Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Fucosyltransferase 4 as a candidate diagnostic biomarker and potential therapeutic target in osteosarcoma: a multi-dataset transcriptomic and structural bioinformatics analysis

Xingxing Sun
Shanghai Sixth People's Hospital
Yan Huang
Shanghai Tenth People's Hospital
Kun Han
Shanghai Sixth People's Hospital
Hongling Zhu
Shanghai Sixth People's Hospital
Daliu Min
Shanghai Sixth People's Hospital
Cong Tian
tiancong1987@163.com

Shanghai Sixth People's Hospital

Research Article

Keywords: osteosarcoma, FUT4, fucosylation, bioinformatics, diagnostic biomarker, immune infiltration, virtual screening

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11194919/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Fucosyltransferase 4 as a candidate diagnostic biomarker and potential therapeutic target in osteosarcoma: a multi-dataset transcriptomic and structural bioinformatics analysis

Xingxing Sun[1]# (ORCID 0009-0000-2964-9477), Yan Huang[2]# (ORCID 0000-0002-6123-4228), Kun Han[1]# (ORCID 0000-0001-9272-5567), Hongling Zhu[1] (ORCID 0000-0002-8196-2767), Daliu Min[1]* (ORCID 0000-0002-7267-7796), Cong Tian[1]* (ORCID 0009-0003-1194-949X)

[1]Department of Oncology, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, P. R. China

[2]Department of Radiation Oncology, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, P. R. China

# These authors contributed equally to this work.

* Corresponding authors:

Daliu Min, Email: mindaliu1969@163.com

Cong Tian, Email: tiancong1987@163.com

Abstract

Background: Osteosarcoma, the most common primary malignant bone tumor in children and young adults, has a poor prognosis in metastatic or recurrent disease. Fucosyltransferase 4 (*FUT4*) mediates fucosylation and has been implicated in several cancers, but its role in osteosarcoma remains unclear.

Methods: Using two public Gene Expression Omnibus microarray cohorts (GSE36001, GSE16088), we assessed *FUT4* expression and diagnostic performance by limma-based differential analysis and ROC curves with AUC estimation, and profiled function through GO, KEGG, GSEA and GSVA. We further constructed protein-protein interaction and

transcription factor/microRNA/drug regulatory networks, characterized immune-cell signatures by ssGSEA, and prioritized candidate binding compounds by DrugCLIP virtual screening followed by AutoDock Vina docking.

Results: *FUT4* transcripts were significantly down-regulated in osteosarcoma in both cohorts (GSE36001, $p < 0.05$; GSE16088, $p < 0.01$), with moderate-to-high discriminatory capacity, although the high AUC in GSE16088 may be confounded by non-osseous control tissues. *FUT4*-associated genes were enriched in ossification, osteoblast differentiation and extracellular matrix-related terms. GSEA linked high *FUT4* to epithelial-mesenchymal transition and inflammatory-response signatures, and exploratory GSVA suggested differences in cytoskeletal, apoptotic and Wnt/beta-catenin signalling between *FUT4*-low and *FUT4*-high subgroups, which also showed distinct immune-cell correlation patterns. Structure-based screening prioritized ST50926609 as the lead candidate *FUT4*-binding compound (Vina score = −7.35 kcal/mol).

Conclusions: Collectively, these computational findings support *FUT4* as a candidate diagnostic biomarker and potential therapeutic target in osteosarcoma, but they remain hypothesis-generating and require validation in clinical cohorts, cell and animal models.

Keywords: osteosarcoma; FUT4; fucosylation; bioinformatics; diagnostic biomarker; immune infiltration; virtual screening

Introduction

Osteosarcoma is a malignant mesenchymal tumor and the most common primary malignant bone tumor, arising predominantly in the metaphyses of long bones in children and adolescents, with a peak incidence between 10 and 20 years of age [1,2]. It is characterized by aggressive local invasion and a high propensity for early distant dissemination. At initial diagnosis, approximately 15–20 % of patients already present with metastatic lesions, most

frequently in the lungs. Although the combination of neoadjuvant chemotherapy and limb-salvage surgery has improved the 5-year survival of patients with localized disease to approximately 60–70 %, the prognosis of patients with metastatic or recurrent osteosarcoma remains extremely poor, with a 5-year survival rate below 20 % [1,2]. The clinical management of osteosarcoma is further hampered by the dose-limiting toxicity and acquired resistance of conventional chemotherapy, as well as by the scarcity of validated molecularly targeted therapies [1,3]. Moreover, the key molecular events underlying the initiation and progression of osteosarcoma have not yet been fully elucidated. In particular, the contributions of glycosylation-related genes to osteosarcoma biology remain incompletely characterized, leaving their diagnostic and therapeutic relevance insufficiently defined. Elucidating the molecular mechanisms underlying osteosarcoma pathogenesis and identifying reliable diagnostic biomarkers and therapeutic targets therefore remain important research priorities [2,3]. Altered protein glycosylation is increasingly recognized as an important feature of cancer. Among its branches, fucosylation—the transfer of fucose residues onto glycoproteins and glycolipids—has attracted particular attention because of its reported involvement in tumor proliferation, metastasis, and immune evasion [4]. Fucosyltransferase 4 (*FUT4*; α-1,3-fucosyltransferase IV) is a member of the α-1,3-fucosyltransferase family and transfers fucose moieties from GDP-fucose to N-acetylglucosamine residues located in the core regions of glycoproteins and glycolipids.

Through this enzymatic activity, *FUT4* catalyzes the synthesis of Lewis X (LeX), sialyl-Lewis X (sLeX), and Lewis Y (LeY) antigens and contributes to cell adhesion, migration, and immune recognition [5]. In gastric cancer, cholangiocarcinoma, glioblastoma, and chordoma, *FUT4* has been implicated in malignant phenotypes such as proliferation, EMT, invasion, and tumor progression [5–8]. For example, *FUT4* cooperates with the Lewis Y antigen to sustain Helicobacter pylori-driven gastric carcinogenesis [5], enhances

EMT and aggressive phenotypes in chordoma cells [6], and sustains glioblastoma progression through a circPARP4/miR-125a-5p/*FUT4* axis [7]. A similar EMT-promoting role has been documented in cholangiocarcinoma [8], and fucosylation as a whole has been linked to metastatic spread and to the efficacy of immune checkpoint blockade in several tumors [9,10].

In osteosarcoma, *FUT4* has only recently received attention, and its biological and clinical relevance remains incompletely characterized. Three experimental studies have shown that *FUT4* knockdown suppresses osteosarcoma cell growth via *FOXO1* activation [11], that miR-371b-5p dampens migration and proliferation by targeting *FUT4* [12], and that the long non-coding RNA *MRUL* promotes osteosarcoma progression through the miR-125a-5p/*FUT4* axis [13]. Notably, although these cell-level knockdown experiments indicate a pro-tumorigenic role for *FUT4* in osteosarcoma, the tissue-level expression pattern of *FUT4* in osteosarcoma—whether it is up- or down-regulated relative to normal bone—remains unknown, creating a potential contradiction between cell-line functional observations and primary-tissue transcriptomic profiles. These prior reports focused primarily on cell-level functional assays, leaving the broader transcriptomic associations, regulatory context, immune features, and druggable potential of *FUT4* in osteosarcoma incompletely characterized. Accordingly, it remains unclear whether *FUT4* dysregulation is reproducible across independent osteosarcoma datasets and how *FUT4* expression relates to disease-associated pathways, regulatory networks, immune-cell features, and candidate small-molecule binding. To address these questions, we integrated two independent osteosarcoma transcriptomic cohorts and (i) profiled *FUT4* expression and discriminatory performance, (ii) characterized *FUT4*-associated biological pathways, (iii) constructed protein–protein and TF/miRNA/drug regulatory networks around *FUT4*, (iv) examined *FUT4*-associated immune-cell signature patterns, and (v) screened candidate *FUT4*-binding small molecules.

This framework provides a basis for subsequent experimental evaluation of *FUT4* as a candidate biomarker and potential therapeutic target in osteosarcoma.

Methods

Data acquisition and preprocessing

Two publicly available osteosarcoma transcriptomic datasets, GSE36001 and GSE16088, were retrieved from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) using the GEOquery R package (version 2.70.0) [14]. All samples retained for downstream analyses were derived from *Homo sapiens*. GSE36001 (platform GPL6102) contains 25 samples: 19 osteosarcoma cell lines and 6 normal bone/osteoblast controls. GSE16088 (platform GPL96, Affymetrix Human Genome U133A) contains 23 samples, all annotated as *Homo sapiens*, of which 14 osteosarcoma and 6 non-tumor controls (kidney, liver, and lymph node) entered the differential expression and ROC analyses. The remaining three samples (H1-012706, H2-012706, and H3-012706; GSM402684–GSM402686) were excluded because they are annotated in GEO as osteosarcoma cell lines and correspond to the U2OS, HOS and MG63 human osteosarcoma cell lines described in the source study (GEO; PMID 20028558), whereas the GSE16088 analysis was restricted to primary osteosarcoma tissue versus non-tumor tissue. Probe-level annotation, log2 transformation, and quantile normalization were carried out with the limma package (version 3.58.1) [15]. In total, 19 osteosarcoma and 6 control samples from GSE36001, and 14 osteosarcoma and 6 control samples from GSE16088, were retained for the downstream analyses (Table 1). Because no batch-effect correction was applied, each dataset was analysed independently and comparisons between cohorts were restricted accordingly (see Limitations).

Identification of differentially expressed genes

Differential expression analysis between osteosarcoma and control groups was performed with limma [15]. For GSE36001, genes with |log2 fold change (logFC)| > 0.5 and $p < 0.05$ were defined as differentially expressed genes (DEGs); this |logFC| > 0.5 threshold is a commonly applied cutoff in microarray-based osteosarcoma and cancer transcriptomic studies to filter out genes with negligible fold-change changes while retaining moderate-effect DEGs; for GSE16088, DEGs were defined solely by significance (adjusted $p < 0.05$; Benjamini–Hochberg correction), with no fold-change cutoff applied; the resulting DEG count is therefore substantially larger than that of GSE36001, and the two counts are not directly comparable. Because the two datasets differed in platform, control composition, and sample size, dataset-specific screening thresholds were applied as described above; cross-dataset interpretation therefore focused primarily on the consistent direction of *FUT4* dysregulation rather than on identical DEG counts. Volcano plots were generated with ggplot2 and heatmaps with pheatmap (version 1.0.12).

Validation of *FUT4* expression and ROC analysis

The expression levels of *FUT4* were compared between osteosarcoma and control samples using grouped comparison plots, and receiver operating characteristic (ROC) curves were generated with the R package pROC (version 1.18.5) [16] to evaluate the discriminatory performance of *FUT4* using the area under the curve (AUC). The 95% confidence intervals (CIs) of AUC were calculated by bootstrap resampling. An AUC between 0.5 and 0.7 was considered low, 0.7–0.9 moderate, and > 0.9 high accuracy. Subsequently, the osteosarcoma samples from GSE36001 (n = 19) were dichotomized into *FUT4*-high (n = 10) and *FUT4*-low (n = 9) subgroups using the median *FUT4* expression value as the cutoff, and differential analysis between these groups was performed as described above. Of note, the *FUT4*-high (n

= 10) and *FUT4*-low (n = 9) subgroups are very small; all downstream subgroup-level analyses (GSEA, GSVA, and ssGSEA) should therefore be regarded as strictly exploratory, hypothesis-generating analyses rather than definitive statistical findings. The top 15 significantly upregulated and top 15 significantly downregulated genes satisfying $|logFC| > 0.5$ and $p < 0.05$ were selected as *FUT4*-associated differentially expressed genes. Together with *FUT4* itself, a total of 31 genes were obtained.

GO and KEGG enrichment analysis

To further explore the relationships of biological process (BP), cellular component (CC), molecular function (MF) and KEGG pathways with osteosarcoma, Gene Ontology (GO) and KEGG enrichment analyses were conducted on the 30 *FUT4*-associated differentially expressed genes together with *FUT4* (31 genes in total) using the R package clusterProfiler (version 4.10.0) [17]. The background gene list for GO/KEGG enrichment was restricted to all genes detectable on the corresponding microarray platform rather than the whole-genome annotation set. GO terms were considered significant at an adjusted $p < 0.05$, and KEGG pathways [18] at a false discovery rate (FDR) $q < 0.25$, both after Benjamini–Hochberg correction, and the results are presented in Table 2.

GSEA and GSVA

For pathway-level analysis, gene set enrichment analysis (GSEA) [19] was performed. For GSEA, genes from the GSE36001 osteosarcoma samples were ranked by logFC between the *FUT4*-high and *FUT4*-low groups and analyzed using clusterProfiler (version 4.10.0) [17] against the Molecular Signatures Database (MSigDB) c2.all.v2024.1.Hs.symbols.gmt collection, with a random seed of 2022, 1,000 permutations, and a gene-set size range of 10–500. GSEA statistical inference was based on these permutations, and gene sets with adjusted $p < 0.05$ and FDR $q < 0.25$ after Benjamini–Hochberg correction were considered

significantly enriched. Sample-level pathway scores were computed using the GSVA package (version 1.50.0) [20] with the c2.cp.v2024.1.Hs.symbols.gmt gene-set collection, and the selected pathway patterns were visualized using a heatmap and grouped comparison plots. GSEA and GSVA used different MSigDB collections (c2.all and c2.cp, respectively) because GSEA requires the full C2 collection whereas GSVA was restricted to canonical pathways; the corresponding releases are stated above.

Differential scores between *FUT4*-high and *FUT4*-low groups were tested with limma [15], producing the logFC, average expression (AveExpr), moderated t-statistic, raw p value, BH-adjusted p value, and the log-odds of differential expression (B) reported in Table 4. Given the limited subgroup sample size, pathway selection was performed on nominal (uncorrected) $p < 0.05$ rather than BH-adjusted p-values (BH-adjusted values are provided in Table 4; see Statistical analysis). Pathways with nominal $p < 0.05$ were ranked by logFC, and the 10 most positively and 10 most negatively shifted pathways were carried forward. These results are strictly hypothesis-generating and should not be interpreted as statistically significant biological findings. The Mann–Whitney U test was used as a non-parametric check on the limma results.

Protein–protein interaction and regulatory network construction

To characterize the protein-interaction context of *FUT4*, a PPI network was retrieved from the STRING database [21] (https://string-db.org/), using a medium-confidence interaction score threshold of 0.40. Transcription factors (TFs) associated with *FUT4* were obtained from ChIPBase v3.0 [22] (https://rnasysu.com/chipbase3/), miRNAs associated with *FUT4* were obtained from ENCORI (formerly starBase) [23] (https://starbase.sysu.edu.cn/), and drug/compound associations were obtained from the Comparative Toxicogenomics Database (CTD) [24] (https://ctdbase.org/). The mRNA–TF, mRNA–miRNA, and mRNA–drug networks were visualized using Cytoscape [25]. Of note, the TF, miRNA, and

drug/compound associations retrieved from ChIPBase, starBase, and CTD simultaneously include experimentally validated interactions and computationally predicted/interpolated associations; the present study did not apply an additional filtering step to separate validated from predicted interactions. The networks should therefore be interpreted as a candidate resource requiring subsequent experimental validation.

Immune infiltration analysis

Single-sample GSEA (ssGSEA) [20,26] quantifies the relative activity of immune cell signature gene sets within individual samples. Marker signatures for 28 human immune cell subtypes were defined according to the published immune-cell signature set [26], and ssGSEA was used to calculate an enrichment score for each immune cell type in each osteosarcoma sample. Spearman correlation analysis was then used to assess pairwise relationships among immune cell signatures within the *FUT4*-high and *FUT4*-low groups, and correlation heatmaps were visualized using pheatmap (version 1.0.12).
These correlation analyses were not corrected for multiple testing, and no cross-validation with alternative deconvolution tools was performed (see Limitations).

Virtual screening and molecular docking

No experimentally resolved crystal structure of *FUT4* exists in the Protein Data Bank, so the AlphaFold2-predicted model AF-P22083-F1 was downloaded from the AlphaFold Protein Structure Database [27]. The predicted local-distance difference test (pLDDT) is generally high (> 70) across the catalytic domain, where pocket C1 is located, whereas low-confidence terminal regions were excluded during pocket prediction. Of note, this AlphaFold-derived structure has not been experimentally validated; all docking outcomes should therefore be treated strictly as computational predictions. After standard preprocessing, binding pockets were predicted with the CB-Dock2 web server [28], and the largest pocket (C1) was used as

the docking site. DrugCLIP [29], a contrastive protein–molecule representation learning method for virtual screening, was then used to rank candidate molecules from the input compound library by predicted affinity. The top-100 ranked candidates were further re-scored using AutoDock Vina. The input molecular library for DrugCLIP virtual screening was a curated commercial compound collection aggregating multiple vendor catalogs (e.g., Vitas-M Laboratory, InterBioScreen, Pharmeks, TargetMol, and TimTec). According to the Vina scoring criteria, scores above −5 kcal/mol indicate weak or negligible binding, scores between −7 and −5 kcal/mol represent moderate binding affinity, and scores below −7 kcal/mol denote strong binding. Of note, the AutoDock Vina score represents a heuristic empirical scoring metric rather than experimentally measured binding free energy.

Statistical analysis

Statistical analyses and R-based data processing were performed using R (version 4.3.3). The Mann–Whitney U test (Wilcoxon rank-sum test) was used as a non-parametric verification of selected GSVA pathway scores; differential expression of genes and GSVA pathway scores was assessed using the moderated t-test implemented in limma [15]; immune-cell relationships were assessed using Spearman rank correlation; and discriminatory performance was evaluated using ROC/AUC, with 95% CIs estimated by bootstrap resampling. Unless otherwise specified, all statistical tests were two-sided. In brief, BH-FDR multiple-testing correction was applied for the DEG, GO, KEGG, and GSEA analyses (adjusted $p < 0.05$); for GSVA, BH multiple-testing correction was computed but pathway selection relied on nominal p-values given the small subgroup sample size; BH-adjusted p-values are provided in Table 4 for reference.

Use of artificial intelligence tools

Generative AI tools were used only for language polishing and grammar revision, and were not involved in experimental design, data analysis or result interpretation.

Results

The overall analytical workflow of this study is illustrated in Fig. 1.

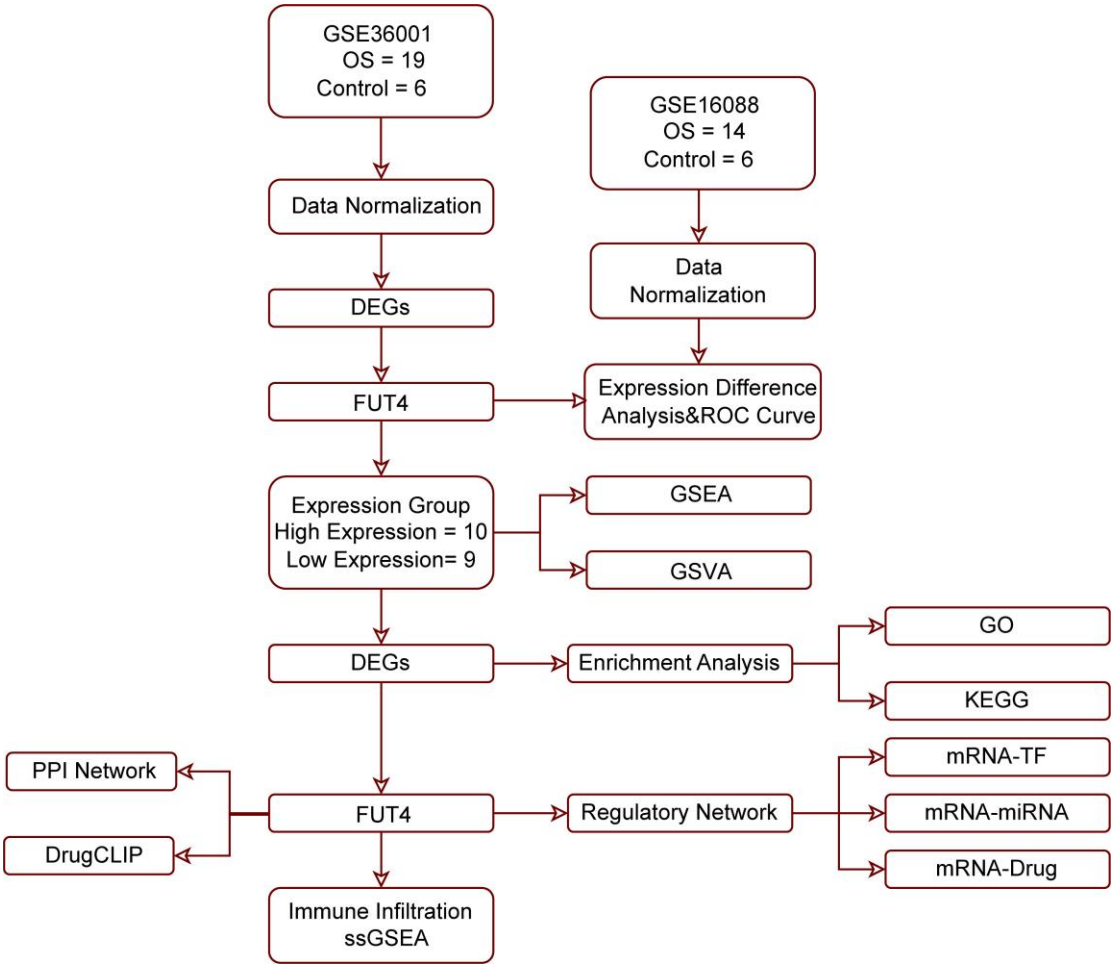


Fig. 1. Flow chart of the comprehensive bioinformatics analysis of *FUT4* in osteosarcoma. OS, osteosarcoma; DEGs, differentially expressed genes; ROC, receiver operating characteristic; GSEA, gene set enrichment analysis; GSVA, gene set variation analysis; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; PPI, protein–protein interaction; TF, transcription factor; ssGSEA, single-sample gene set enrichment analysis.

Differential expression and diagnostic value of *FUT4*

After preprocessing and normalization, the expression distributions became more comparable across samples (Fig. 2), and differential expression analysis identified 2,159 DEGs in GSE36001 (1,023 upregulated and 1,136 downregulated) and 10,884 DEGs in GSE16088

(4,311 upregulated and 6,573 downregulated) (Fig. 3); the larger DEG count in GSE16088 reflects the absence of a fold-change cutoff and the use of non-osseous control tissues in that dataset, and the two counts are therefore not directly comparable. *FUT4* transcript levels were consistently lower in osteosarcoma than in control samples in both cohorts (GSE36001, $p < 0.05$; GSE16088, $p < 0.01$) (Fig. 4A–B). ROC analysis showed an AUC of 0.789 (95% CI 0.590 – 0.989) in GSE36001 and 0.905 (95% CI 0.765 – 1.000) in GSE16088, indicating moderate and high discriminatory performance, respectively (Fig. 4C–D). Note that the control samples in GSE16088 are non-osseous tissues; the high AUC from this dataset may be confounded by tissue-origin differences, and cross-dataset validation focuses on the consistent direction of *FUT4* dysregulation.

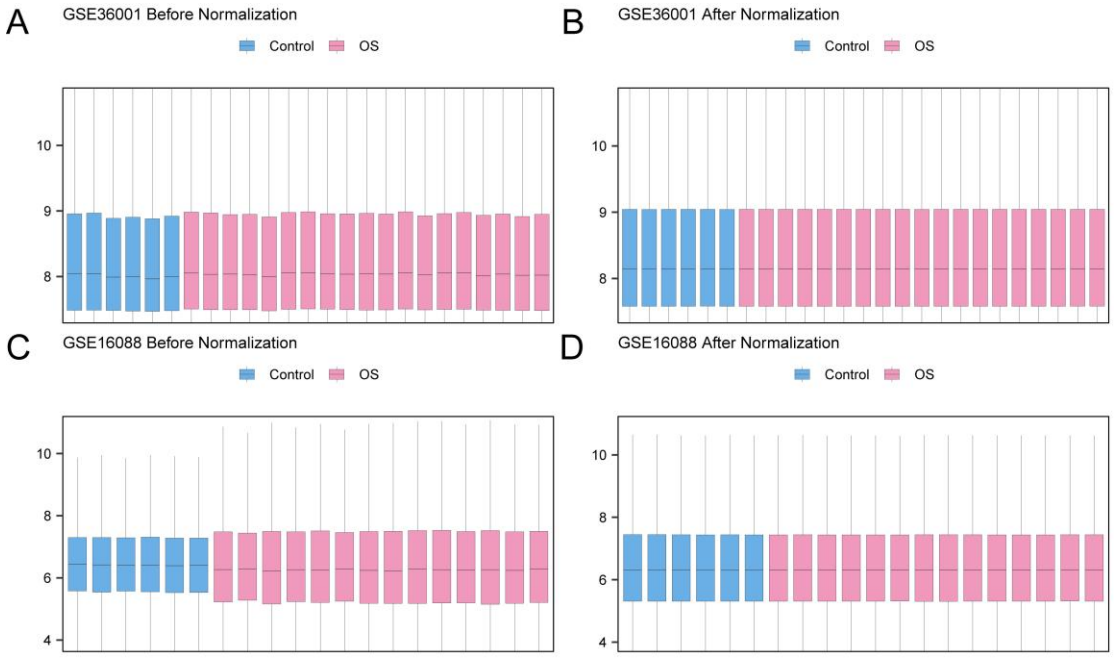


Fig. 2. Expression distributions of the two datasets before and after normalization. Panels show GSE36001 before (A) and after (B) normalization and GSE16088 before (C) and after (D) normalization. Pink bars correspond to osteosarcoma specimens, and blue bars indicate control samples.

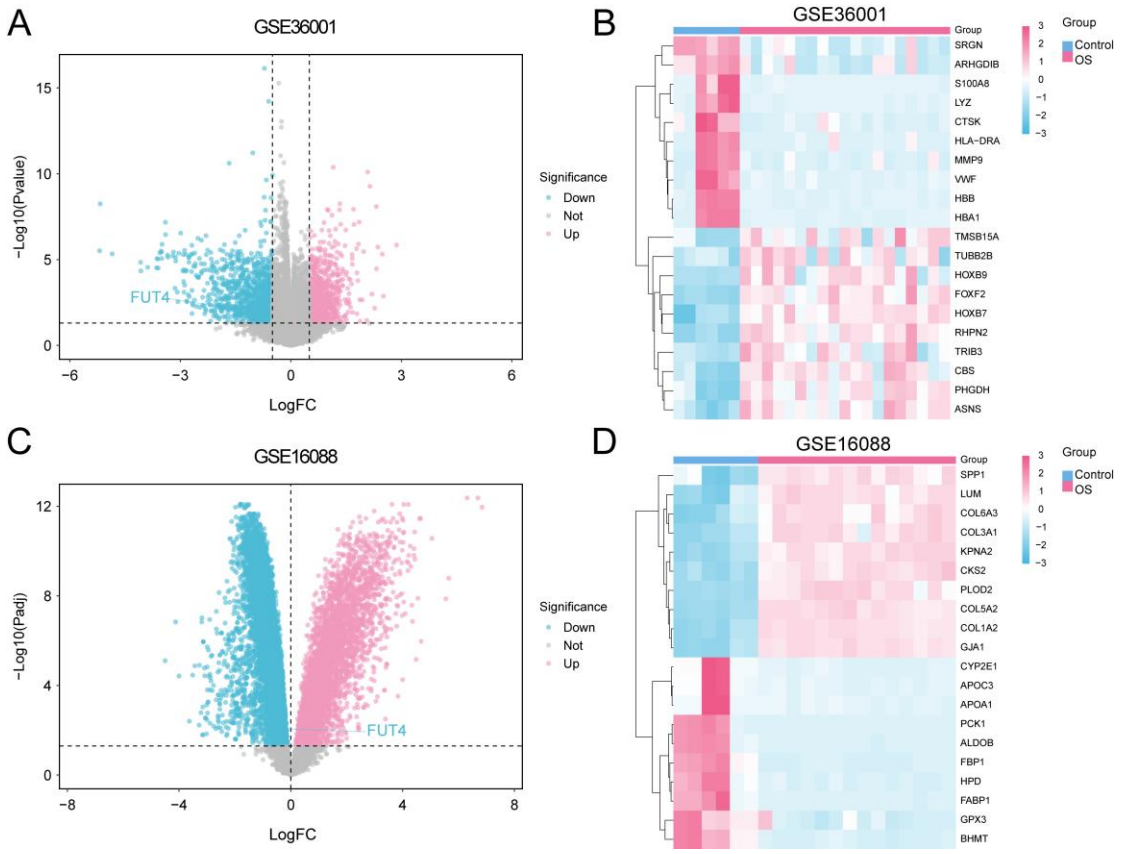


Fig. 3. (A) Volcano plot of differentially expressed genes (DEGs) between osteosarcoma and control groups in GSE36001. (B) Heatmap of the top 10 upregulated and top 10 downregulated DEGs in GSE36001. (C) Volcano plot of DEGs in GSE16088. (D) The corresponding heatmap for GSE16088. On heatmaps, blue stands for control samples and pink for osteosarcoma samples.

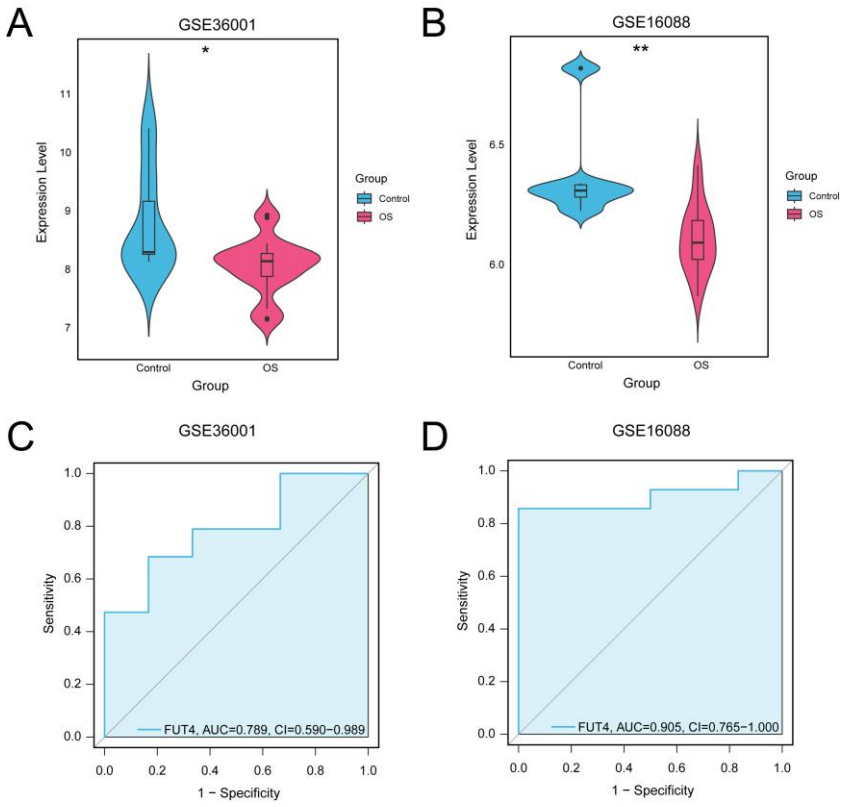


Fig. 4. (A, B) Group comparison plots of *FUT4* expression in GSE36001 (A) and GSE16088 (B). (C, D) ROC curves of *FUT4* in GSE36001 (C) and GSE16088 (D). Group comparisons were performed using the Mann–Whitney U test. Statistical significance: * $p < 0.05$; ** $p < 0.01$. ROC, receiver operating characteristic; AUC, area under the curve.

Identification of *FUT4*-associated differentially expressed genes and functional enrichment

GSE36001 osteosarcoma samples were split at the median *FUT4* level into *FUT4*-high and *FUT4*-low groups, yielding 297 DEGs (147 up, 150 down) between them (Fig. 5; Table S1). Applying these criteria, the 15 most upregulated and 15 most downregulated DEGs (30 genes in total) were then designated *FUT4*-associated differentially expressed genes. The 30 *FUT4*-associated differentially expressed genes together with *FUT4* (31 genes in total) were subjected to GO and KEGG enrichment analysis to further explore their relationships with osteosarcoma. GO enrichment analysis showed significant enrichment of this gene set in

biological processes including ossification, osteoblast differentiation, biomineral tissue development, and neuron migration; cellular components including the collagen-containing extracellular matrix, endoplasmic reticulum lumen, and basement membrane; and the molecular function term extracellular matrix structural constituent (Table 2; Fig. 6A). KEGG analysis further showed enrichment of this gene set in extracellular matrix (ECM)–receptor interaction, focal adhesion, and phagocytosis pathways (Fig. 6A). The corresponding networks of the significantly enriched BP, CC, MF, and KEGG terms are shown in Fig. 6B–E.

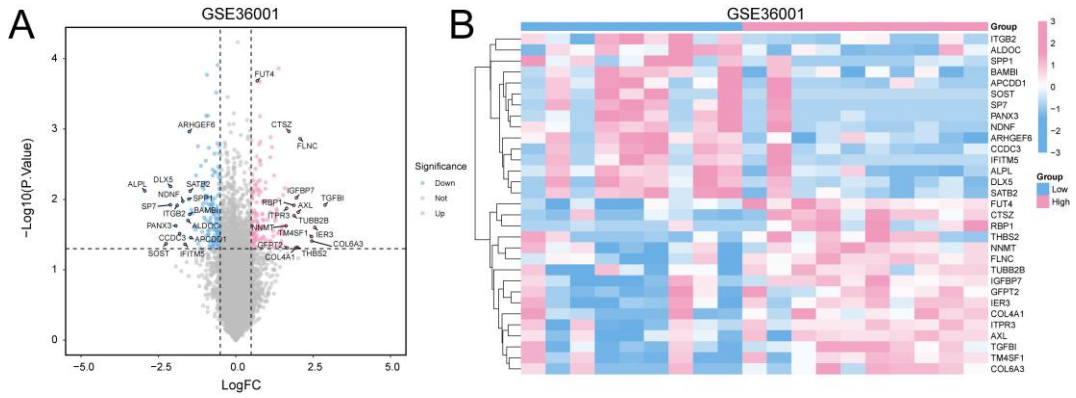


Fig. 5. Differential gene expression between *FUT4*-high and *FUT4*-low groups in GSE36001. (A) Volcano plot of differentially expressed genes (DEGs). (B) Heatmap of the 31 selected genes (15 up, 15 down, plus *FUT4*).

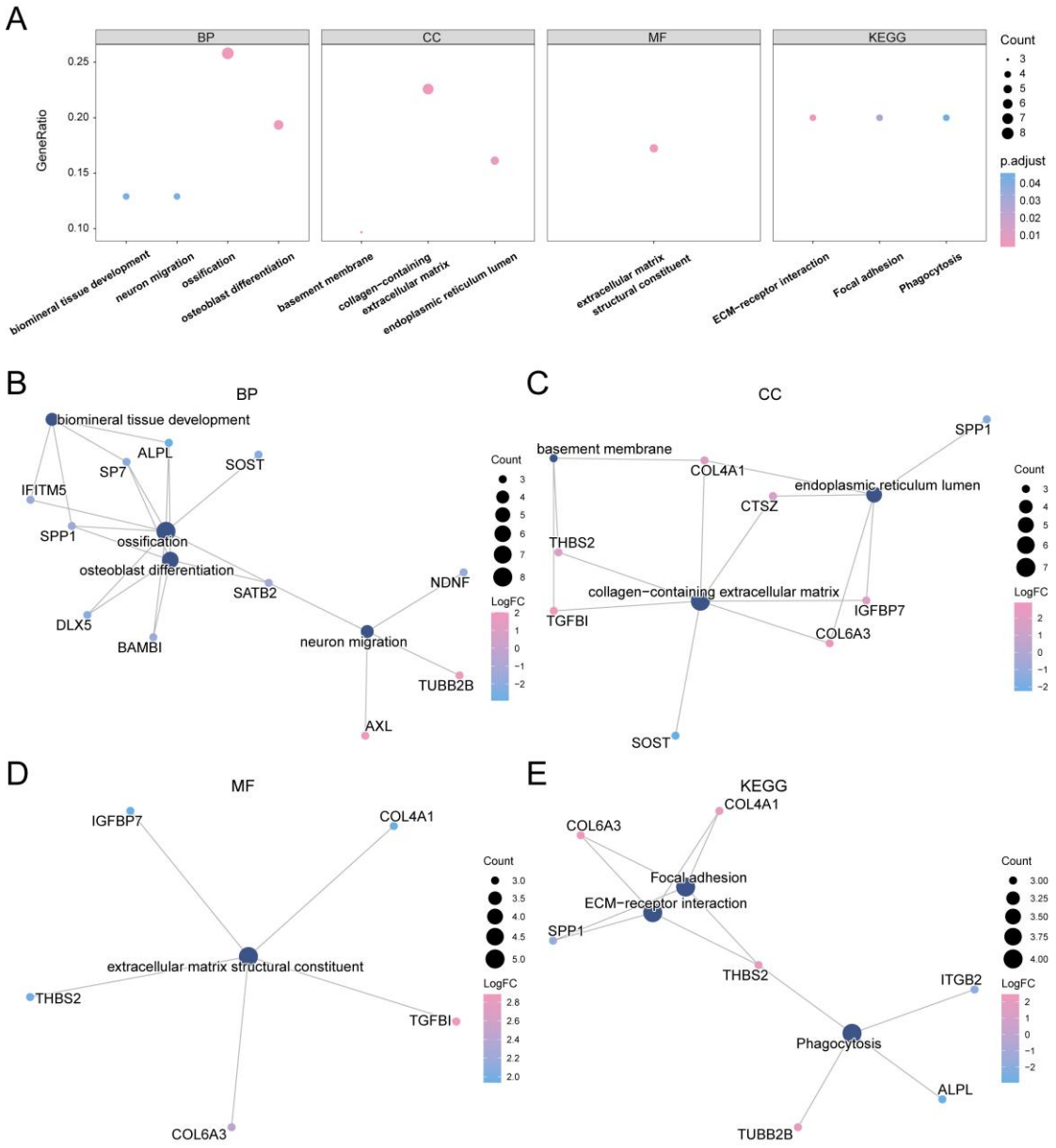


Fig. 6. GO and KEGG enrichment analysis of *FUT4*-associated genes. (A) Bubble plot of biological process (BP), cellular component (CC), molecular function (MF), and KEGG pathways. (B–E) Network diagrams of BP (B), CC (C), MF (D), and KEGG (E). Enrichment thresholds: adjusted $p < 0.05$ for GO terms (BP, CC, MF) and FDR (q value) < 0.25 for KEGG pathways, both after Benjamini–Hochberg correction.

GSEA and GSVA reveal *FUT4*-associated pathways

Given the very small *FUT4*-high (n = 10) and *FUT4*-low (n = 9) subgroups, both GSEA and GSVA results should be interpreted as exploratory, hypothesis-generating observations. Of

note, the two gene sets with the highest normalized enrichment scores in Table 3, FOROUTAN_PRODRANK_TGFB_EMT_UP (NES = 2.07) and PLASARI_TGFB1_TARGETS_10HR_UP (NES = 2.03), describe closely related TGF-β/EMT transcriptional programs; the Foroutan Integrated TGFβ EMT Up, Sarrio Epithelial Mesenchymal Transition Up, and Zhou Inflammatory Response Live Up sets were selected as representative examples for visualization because, among the significantly enriched C2:CGP sets, many large immune- and metabolic-oriented collections are less directly relevant to osteosarcoma biology, whereas these three sets represent two complementary osteosarcoma-relevant programs (TGF-β/EMT transition and inflammatory activation). GSEA showed positive enrichment of Foroutan Integrated TGFβ EMT Up (NES = 2.01, adjusted $p = 4.29 \times 10^{-6}$), Sarrio Epithelial Mesenchymal Transition Up (NES = 1.93, adjusted $p = 2.31 \times 10^{-5}$), and Zhou Inflammatory Response Live Up (NES = 1.40, adjusted p = 0.0272) in the *FUT4*-high group, whereas Reactome Iron Uptake and Transport was negatively enriched (NES = −1.76, adjusted p = 0.0137) (Table 3; Fig. 7). GSVA prioritized 20 pathways with nominally altered activity (nominal $p < 0.05$), including pathways related to actin-cytoskeleton regulation, apoptotic signaling, extracellular-matrix interactions, immune response, and Wnt/β-catenin-related transcriptional regulation (Table 4; Fig. 8). Although BH-adjusted p-values were calculated for all GSVA pathways, only one pathway remained statistically significant after multiple-testing correction; these outputs are therefore discussed as exploratory only.

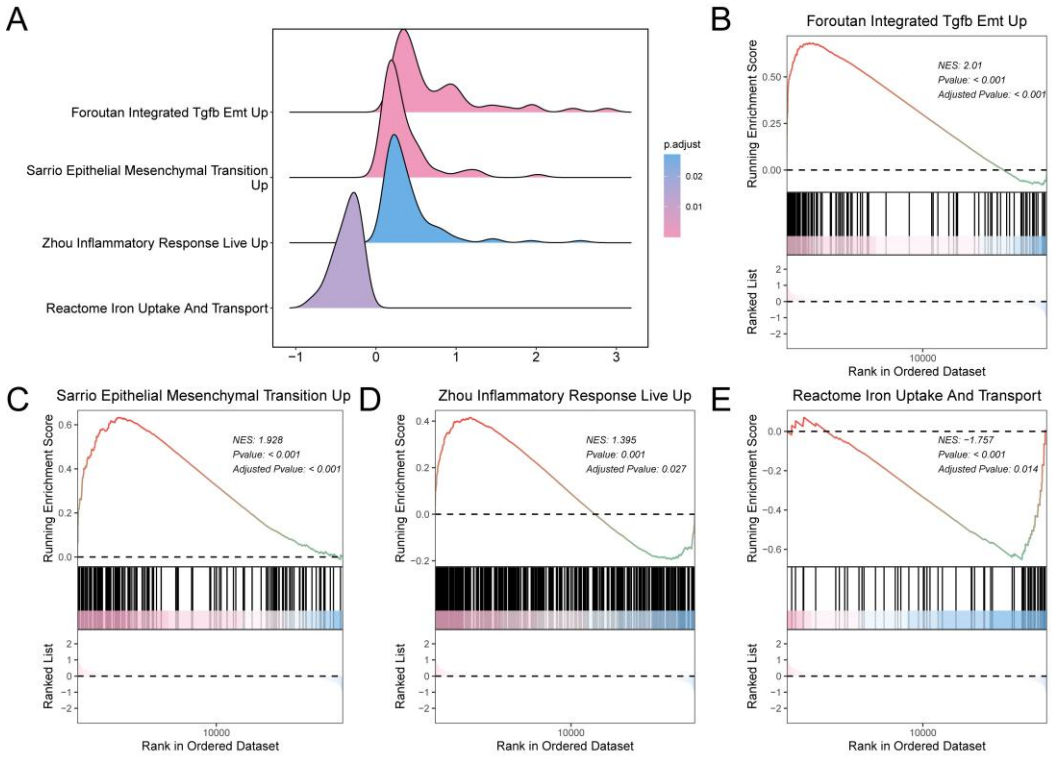


Fig. 7. GSEA for *FUT4* expression groups. (A) Overview ridge plot of the significantly enriched gene sets. (B–E) Gene set enrichment analysis (GSEA) enrichment plots of Foroutan Integrated TGFβ EMT Up (B), Sarrio Epithelial Mesenchymal Transition Up (C), Zhou Inflammatory Response Live Up (D), and Reactome Iron Uptake and Transport (E). Statistical significance thresholds are as defined in Methods.

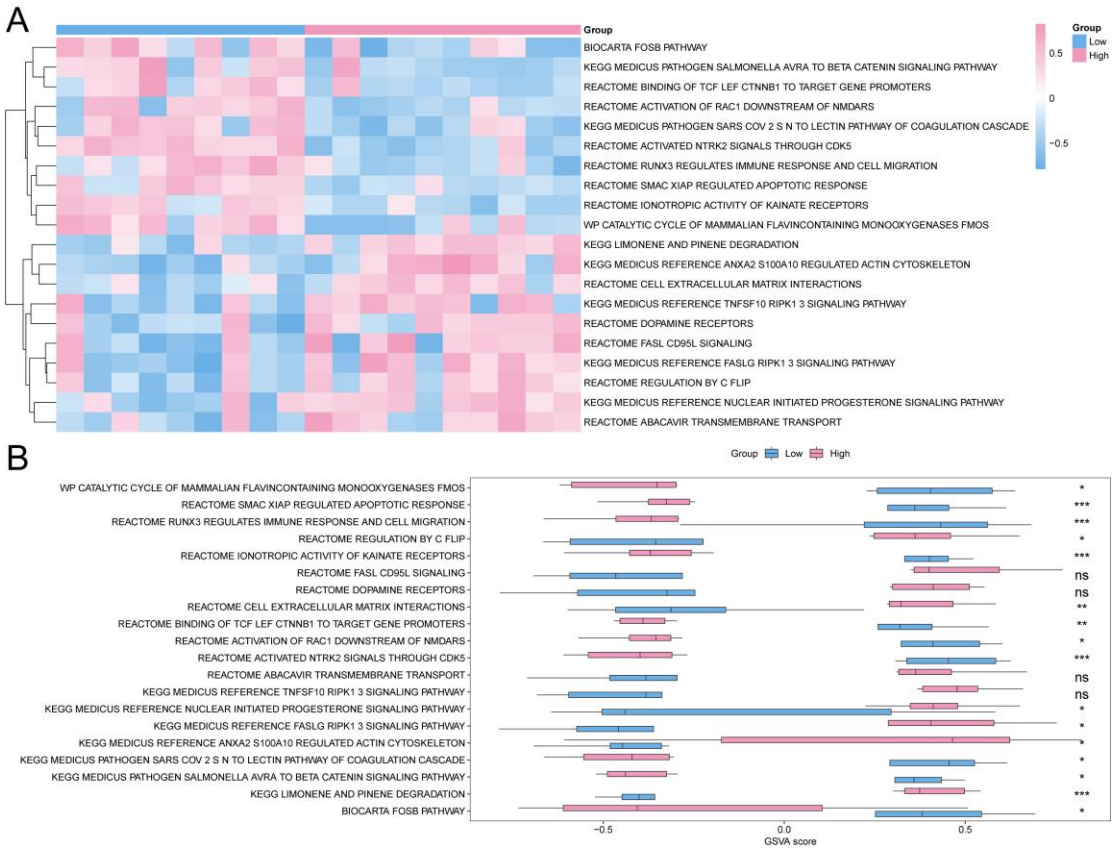
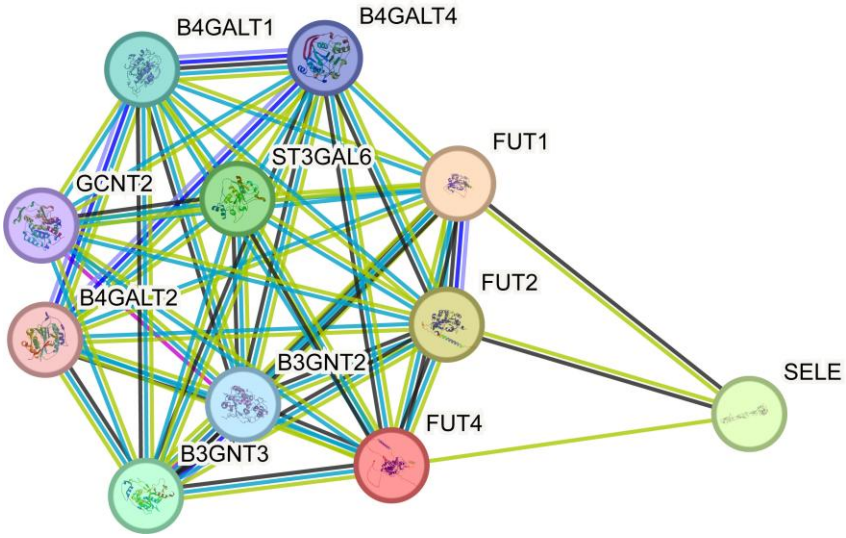


Fig. 8. GSVA between *FUT4*-low and *FUT4*-high groups. (A) Heatmap of the top 20 pathways with nominally altered activity. (B) Grouped comparison plot. ns: $p \geq 0.05$ (non-significant); * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Pathways shown in the heatmap comprise the 10 most positively and 10 most negatively logFC-shifted pathways among those with nominal $p < 0.05$; BH-adjusted p-values were also calculated and are reported in Table 4 for reference. GSVA, gene set variation analysis.

PPI and regulatory networks of *FUT4*

At the STRING medium-confidence interaction score threshold of 0.40, *FUT4* had 187 direct interaction partners, and the 10 with the highest combined scores were retained: *B3GNT2*, *B3GNT3*, *B4GALT1*, *B4GALT2*, *B4GALT4*, *FUT1*, *FUT2*, *GCNT2*, *ST3GAL6*, and *SELE*. Together with *FUT4*, these 10 partners formed an 11-node PPI network (Fig. 9). Only direct interaction partners were included in this network, without secondary interactors. ChIPBase, ENCORI, and CTD identified 51 *FUT4*-associated transcription factors (Table S2), 57

352 miRNAs (Table S3), and 46 drug/compound associations (Table S4), respectively, which
353 were summarized in the corresponding regulatory networks in Fig. 10.



354

355 Fig. 9. STRING-derived protein–protein interaction (PPI) network of *FUT4* (interaction score
356 ≥ 0.40). Nodes represent proteins and edges represent STRING-supported protein
357 associations; the 10 top-scoring direct interaction partners are listed in the Results.

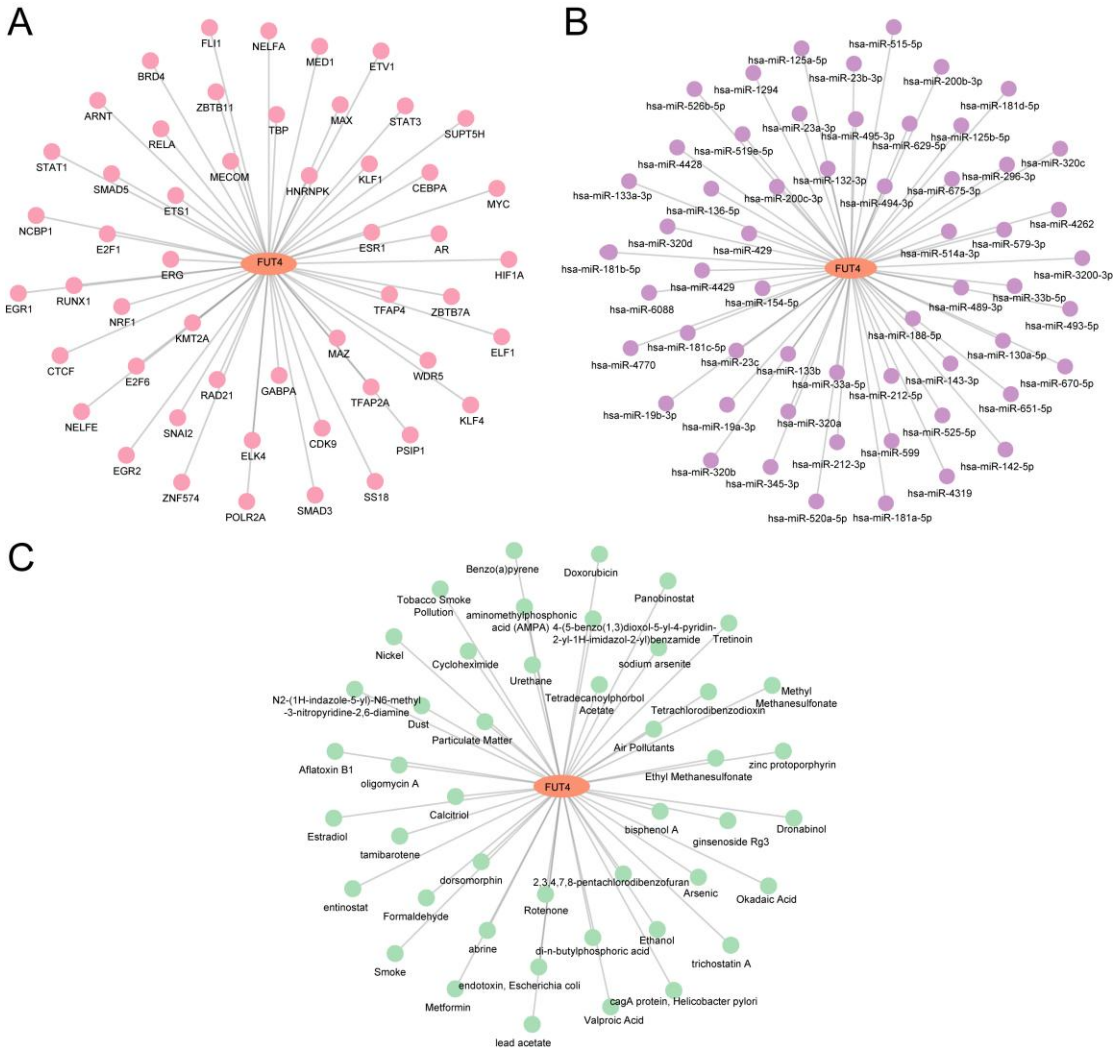


Fig. 10. *FUT4*-centered regulatory networks. (A) mRNA–TF network comprising *FUT4* and 51 transcription factors. (B) mRNA–miRNA network comprising *FUT4* and 57 miRNAs. (C) mRNA–drug/compound network comprising *FUT4* and 46 drug/compound associations. Orange denotes mRNA, pink denotes transcription factors, purple denotes miRNAs, and green denotes drugs/compounds. TF, transcription factor.

Immune infiltration landscape of *FUT4* expression groups

ssGSEA was applied to score 28 immune cell subsets in each osteosarcoma sample (Fig. 11). Correlation heatmaps revealed distinct descriptive immune-cell signature association patterns in the two subgroups. In the *FUT4*-low group, natural killer T cells correlated most strongly with effector memory CD4 T cells ($r = 0.967$, $p < 0.05$); in the *FUT4*-high group, immature

dendritic cells and CD56bright natural killer cells exhibited the highest positive correlation (r = 0.903, $p < 0.05$).

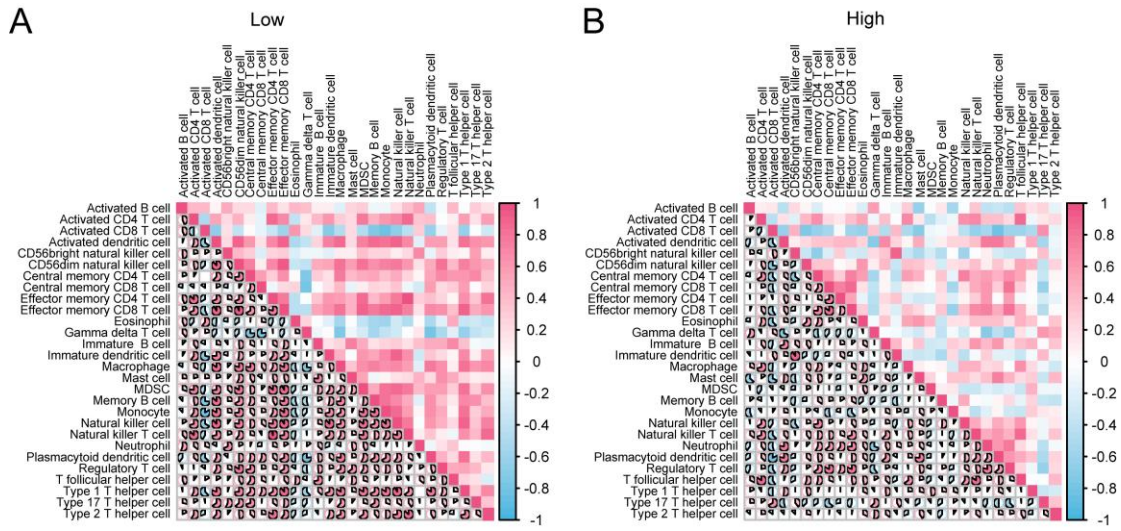


Fig. 11. Immune infiltration analysis by ssGSEA. (A) *FUT4*-low group; (B) *FUT4*-high group. Correlation strength criteria: $|r| < 0.3$, negligible; $0.3 \leqslant |r| < 0.5$, weak; $0.5 \leqslant |r| < 0.8$, moderate; $|r| \geqslant 0.8$, strong. ssGSEA, single-sample gene set enrichment analysis. These correlations were not corrected for multiple testing (see Methods and Limitations).

Virtual screening and molecular docking of candidate small molecules against *FUT4*

DrugCLIP ranked the top 100 compounds with the highest predicted *FUT4* affinity (Table S5). The largest predicted binding pocket, C1, occupied a volume of 3,477 Å^3 and was centered at (3, −12, −3) with dimensions of roughly 30 × 19 × 23 Å. ST50926609 ranked first by DrugCLIP predicted affinity among the screened candidates, with a Vina score of −7.35 kcal/mol (Fig. 12). The predicted docking pose showed hydrogen-bond, hydrophobic, and halogen-bond interactions between *FUT4* and ST50926609, supporting its prioritization as a candidate *FUT4*-binding compound (Fig. 12).

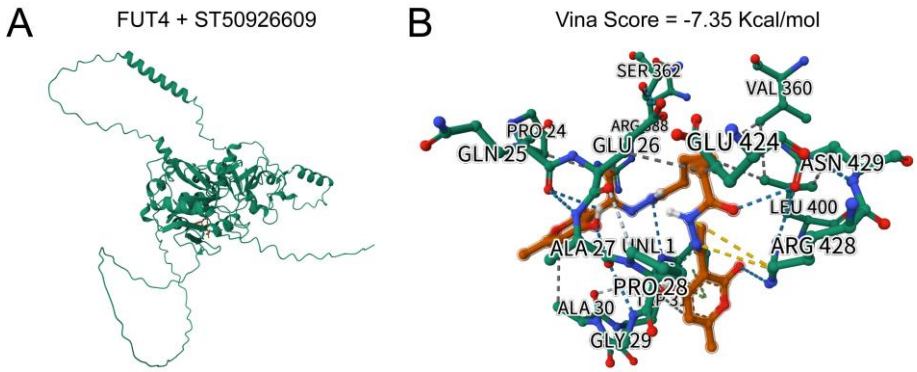


Fig. 12. Molecular docking of *FUT4* with the candidate compound ST50926609. ST50926609 is a computationally predicted binder, not an experimentally validated inhibitor; docking used the AlphaFold2-predicted *FUT4* model described in Methods. (A) Global docking map. (B) Interaction force map. Blue dashed lines indicate hydrogen bonds, gray dashed lines denote hydrophobic interactions, and cyan dashed lines indicate halogen bonds.

Discussion

Osteosarcoma, the most common primary malignant bone tumor, predominantly affects children and adolescents and remains challenging to treat. Current therapeutic options are limited, highlighting an unmet clinical need for reliable biomarkers and actionable therapeutic targets. In this work, we evaluated the biological and clinical relevance of fucosyltransferase 4 (*FUT4*) in osteosarcoma through an integrative bioinformatics framework that combined cross-dataset validation across two independent GEO datasets (GSE36001 and GSE16088), multidimensional functional enrichment, multilayer regulatory network analysis, immune infiltration assessment, and in-silico virtual screening of small-molecule candidates. A key strength of the present study lies in cross-validation using two independent transcriptomic datasets and a multi-layer analytical workflow covering gene-level expression signatures, putative regulatory mechanisms, and potential drug-target discovery.

Several observations from our computational analyses merit further discussion, namely the direction of *FUT4* dysregulation relative to its reported cell-level functions, the biological processes associated with *FUT4* expression status, the regulatory and immune-cell features recovered from the network analyses, and the candidate *FUT4*-binding compound identified by structure-based screening. Each is considered below.

The first point concerns the direction of *FUT4* dysregulation. Although elevated *FUT4* expression has been reported in gastric and biliary tract cancers, glioblastoma, and chordoma [5–8], our transcriptomic analysis showed significantly lower *FUT4* expression in osteosarcoma than in the included control samples in both datasets (GSE36001, $p < 0.05$; GSE16088, $p < 0.01$). ROC analyses showed that *FUT4* could distinguish osteosarcoma specimens from control specimens within the included datasets, with AUCs of 0.789 in GSE36001 and 0.905 in GSE16088 (95% CIs in Fig. 4C–D). The consistent direction of change across two independent microarray platforms argues against a purely technical artefact. However, this direction differs from that expected from the reported pro-tumorigenic role of *FUT4*: *FUT4* knockdown has been shown to suppress osteosarcoma cell growth via *FOXO1* activation, and miR-371b-5p exerts suppressive effects by targeting *FUT4* in osteosarcoma [11–13]. Several non-exclusive explanations may reconcile these observations: (i) the GSE36001 osteosarcoma group consists of cell lines, whose *FUT4* levels may not reflect those of primary tumors; (ii) the GSE16088 controls are non-osseous tissues, so between-group differences may in part reflect tissue of origin; and (iii) within osteosarcoma, the *FUT4*-high subgroup nonetheless showed enrichment of EMT- and inflammation-related programmes, which is consistent with a pro-tumorigenic role for *FUT4* in an expression-range-dependent manner. These discrepancies underscore that the direction of *FUT4* dysregulation in osteosarcoma requires validation in matched primary tumor and normal-bone specimens.

Nevertheless, important caveats must be considered when interpreting the ROC results. The two datasets employ dissimilar control tissues: GSE36001 uses normal bone and osteoblast controls, whereas GSE16088 uses kidney, liver, and lymph-node tissues as non-tumor controls. Accordingly, the high AUC observed in GSE16088 may partly reflect inter-tissue transcriptomic differences, and the clinical discriminatory performance of *FUT4* requires validation in cohorts with appropriately matched normal-bone controls. Our regulatory-network analysis suggests that transcriptional regulation may represent one potential contributor to *FUT4* dysregulation; however, other possible mechanisms, including copy-number and epigenetic alterations, were not evaluated in the present study. Our regulatory-network analysis predicted 51 transcription factors potentially binding to the *FUT4* locus, including well-known oncogenic regulators such as *MYC*, *STAT3* and *HIF1A*; however, these are purely computational predictions and require epigenetic and chromatin-based experimental validation.

Furthermore, owing to incomplete clinical metadata in the public GEO datasets, we could not assess correlations between *FUT4* expression and clinicopathological parameters such as tumor grade, clinical stage, or metastatic status. The transcriptomic analyses were based on publicly available expression datasets and provide no evidence that *FUT4* can serve as a circulating serum-based biomarker. Future investigations should benchmark *FUT4* against established osteosarcoma biomarkers and validate its sensitivity, specificity, and prognostic value using well-annotated clinical patient cohorts.

Functional enrichment of genes differentially expressed between the *FUT4*-high and *FUT4*-low groups highlighted biological processes including ossification, osteoblast differentiation, biomineral tissue development, and neuron migration. The most prominent cellular-component terms pointed toward collagen-containing extracellular matrix, endoplasmic reticulum lumen and basement membrane, while “extracellular matrix structural

constituent" represented the top molecular-function term. KEGG results indicated enrichment of ECM-receptor interaction, focal adhesion and phagocytosis. In addition, GSEA revealed that *FUT4*-high samples displayed enriched gene sets for TGF-β-driven EMT and inflammatory-response signatures.

Osteosarcoma originates from osteoblast-lineage progenitor cells. These enrichment results raise the hypothesis that *FUT4* dysregulation, if accompanied by altered fucosylation activity, may be linked to osteogenic differentiation and ECM–receptor signaling during osteosarcoma progression. Aberrant fucosylation may also be associated with EMT- and inflammation-related processes, providing a potential link between *FUT4* dysregulation and malignant osteosarcoma phenotypes. These interpretations remain speculative, and wet-laboratory experiments are essential to clarify how *FUT4*-dependent fucosylation modulates osteoblast maturation, matrix mineralization, EMT-associated transcription programs, and inflammatory signaling within the osteosarcoma cellular ecosystem.

GSVA prioritized 20 pathways exhibiting nominally altered activity between *FUT4*-high and *FUT4*-low subgroups (nominal $p < 0.05$, Table 4), encompassing actin-cytoskeleton signaling, apoptotic cascades, and the Wnt/β-catenin axis. Although Benjamini–Hochberg multiple-testing correction was computed for all GSVA outputs, only one pathway remained statistically significant after adjustment, an expected consequence of the substantial statistical noise inherent in high-dimensional GSVA scores estimated from such small subgroups. Therefore, these GSVA outputs should be interpreted strictly as hypothesis-generating observations rather than definitive biological findings. This exploratory multi-pathway pattern suggests that *FUT4* expression may be associated with several biological modules related to cytoskeletal regulation, apoptosis, and Wnt/β-catenin signaling, providing candidate pathways for further experimental investigation. Mechanistically, fucosylation modifications mediated by *FUT4* could alter actin polymerization and cytoskeletal dynamics,

thereby influencing tumor-cell motility. *FUT4*-associated glycosylation may also be linked to apoptotic signaling and tumor-cell survival, although this possibility requires direct functional validation. Moreover, crosstalk between aberrant fucosylation and canonical Wnt/β-catenin signaling merits dedicated experimental exploration. Collectively, the transcriptomic analyses link *FUT4* expression to EMT, inflammatory, apoptosis-related, and cytoskeletal pathway signatures in osteosarcoma, although the functional relationships among these processes remain to be established. Again, these mechanistic models require direct experimental validation.

STRING-based PPI analysis identified ten *FUT4*-associated proteins, most of which belong to glycosyltransferase or glycan-binding protein families.These STRING associations place *FUT4* within a broader network containing several glycosylation-related proteins, although the functional relationships among these proteins in osteosarcoma remain to be validated experimentally. The oncogenic roles of these cooperating glycosylation-related partners also deserve further experimental characterization.

Beyond the PPI network, the in-silico analyses identified 51 *FUT4*-associated transcription factors, 57 miRNAs, and 46 drug/compound associations, forming complementary regulatory and compound-interaction networks. Among the predicted miRNA associations, miR-371b-5p and miR-125a-5p have been functionally linked to *FUT4* in osteosarcoma [12,13], and miR-125a-5p to *FUT4* in glioblastoma [7], providing independent literature support for part of the miRNA–*FUT4* network identified here. These regulatory networks provide a rich resource for future mechanistic research investigating transcriptional/post-transcriptional control of *FUT4* and prospective drug repurposing opportunities.

ssGSEA analysis of bulk transcriptomes revealed different correlation patterns among 28 immune-cell signature scores in the *FUT4*-low and *FUT4*-high groups. Within the *FUT4*-low subgroup the strongest positive correlation involved natural killer T cells and effector

503 memory CD4+ T cells, whereas in the *FUT4*-high subgroup it involved immature dendritic
504 cells and CD56bright natural-killer cells (Fig. 11A–B).These findings extend the
505 transcriptomic characterization of *FUT4* in osteosarcoma by identifying *FUT4* expression-
506 associated immune-cell signature correlation patterns. These correlations are purely
507 descriptive and hypothesis-generating and do not imply a causal role for *FUT4* in immune
508 regulation.
509 In other tumor systems, dysregulated fucosylation (e.g., elevated sialyl-Lewis X) impairs
510 immune recognition and fosters immune evasion, and modulates responses to
511 immune-checkpoint-blockade therapy [9,26]. Given the reported links between fucosylation
512 and tumor immunity in other cancers [9,26], the relationship between *FUT4* dysregulation,
513 immune-cell states, and immunotherapy response in osteosarcoma warrants investigation in
514 appropriately designed experimental and clinical cohorts. Immunotherapy is still an emerging
515 therapeutic modality for osteosarcoma. If experimentally validated, *FUT4* may in the future
516 assist patient stratification or inform combination-regimen design. Nevertheless, all
517 above-mentioned immune-related hypotheses require rigorous in-vitro, in-vivo and
518 clinical-cohort validation before any translational consideration.
519 Combining DrugCLIP virtual screening and AutoDock-Vina re-scoring, ST50926609 was
520 selected as the lead candidate compound targeting the largest predicted binding pocket C1
521 (volume = 3,477 Å^3) of the AlphaFold-predicted *FUT4* structure. It achieved a Vina score of
522 −7.35 kcal/mol, and the predicted docking pose contained hydrogen-bond, hydrophobic, and
523 halogen-bond interactions with *FUT4*. This molecule offers a valuable chemical scaffold for
524 future rational design of *FUT4*-targeted small-molecule agents. ST50926609 therefore
525 represents a prioritized computational candidate for subsequent biochemical binding, activity,
526 and cellular validation.

Limitations

Several caveats of this bioinformatics-based study should be acknowledged.

- The principal biological findings are computational and have not been validated experimentally; the functional and clinical relevance of *FUT4* in osteosarcoma requires further evaluation in cell models, animal models, and patient-derived specimens.
- The sample sizes of the analyzed datasets are relatively small, and clinical follow-up and prognostic metadata are largely absent, prohibiting survival-related analyses. In particular, although BH-adjusted p-values were calculated for all GSVA outputs, pathway prioritization was based on nominal $p < 0.05$ because of the limited subgroup sample size, and most selected pathways did not remain significant after multiple-testing correction.
- Despite expression preprocessing and normalization, inter-dataset heterogeneity related to different microarray platforms and sample sources may still influence the analytical results.
- GO, KEGG and GSEA depend on existing public annotation databases and may miss unannotated genes or poorly characterized biological pathways.
- ssGSEA-derived immune-cell signature enrichment scores are computational proxies rather than direct measurements of cell abundance and cannot fully recapitulate findings from immunohistochemistry, flow cytometry, or single-cell sequencing.
- Virtual screening and molecular-docking outputs constitute purely in-silico predictions; only one candidate compound was subjected to detailed docking analysis, and biochemical validation is absent.
- The GSE36001 osteosarcoma samples consist of cell-line specimens rather than primary patient tumors, whereas only primary osteosarcoma specimens were retained from GSE16088, and the control samples in GSE16088 are derived from non-osseous tissues;

because the two datasets differ in tumor as well as control specimen type, cross-dataset comparison was restricted to the direction of *FUT4* dysregulation, and no between-dataset effect sizes or DEG counts were compared. These sample-source differences also limit direct extrapolation to primary osteosarcoma biology and clinical diagnostic performance.

- Each *FUT4* expression subgroup comprised only 10 and 9 samples, so the 28 × 27/2 = 378 pairwise immune-signature correlations are statistically unstable and should be regarded as exploratory observations rather than robust estimates.
- In addition, no cross-dataset batch-effect correction was applied; datasets were analyzed independently, and cross-dataset comparisons were restricted only to the direction of *FUT4* expression alteration. Pair-wise Spearman correlation analyses of immune-cell signatures were performed without multiple-testing correction owing to limited sample size; these correlation patterns remain exploratory. We also did not conduct cross-validation of immune cell estimates using alternative deconvolution tools such as CIBERSORT or xCell. Functional enrichment was performed against a platform-restricted background, which may omit genes absent from the array annotation.
- The TF, miRNA, and drug/compound regulatory networks retrieved from ChIPBase, ENCORI, and CTD combine experimentally validated interactions with computationally predicted associations without additional filtering to distinguish between the two categories; thus, only a subset of these network edges (e.g., miR-371b-5p and miR-125a-5p, supported by prior functional studies) has direct experimental support.
- The 31-gene diagnostic panel was derived from a single database screen and has not been evaluated prospectively; its diagnostic performance therefore requires independent validation in external cohorts before any clinical use.

Future studies should therefore validate *FUT4* in independent primary osteosarcoma cohorts with matched normal-bone controls, investigate its biological functions using cell and animal models, and experimentally evaluate the predicted immune associations and *FUT4*–ST50926609 interaction.

Conclusions

Taken together, our multi-dataset bioinformatic analysis supports further evaluation of *FUT4* as a candidate osteosarcoma biomarker and a putative target for small-molecule intervention. *FUT4* showed consistently lower expression in osteosarcoma than in control samples and showed moderate-to-high discriminatory performance across the two datasets, although the estimate in GSE16088 is influenced by the use of non-osseous control tissues. *FUT4*-associated transcriptomic patterns involve bone-related development and ECM-related processes, together with EMT, inflammatory, cytoskeletal, apoptosis-related, and Wnt/β-catenin pathway signatures. PPI and multilayer regulatory analyses identified multiple *FUT4*-associated molecular networks, while ssGSEA revealed distinct immune-cell signature correlation patterns between *FUT4* expression groups rather than direct changes in absolute immune-cell abundance. Structure-based virtual screening prioritized ST50926609 as the lead candidate *FUT4*-binding compound. These computational findings support further investigation of *FUT4* in osteosarcoma, while experimental studies in cell and animal models and validation in well-characterized clinical cohorts are required before diagnostic or therapeutic translation.

Abbreviations

AUC, area under the curve; BP, biological process; BH, Benjamini–Hochberg; CC, cellular component; CTD, Comparative Toxicogenomics Database; CI, confidence interval; DEG, differentially expressed gene; ECM, extracellular matrix; EMT, epithelial–mesenchymal

transition; FDR, false discovery rate; *FUT4*, fucosyltransferase 4; GEO, Gene Expression Omnibus; GO, Gene Ontology; GSEA, gene set enrichment analysis; GSVA, gene set variation analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; logFC, log2 fold change; MF, molecular function; miRNA, microRNA; MSigDB, Molecular Signatures Database; NES, normalized enrichment score; OS, osteosarcoma; pLDDT, predicted local-distance difference test; PPI, protein–protein interaction; ROC, receiver operating characteristic; ssGSEA, single-sample gene set enrichment analysis; TF, transcription factor.

Declarations

Ethics approval and consent to participate

Not applicable. Only de-identified public transcriptomic datasets retrieved from the GEO database were re-analyzed in this study; no human or animal specimens were involved.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analyzed in this study are available in the Gene Expression Omnibus (GEO) repository under accession numbers GSE36001 and GSE16088. Supplementary Tables S1–S5 are supplied as supporting information files. Custom R scripts for bioinformatic analysis are included in the supplementary materials.

Competing interests

The authors declare no competing interests.

Funding

This research was financially supported by the Pudong New Area Science & Technology Development Fund (PKJ2021-Y49), the Hospital-Level Basic Scientific Research Youth

Cultivation Project (Lingang Special) of Shanghai Sixth People's Hospital (ynqnlg202401), and the General Program of the National Natural Science Foundation of China (82473103).

Authors' contributions

Per the CRediT taxonomy: XS: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. YH: Methodology, Software, Visualization, Writing – review & editing. KH: Data curation, Validation, Resources, Investigation. HZ: Formal analysis, Investigation. DM: Supervision, Funding acquisition, Writing – review & editing. CT: Conceptualization, Supervision, Funding acquisition, Writing – review & editing. All authors have read and approved the final manuscript.

Acknowledgements

We thank the researchers who deposited the GSE36001 and GSE16088 datasets in the Gene Expression Omnibus (GEO) database, as well as the developers of the open-source R packages and bioinformatics tools used in this study. We are also grateful to the staff of Shanghai Jiao Tong University Affiliated Sixth People's Hospital for their administrative support.

References

1. Gill J, Gorlick R. Advancing therapy for osteosarcoma. Nat Rev Clin Oncol. 2021;18(10):609–624. https://doi.org/10.1038/s41571-021-00519-8.

2. Meltzer PS, Helman LJ. New horizons in the treatment of osteosarcoma. N Engl J Med. 2021;385(22):2066–2076. https://doi.org/10.1056/NEJMra2103423.

3. Yu S, Yao X. Advances on immunotherapy for osteosarcoma. Mol Cancer. 2024;23(1):192. https://doi.org/10.1186/s12943-024-02105-9.

4. Xu X, Peng Q, Jiang X, Tan S, Yang W, Han Y, et al. Altered glycosylation in cancer: molecular functions and therapeutic potential. Cancer Commun. 2024;44(11):1316–1336. https://doi.org/10.1002/cac2.12610.

5. Aziz F, Khan I, Shukla S, Dey DK, Yan Q, Chakraborty A, et al. Partners in crime: the Lewis Y antigen and fucosyltransferase IV in Helicobacter pylori-induced gastric cancer. Pharmacol Ther. 2022;232:107994. https://doi.org/10.1016/j.pharmthera.2021.107994.

6. Yang M, Liu SC, Hao DJ, Yan L, Liu ZK, Yin XH. RP11-867G2.8 promotes EMT and chordoma malignant phenotypes by enhancing FUT4 mRNA stability and translation. Am J Cancer Res. 2022;12(3):1264 – 1281. PMID: 35411246; PMCID: PMC8984897.

7. Zhou J, Wang H, Hong F, Hu S, Su X, Chen J, Chu J. CircularRNA circPARP4 promotes glioblastoma progression through sponging miR-125a-5p and regulating FUT4. Am J Cancer Res. 2021;11(1):138 – 156. PMID: 33520365; PMCID: PMC7840713.

8. Liu E, Qian X, He Y, Chen K. FUT4 promotes the progression of cholangiocarcinoma by modulating epithelial–mesenchymal transition. Cell Cycle. 2024;23(2):218–231. https://doi.org/10.1080/15384101.2024.2318949.

9. Nie H, Saini P, Miyamoto T, Liao L, Zielinski RJ, Liu H, et al. Targeting branched N-glycans and fucosylation sensitizes ovarian tumors to immune checkpoint blockade. Nat Commun. 2024;15(1):2853. https://doi.org/10.1038/s41467-024-47069-y.

10. Liu Q, Adhikari E, Lester DK, Fang B, Johnson JO, Tian Y, et al. Androgen drives melanoma invasiveness and metastatic spread by inducing tumorigenic fucosylation. Nat Commun. 2024;15(1):1148. https://doi.org/10.1038/s41467-024-45324-w.

11. Yang Y, Yan X, Chen Y, Liu J, Xue J, Sheng X, et al. Silencing FUT4 inhibits the progression of osteosarcoma through activation of FOXO1. Curr Pharm Des. 2024;30(6):440–447. https://doi.org/10.2174/0113816128269432240103052108.

669 12. Xue Q, Ma R, Chen Y, Yan X, Liu J, Xue J, Yang Y, Liu X. miR-371b-5p reduces
670 osteosarcoma cell migration and proliferation to induce apoptosis by targeting FUT4. J
671 Cancer. 2025;16(7):2388–2400. https://doi.org/10.7150/jca.103286.
672 13. Tian C, Sun X, Han K, Zhu H, Min D, Lin S. Long non-coding RNA MRUL contributes
673 to osteosarcoma progression through the miR-125a-5p/FUT4 axis. Front Genet. 2020;11:672.
674 https://doi.org/10.3389/fgene.2020.00672.
675 14. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO)
676 and BioConductor. Bioinformatics. 2007;23(14):1846–1847.
677 https://doi.org/10.1093/bioinformatics/btm254.
678 15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential
679 expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res.
680 2015;43(7):e47. https://doi.org/10.1093/nar/gkv007.
681 16. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-
682 source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics.
683 2011;12:77. https://doi.org/10.1186/1471-2105-12-77.
684 17. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological
685 themes among gene clusters. OMICS. 2012;16(5):284–287.
686 https://doi.org/10.1089/omi.2011.0118.
687 18. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids
688 Res. 2000;28(1):27–30. https://doi.org/10.1093/nar/28.1.27.
689 19. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al.
690 Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide
691 expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545–15550.
692 https://doi.org/10.1073/pnas.0506580102.

693 20. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray
694 and RNA-seq data. BMC Bioinformatics. 2013;14:7. https://doi.org/10.1186/1471-2105-14-
695 7.
696 21. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING
697 v11: protein-protein association networks with increased coverage, supporting functional
698 discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607–
699 D613. https://doi.org/10.1093/nar/gky1131.
700 22. Huang J, Zheng W, Zhang P, Lin Q, Chen Z, Xuan J, et al. ChIPBase v3.0: the
701 encyclopedia of transcriptional regulations of non-coding RNAs and protein-coding genes.
702 Nucleic Acids Res. 2023;51(D1):D46–D56. https://doi.org/10.1093/nar/gkac1067.
703 23. Zhou K, Huang J, Liu S, Zheng WJ, Liu SR, Xing Y, et al. An encyclopedic regulatory
704 and functional atlas of RNA interactomes. Nat Methods. 2026;23(6):1109-1114.
705 https://doi.org/10.1038/s41592-026-03105-x.
706 24. Davis AP, Wiegers TC, Johnson RJ, Sciaky D, Wiegers J, Mattingly CJ. Comparative
707 Toxicogenomics Database (CTD): update 2023. Nucleic Acids Res. 2023;51(D1):D1257–
708 D1262. https://doi.org/10.1093/nar/gkac833.
709 25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a
710 software environment for integrated models of biomolecular interaction networks. Genome
711 Res. 2003;13(11):2498–2504. https://doi.org/10.1101/gr.1239303.
712 26. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-
713 cancer immunogenomic analyses reveal genotype-immunophenotype relationships and
714 predictors of response to checkpoint blockade. Cell Rep. 2017;18(1):248–262.
715 https://doi.org/10.1016/j.celrep.2016.12.019.
716 27. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold
717 Protein Structure Database: massively expanding the structural coverage of protein-sequence

718 space with high-accuracy models. Nucleic Acids Res. 2022;50(D1):D439–D444.
719 https://doi.org/10.1093/nar/gkab1061.
720 28. Liu Y, Yang X, Gan J, Chen S, Xiao ZX, Cao Y. CB-Dock2: improved protein-ligand
721 blind docking by integrating cavity detection, docking and homologous template fitting.
722 Nucleic Acids Res. 2022;50(W1):W159–W164. https://doi.org/10.1093/nar/gkac394.
723 29. Gao B, Qiang B, Tan H, Ren M, Jia Y, Lu M, et al. DrugCLIP: contrastive protein-
724 molecule representation learning for virtual screening. Adv Neural Inf Process Syst.
725 2023;36:44595–44614. https://doi.org/10.52202/075280-1932.

726 # Tables

727 **Table 1. GEO microarray chip information**

	GSE36001	GSE16088
Platform	GPL6102	GPL96
Species	*Homo sapiens*	*Homo sapiens*
Tissue	Normal bone, Normal osteoblast, Osteosarcoma cell line	Kidney, Liver, Lymph, Human Osteosarcoma
Samples in OS group	19	14
Samples in Control group	6	6
Reference	PMID: 23144859	PMID: 20028558

728 GEO, Gene Expression Omnibus; OS, osteosarcoma.

729 **Table 2. GO and KEGG enrichment analysis of the 30 *FUT4*-associated DEGs together**
730 **with *FUT4***

ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
BP	GO:0001503	ossification	8/31	444/18870	4.33E-07	3.45E-04	2.74E-04
BP	GO:0001649	osteoblast differentiation	6/31	257/18870	3.33E-06	1.33E-03	1.05E-03
BP	GO:0031214	biomineral tissue development	4/31	175/18870	1.85E-04	3.85E-02	3.05E-02
BP	GO:0001764	neuron migration	4/31	177/18870	1.93E-04	3.85E-02	3.05E-02
CC	GO:006	collagen-containing	7/31	429/19	3.48E-06	2.57E-04	2.09E-04

	2023	extracellular matrix		886			
CC	GO:0005788	endoplasmic reticulum lumen	5/31	313/19886	1.14E-04	4.20E-03	3.41E-03
CC	GO:0005604	basement membrane	3/31	90/19886	3.68E-04	9.07E-03	7.35E-03
MF	GO:0005201	extracellular matrix structural constituent	5/29	166/18496	5.47E-06	6.02E-04	4.37E-04
KEGG	hsa04512	ECM-receptor interaction	4/20	89/9423	3.21E-05	1.77E-03	1.55E-03
KEGG	hsa04510	Focal adhesion	4/20	203/9423	7.73E-04	2.83E-02	2.49E-02
KEGG	hsa04145	Phagocytosis	4/20	260/9423	1.94E-03	4.26E-02	3.75E-02

GeneRatio denominators reflect the number of input genes with valid annotations in each ontology (31 for BP and CC, 29 for MF, and 20 for KEGG). GO, Gene Ontology; BP, Biological Process; CC, Cellular Component; MF, Molecular Function; KEGG, Kyoto Encyclopedia of Genes and Genomes. P values are shown in scientific notation.

Table 3. GSEA results for *FUT4* expression groups

ID	Set Size	Enrichment Score	NES	pvalue	p.adjust	qvalue
FOROUTAN_PRODRANK_TGFB_EMT_UP	183	6.68E-01	2.07E+00	1.00E-10	6.80E-08	5.86E-08
PLASARI_TGFB1_TARGETS_10HR_UP	187	6.50E-01	2.03E+00	1.00E-10	6.80E-08	5.86E-08
FOROUTAN_INTEGRATED_TGFB_EMT_UP	120	6.83E-01	2.01E+00	1.26E-08	4.29E-06	3.70E-06
FOROUTAN_TGFB_EMT_UP	190	6.44E-01	2.01E+00	3.77E-10	2.10E-07	1.81E-07
GROSS_HYPOXIA_VIA_HIF1A_DN	108	6.66E-01	1.94E+00	1.06E-06	1.47E-04	1.27E-04
SARRIO_EPITHELIAL_MESENCHYMAL_TRANSITION_UP	158	6.34E-01	1.93E+00	8.23E-08	2.31E-05	1.99E-05
GROSS_HYPOXIA_VIA_ELK3_DN	149	6.23E-01	1.88E+00	3.44E-07	6.58E-05	5.67E-05
MURATA_VIRULENCE_OF_H_PILORI	22	8.16E-01	1.85E+00	3.65E-04	1.21E-02	1.05E-02
BIOCARTA_IL1R_PATHWAY	30	7.71E-01	1.84E+00	2.42E-04	9.03E-03	7.78E-03
GROSS_HYPOXIA_VIA_ELK3_AND_HIF1A_UP	138	6.12E-01	1.84E+00	5.79E-06	5.91E-04	5.09E-04

ALTEMEIER_RESPONSE_TO_LPS_WITH_MECHANICAL_VENTILATION	119	5.24E-01	1.54E+00	3.01E-03	4.79E-02	4.13E-02
MARZEC_IL2_SIGNALING_UP	114	5.25E-01	1.53E+00	3.10E-03	4.89E-02	4.22E-02
BLANCO_MELO_BRONCHIAL_EPITHELIAL_CELLS_INFLUENZA_A_DEL_NS1_INFECTION_DN	173	4.86E-01	1.50E+00	3.13E-03	4.93E-02	4.25E-02
ZHOU_INFLAMMATORY_RESPONSE_LIVE_UP	408	4.16E-01	1.40E+00	1.24E-03	2.72E-02	2.34E-02
GENTILE_UV_HIGH_DOSE_DN	281	4.26E-01	1.39E+00	2.90E-03	4.68E-02	4.03E-02
FULCHER_INFLAMMATORY_RESPONSE_LECTIN_VS_LPS_DN	432	-3.95E-01	-1.37E+00	1.56E-03	3.09E-02	2.66E-02
REACTOME_IRON_UPTAKE_AND_TRANSPORT	57	-6.53E-01	-1.76E+00	4.30E-04	1.37E-02	1.18E-02
WP_CHOLESTEROL_METABOLISM	71	-6.27E-01	-1.77E+00	2.40E-04	9.00E-03	7.76E-03
REACTOME_REGULATION_OF_MITF_M_DEPENDENT_GENES_INVOLVED_IN_LYSOSOME_BIOGENESIS_AND_AUTOPHAGY	17	-8.31E-01	-1.79E+00	3.35E-04	1.14E-02	9.83E-03
WP_ENTEROCYTE_CHOLESTEROL_METABOLISM	36	-7.31E-01	-1.82E+00	1.69E-04	7.17E-03	6.18E-03

736 GSEA, Gene Set Enrichment Analysis; NES, normalized enrichment score. P values are
737 shown in scientific notation.

738 **Table 4. GSVA results for *FUT4* expression groups**

Pathway	logFC	Ave Expr	t	P.Value	adj.P.Val	B
KEGG_MEDICUS_REFERENCE_ANXA2_S100A10_REGULATED_ACTIN_CYTOSKELETON	6.44E-01	-5.65E-02	3.59E+00	1.00E-03	4.28E-01	-1.12E+00
KEGG_LIMONENE_AND_PINENE_DEGRADATION	6.38E-01	3.74E-02	4.43E+00	8.99E-05	1.52E-01	5.41E-01
KEGG_MEDICUS_REFERENCE_FASLG_RIPK1_3_SIGNALING_PATHWAY	5.89E-01	7.48E-03	3.22E+00	2.76E-03	6.35E-01	-1.82E+00
KEGG_MEDICUS_REFERENCE_TNFSF10_RIPK1_3_SIGNALING_PATHWAY	5.76E-01	2.98E-02	3.14E+00	3.42E-03	7.04E-	-1.97

					01	E+00
REACTOME_FASL_CD95L_SIGNALING	5.38E-01	1.96E-02	2.72E+00	1.02E-02	9.99E-01	-2.73E+00
REACTOME_ABACAVIR_TRANSMEMBRANE_TRANSPORT	5.32E-01	2.56E-02	3.08E+00	3.98E-03	7.09E-01	-2.08E+00
KEGG_MEDICUS_REFERENCE_NUCLEAR_INITIATED_PROGESTERONE_SIGNALING_PATHWAY	5.30E-01	9.31E-02	2.97E+00	5.36E-03	7.50E-01	-2.28E+00
REACTOME_DOPAMINE_RECEPTORS	5.28E-01	2.87E-02	3.11E+00	3.68E-03	7.09E-01	-2.02E+00
REACTOME_REGULATION_BY_C_FLIP	5.10E-01	7.82E-03	3.09E+00	3.90E-03	7.09E-01	-2.06E+00
REACTOME_CELL_EXTRACELLULAR_MATRIX_INTERACTIONS	5.08E-01	1.23E-02	3.42E+00	1.60E-03	4.82E-01	-1.44E+00
BIOCARTA_FOSB_PATHWAY	-5.37E-01	-5.59E-03	-3.03E+00	4.63E-03	7.09E-01	-2.18E+00
REACTOME_SMAC_XIAP_REGULATED_APOPTOTIC_RESPONSE	-5.48E-01	-2.05E-02	-4.01E+00	3.02E-04	2.96E-01	-2.89E-01
REACTOME_BINDING_OF_TCF_LEF_CTNNB1_TO_TARGET_GENE_PROMOTERS	-5.50E-01	-4.00E-02	-3.56E+00	1.09E-03	4.28E-01	-1.18E+00
KEGG_MEDICUS_PATHOGEN_SALMONELLA_AVRA_TO_BETA_CATENIN_SIGNALING_PATHWAY	-5.70E-01	-4.34E-02	-3.43E+00	1.57E-03	4.82E-01	-1.43E+00
REACTOME_RUNX3_REGULATES_IMMUNE_RESPONSE_AND_CELL_MIGRATION	-5.71E-01	-4.89E-03	-3.73E+00	6.71E-04	4.28E-01	-8.41E-01
KEGG_MEDICUS_PATHOGEN_SARS_COV_2_S_N_TO_LECTIN_PATHWAY_OF_COAGULATION_CASCADE	-5.77E-01	-4.17E-02	-3.43E+00	1.57E-03	4.82E-01	-1.43E+00
REACTOME_ACTIVATION_OF_RAC1_DOWNSTREAM_OF_NMDARS	-5.77E-01	-5.75E-02	-3.56E+00	1.09E-03	4.28E-01	-1.18E+00
REACTOME_IONOTROPIC_ACTIVITY_OF_KAINATE_RECEPTORS	-5.97E-01	-3.68E-02	-4.34E+00	1.17E-04	1.52E-01	3.63E-01
WP_CATALYTIC_CYCLE_OF_MAMMALIAN_FLAVINCONTAINING_MONOOXYGENASES_FMOS	-6.27E-01	2.12E-02	-3.89E+00	4.26E-04	3.34E-01	-5.26E-01
REACTOME_ACTIVATED_NTRK2_SIGNALS_THROUGH_CDK5	-8.12	2.63E-02	-6.22	4.02E-07	1.57E-	4.09E+0

	E-01	E+00	03	0

GSVA, Gene Set Variation Analysis. B, log-odds that the pathway is differentially expressed (limma). Nominal p-values (P) were used for pathway filtering; BH-adjusted p-values (adj.P) are provided for reference only. No multiple-testing correction was applied for pathway selection owing to the limited subgroup sample size. P values are shown in scientific notation.

Supplementary Table S1. Differential expression analysis between FUT4-high and FUT4-low osteosarcoma groups in GSE36001 (limma)

Complete limma output for all genes, including logFC, average expression (AveExpr), moderated t statistic, nominal P value (P.Value), Benjamini–Hochberg adjusted P value (adj.P.Val), and B statistic. Genes were classified as Up (n = 147), Down (n = 150), or Not significant using the thresholds $|\log FC| \geq 0.5$ and $P < 0.05$, corresponding to the 297 FUT4-associated DEGs reported in the Results.

Supplementary Table S2. Transcription factors associated with FUT4 identified from ChIPBase v3.0 (mRNA–TF pairs)

The retrieved associations include both experimentally validated and computationally predicted interactions, consistent with the Methods section.

Supplementary Table S3. MicroRNAs associated with FUT4 identified from ENCORI (formerly starBase) (mRNA–miRNA pairs)

The retrieved associations include both experimentally validated and computationally predicted interactions, consistent with the Methods section.

Supplementary Table S4. Drugs and compounds associated with FUT4 retrieved from the Comparative Toxicogenomics Database (CTD) (mRNA–drug pairs)

The retrieved associations include both experimentally validated and computationally inferred links, consistent with the Methods section.

Supplementary Table S5. Top 100 candidate small-molecule ligands targeting FUT4 ranked by DrugCLIP virtual screening

For each ligand, the DrugCLIP affinity prediction score (drugclip_score) and the AutoDock Vina docking score (vina_docking_score) are provided; lower Vina scores indicate more favorable predicted binding.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- README.rar
- TableS1DiffAnalysislogFC0.5pvalue0.05.xlsx
- TableS2mRNATF.xlsx
- TableS3mRNAmiRNA.xlsx
- TableS4mRNADrug.xlsx
- TableS5selectedligands.xlsx