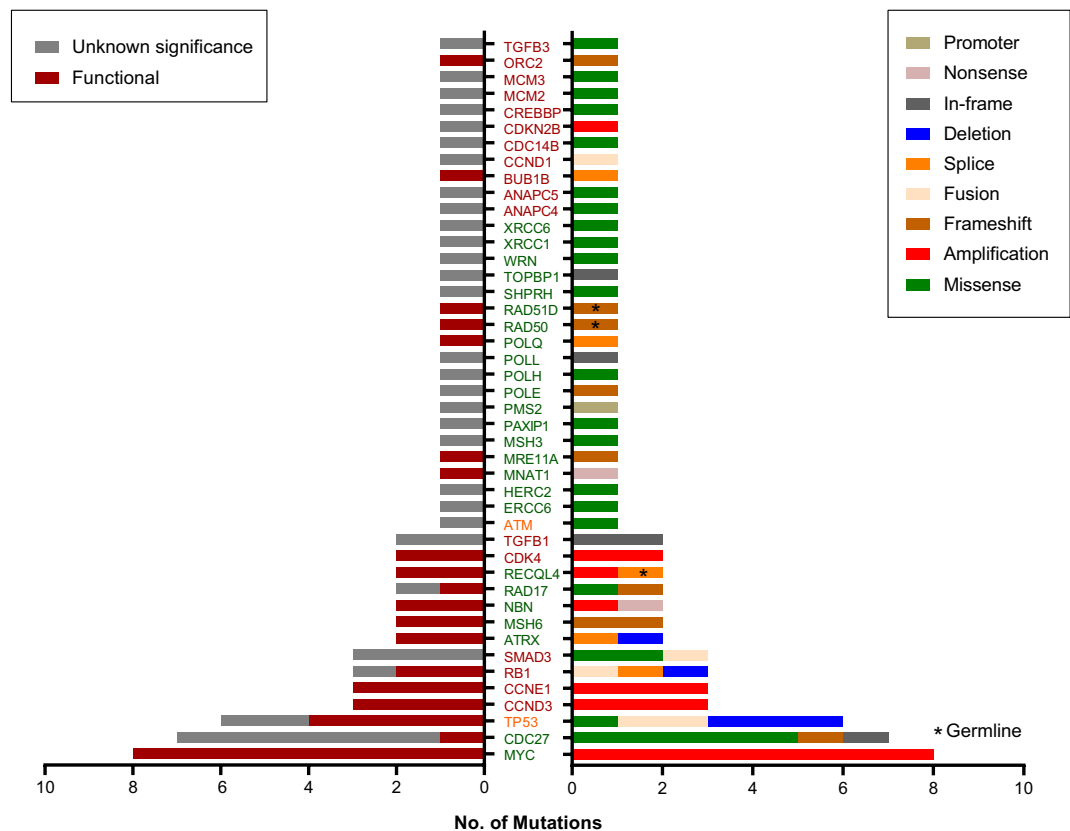


Figure 2. Mutations in DNA damage repair and cell cycle pathways. Genes belonging to DNA damage repair, cell cycle, and both of them were represented using green, red and orange color, respectively.

ligand-receptor interaction pathway was the only significant KEGG pathway with the adjusted p value smaller than 0.05 (Fig. 5B). Genes involved in these pathways were summarized in Supplementary Table S2. The PPI analysis was performed using STRING database and visualized by Cytoscape (Supplementary Fig. 1). A total of 709 nodes and 1888 edges were included in the network. *ACTN2*, *TTN* and *MYH6* were considered as potential hub genes with the node degree score of 45, 43 and 39, respectively. *ACTN2*, *TTN* and *MYH6* were overexpressed in chemotherapy responders with the $|\log_2FC|$ of 1.8487, 3.6915, 2.1881 and p value of 0.0103, 0.0001, 0.0200, respectively (data not shown). In addition, all of them were involved in pathways enriched in chemotherapy responders (Supplementary Table S2).

Model validation to predict chemotherapy response

Zeng et al.¹² developed a robust chemoresistant risk model using the expression levels of 10 genes. The performance of this model to discriminate chemotherapy response was validated using our data. The chemoresistant risk score was calculated using the expression level of 10 genes according to the formula: Risk score = $-36.36 + 0.110 * ADAMTS2 + 0.042 * SPAG16 + 0.124 * CGREF1 + 0.328 * JTB + 0.083 * ENPP2 + 0.528 * ACP1 + 1.485 * NPM1 + 0.759 * CTSF + 0.045 * MPP6 + 1.109 * PARD6G$. The model had the area under the curve of 0.796 (95% confidence interval, 0.583–1.000) (Fig. 6), suggesting the acceptable performance.

Discussion

Osteosarcoma is a heterogeneous disease with regard to its histology, chemotherapy response and clinical outcomes^{17,18}. In this study, whole exome and transcriptome sequencing were conducted in 25 patients with osteosarcoma. Genomic and transcriptomic features relevant to chemotherapy response were explored and a previous chemoresistant model was validated. To our knowledge, this is the first study to comprehensively investigate the clinical and molecular characteristics of chemotherapy response in osteosarcoma using multi-omics techniques, which expands our knowledge of this complex disease.

In multiple previous studies^{19–21}, osteosarcoma was mainly characterized by diverse variants, recurrent structural variants, high frequencies of *TP53* and *RB1* mutations. Similarly, our study revealed that the genetic variants in osteosarcoma were dispersedly distributed. In terms of mutation type, single nucleotide variants and small insertions and deletions were atypical events, whereas CNVs, especially amplification, were common in osteosarcoma. *PHOX2B* mutations were firstly identified in osteosarcoma in our study, which encodes a transcription factor participating in the development of the peripheral nervous system and is known to be related to neuroblastoma and congenital central hypoventilation syndrome²². In our study, all *PHOX2B* mutations

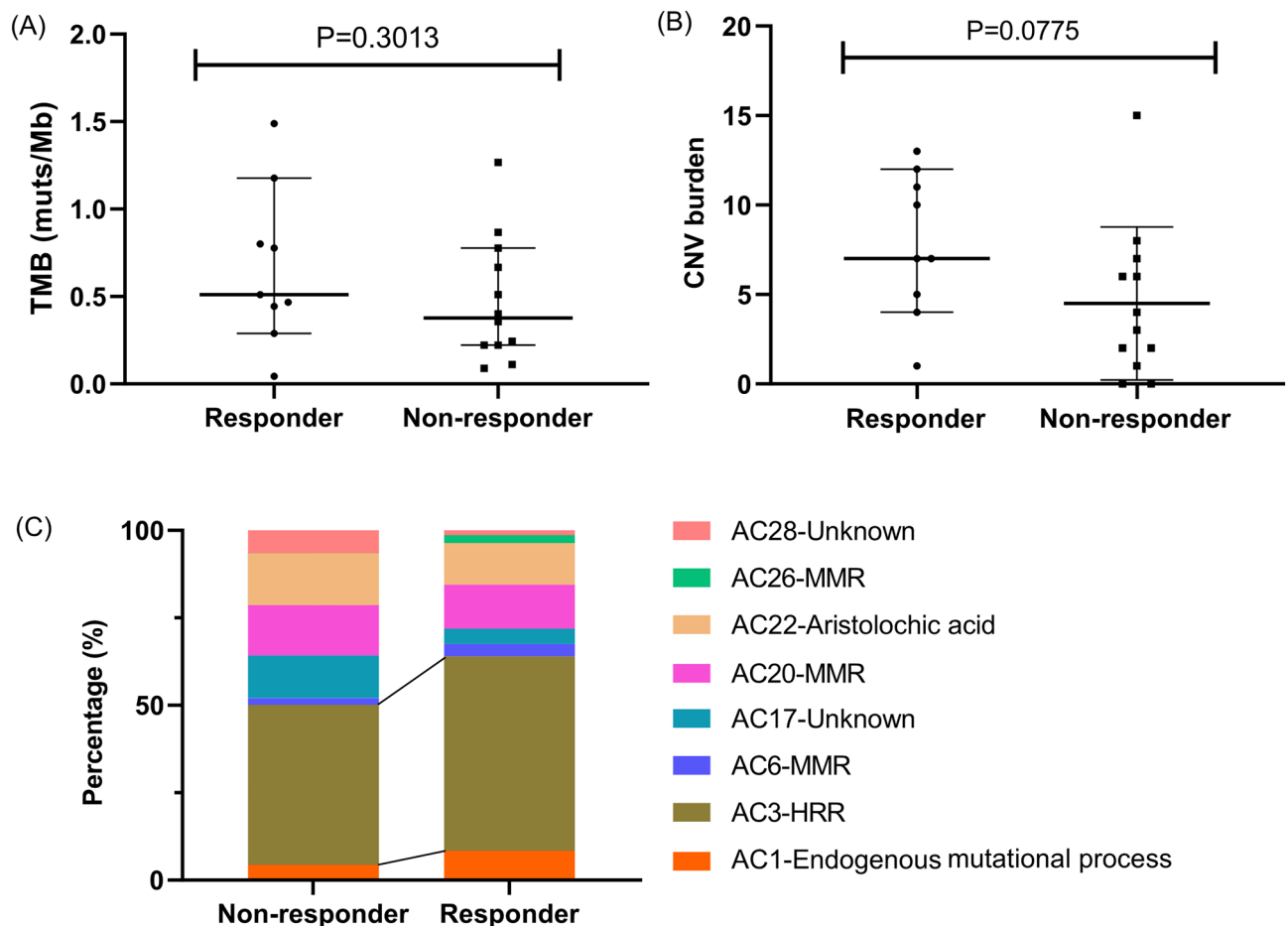


Figure 3. The relationship between genomic features and chemotherapy response. **(A)** Comparison of TMB between chemotherapy responders and non-responders; **(B)** comparison of CNV burden between chemotherapy responders and non-responders; **(C)** comparison of the Catalogue of Somatic Mutations in Cancer (COSMIC) mutational signatures composition between chemotherapy responders and non-responders. CNV, copy number variant; HRR, homologous recombination repair; MMR, mismatch repair; TMB, tumor mutation burden.

were in-frame deletions occurred in the 20 polyalanine region of *PHOX2B* (i.e. polyAla contractions) and evaluated as benign/likely benign variants by ClinVar (Variation ID: 227016). Therefore, they are more likely to be passenger mutations. Our study also found that osteosarcoma had the low level of TMB, which was consistent with previous studies showing the median TMB ranged from 0.1 to 2.5 muts/Mb^{13,23–27}. In addition, osteosarcoma had significant amount of chromosomal CNV. It seemed that patients with high chromosomal CNV burden had a nonsignificantly better response to chemotherapy, which required to be verified in further studies. Consistent with the previous study²⁰, mutations in osteosarcoma were mainly caused by the dysfunctional DNA damage repair in this study. Moreover, this study suggested that chemotherapy responders had more homologous recombination repair deficiency—induced mutations, which may be explained by the administration of platinum-based chemotherapy regimens in osteosarcoma and the favorable response to platinum in defective homologous recombination repair ovarian cancer²⁸.

In this study, 11 genes were significantly up-regulated in responders and 18 genes were significantly up-regulated in non-responders. Both KEGG and GSEA enrichment analyses suggested that 4 pathways related to cardiomyopathy were significantly enriched in responders and neuroactive ligand-receptor interaction pathway was significant in non-responders. So far, no direct evidence has been found to support the relationship between cardiomyopathy and chemotherapy response. In a previous study, transcriptomic analysis was performed in breast cancer to investigate biomarkers related to chemotherapy sensitivity. KEGG analysis indicated that a great number of DEGs were enriched in neuroactive ligand-receptor interaction pathway. In addition, the number of DEGs in this pathway were more likely to be subexpressed in chemotherapy resistant group. However, whether the entire pathway was up-regulated or down-regulated was not analyzed²⁹. In another study in breast cancer, the relationship between neuroactive ligand-receptor interaction pathway and pathological complete response to chemotherapy was not clearly elucidated³⁰. One study used the paclitaxel-induced peripheral neuropathy rat model to understand the transcriptomic level of the dorsal root ganglia neurons and found that neuroactive ligand-receptor interaction was majorly involved in sensory neurons of rats with paclitaxel-induced peripheral neuropathy³¹. PPI analysis identified *ACTN2*, *TTN* and *MYH6* as hub genes resulting in

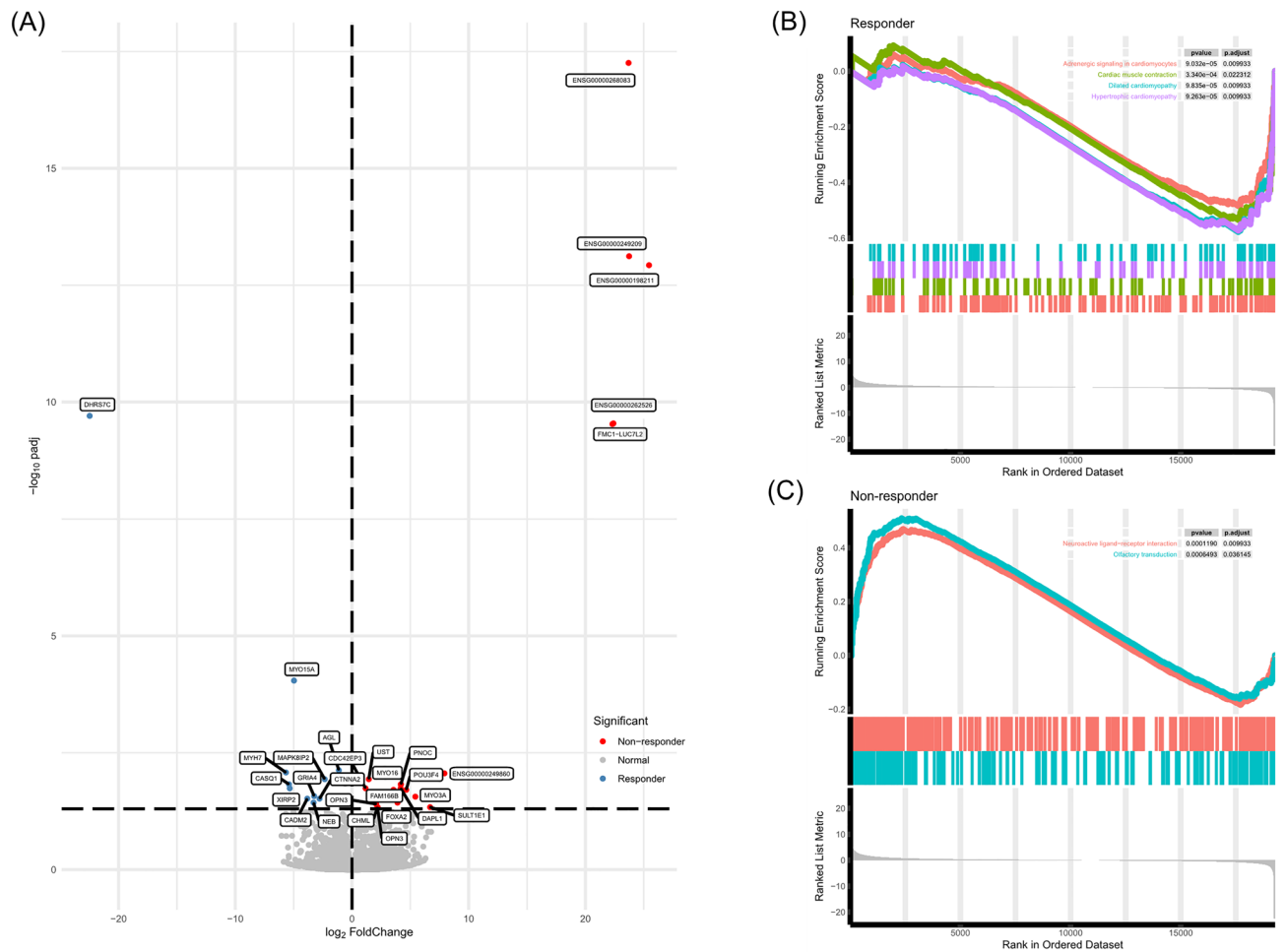


Figure 4. Differentially expressed genes and enrichment analysis between chemotherapy responders and non-responders. **(A)** Volcano plot showing the differentially expressed genes between groups; GSEA analyses of gene sets for responders **(B)** and non-responders **(C)**.

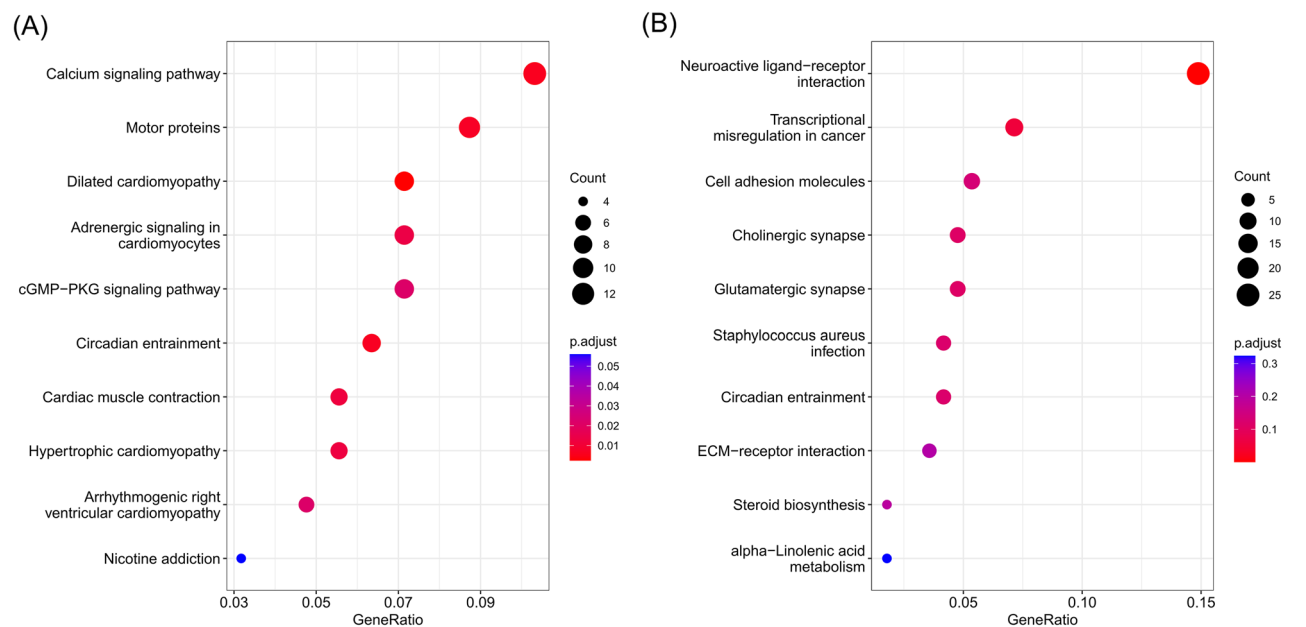


Figure 5. Dotplots showing the KEGG pathway enrichment analysis of genes differentially expressed in **(A)** chemotherapy responders and **(B)** non-responders.

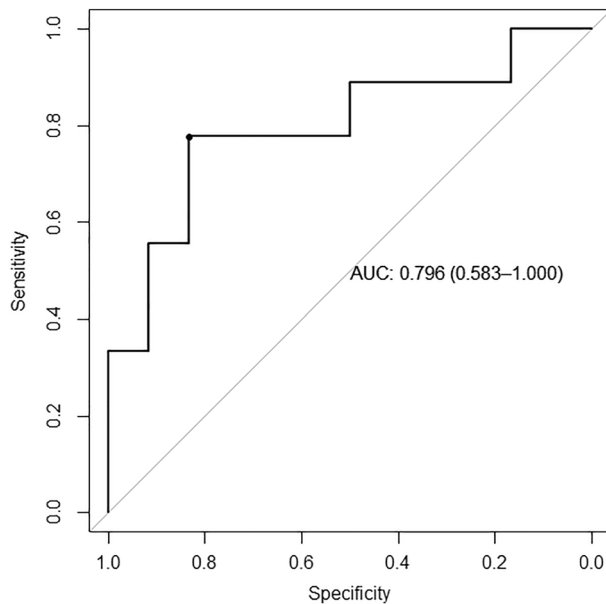


Figure 6. Receiver operator characteristic curve of the published model to predict chemotherapy response using our dataset.

chemotherapy responsiveness. *ACTN2* encoded protein links the anti-parallel actin filaments, contributes to sarcomere stability and is related to cardiomyopathy³². *TTN* encoded protein plays an important role in skeletal and in heart muscles³³. *MYH6* gene provides instructions for making a protein known as the cardiac alpha (α)-myosin heavy chain, which forms part of type II myosin in cardiac muscles³⁴. According to The Human Protein Atlas database, *ACTN2*, *TTN* and *MYH6* are highly expressed in muscle cells. Their higher expression in the chemotherapy responders may indicate that the specimen was derived from an extra-osseous lesion that had invaded muscle, and tumors invading the surrounding muscle may be more sensitive to chemotherapy. However, this hypothesis can not be verified using histological or immunohistochemistry analysis as no sufficient tissue can be obtained, which requires further investigations.

In clinics, multiple methods are used to evaluate the neoadjuvant chemotherapy response in osteosarcoma, including symptoms and signs, laboratory tests, imaging examinations and evaluation of tumor necrosis rate. Zeng et al.¹² developed a chemoresistant risk model with the area under the curve of 0.82 in the TARGET-OS training cohort and 0.84 in the GSE33382 validation cohort. Our study revealed an acceptable performance of this model with the AUC of 0.796. A potential limitation of this study was the small amount number of participants, which may reduce the representativeness of certain findings and need to be further verified in large-scale studies.

In conclusion, multi-omics sequencing techniques help us better understand the molecular characteristics of osteosarcoma. Osteosarcoma is a highly heterogeneous disease with frequent occurrence of CNVs at the genomic level. Transcriptomic analysis identified several signaling pathways associated with chemotherapy responsiveness to osteosarcoma, including pathways related to cardiomyopathy and neuroactive ligand-receptor interaction. Additionally, a model based on transcriptomic characteristics of osteosarcoma may be used for predicting the chemotherapy response in clinics..

Materials and methods

Patients and samples

A total of 25 patients with the definite diagnosis of osteosarcoma in Beijing Jishuitan Hospital (Beijing, China) from December 2021 to November 2022 were prospectively enrolled in this study. Clinical information including demographics, pathologic diagnoses, treatment history and imaging examinations were collected. Tumor tissue samples (including 6 fresh tissues, 18 formalin immersed tissues and 1 formalin fixed paraffin-embedded tissue) and 2 ml matched peripheral blood were collected from all participants to perform the whole exome and transcriptome sequencing. All procedures were conducted in accordance with the Declaration of Helsinki. This study was approved by the Ethics Committee of Beijing Jishuitan Hospital (Approval No. K2023013-00) and written informed consent was obtained from all participants.

Evaluation of chemotherapy response

The evaluation of chemotherapy response mainly depended on tumor necrosis rate. The tumor cell necrosis rate was analyzed using the tumor tissues collected from tumor resection and separated into two groups: greater than 90% and less than 90%. As the lesion was suspected to be benign before surgery, one patient (P21) underwent surgery directly without preoperative chemotherapy. Three patients (P12, P15, P18) underwent unplanned surgery for other reasons and did not receive preoperative chemotherapy. Among 21 patients treated with

chemotherapy, patients with tumor necrosis rates greater than 90% were considered as chemotherapy responders (N=9), and those with tumor necrosis rates smaller than 90% were non-responders (N=10). However, there were a few exceptions for the remaining two patients. P24 was initially diagnosed as AJCC stage IV osteosarcoma with lung metastases and not suitable for surgical resection. The tumor necrosis rate could not be assessed as the post-treatment tumor tissues were not available. Considering of the increase of the circumference difference of the tumor site between affected limb and healthy limb after chemotherapy, he was considered as a chemotherapy non-responder. For P14, the tumor necrosis rate assessed using small pieces of tumor tissues was greater than 90%. However, the clinical and imaging findings (including post-treatment increase of the circumference difference by 4.5 cm, unclear margin, increase of ossification and blood supply) suggested that the patient did not respond to chemotherapy. Therefore, the patient was finally evaluated as a non-responder by investigators. Taken together, 9 patients responded to chemotherapy, while 12 patients were non-responders.

DNA- and RNA-based next-generation sequencing

Before sequencing, all tumor samples were treated with formalin to ascertain the tumor cell content. Whole exome and transcriptome sequencing were performed in Geneplus-Beijing (Beijing, China) as previously described^{27–29}. Briefly, genomic DNA and RNA were extracted from the tumor samples using FirePure™ FFPE gDNA Extraction Kit for genomic DNA and RNeasy FFPE Kit for RNA or AllPrep DNA/RNA FFPE Kit for both DNA and RNA (QIAGEN, Hilden, Germany). Genomic DNA from leukocytes was extracted using the CWE9600 Blood DNA Kit (Cwbiotech, Taizhou, China). Sequencing libraries of genomic DNA and mRNA were prepared using KAPA DNA Library Preparation Kit (Kapa Biosystems, Wilmington, MA, USA) and NEB Next Ultra™ RNA Library Prep Kit (Illumina, Inc., San Diego, CA, USA), respectively. The DNA and RNA sequencing were performed using the DNBSEQ-T7RS High-throughput Sequencing platform (MGI, Shenzhen, China), whose performance was comparable to Illumina platform^{35,36}. All experimental procedures followed the manufacturer's instructions. Whole transcriptome sequencing cannot be performed for one patient due to the severe degradation of RNA. The detailed quality control data of DNA and RNA sequencing was provided in Supplementary Table S1.

Bioinformatics analysis

After removal of terminal adaptor sequences and low-quality reads, the clean sequencing reads were aligned to the reference human genome (hg19) using BWA (version 0.7.10) and HISAT (version 2.0.4) for DNA and RNA sequencing, respectively. Genomic single nucleotide variants, small insertions and deletions, copy number variants (CNVs) and structural variants were detected using MuTect (version 1.1.4)/NChot, GATK (version 3.4–46-gbc02625) and CONTRA (version 2.0.8), respectively. Transcript assembly was performed using StringTie (version 1.2.3).

Tumor mutation burden was evaluated as the number of non-synonymous variants with the mutant allele frequency greater than 5% per megabase in the coding region. Chromosomal CNV burden represented the total level of amplifications or deletions at the chromosome level, which was calculated as previously described³⁰. Gistic 2.0 was used to detect the significantly recurrent regions with amplification or deletion. The mutational landscape was portrayed using R package 'maftools' (version 2.14.0). To infer the composition of known Catalogue Of Somatic Mutations In Cancer (COSMIC) mutational signatures in osteosarcoma, R package 'yapsa' (version 1.24.0) was performed using the COSMIC mutational signatures version 2 (https://cancer.sanger.ac.uk/signatures/signatures_v2). Differentially expressed genes were analyzed using R package 'DESeq2' (version 1.38.3). R package 'clusterProfiler' (version 4.7.1.3) was used to perform gene set enrichment analysis (GSEA) enrichment analysis and conventional Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. An online tool (STRING, <https://string-db.org/>) was used to analyze protein–protein interaction (PPI). The default confidence score > 0.40 was used to screen the PPI pairs. Cytoscape (version 3.9.1) was applied to visualize the PPI network. R package 'pROC' (version 1.18.0) was used to plot the receiver operator characteristic curve and calculate the area under the curve.

Statistical analysis

Z-score normalization of gene expression data was performed using the scale function in R package 'base' (version 4.2.2) taking the raw count matrix as an input. Spearman correlation analysis was conducted to study the correlation between copy number and Z-score of CNVs. Differences of variables between groups were assessed using Mann–Whitney test for continuous variables and Fisher's exact test or Chi-square test for categorical variables, with $P < 0.05$ considered as statistically significant.

Ethics statement

All procedures were conducted in accordance with the Declaration of Helsinki. All the procedures involving human subjects were conducted following the ethical guidelines approved by the ethical committee of Beijing Jishuitan Hospital (Approval No. K2023013-00). All the participants have signed the informed consent for sequencing.

Data availability

The datasets generated and/or analysed during the current study are available in the Genome Sequence Archive (GSA) repository, [ACCESSION NUMBER: HRA005488].

Received: 5 July 2023; Accepted: 6 November 2023

Published online: 27 November 2023

References

- Cole, S., Gianferante, D. M., Zhu, B. & Mirabello, L. Osteosarcoma: A surveillance, epidemiology, and end results program-based analysis from 1975 to 2017. *Cancer* **128**, 2107–2118. <https://doi.org/10.1002/cncr.34163> (2022).
- Ritter, J. & Bielack, S. S. Osteosarcoma. *Ann. Oncol.* **21**(7), 320–325. <https://doi.org/10.1093/annonc/mdq276> (2010).
- Bacci, G. *et al.* Prognostic factors for osteosarcoma of the extremity treated with neoadjuvant chemotherapy: 15-year experience in 789 patients treated at a single institution. *Cancer* **106**, 1154–1161. <https://doi.org/10.1002/cncr.21724> (2006).
- Bishop, M. W. *et al.* Assessing the prognostic significance of histologic response in osteosarcoma: A comparison of outcomes on CCG-782 and INT0133-A report from the children's oncology group bone tumor committee. *Pediatr. Blood Cancer* **63**, 1737–1743. <https://doi.org/10.1002/psc.26034> (2016).
- Huvos, A. G., Rosen, G. & Marcove, R. C. Primary osteogenic sarcoma: Pathologic aspects in 20 patients after treatment with chemotherapy en bloc resection, and prosthetic bone replacement. *Arch. Pathol. Lab. Med.* **101**, 14–18 (1977).
- Miller, M. & Hanna, N. Advances in systemic therapy for non-small cell lung cancer. *BMJ (Clin. Res. Ed.)* **375**, n2363. <https://doi.org/10.1136/bmj.n2363> (2021).
- Biller, L. H. & Schrag, D. Diagnosis and treatment of metastatic colorectal cancer: A review. *Jama* **325**, 669–685. <https://doi.org/10.1001/jama.2021.0106> (2021).
- Chen, M. & Zhao, H. Next-generation sequencing in liquid biopsy: Cancer screening and early detection. *Hum. Genom.* **13**, 34. <https://doi.org/10.1186/s40246-019-0220-8> (2019).
- Mirabello, L. *et al.* Frequency of pathogenic germline variants in cancer-susceptibility genes in patients with osteosarcoma. *JAMA Oncol.* **6**, 724–734. <https://doi.org/10.1001/jamaoncol.2020.0197> (2020).
- Zhang, J. T. *et al.* Longitudinal undetectable molecular residual disease defines potentially cured population in localized non-small cell lung cancer. *Cancer Discov.* **12**, 1690–1701. <https://doi.org/10.1158/2159-8290.cd-21-1486> (2022).
- Yi, Z. *et al.* The molecular tumor burden index as a response evaluation criterion in breast cancer. *Signal Transd. Target. Therapy* **6**, 251. <https://doi.org/10.1038/s41392-021-00662-9> (2021).
- Zeng, Z. *et al.* Development of a chemoresistant risk scoring model for prechemotherapy osteosarcoma using single-cell sequencing. *Front. Oncology* **12**, 893282. <https://doi.org/10.3389/fonc.2022.893282> (2022).
- Chalmers, Z. R. *et al.* Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genom. Med.* **9**, 34. <https://doi.org/10.1186/s13073-017-0424-2> (2017).
- Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucl. Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
- Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951. <https://doi.org/10.1002/pro.3715> (2019).
- Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucl. Acids Res.* **51**, D587–d592. <https://doi.org/10.1093/nar/gkac963> (2023).
- Whelan, J. S. & Davis, L. E. Osteosarcoma, chondrosarcoma, and chordoma. *J. Clin. Oncol.* **36**, 188–193. <https://doi.org/10.1200/jco.2017.75.1743> (2018).
- Lilienthal, I. & Herold, N. Targeting molecular mechanisms underlying treatment efficacy and resistance in osteosarcoma: A review of current and future strategies. *Int. J. Mol. Sci.* **21**, 748. <https://doi.org/10.3390/ijms21186885> (2020).
- Bousquet, M. *et al.* Whole-exome sequencing in osteosarcoma reveals important heterogeneity of genetic alterations. *Ann. Oncol.* **27**, 738–744. <https://doi.org/10.1093/annonc/mdw009> (2016).
- Jiang, Y. *et al.* Multi-omics analysis identifies osteosarcoma subtypes with distinct prognosis indicating stratified treatment. *Nat. Commun.* **13**, 7207. <https://doi.org/10.1038/s41467-022-34689-5> (2022).
- Chen, X. *et al.* Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell Rep.* **7**, 104–112. <https://doi.org/10.1016/j.celrep.2014.03.003> (2014).
- Raabe, E. H. *et al.* Prevalence and functional consequence of PHOX2B mutations in neuroblastoma. *Oncogene* **27**, 469–476. <https://doi.org/10.1038/sj.onc.1210659> (2008).
- Xu, H. *et al.* Genetic and clonal dissection of osteosarcoma progression and lung metastasis. *Int. J. Cancer* **143**, 1134–1142. <https://doi.org/10.1002/ijc.31389> (2018).
- Reed, D. E. & Shokat, K. M. Targeting osteosarcoma. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18100–18101. <https://doi.org/10.1073/pnas.1420596111> (2014).
- Sayles, L. C. *et al.* Genome-informed targeted therapy for osteosarcoma. *Cancer Discov.* **9**, 46–63. <https://doi.org/10.1158/2159-8290.cd-17-1152> (2019).
- Liu, W. *et al.* Whole-exome sequencing in osteosarcoma with distinct prognosis reveals disparate genetic heterogeneity. *Cancer Genet.* **256–257**, 149–157. <https://doi.org/10.1016/j.cancergen.2021.05.013> (2021).
- Xie, L. *et al.* The clinical implications of tumor mutational burden in osteosarcoma. *Front. Oncol.* **10**, 595527. <https://doi.org/10.3389/fonc.2020.595527> (2020).
- Wen, H. *et al.* Homologous recombination deficiency in diverse cancer types and its correlation with platinum chemotherapy efficiency in ovarian cancer. *BMC Cancer* **22**, 550. <https://doi.org/10.1186/s12885-022-09602-4> (2022).
- Barrón-Gallardo, C. A. *et al.* Transcriptomic analysis of breast cancer patients sensitive and resistant to chemotherapy: Looking for overall survival and drug resistance biomarkers. *Technol. Cancer Res. Treatment* **21**, 15330338211068964. <https://doi.org/10.1177/15330338211068965> (2022).
- Li, Y., Liu, X., Tang, H., Yang, H. & Meng, X. RNA sequencing uncovers molecular mechanisms underlying pathological complete response to chemotherapy in patients with operable breast cancer. *Med. Sci. Monit.* **23**, 4321–4327. <https://doi.org/10.12659/msm.903272> (2017).
- Sun, W. *et al.* Transcriptome analysis reveals dysregulation of inflammatory and neuronal function in dorsal root ganglion of paclitaxel-induced peripheral neuropathy rats. *Mol. Pain* **19**, 17448069221106168. <https://doi.org/10.1177/17448069221106167> (2023).
- Ranta-Aho, J. *et al.* Mutation update for the ACTN2 gene. *Hum. Mut.* **43**, 1745–1756. <https://doi.org/10.1002/humu.24470> (2022).
- Chauveau, C., Rowell, J. & Ferreira, A. A rising titan: TTN review and mutation update. *Hum. Mutat.* **35**, 1046–1059. <https://doi.org/10.1002/humu.22611> (2014).
- Carniel, E. *et al.* Alpha-myosin heavy chain: A sarcomeric gene associated with dilated and hypertrophic phenotypes of cardiomyopathy. *Circulation* **112**, 54–59. <https://doi.org/10.1161/circulationaha.104.507699> (2005).
- Jeon, S. A. *et al.* Comparison between MGI and Illumina sequencing platforms for whole genome sequencing. *Genes Genom.* **43**, 713–724. <https://doi.org/10.1007/s13258-021-01096-x> (2021).
- Lang, J. *et al.* Evaluation of the MGISEQ-2000 sequencing platform for illumina target capture sequencing libraries. *Front. Genet.* **12**, 730519. <https://doi.org/10.3389/fgene.2021.730519> (2021).

Author contributions

Y.Y.: data curation, writing - original draft, writing—reviewing & editing; Z.H.: data curation, writing—reviewing & editing; M.Y.: formal analysis, methodology, writing—review & editing; J.R.: writing—review & editing; R.C.:

supervision, writing—review & editing; T.J.: writing—review & editing; Y.S.: writing—review & editing; Z.D.: writing—review & editing; H.S.: writing – review & editing; X.N.: writing—review & editing, supervision; W.L.: writing—review & editing, conceptualization, supervision.

Funding

This study was funded by National Key R&D Program of China (grant No. 2023YFB4706300), Beijing Natural Science Foundation (grant No. L212042), Beijing Municipal Administration of Hospitals Incubating Program (grant No. PX2021015), Beijing Hospitals Authority Youth Programme (grant No. QML20200403), Beijing Jishuitan Hospital Elite Young Scholar Programme (grant No. XKGG202105), Beijing JST Research Funding (grant No. ZR[1]202104) and National Key R&D Program of China (grant No. 2021YFC2400500).

Competing interests

Conflict of Interest: M.Y. J.R. and R.C. are employees of Geneplus-Beijing (Beijing, China). The remaining authors report no conflict of interest.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-46857-8>.

Correspondence and requests for materials should be addressed to W.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023