Methods

Data collection

Analyses were performed on two bulk RNA-seq osteosarcoma datasets, TARGET-OS Osteosarcoma and GSE87686. Data was collected from TARGET-OS using GDC Data Transfer Tool UI (v1.0.0), returning 19493 protein coding genes and 88 samples for TARGET-OS, containing both raw data and TPM data. GSE87686 data was obtained through the lab's previously pre-processed kallisto files, downloaded through SRA Run Selector to obtain SRA run files. Data was then imported from kallisto files via *tximport (v1.22.0)* R package. Genes were converted to ENST and ENSG and finally HUGO gene symbols through *biomaRt (v2.50.3)*.

Normalization

Raw data was converted to TPM as it is the best performing normalization method than FPKM or RPKM, based on its preservation of biological signal as compared to other methods [@Abrams.2019]. Calculation was performed using counts and lengths for each gene, returned from the tximport to kallisto process.

Normalization by Z-score with the *scale* function was used to normalize the data for each gene, in order to be able to visualize the data in corresponding heatmaps.

Z-score of the mean of TPM data was used to construct grouped heatmaps for a given gene or gene signature.

Inflammatory groups characterizing the intensity of inflammatory status in tumors were created by choosing the lowest and highest mean of Z-score of the Hallmark Inflammatory Response signature from MSigDB, containing 200 genes. ICAM4 was notably not present in the dataset in TARGET-OS cohort. The groups were cut off evenly using the *ntile* function in *dplyr (v1.0.8)* R package.

Construction of gene signatures

Gene signatures relevant to the research topic were obtained from MSigDB via *msigdbr* (v7.5.1)
HAY_BONE_MARROW_NEUTROPHIL.v7.5.1

Canonical markers for osteosarcoma markers were adapted from @Zhou.202005 's analysis and their canonical markers generated from the literature in their Supplementary Table 2.

Construction of inflammatory groups

The groups were cut off evenly using the *ntile* function in *dplyr* R package.

MDS Clustering

Determination of cell population abundance

MCP Counter

Estimation of tumor immune infiltration using bulk RNA-seq can be estimated using Microenvironment Cell Populations-counter (MCP-counter) [@Becht.2016].

CIBERSORTx.

CIBERSORTx algorithm was also used for immune cell deconvolution, from the CIBERSORTx platform (https://cibersortx.stanford.edu/) to generate abundance for 22 immune cells. CIBERSORTx is run in absolute mode, with batch correction, no quantile normalization. # normalization

Identification of Differentially Expressed Genes (DEG)

ESeq2 (version 1.26.0) was used to perform normalization and differential gene expression analysis (Cluster2/3 vs Cluster1), and the obtained differential gene expression (| log2FC | >1 and adjust P-value <0.05) was used for visualization and subsequent analysis (46). CIBERSORT (version 1.03) was used for estimating the abundance profiles of the osteoblastic OS cells in the 85 samples. After the calculations had been made, samples with $P < 0.05$ were used for subsequent analysis and figure display (Figure S1D).

Using *DESeq2 (v1.34.0)* [@Love.2014], standard DEG pipeline using raw data was performed between inflammatory groups (Low, Medium, High). Fold Change (FC) = was calculated between Low and High group. DEGs were identified with adjusted p.value (or False Discovery Rate) ≤ 0.05 and $\log_2(FC) \geq 1$ or $\log_2(FC) \leq -1$. An interpretation of the FDR/Benjamini-Hochberg method for controlling the FDR is implemented in DESeq2 in which we rank the genes by p-value, then multiply each ranked p-value by m/rank (m = total number of tests).

Over Representation Analysis (ORA)

enrichR v(3.0) was used [@R-enrichR; @Xie.2021] for functional gene enrichment/pathway analysis. Following database were queried : *"GO_Molecular_Function_2021", "Human_Gene_Atlas", "BioPlanet_2019", "GO_Biological_Process_2021", "GO_Cellular_Component_2021"*

Statistical Testing

Statistical testing was performed using *R software 4.2.1 (2022-06-23)* and the *rstatix* and *car* package.

ANOVA was performed for multiple comparison testing, only if the data is normally distributed and has homoscedasticity, verified through shapiro's test and levene's test, respectively. Tukey's honestly significant different testing was effected when appropriate.

Kruskal-Wallis testing was performed for non-metric comparative analysis between groups when. Post-hoc analysis was performed using Dunn's test as opposed to Wilcoxon due to the test taking into account Kruskal-Wallis's rank. Post-hoc dunn's test was done when appropriate.

$P < 0.05$ and $P.adj < 0.05$ was considered statistically significant.

Gene Set Enrichment Analysis (GSEA)

GSEA was computed through *clusterprofiler* which uses *fgsea* as a backend. $P < 0.10$ was considered statistically significant. padjust is calculated using the False Discovery Rate from Benjamini-Hochberg. Pathways in the MSigDB databases (<https:// www.gsea-msigdb.org/gsea/msigdb>) were used for the GSEA analysis.