Methods

Data collection

Analyses were performed on two osteosarcoma bulk RNA-seq 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* R package (v1.22.0).

Genes were converted to ENST and ENSG and finally HUGO gene symbols through biomaRt (v2.50.3).

Protocol similar to SkeletalVis(Soul et al., 2019)

Data normalization via TPM method by Kallisto (?), TPM is the best normalization method (Abrams et al., 2019).

DESeq2, differential expression gene analysis, representation through heatmap and volcano plot

Citer : Love, M.I., Huber, W., Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15:550. 10.1186/s13059-014-0550-8

Data visualization with UMAP, k-means clustering,

Using EnrichR package for functional gene enrichment/pathway analysis, querying "GO_Molecular_Function_2021", "Human_Gene_Atlas", "BioPlanet_2019", "GO_Biological_Process_2021", "GO_Cellular_Component_2021"

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.2020ldd]

MDS Clustering

Differential Expression Gene analysis

Differentially Expressed Genes (DEG)