



OPEN Feasibility of machine learning–based modeling and prediction to assess osteosarcoma outcomes

Qinfei Zhao^{1,6}, Weiquan Hu^{2,6}, Yu Xia³, Shengyun Dai⁴, Xiangsheng Wu¹, Jing Chen¹, Xiaoying Yuan⁵, Tianyu Zhong¹✉, Xuxiang Xi¹✉ & Qi Wang⁵✉

Osteosarcoma, an aggressive bone malignancy predominantly affecting children and adolescents, is characterized by a poor prognosis and high mortality rates. The development of reliable prognostic tools is critical for advancing personalized treatment strategies. However, identifying robust gene signatures to predict osteosarcoma outcomes remains a significant challenge. In this study, we analyzed gene expression data from 138 osteosarcoma samples across two multicenter cohorts and identified 14 consensus prognosis-associated genes via univariate Cox regression analysis. Using 66 combinations of 10 machine learning (ML) algorithms, we developed a machine learning-derived prognostic signature (MLDPS) optimized by the average C-index across TARGET, GSE21257, and merged cohorts. The MLDPS effectively stratified osteosarcoma patients into high- and low-risk score groups, achieving strong predictive performance for 1-, 3-, and 5-year overall survival (AUC range: 0.852 — 0.963). The MLDPS, comprising seven genes (CTNNBIP1, CORT, DLX2, TERT, BBS4, SLC7A1, NKX2-3), exhibited superior predictive accuracy compared to 10 established gene signatures. The findings of the MLDPS carry significant clinical implications for osteosarcoma treatment. Patients with a high-risk score demonstrated worse prognosis, increased metastasis risk, reduced immune infiltrations, and greater sensitivity to immunotherapy. Conversely, low-risk patients exhibited prolonged survival and distinct drug sensitivities. These findings underscore the potential of MLDPS to guide risk stratification, inform personalized therapeutic strategies, and improve clinical management in osteosarcoma.

Keywords Osteosarcoma, Machine learning, Tumor immunotherapy, Risk score, Prognosis

Osteosarcoma, a rare yet aggressive primary malignant bone tumor, predominantly affects pediatric and adolescent populations^{1,2}. Despite advancements in surgical resection, chemotherapy, and immunotherapy, survival outcomes remain poor, largely due to the high incidence of metastasis and chemotherapy resistance^{3,4}. The five-year survival rate for osteosarcoma patients exhibiting chemotherapy resistance or metastasis is a mere 20–30%^{5–7}. Although the TNM staging system provides basic prognostic insights⁸, it has limited capacity to capture the molecular complexity and heterogeneity of osteosarcoma. Consequently, researchers are prioritizing the discovery of novel molecular biomarkers to improve clinical outcome prediction and optimize therapeutic strategies^{9–11}.

The treatment landscape for sarcomas, including osteosarcoma, is evolving through the integration of molecular and genomic insights. Immune checkpoint inhibitors, tumor microenvironment-targeted therapies, and combination regimens show promise in sarcoma management; however, their application in osteosarcoma remains exploratory. For example, genetic profiling of uterine leiomyosarcoma has uncovered frequent mutations in tumor suppressor genes such as TP53, RB1, and PTEN, pointing towards potential therapeutic targets¹². Nonetheless, adjuvant chemotherapy yields inconsistent results across early-stage uterine leiomyosarcoma subtypes^{13,14}. Similarly, while immune checkpoint inhibitors demonstrate efficacy in advanced patients with cancers^{15–20}, their role in osteosarcoma requires further validation. However, accurately forecasting prognosis has proved to be challenging due to the variations in the efficacy of treatment approaches^{21–24}. These

¹Department of Laboratory Medicine, First Affiliated Hospital of Gannan Medical University, Ganzhou 341000, Jiangxi, China. ²Department of Joint Surgery, Ganzhou People's Hospital, Ganzhou 341000, Jiangxi, China. ³Department of Orthopaedics, The Second Xiangya Hospital, Central South University, 139 Renmin Road, Changsha 410011, Hunan, China. ⁴National Institutes for Food and Drug Control, Beijing, China. ⁵The First School of Clinical Medicine, Gannan Medical University, Ganzhou 341000, Jiangxi, China. ⁶Qinfei Zhao and Weiquan Hu contributed equally to this work. ✉email: zhongtianyu@gmail.com; 450231625@qq.com; qwangk@naver.com

findings emphasize the difficulty of establishing universal adjuvant therapies for diverse sarcoma subtypes and accentuate the necessity for individualized approaches grounded in tumor-specific biology. To significantly improve osteosarcoma survival rates, it is important to ascertain innovative biomarkers capable of accurately predicting clinical outcomes and therapeutic responsiveness. Therefore, high-confidence prognostic gene signatures are essential for guiding prognosis and treatment regimens.

As high-throughput sequencing technology advances, researchers are increasingly confronted with vast amounts of biological and clinical data, collectively termed “Big Data”^{25–27}. These datasets enable the systematic identification of gene signatures and molecular pathways driving cancer progression²⁸, including osteosarcoma^{29–31}. Previous studies have constructed prognostic signatures associated with hypoxia³², apoptosis³³, energy metabolism²², and the immune tumor microenvironment (TME)³⁴. Nevertheless, the absence of reliable biomarkers often leads to either overtreatment or undertreatment, causing significant financial burden, severe adverse effects, or accelerated disease progression^{35,36}.

To address this gap, machine learning (ML), a subfield of artificial intelligence, has emerged as a powerful tool for constructing predictive models from large-scale datasets^{37,38}. ML algorithms have proven instrumental in biomedical research, facilitating biomarker discovery, therapeutic target identification, and mechanistic insights across diseases^{39–41}. For instance, Zhang et al.⁴² leveraged ML to develop a six-gene oxidative stress signature for predicting acute myeloid leukemia outcomes. Similarly, recent studies have employed ML to identify diagnostic and prognostic biomarkers in osteosarcoma^{43,44}, such as Shi et al.’s six-gene risk model (MYC, COL13A1, UHRF2, MT1A, ACTB, GBP1)²³. Nevertheless, integrating ML with bioinformatics to link prognostic signatures to diagnostic biomarkers and elucidate their clinical significance remains underexplored in osteosarcoma.

To identify an optimal biomarker, we constructed an ML-derived prognostic signature (MLDPS) by implementing 66 ML algorithm combinations based on consensus prognostic signature genes (CPSGs). The MLDPS demonstrated robust predictive capabilities for overall survival (OS), metastasis, immunotherapy response, and drug treatment efficacy in osteosarcoma cohorts. After integrating key clinicopathological characteristics and 10 established osteosarcoma signatures, MLDPS exhibited consistent and significantly improved predictive performance. Overall, our research provides a valuable resource for guiding early diagnosis, prognostic evaluation, stratified treatment, personalized therapy, and post-treatment monitoring of osteosarcoma in clinical settings.

Materials and methods

Public osteosarcoma dataset collection and preprocessing

The TARGET-osteosarcoma cohort (<https://ocg.cancer.gov/programs/target>) consists of 85 samples with survival data and follow-up periods exceeding zero. We also acquired the GSE21257 dataset from the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>), which includes 53 osteosarcoma samples with survival information and follow-up periods greater than zero for all patients. In total, 138 samples were included in this study, encompassing 14,174 genes from the TARGET and GSE21257 cohorts.

Univariate Cox regression analysis

Univariate Cox regression analysis was conducted on the GSE21257 and TARGET cohorts using intersecting genes. CPSGs were selected for further investigation based on the criteria of $P < 0.01$ and consistent hazard ratios (HRs) that were either above or below 1 in both cohorts.

Functional enrichment analysis

The CPSGs were used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. The R package ‘clusterProfiler’ was employed for Gene Set Enrichment Analysis (GSEA) along with the GO and KEGG analyses^{45–49}. Gene Set Variations Analysis (GSVA) was conducted using the R package ‘GSVA’⁵⁰. The biocarta, KEGG, Reactome, wikipathways, and Hallmark gene sets used in GSVA were downloaded from the Molecular Signatures Database (MSigDB; <http://software.broadinstitute.org/gsea/msigdb/index.jsp>) version 6.2.

Development of MLDPS

To construct the MLDPS and ensure high accuracy and generalizability, we integrated 10 widely used ML algorithms: gradient boosting machine (GBM), survival support vector machine (Survival-SVM), supervised principal components (SuperPC), ridge regression, partial least squares regression for Cox (plsRcox), elastic network (Enet), random forest (RSF), least absolute shrinkage and selection operator (LASSO), CoxBoost, and Stepwise Cox. Notably, RSF, LASSO, CoxBoost, and Stepwise Cox effectively reduce dimensions and screen variables.

The development pipeline of the MLDPS is outlined as follows:

- Univariate Cox analysis was performed on the merged cohort (created by removing batch effects and redundancies from the TARGET and GSE21257 cohorts) to identify CPSGs, defined as genes with a significance level of $P < 0.01$ and consistent HR orientation across the two cohorts.
- Ten ML algorithms were used to create 66 algorithms combinations to develop the most predictive MLDPS with superior C-index performance.
- After establishing the model on the training cohort, we evaluated it on two separate testing cohorts. The average C-index was calculated for each model, with the model exhibiting the highest value being deemed optimal.

Prognostic value of MLDPS and potential clinical application

Patients from the TARGET, GSE21257, and merged cohorts were stratified into high- and low-risk score groups based on the median MLDPS value. The prediction accuracy of the MLDPS was assessed using receiver operating characteristic (ROC) curves and by calculating the time-dependent area under the ROC curve (AUC). The prognostic significance of the MLDPS was evaluated using Kaplan–Meier (KM) curves.

Collection and calculation of osteosarcoma's published signatures

Numerous prognostic signatures have been developed for osteosarcoma, focusing on stratified management and precise treatment, including those based on glycolysis- and cholesterol synthesis-related genes²², M1 macrophage-related genes²⁴, and apoptosis-associated genes³³. To compare the predictive efficacy of the MLDPS with established prognostic signatures, we systematically reviewed 10 published articles on prognostic models. Subsequently, we computed the risk scores for each signature in the three cohorts using the genes and coefficients outlined in the respective articles. In addition, we conducted a comparative analysis of key clinical characteristics, such as gender, age, and metastasis, between the high- and low-risk score groups in the merged cohort.

Assessment of immune infiltrations

Immune infiltrations were evaluated in the merged cohort using the R 'TOBR' package⁵¹ and various algorithms, including MCPcounter⁵², EPIC⁵³, xCell⁵⁴, CIBERSORT⁵⁵, IPS, quantIseq⁵⁶, ESTIMATE⁵⁷, and TIMER⁵⁸. We compared microenvironment-related scores, such as ESTIMATEScore, ImmuneScore, MicroenvironmentScore, and StromalScore, between the high- and low-risk score groups in the merged cohort. The correlation between the MLDPS risk score and chemokines, as well as chemokine receptors, in the combined cohort was determined utilizing the Pearson method to calculate the correlation coefficient and P-value. Correlation coefficients greater than 0 signify a positive correlation, while those less than 0 indicate a negative correlation. A P-value below 0.05 was considered statistically significant.

Response to immunotherapy

Tumor Immune Dysfunction and Exclusion (TIDE) (<http://tide.dfci.harvard.edu/>)⁵⁹ is an integrated web tool used to forecast patients' responses to immunotherapy. TIDE scores were calculated for each patient in the merged cohort. Subsequently, immunotherapeutic responses, represented by dysfunction scores and TIDE scores, were compared between high- and low-risk score groups.

Prediction of drug sensitivity

To further compare differences between high- and low-risk score groups, drug sensitivity analysis was performed to calculate the IC50 values of drugs for each sample using the 'oncoPredict' software package. Higher IC50 values indicate decreased sensitivity to treatment.

Statistical analysis

All statistical analyses were conducted using R software (version 4.2.2). AUC analysis was performed using the 'timeROC' package. The KM method, combined with the log-rank test, was employed to assess differences in survival rates among various patient subgroups.

Results

Development of osteosarcoma CPSGs using univariate Cox regression

Prior to biological data analysis, the collected datasets were assessed for batch effects, which were found to be present in both datasets (Fig. 1A). To address this, the datasets were merged for downstream analysis, using the R packages 'limma' and 'sva' to remove batch effects and reduce dataset size (Fig. 1B). Next, CPSGs were screened from the merged dataset consisting of 138 samples and 14,174 genes using univariate Cox regression analysis. The robust MLDPS was subsequently constructed using the 14 CPSGs, which showed consistent prognostic outcomes in both the GSE21257 and TARGET cohorts (Fig. 1C). The 14 CPSGs comprised 12 risky genes (P-value < 0.01, HR > 1) and two protective genes (P-value < 0.01, HR < 1) (Fig. 1C).

GO enrichment analysis revealed that the 14 CPSGs were predominantly enriched in the biological process (BP) of "pattern specification process" (Fig. 2A), the cell component (CC) of "mitochondrial inner membrane" (Fig. 2B), and the molecular function (MF) of "transcription coactivator binding," "telomeric DNA binding," and "telomerase RNA binding" (Fig. 2C). KEGG analysis identified the "Arginine and proline metabolism" as the most significant signaling pathway (Fig. 2D). These findings lay a groundwork for deeper investigations into the mechanisms of osteosarcoma and uncover pivotal biological processes and pathways that may influence its prognosis.

Construction of MLDPS

Risk prediction models were developed in the TARGET training cohort using a combination of 10 well-established ML algorithms. The predictive performance of all models was evaluated by calculating the average C-index across the TARGET, GSE21257 and merged cohorts. The optimal model (MLDPS) was built by selecting the combination of LASSO and RSF with the highest average C-index (0.866) (Fig. 3A). The LASSO algorithm effectively identified the most critical CPSGs, while the RSF algorithm filtered the most valuable model. Finally, seven potential gene signatures (CTNNBIP1, CORT, DLX2, TERT, BBS4, SLC7A1, and NKX2-3) were identified, as illustrated in Fig. 3B.

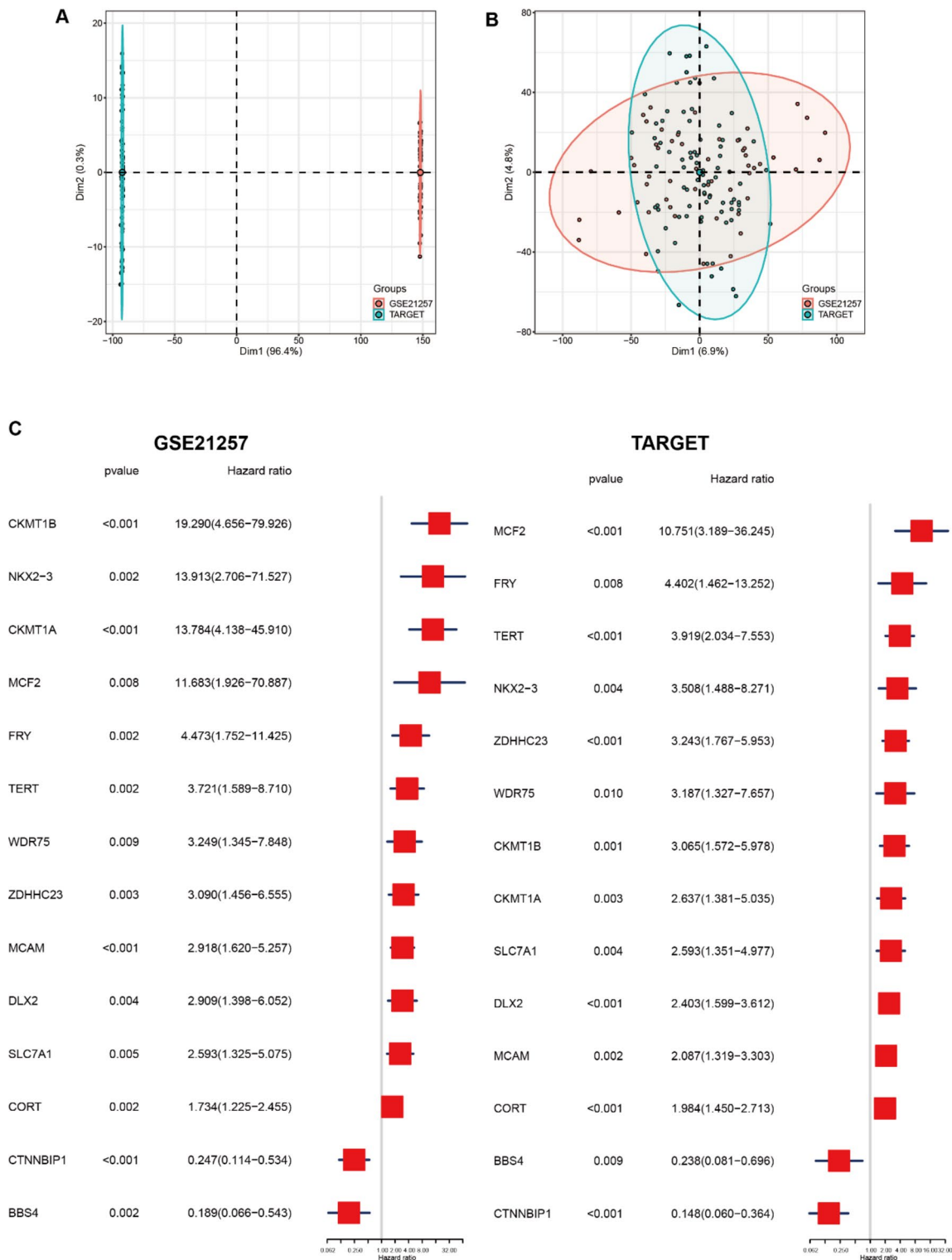


Fig. 1. Removal of batch effects and univariate regression analysis in osteosarcoma. (A) Principle-component Analysis (PCA) plot before batch effect removal. (B) PCA plot after batch effect removal. (C) Univariate Cox regression analysis results for 14 CPSPGs in the GSE21257 and TARGET cohorts.

Robust prediction performance of MLDPS

To estimate the discriminatory capability and clinical utility of the MLDPS, we plotted ROC and survival curves. As shown in Fig. 3C and D, the AUCs for OS in the TARGET cohort at 1, 3, and 5 years were 0.949, 0.957, and 0.963, respectively. Similarly, the model exhibited robust performance in the GSE21257 cohort, achieving AUCs of 0.954, 0.869, and 0.852 (Fig. 4A and B). The merged cohort had AUCs of 0.935, 0.913, and 0.912 (Fig. 4D and

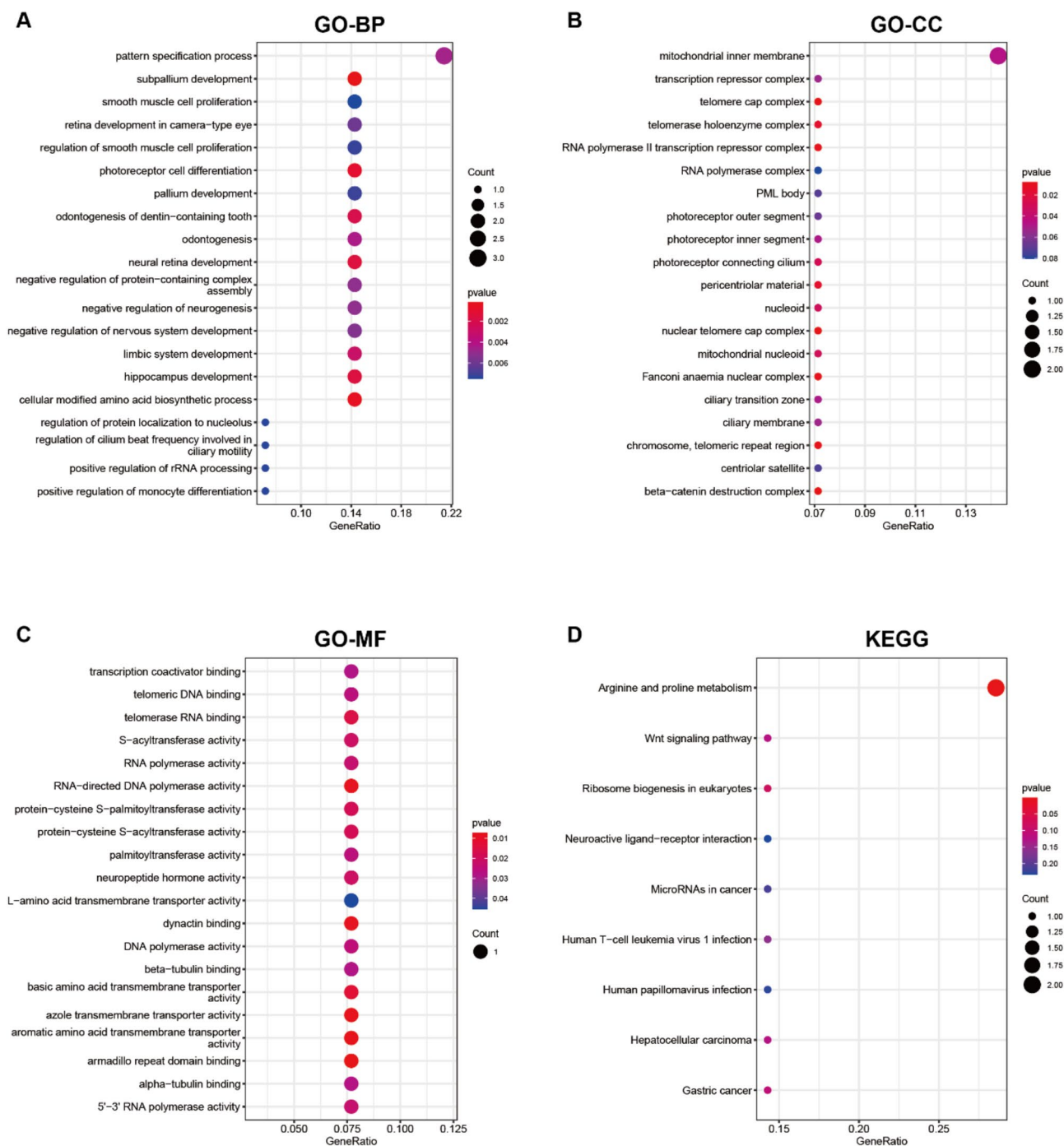


Fig. 2. GO and KEGG pathway analysis of CPSGs. The GO enrichment analysis of 14 CPSGs. The enriched BPs (A), CCs (B), and MFs (C) are noted. (D) The KEGG pathway enrichment analysis of 14 CPSGs.

E). These findings indicate the robust prognostic capability of the MLDPS model for osteosarcoma patients, with AUCs ranging from 0.852 to 0.963 for OS at 1, 3, and 5 years across all three cohorts. Furthermore, KM analysis confirmed that the MLDPS could effectively stratify osteosarcoma patients into distinct high- and low-risk score groups. Patients in the high-risk score group, exhibited significantly worse survival outcomes compared to the low-risk score group across all three cohorts ($P < 0.0001$ for all cohorts) (refer to Figs. 3E and 4C, and Fig. 4F).

Re-evaluation of 10 previously published signatures in osteosarcoma

The expeditious advancement of high-throughput sequencing has boosted the stratified management and precise therapeutic approaches for tumors⁶⁰. Recently, numerous prognostic signatures for osteosarcoma have been developed using ML algorithms, including RSF and CoxBoost, leveraging extensive and reliable datasets^{23,61,62}. A total of 10 prognostic signatures were collected from published sources to perform a comparative assessment

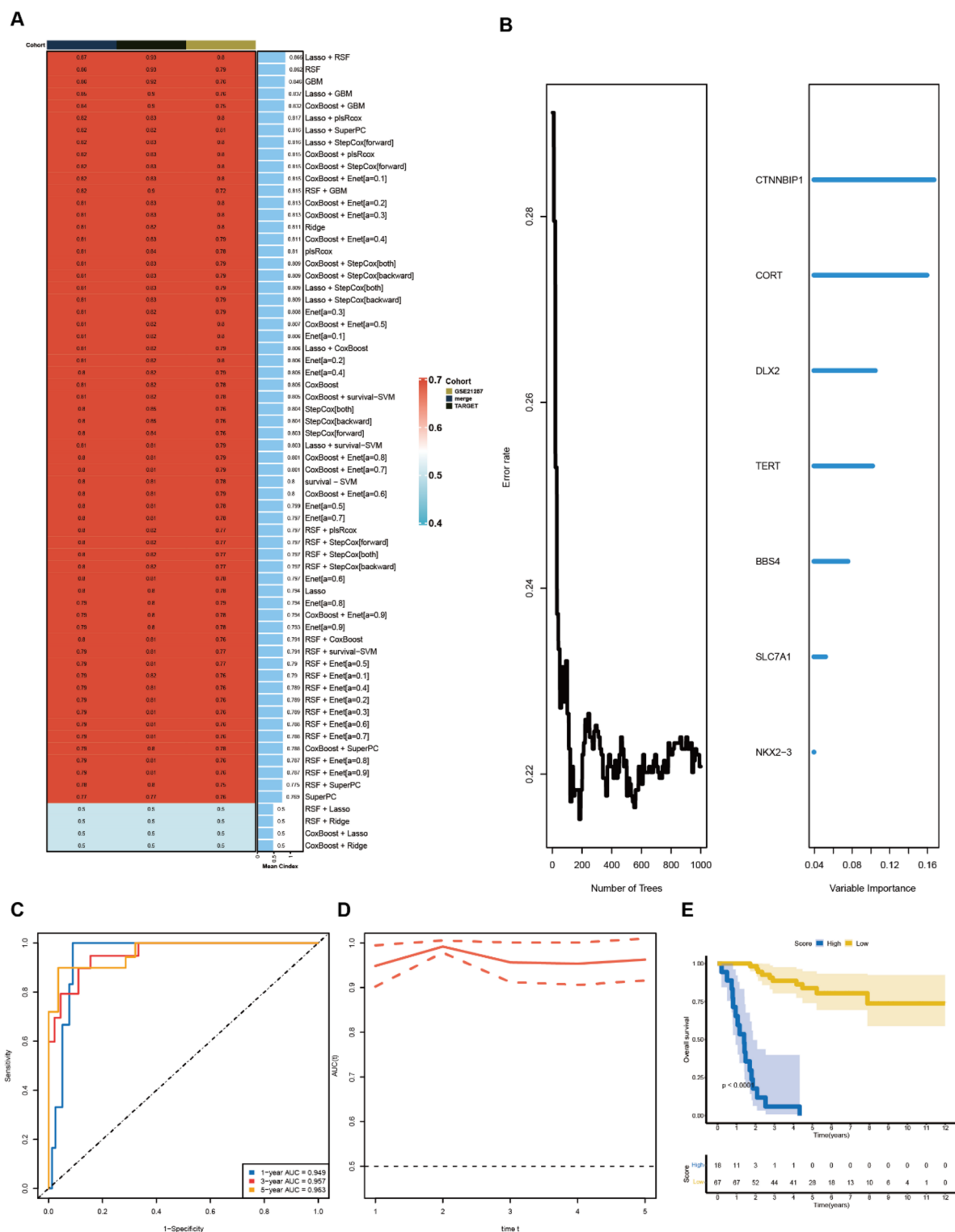


Fig. 3. Construction and evaluation of the MLDPS effectiveness. **(A)** The C-indexes values of 10 well-established ML algorithm combinations across three cohorts. **(B)** Seven CPSGs selected for the final model using the LASSO algorithm. **(C)** Time-dependent ROC analysis for predicting OS in the TARGET cohort. **(D)** AUC values of the time-dependent ROC curves in the TARGET cohort. **(E)** KM curves for OS in the TARGET cohort.

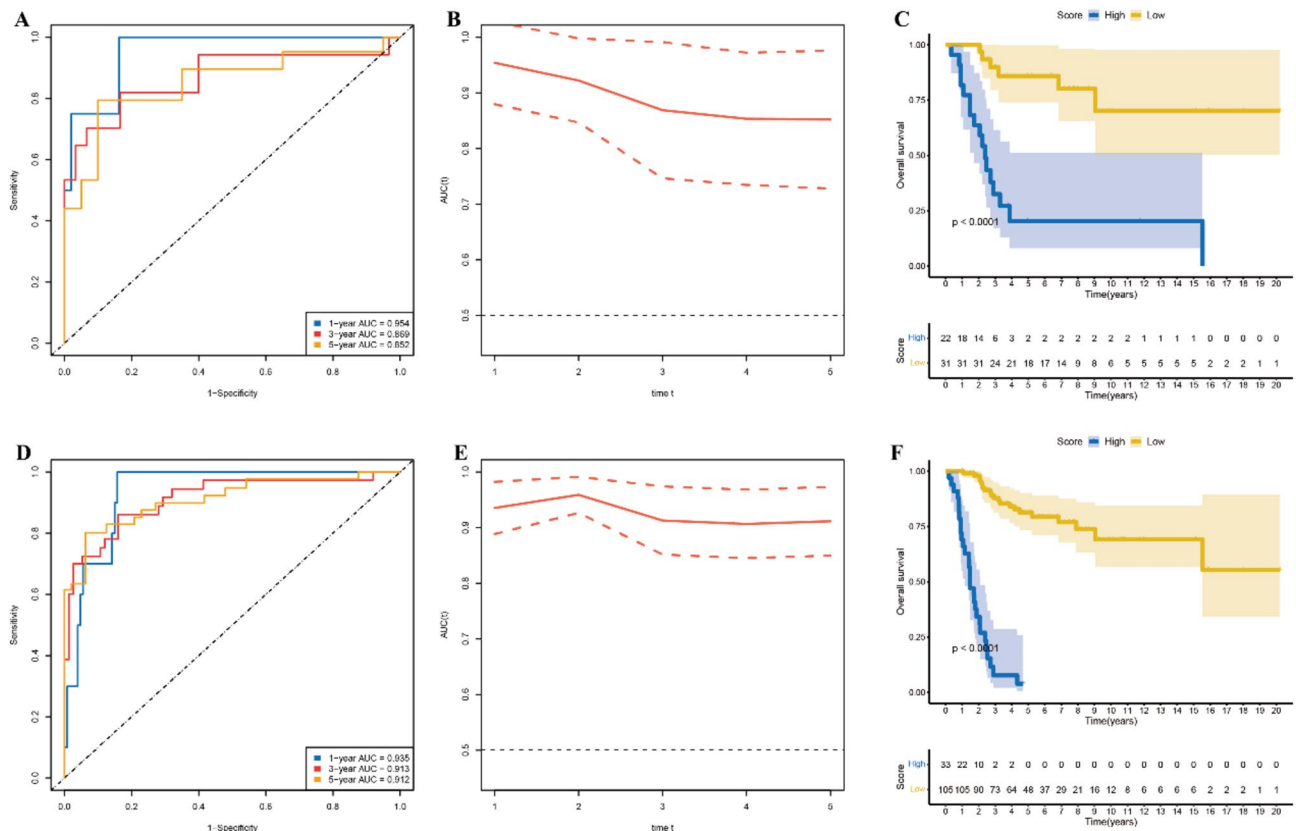


Fig. 4. Prediction performance of MLDPS in the GSE21257 and merged cohorts. Time-dependent ROC analysis for predicting OS in the (A) GSE21257 and (D) merged cohorts. AUC values of the time-dependent ROC curves in the (B) GSE21257 and (E) merged cohorts. KM curves for OS in the (C) GSE21257 and (F) merged cohorts.

of the MLDPS' predictive efficacy relative to these models. The evaluation was carried out using the C-index metric across three distinct cohorts. These results demonstrated that the MLDPS achieved significantly higher predictive accuracy than the other models across all three cohorts, consistently ranking first (Fig. 5). This underscores the robustness and reliability of the MLDPS. Notably, certain prognostic signatures, such as those reported in PMID34894177²³, PMID37234254⁹, and PMID34337075³², exhibited higher C-index values in one or two cohorts but performed poorly in the others. This discrepancy may stem from impaired generalization ability caused by overfitting, as plotted in Fig. 5.

The clinical signature of MLDPS

Univariate Cox regression analysis revealed that the MLDPS, published signatures (excluding PMID35598382³³), and metastasis were statistically significant within the merged cohort (Fig. 6A). However, no statistically significant differences were observed for gender or age (Fig. 6A). Further analysis was conducted to compare OS and metastasis between high- and low-risk score groups. Compared to the low-risk score group, the high-risk score group exhibited significantly inferior OS (Fig. 6B). Patients in the high-risk score group showed a greater propensity for developing metastasis, potentially contributing to their unfavorable prognosis (Fig. 6C).

Collectively, the timeROC curve, KM survival analysis, and Cox regression analysis outcomes consistently demonstrated that the MLDPS accurately forecasted the prognosis of osteosarcoma patients across all three cohorts. These findings suggest that the MLDPS holds promise as a valuable tool for clinical application.

The underlying biological mechanisms of MLDPS

Heatmaps (Fig. 7A and B) depicted the top 50 genes showing the most significant positive and negative correlations with the risk score. GO enrichment analysis of co-expressed genes positively associated with the MLDPS exhibited significant enrichment in various BPs, namely "ossification," "ncRNA processing," "embryonic organ morphogenesis," "ribosome biogenesis," "rRNA metabolic process," and "rRNA processing" (Fig. 7C). Enrichment was observed in CCs such as "ribosome," "mitochondrial matrix," "ribosomal subunit," "asymmetric synapse," and "cytosolic ribosome" (Fig. 7D). With respect to MFs, "structural constituent of ribosome" (Fig. 7E) was significantly enriched. Notably, KEGG analysis showed "Ribosome" as the only pathway with significant enrichment (Fig. 7F).

To understand the biological implications of MLDPS, we performed a GSEA analysis. The findings indicated a positive correlation between a high-risk score and the upregulation of key gene sets involved in translation,

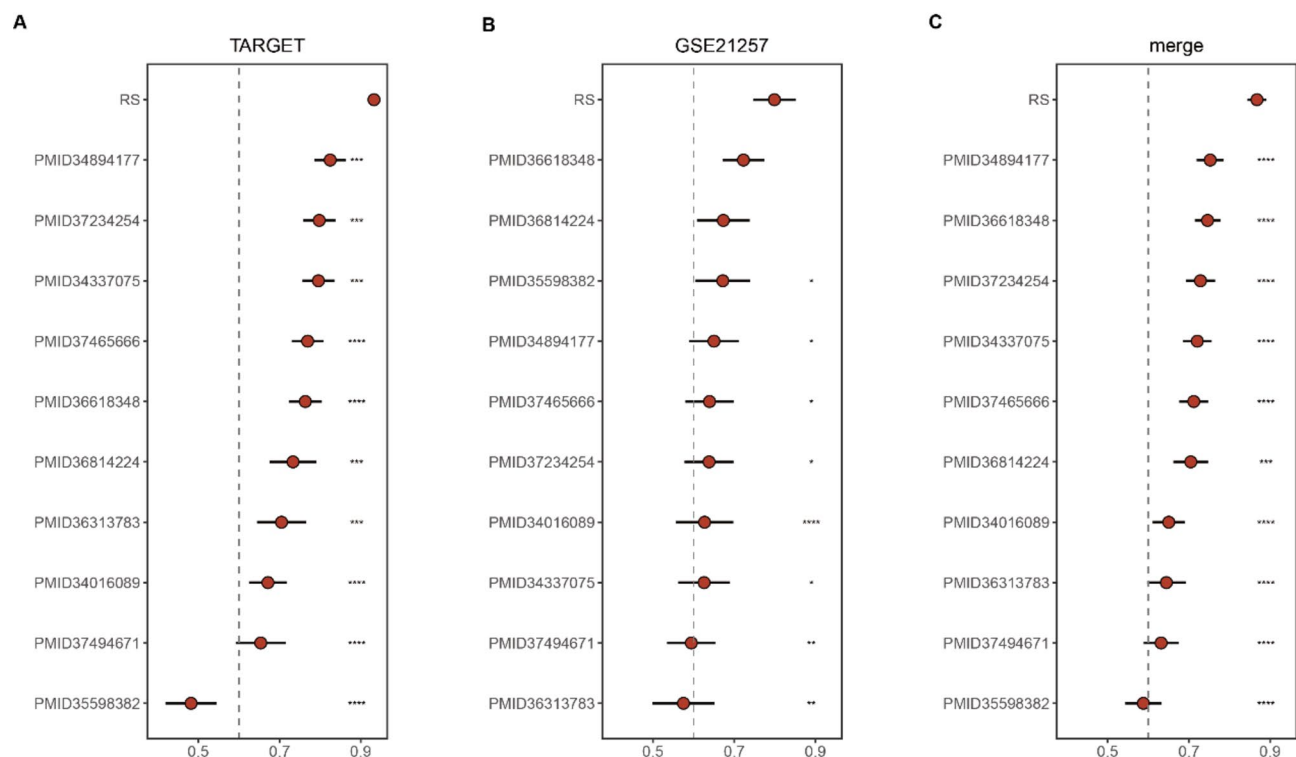


Fig. 5. Comparisons of C-indexes values between MLDPS and 10 expression-based signatures. C-indexes values of MLDPS and 10 published signatures in the TARGET, GSE21257, and merged cohorts. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, and **** $P < 0.0001$.

rRNA processing in the nucleus and cytosol, the major pathway of rRNA processing in the nucleolus and cytosol, and other biological pathways (Supplementary Fig. 1A). GSVA was utilized to analyze the effects of MLDPS on biological pathways in osteosarcoma. A heatmap was used to visualize the top 20 biological processes with statistically significant differences, ranked in ascending order of their P-values. As depicted in Supplementary Fig. 1B, significant pathway differences were observed between the high- and low-risk score groups. These observations may offer a potential explanation for the unfavorable prognosis associated with the high-risk score group.

Immune landscape of MLDPS

A thorough analysis of the TME in osteosarcoma was conducted using the 'IOBR' R package⁵¹. The heatmap exhibited variations in the immunological composition between high- and low-risk score groups (Fig. 8A). As shown in Fig. 8B, the high-risk score group demonstrated notably diminished ESTIMATEScore, ImmuneScore, MicroenvironmentScore, and StromalScore, suggesting an immunosuppressive condition.

Further analysis identified a negative correlation between the risk score and the gene expression levels of most chemokines and chemokine receptors (Fig. 9A). Specifically, the correlation coefficients between the risk score and the expression levels of CCL8, CSF1, CCR1, CCR2, CCR4, and CCR7 were all below -0.2 , with corresponding P-values less than 0.05 (Fig. 9B).

TIDE algorithms, widely used for predicting responses to cancer immunotherapy, indicate suboptimal immunotherapy outcomes when TIDE values are elevated⁵⁹. In the merged cohort, TIDE algorithms predicted responses to immunotherapy for 138 patients. Of these, 46 were identified as responders (TIDE score < 0), whereas the remaining 92 were classified as non-responders (Fig. 10A). To evaluate the predictive power of the MLDPS for cancer immunotherapy response, dysfunction and TIDE scores were calculated. Osteosarcoma patients with elevated MLDPS risk scores exhibited decreased dysfunction and TIDE scores, indicating a higher likelihood of responding to immunotherapy in the merged cohort (Fig. 10B–D). Overall, these findings indicate that individuals in the high-risk score group are more likely to benefit from immunotherapy.

Comparative analysis of drug sensitivity between high- and low-risk score groups

Precision medicine necessitates clinicians to promptly recognize patients sensitive to different treatments in order to provide tailored and personalized interventions^{63,64}. Hence, the drug sensitivity between the two risk score groups was evaluated by the 'oncoPredict' R package. Drug sensitivity analysis demonstrated a notable discrepancy in the drug response between the two risk score groups. Specifically, the high-risk score group exhibited a correlation with decreased IC50 values for drugs such as ABT737_1910, AZ6102_2109, AZD5153_1706, AZD5991_1720, AZD6738_1917, and BDP00009066_1866. Conversely, the low-risk score group was correlated with decreased IC50 values for drugs such as AZ960_1250, BMS.754807_2171,

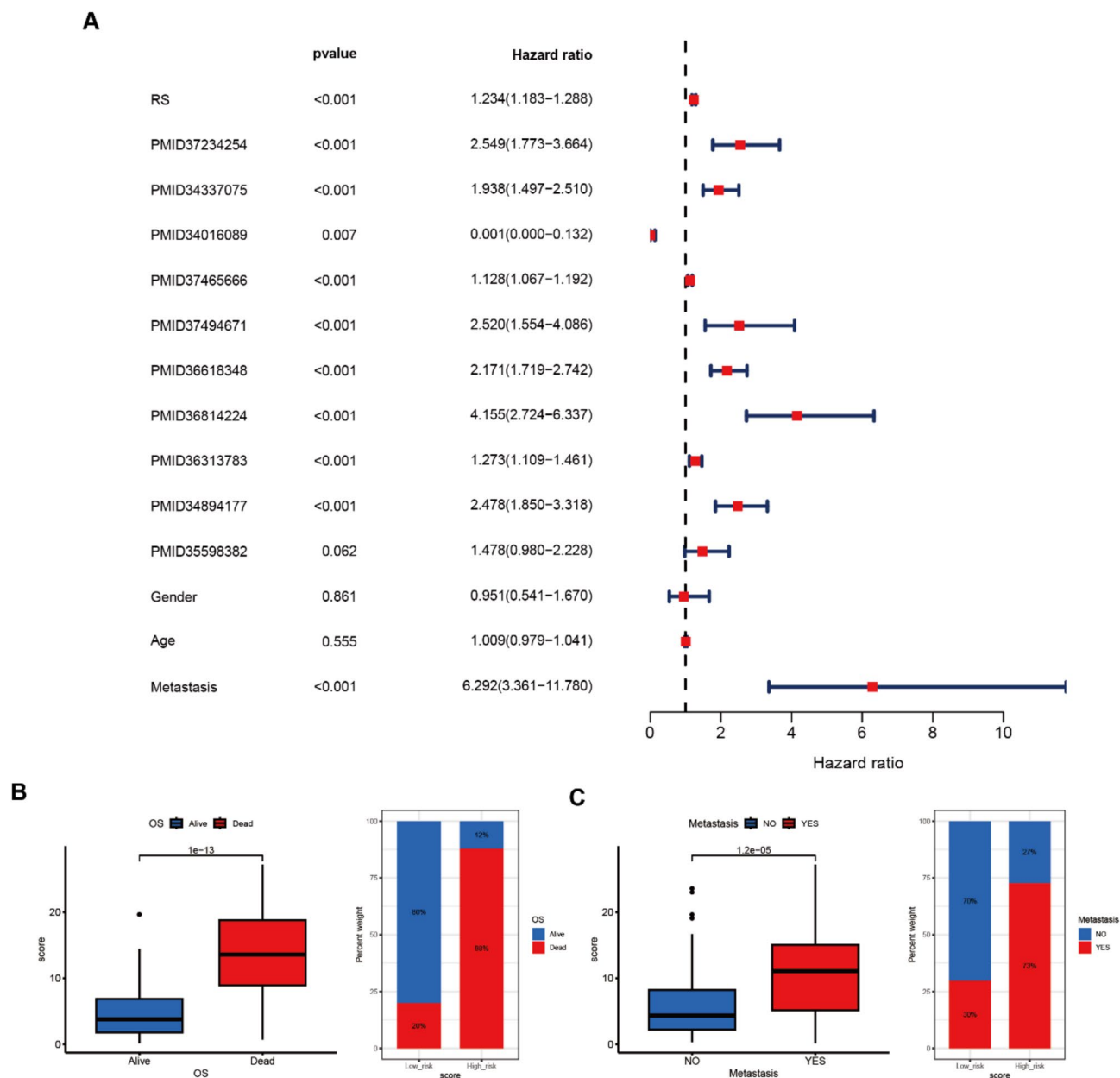


Fig. 6. The clinical signature of MLDPS and other signatures in the merged cohort. **(A)** Univariate Cox regression analysis of MLDPS and other signatures in the merged cohort. **(B)** Comparison of osteosarcoma between the high- and low-risk score groups. Left: Differences in the MLDPS risk scores between the OS groups. Right: Composition percentage of the high- and low-risk score groups in osteosarcoma. **(C)** Comparison of metastasis between the high- and low-risk score groups. Left: Differences in the MLDPS risk scores between the metastasis groups. Right: Composition percentage of the high- and low-risk score groups in metastasis.

Entospletinib_1630, IAP_5620_1428, PF4708671_1129, and Ribociclib_1632 (Fig. 11). Precise prediction of drug sensitivity may have the potential to enhance patient treatment outcomes.

Discussion

Osteosarcoma is a rare malignant tumor originating in bone tissue that predominantly afflicts adolescents and young adults^{65,66}. Clinicians and researchers face critical challenges due to the lack of readily available biomarkers for screening, stratified treatment, and prognostic follow-up, potentially resulting in inadequate treatment⁶⁷. To bridge these gaps, novel methodologies are needed to identify osteosarcoma biomarkers that accurately reflect the biological characteristics of the disease.

Recent studies have developed various prognostic signatures for osteosarcoma, many of which focus on specific biological pathways such as glycolysis, ferroptosis, and immunity^{68–70}. However, this focused approach

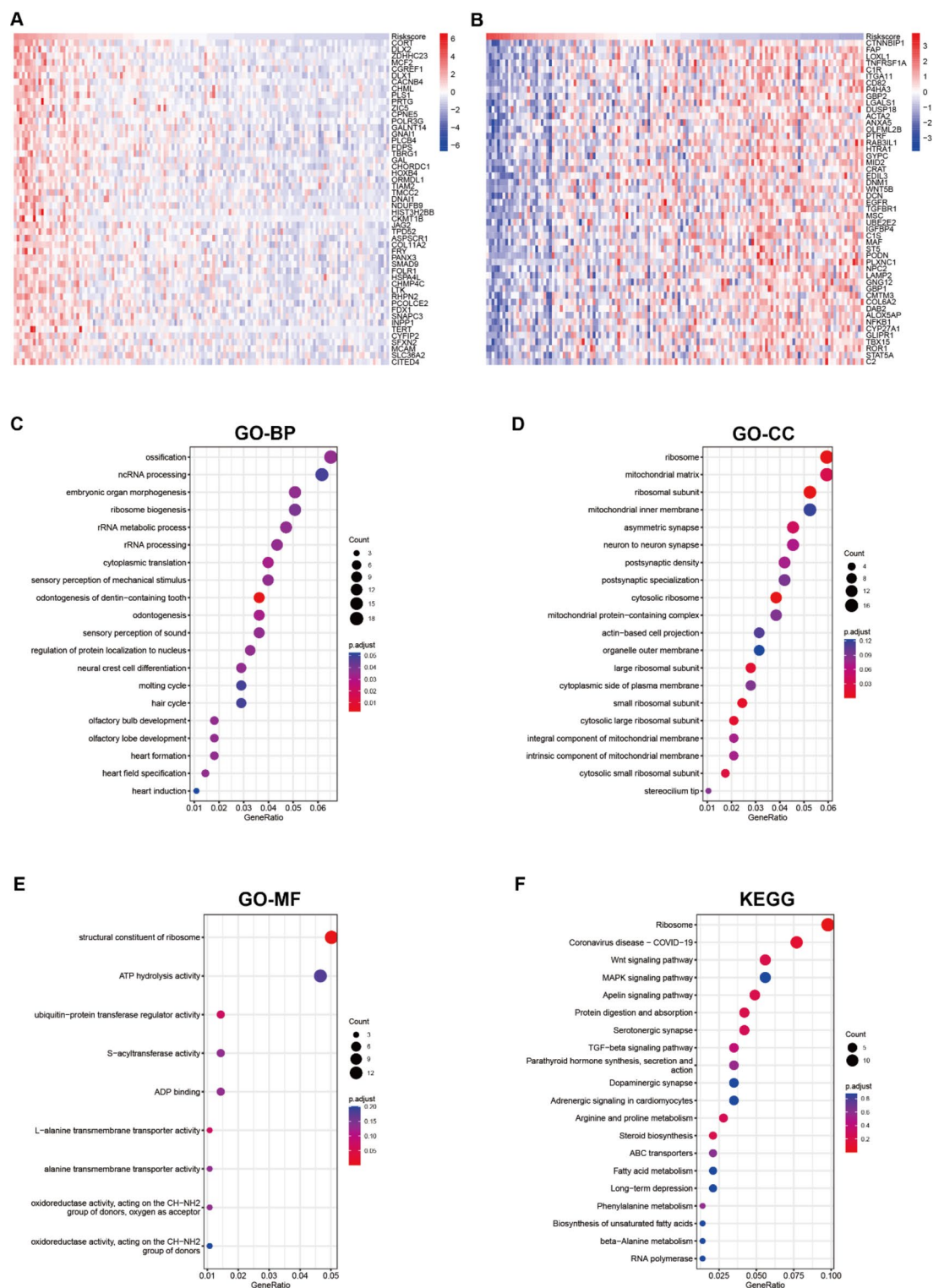


Fig. 7. Function and pathway analysis of MLDPS. (A) Heatmap of the top 50 positive correlation genes with the MLDPS risk score; (B) Heatmap of the top 50 negative correlation genes with the MLDPS risk score; The GO enrichment analysis of the top 500 positive correlation genes, including BP (C), CC (D) and MF (E); (F) The KEGG pathways enrichment analysis of the top 500 positive correlation genes.

neglects crucial biological processes that may also contribute to the oncogenesis and progression of osteosarcoma. Moreover, existing research indicates that researchers tend to select modeling algorithms based on their knowledge limitations and personal preferences^{23,61,71}. To compensate for these shortcomings, we devised an effective computational strategy for identifying optimal prognostic genes and constructed a highly accurate risk

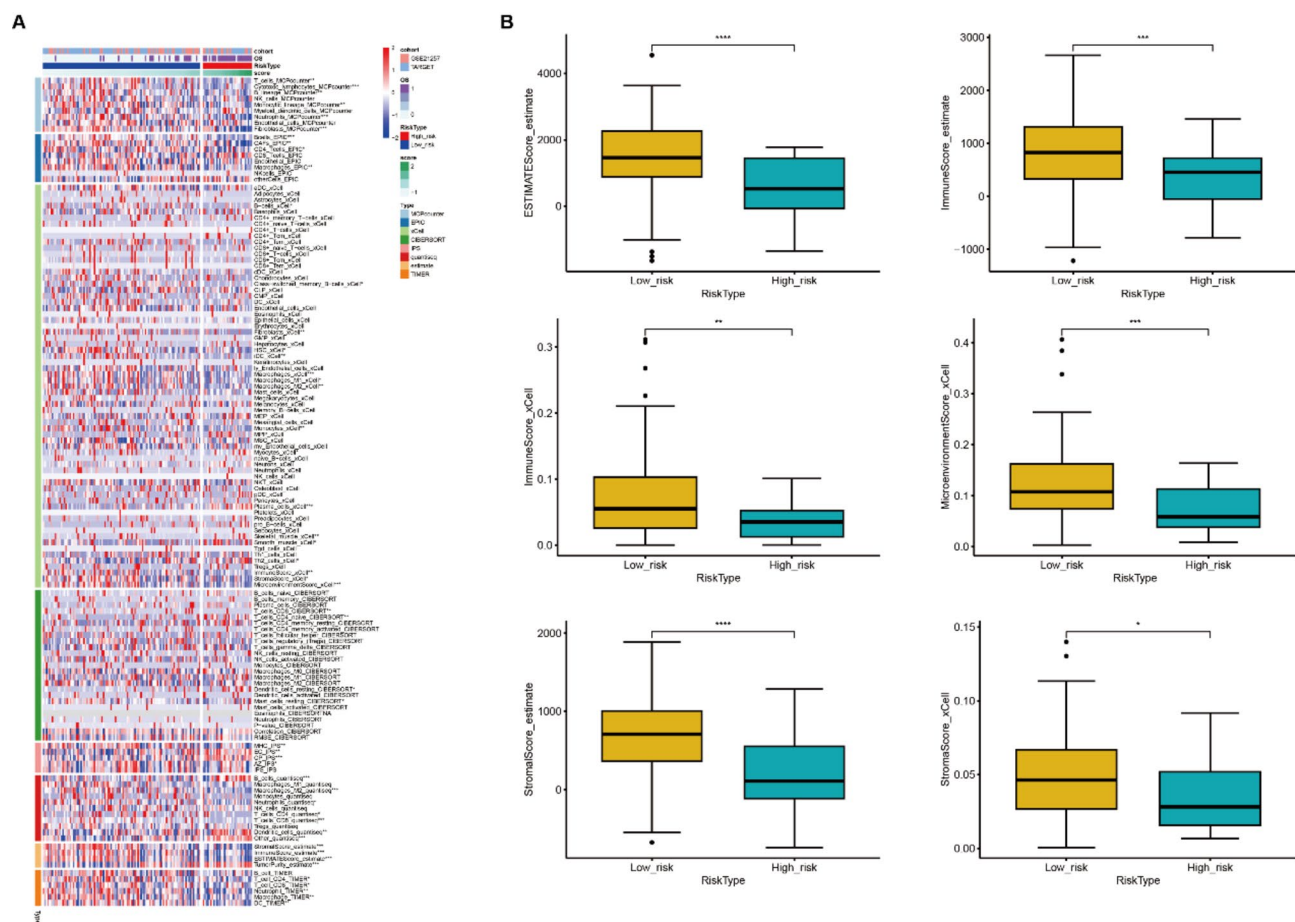


Fig. 8. Landscapes of the TME between the high- and low-risk score groups. (A) Heatmap indicates the difference in the TME between the high- and low-risk score groups. (B) The diversities of the ESTIMATEScore, ImmuneScore, MicroenvironmentScore, and StromalScore between the high- and low-risk score groups. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

prediction model for stratifying risk levels. First, 14 CPSGs were identified by univariate Cox regression analysis from the 14,174 intersection genes in the TARGET and GSE21257 cohorts. Enrichment analysis revealed that the identified CPSGs were mainly enriched in “pattern specification process,” “mitochondrial inner membrane,” “transcription coactivator binding,” “telomeric DNA binding,” “telomerase RNA binding,” and “Arginine and proline metabolism.” These results uncover the biological functions of genes linked to osteosarcoma prognosis and offer valuable insights into their roles in the molecular dynamics of the disease, paving the way for further targeted research.

A total of 66 combinations of 10 widely used ML algorithms were generated to develop prognostic signatures. Among the 10 acknowledged ML algorithms, RSF, LASSO, CoxBoost, and Stepwise Cox possess the capability for variable screening and reduction of data dimensionality. Ultimately, we built the optimal model (MLDPS) by selecting the combination of LASSO and RSF based on their highest average C-index (0.866) across all three cohorts. Overfitting is a common challenge in ML, often causing models to perform well on training dataset but poorly on external validation datasets^{37,72}. After reducing redundant information with LASSO, a seven-gene signature referred to as MLDPS was ultimately derived through RSF.

The timeROC curve analysis, KM survival analysis, and Cox regression analysis consistently indicated that MLDPS accurately and reliably predicted the prognosis of osteosarcoma patients across all three cohorts. The immune-related six-gene signature²³, TGF β -related two-gene signature⁹, and hypoxia-related four-gene signature³² achieved higher C-index values in one or two cohorts but exhibited suboptimal performance in others, likely due to limited generalization caused by overfitting⁷³. In contrary, our MLDPS consistently ranked first in C-index values and exhibited superior predictive efficacy across all cohorts. Stratification of osteosarcoma patients into high- and low-risk score groups indicated that individuals in the high-risk score group had a notably poorer OS in comparison to those in the low-risk score group. Additionally, patients with high-risk score exhibited an increased likelihood of developing metastasis, contributing to their unfavorable prognosis.

The MLDPS model comprised seven potential gene signatures: CTNNBIP1, CORT, DLX2, TERT, BBS4, SLC7A1, and NKX2-3. The seven novel genes discovered in our study have been validated in previous research as being linked to tumor initiation and progression. CTNNBIP1 has been suggested as a potential tumor suppressor in colorectal⁷⁴, breast⁷⁵, cervical⁷⁶, and gastric cancers⁷⁷. CORT has been identified as a regulator of the growth

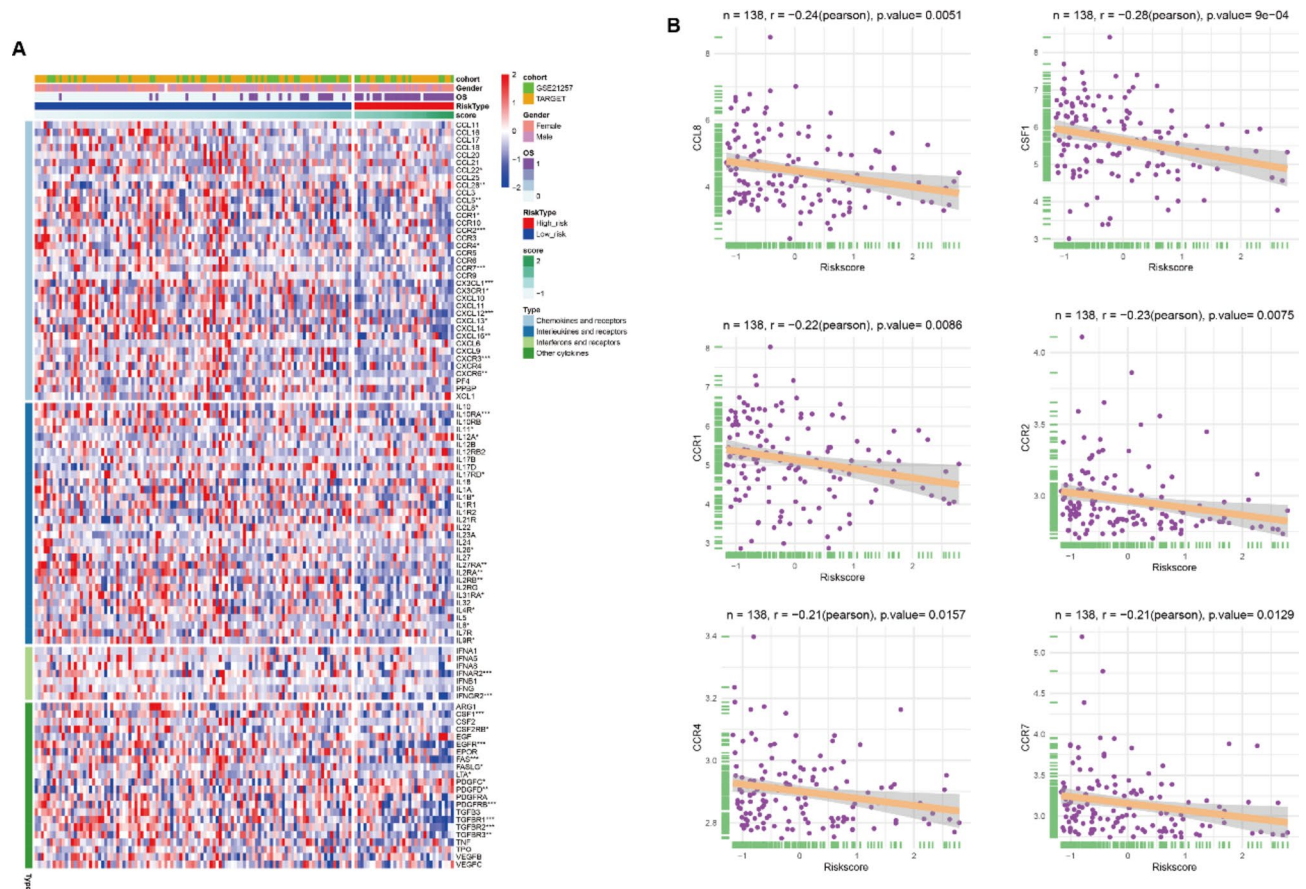


Fig. 9. The correlations between MLDPs and gene expression of chemokines and their receptors. **(A)** Correlation heatmap between MLDPs and expression of chemokines and their receptors. **(B)** Correlation analysis of MLDPs and CCL8, CSF1, CCR1, CCR2, CCR4 and CCR7 gene expression. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

and metastasis in lung and thyroid cancers^{78,79}. Additionally, elevated CORT expression has been correlated with a high-risk subgroup of osteosarcoma patients⁸⁰, consistent with our findings. DLX2 has been identified as an independent prognostic factor for colorectal cancer⁸¹. Recent research has provided increasing support for the safety of TERT-targeted immunotherapy^{82,83}. SLC7A1, an arginine transporter, has been suggested to be associated with liver cancer⁸⁴. NKX2-3 has been confirmed as an oncogenic driver in marginal-zone B-cell lymphomas⁸⁵. However, the functional roles of the BBS gene remain unclear, nor has it been linked to cancer progression.

Functional and pathway analyses were performed to elucidate the biological mechanisms underlying MLDPS. Enrichment analysis of the co-expressed genes correlated with MLDPS revealed significant enrichment in various biological processes and cellular components, namely “ossification,” “ncRNA processing,” “embryonic organ morphogenesis,” “ribosome biogenesis,” “rRNA metabolic process,” “rRNA processing,” “ribosome,” “mitochondrial matrix,” “ribosomal subunit,” “asymmetric synapse,” “cytosolic ribosome,” “structural constituent of ribosome,” and “Ribosome”. GSEA revealed a positive correlation between the high-risk score group and the upregulation of key gene sets involved in translation, rRNA processing in the nucleus and cytosol, the primary pathway of rRNA processing in the nucleolus and cytosol, and other biological pathways. The activation of the pathways associated with MLDPS was further validated using GSVA. The findings partially elucidate the mechanisms underlying metastasis and poor prognosis in the high-risk score group.

Next, correlation analysis was conducted to explore the differences in the immune landscape between high- and low-risk score groups. The findings of the TME indicated a negative correlation between MLDPs risk scores and ESTIMATEScore, ImmuneScore, MicroenvironmentScore, and StromalScore, and the gene expression levels of many chemokines and chemokine receptors. These observations suggest that MLDPs can differentiate the TME, and individuals in the high-risk score group may experience an immunosuppressive state. However, lower TIDE scores suggested that individuals with high MLDPs risk scores are more likely to respond to immunotherapy. These observations underscore the necessity for further exploration of immunotherapeutic approaches and indicate the existence of unidentified immune evasion mechanisms that warrant investigation^{86,87}. Our findings indicate that MLDPs might contribute to resistance against small-molecule drugs in human cancer cell lines. MLDPs has potential as both a prognostic indicator and a predictor of immunotherapy response, representing a significant advancement in osteosarcoma treatment.

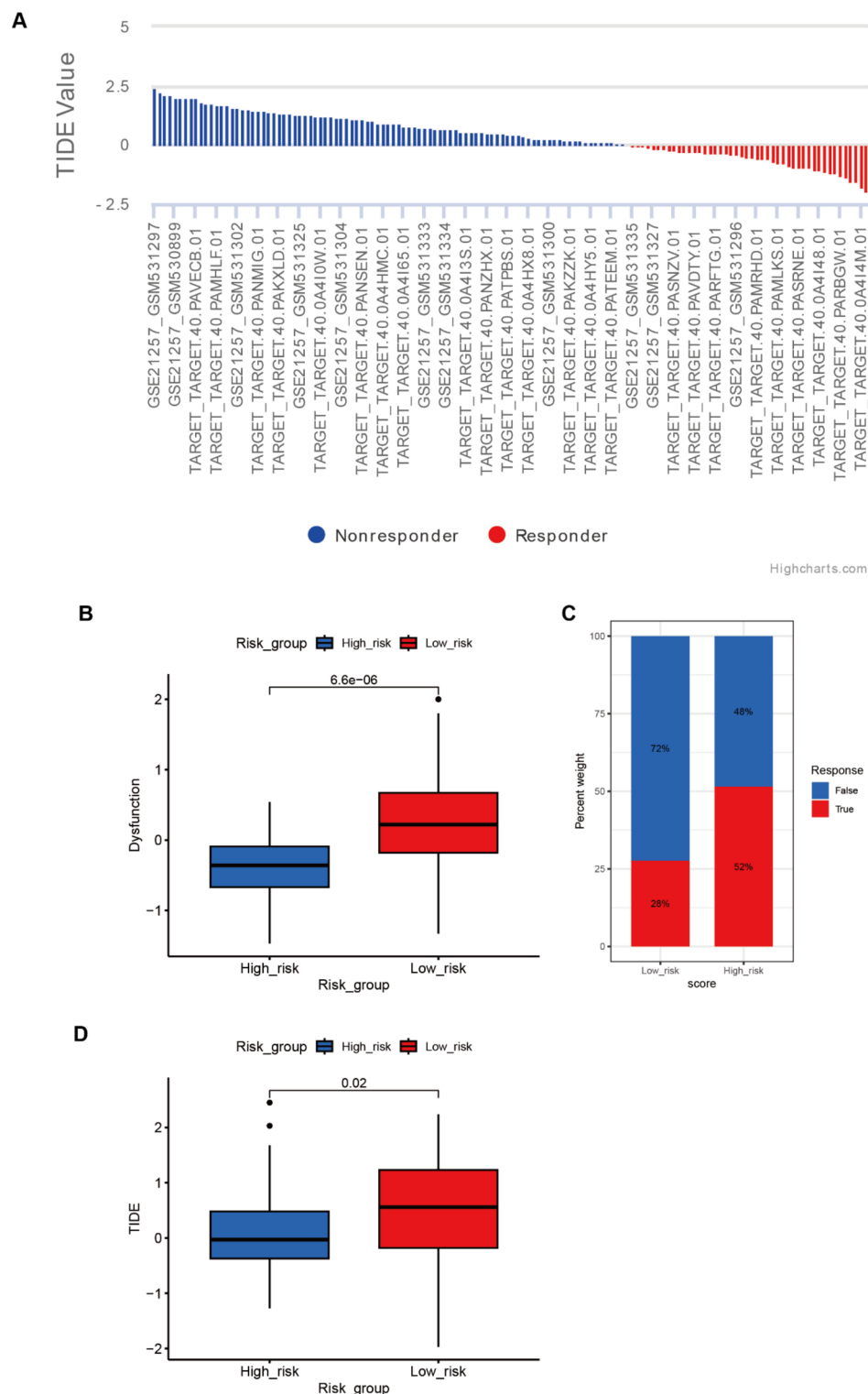


Fig. 10. Immunotherapeutic response prediction by TIDE. **(A)** Patients were classified as potential responders or non-responders to immunotherapy based on their TIDE values in the merged cohort. **(B)** Differences in Dysfunction scores between high- and low-risk score groups ($P < 0.001$). **(C)** Percent weight of clinical response to immunotherapeutic in high- and low-risk score groups. **(D)** Differences in TIDE scores between high- and low-risk score groups ($P = 0.02$).

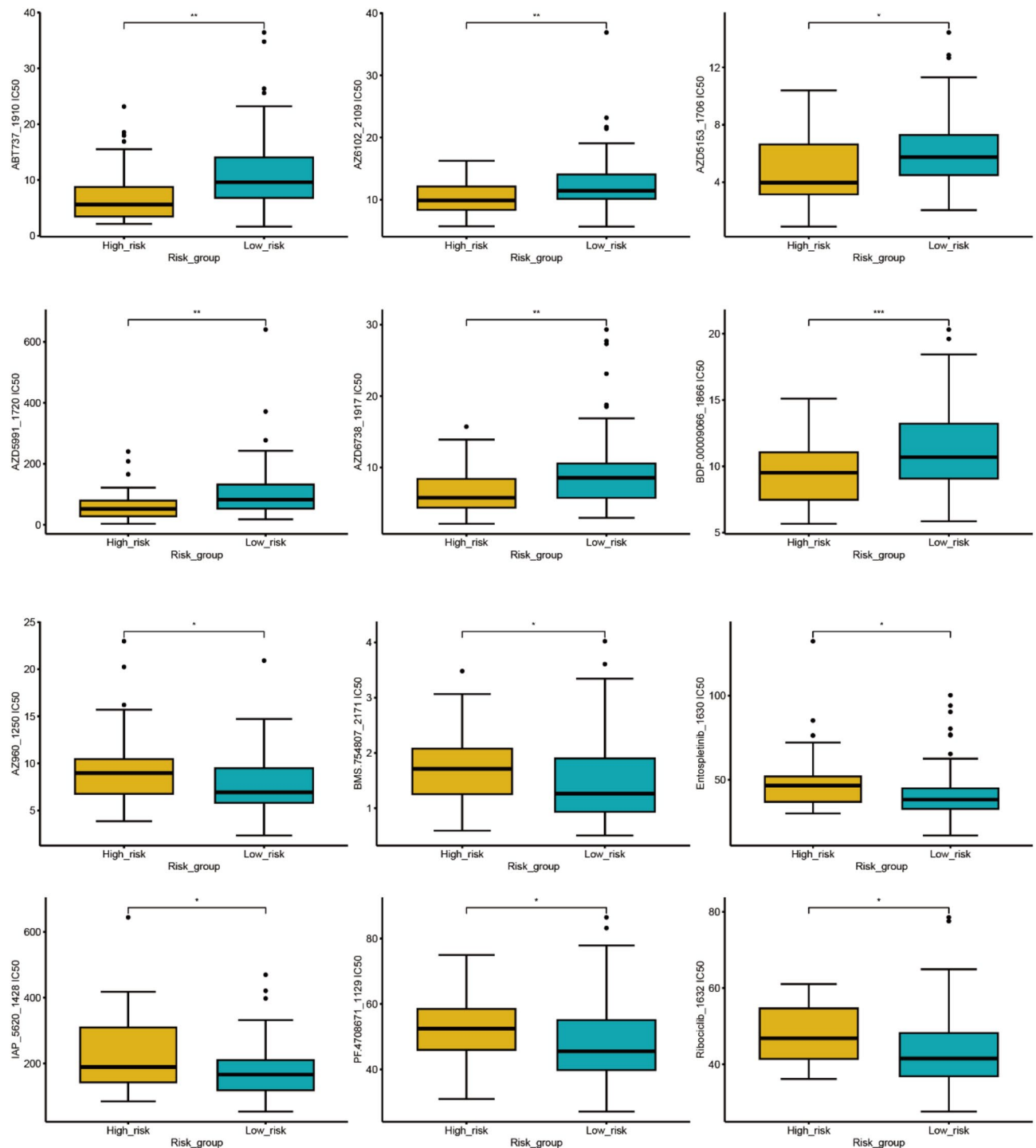


Fig. 11. Association between the risk model and drug sensitivity in osteosarcoma. Drug sensitivity of osteosarcoma patients with high- and low-risk score. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

To our knowledge, no prior studies have applied these feature extraction methods in the creation of a prognostic risk model for osteosarcoma. First, in contrast to existing prognostic models for a specific pathway, our model incorporated factors from CPSGs obtained from diverse cohorts, improving the robustness and prognostic relevance of the included genes. Second, to minimize bias from subjective preferences in modeling, we integrated 10 established ML algorithms into 66 combinations and identified the optimal model based on accuracy evaluation. Third, we methodically gathered data from two multicenter cohorts and selected the algorithm with the highest average C-index from the TARGET, GSE21257, and merged cohorts to construct the MLDPS. The seven-gene signature elucidated in this study empowers clinicians to better identify high-risk patients who may benefit from more rigorous surveillance and treatment strategies. By correlating MLDPS scores

with drug sensitivity and immune microenvironment characteristics, our model offers valuable insights for personalized treatment strategies. Enrichment analyses illuminate molecular processes involved in osteosarcoma progression, including pathways related to ribosome biogenesis, rRNA processing, and immune suppression. These findings establish a foundation for further mechanistic studies and targeted therapeutic development.

While promising, our study has certain limitations. Despite presenting robust computational findings, experimental validation is necessary to affirm the biological roles of the identified genes in osteosarcoma progression and metastasis. Functional studies employing both in vitro and in vivo models may provide profound insights into the mechanisms underlying the MLDPS. Our results suggest an immunosuppressive tumor microenvironment in patients with high MLDPS scores, yet the exact mechanisms are unknown. Further studies are needed to investigate how these patients evade immune responses and to determine whether combination strategies could overcome immunotherapy resistance. Additionally, despite correction algorithms, the analyzed cohorts exhibited variations in size and sequencing platform. Therefore, larger prospective multicenter cohort studies are warranted to validate the clinical utility of MLDPS.

Over the next five years, we expect major progress in integrating ML with omics data for cancer research, particularly in osteosarcoma. The combination of genomic, transcriptomic, proteomic, and metabolomic data could lead to better prognostic models. With growing access to electronic health records and patient-specific omics data, ML-based tools like MLDPS could enhance clinical decision-making platforms by enabling real-time risk assessment and therapy optimization. As immunotherapy advances, the identification of predictive biomarkers for patient response will gain increasing importance. Our findings establish a preliminary framework for linking risk stratification with immunotherapy response, potentially guiding future clinical trials. Collaboration among researchers, clinicians, and bioinformaticians is crucial to tackle osteosarcoma data variability and to create standardized protocols for data processing and model validation.

Conclusion

In summary, leveraging 14 CPSGs from the GSE21257 and TARGET cohorts, we developed and validated a robust MLDPS model based on 66 ML algorithm combinations. The MLDPS exhibited enhanced prediction accuracy for patient prognosis across diverse cohorts and showed significant associations with immunotherapy response and chemotherapy efficacy. Consequently, our MLDPS has significant potential as a tool to guide clinical decision-making and customize treatment strategies for osteosarcoma patients. However, additional research is needed to fully evaluate the capabilities of MLDPS and clarify its clinical relevance.

Data availability

The datasets used and/or analysed during this study are available from the corresponding author on request.

Received: 3 May 2024; Accepted: 25 April 2025

Published online: 19 May 2025

References

1. Bielack, S. S. et al. Prognostic factors in high-grade osteosarcoma of the extremities or trunk: an analysis of 1,702 patients treated on neoadjuvant cooperative osteosarcoma study group protocols. *J. Clin. Oncol.* **20** (3), 776–790 (2002).
2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **70** (1), 7–30 (2020).
3. Marchandet, L. et al. Mechanisms of resistance to conventional therapies for osteosarcoma. *Cancers* **13** (4), 683 (2021).
4. Rainusso, N., Wang, L. L. & Yustein, J. T. The adolescent and young adult with cancer: state of the Art -- bone tumors. *Curr. Oncol. Rep.* **15** (4), 296–307 (2013).
5. Mirabello, L., Troisi, R. J. & Savage, S. A. Osteosarcoma incidence and survival rates from 1973 to 2004: data from the surveillance, epidemiology, and end results program. *Cancer* **115** (7), 1531–1543 (2009).
6. Adamopoulos, C., Gargalionis, A. N., Basdra, E. K. & Papavassiliou, A. G. Deciphering signaling networks in osteosarcoma pathobiology. *Exp. Biol. Med. (Maywood)*. **241** (12), 1296–1305 (2016).
7. Kansara, M., Teng, M. W., Smyth, M. J. & Thomas, D. M. Translational biology of osteosarcoma. *Nat. Rev. Cancer*. **14** (11), 722–735 (2014).
8. Jo, V. Y. & Fletcher, C. D. WHO classification of soft tissue tumours: an update based on the 2013 (4th) edition. *Pathology* **46** (2), 95–104 (2014).
9. Liu, D. et al. Comprehensive landscape of TGFβ-related signature in osteosarcoma for predicting prognosis, immune characteristics, and therapeutic response. *J. Bone Oncol.* **40**, 100484 (2023).
10. Xie, L., Zeng, J. & He, M. Identification and verification of a BMPs-related gene signature for osteosarcoma prognosis prediction. *BMC Cancer*. **23** (1), 181 (2023).
11. Zhong, Y. et al. Dissecting the effect of sphingolipid metabolism gene in progression and microenvironment of osteosarcoma to develop a prognostic signature. *Front. Endocrinol. (Lausanne)*. **13**, 1030655 (2022).
12. Astolfi, A. et al. Genomic database analysis of uterine leiomyosarcoma mutational profile. *Cancers (Basel)*. **12** (8), 2126 (2020).
13. Rizzo, A. et al. Impact of chemotherapy in the adjuvant setting of early stage uterine leiomyosarcoma: A systematic review and updated Meta-Analysis. *Cancers (Basel)*. **12** (7), 1899 (2020).
14. Rizzo, A. et al. Current status of the adjuvant therapy in uterine sarcoma: A literature review. *World J. Clin. Cases*. **7** (14), 1753–1763 (2019).
15. Guven, D. C. et al. The association between albumin levels and survival in patients treated with immune checkpoint inhibitors: A systematic review and meta-analysis. *Front. Mol. Biosci.* **9**, 1039121 (2022).
16. Dall'Olio, F. G. et al. Immortal time bias in the association between toxicity and response for immune checkpoint inhibitors: a meta-analysis. *Immunotherapy* **13** (3), 257–270 (2021).
17. Rizzo, A. et al. Peripheral neuropathy and headache in cancer patients treated with immunotherapy and immuno-oncology combinations: the MOUSEION-02 study. *Expert Opin. Drug Metab. Toxicol.* **17** (12), 1455–1466 (2021).
18. Rizzo, A. Identifying optimal first-line treatment for advanced non-small cell lung carcinoma with high PD-L1 expression: a matter of debate. *Br. J. Cancer*. **127** (8), 1381–1382 (2022).
19. Federico, A. D. et al. Immunotherapy in pancreatic cancer: why do we keep failing?? A focus on tumor immune microenvironment, predictive biomarkers and treatment outcomes. *Cancers (Basel)*. **14** (10), 2429 (2022).

20. Ricci, A. D., Rizzo, A. & Brandi, G. Immunotherapy in biliary tract cancer: worthy of a second look. *Cancer Control*. **27** (3), 1073274820948047 (2020).
21. Sahin, T. K., Rizzo, A., Aksoy, S. & Guven, D. C. Prognostic significance of the Royal marsden hospital (RMH) score in patients with cancer: A systematic review and Meta-Analysis. *Cancers (Basel)*. **16** (10), 1835 (2024).
22. Xu, F., Yan, J., Peng, Z., Liu, J. & Li, Z. Comprehensive analysis of a Glycolysis and cholesterol synthesis-related genes signature for predicting prognosis and immune landscape in osteosarcoma. *Front. Immunol.* **13**, 1096009 (2022).
23. Shi, D. et al. Integrative analysis of immune-related multi-omics profiles identifies distinct prognosis and tumor microenvironment patterns in osteosarcoma. *Mol. Oncol.* **16** (11), 2174–2194 (2022).
24. Mao, X. et al. Prognostic and immunological significance of an M1 macrophage-related gene signature in osteosarcoma. *Front. Immunol.* **14**, 1202725 (2023).
25. Qiu, Y. L., Zheng, H., Devos, A. & Selby, H., Gevaert, O. A meta-learning approach for genomic survival analysis. *Nat. Commun.* **11** (1), 6350 (2020).
26. Narwade, N. et al. Mapping of scaffold/matrix attachment regions in human genome: a data mining exercise. *Nucleic Acids Res.* **47** (14), 7247–7261 (2019).
27. Raimondi, D. et al. DEOGEN2: prediction and interactive visualization of single amino acid variant deleteriousness in human proteins. *Nucleic Acids Res.* **45** (W1), W201–w206 (2017).
28. Sahin, T. K. et al. Prognostic value of Neutrophil-to-Eosinophil ratio (NER) in cancer: A systematic review and Meta-Analysis. *Cancers (Basel)*. **16** (21), 3689 (2024).
29. Le, D. T. et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 Blockade. *Sci. (New York N Y)*. **357** (6349), 409–413 (2017).
30. Doebele, R. C. et al. Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1–2 trials. *Lancet Oncol.* **21** (2), 271–282 (2020).
31. Gao, S. et al. Identification and construction of combinatory Cancer Hallmark-Based gene signature sets to predict recurrence and chemotherapy benefit in stage II colorectal Cancer. *JAMA Oncol.* **2** (1), 37–45 (2016).
32. Jiang, F. et al. A hypoxia Gene-Based signature to predict the survival and affect the tumor immune microenvironment of osteosarcoma in children. *J. Immunol. Res.* **2021**(2021), 5523832 (2021).
33. Fu, Y. et al. Construction and validation of a novel apoptosis-associated prognostic signature related to osteosarcoma metastasis and immune infiltration. *Transl. Oncol.* **22**(2022), 101452 (2022).
34. He, L., Yang, H. & Huang, J. The tumor immune microenvironment and immune-related signature predict the chemotherapy response in patients with osteosarcoma. *BMC cancer*. **21** (1), 581 (2021).
35. Chen, Y. et al. Advances in targeted therapy for osteosarcoma based on molecular classification. *Pharmacol. Res.* **169**(2021), 105684 (2021).
36. Zhao, J., Dean, D. C., Hornicek, F. J., Yu, X. & Duan, Z. Emerging next-generation sequencing-based discoveries for targeted osteosarcoma therapy. *Cancer Lett.* **474**(2020), 158–167 (2020).
37. Deo, R. C. Machine learning in medicine. *Circulation* **132** (20), 1920–1930 (2015).
38. Greener, J. G., Kandathil, S. M., Moffat, L. & Jones, D. T. A guide to machine learning for biologists. *Nat. Rev. Mol. Cell. Biol.* **23** (1), 40–55 (2022).
39. Bigorra, L., Larriba, I. & Gutiérrez-Gallego, R. Machine learning algorithms for accurate differential diagnosis of lymphocytosis based on cell population data. *Br. J. Haematol.* **184** (6), 1035–1037 (2019).
40. Chen, Z. et al. A machine learning model to predict the triple negative breast Cancer immune subtype. *Front. Immunol.* **12**, 749459 (2021).
41. Liu, L. P. et al. Identification and Validation of the Pyroptosis-Related Molecular Subtypes of Lung Adenocarcinoma by Bioinformatics and Machine Learning. *Front. Cell Dev. Biol.* **9**(2021), 756340 (2021).
42. Zhang, J. et al. Machine learning assistants construct oxidative Stress-Related gene signature and discover potential therapy targets for acute myeloid leukemia. *Oxid. Med. Cell. Longev.* **2022**(2022), 1507690 (2022).
43. Li, W. et al. Machine Learning-Based Prediction of Lymph Node Metastasis Among Osteosarcoma Patients. *Front. Oncol.* **12**(2022), 797103 (2022).
44. Huang, H. et al. Employing machine learning using ferroptosis-related genes to construct a prognosis model for patients with osteosarcoma. *Front. Genet.* **14**(2023), 1099272 (2023).
45. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16** (5), 284–287 (2012).
46. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A.* **102** (43), 15545–15550 (2005).
47. Kanehisa, M. & Goto, S. K. E. G. G. Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28** (1), 27–30 (2000).
48. Kanehisa, M. Toward Understanding the origin and evolution of cellular organisms. *Protein Sci.* **28** (11), 1947–1951 (2019).
49. Kanehisa, M. et al. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51** (D1), D587–D592 (2023).
50. Hänzelmann, S., Castelo, R. & Guinney, J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC. Bioinformatics* **14**(2013), 7 (2013).
51. Zeng, D. et al. IOBR: Multi-Omics Immuno-Oncology Biological Research to Decode Tumor Microenvironment and Signatures. *Front. Immunol.* **12**(2021), 687975 (2021).
52. Becht, E. et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17** (1), 218 (2016).
53. Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E. & Gfeller, D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife* **6**, 26476 (2017).
54. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18** (1), 220 (2017).
55. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37** (7), 773–782 (2019).
56. Finotello, F. et al. Molecular and Pharmacological modulators of the tumor immune contexture revealed by Deconvolution of RNA-seq data. *Genome Med.* **11** (1), 34 (2019).
57. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**(2013), 2612 (2013).
58. Li, T. et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **48** (W1), W509–w514 (2020).
59. Jiang, P. et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24** (10), 1550–1558 (2018).
60. Sia, D. et al. Integrative molecular analysis of intrahepatic cholangiocarcinoma reveals 2 classes that have different outcomes. *Gastroenterology* **144** (4), 829–840 (2013).
61. Li, W. et al. A deep belief network-based clinical decision system for patients with osteosarcoma. *Front. Immunol.* **13**, 1003347 (2022).
62. Sun, J. et al. Exploring the role of lactylation-related genes in osteosarcoma: A deep dive into prognostic significance and therapeutic potential. *Environ. Toxicol.* **39** (2), 1001–1017 (2024).

63. Benech, N. & Sokol, H. Fecal microbiota transplantation in Gastrointestinal disorders: time for precision medicine. *Genome Med.* **12** (1), 58 (2020).
64. Ong, E. et al. Modelling kidney disease using ontology: insights from the kidney precision medicine project. *Nat. Rev. Nephrol.* **16** (11), 686–696 (2020).
65. Gill, J. & Gorlick, R. Advancing therapy for osteosarcoma. *Nat. Rev. Clin. Oncol.* **18** (10), 609–624 (2021).
66. Meltzer, P. S. & Helman, L. J. New horizons in the treatment of osteosarcoma. *N Engl. J. Med.* **385** (22), 2066–2076 (2021).
67. Somarelli, J. A. et al. A comparative oncology drug discovery pipeline to identify and validate new treatments for osteosarcoma. *Cancers (Basel).* **12** (11), 3335 (2020).
68. Huang, W., Xiao, Y., Wang, H., Chen, G. & Li, K. Identification of risk model based on glycolysis-related genes in the metastasis of osteosarcoma. *Front. Endocrinol. (Lausanne)* **13**(2022), 1047433 (2022).
69. Wang, X. et al. A ferroptosis-related gene signature associated with immune landscape and therapeutic response in osteosarcoma. *Front. Oncol.* **12**(2022), 1024915 (2022).
70. Yang, W. et al. A cuproptosis-related genes signature associated with prognosis and immune cell infiltration in osteosarcoma. *Front. Oncol.* **12**(2022), 1015094 (2022).
71. Liu, Z. et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat. Commun.* **13** (1), 816 (2022).
72. Chandrasekaran, S. N., Ceulemans, H., Boyd, J. D. & Carpenter, A. E. Image-based profiling for drug discovery: due for a machine-learning upgrade? *Nat. Rev. Drug Discov.* **20** (2), 145–159 (2021).
73. Eraslan, G., Simon, L. M., Mircea, M., Mueller, N. S. & Theis, F. J. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat. Commun.* **10** (1), 390 (2019).
74. Sekiya, T. et al. Overexpression of Icat induces G(2) arrest and cell death in tumor cell mutants for adenomatous polyposis coli, beta-catenin, or Axin. *Cancer Res.* **62** (11), 3322–3326 (2002).
75. Mukerjee, A., Ranjan, A. P. & Vishwanatha, J. K. Targeted Nanocurcumin therapy using Annexin A2 Antibody improves tumor accumulation and therapeutic efficacy against highly metastatic breast Cancer. *J. Biomed. Nanotechnol.* **12** (7), 1374–1392 (2016).
76. Jiang, Y. et al. Inhibitor of β -catenin and TCF (ICAT) promotes cervical cancer growth and metastasis by disrupting E-cadherin/ β -catenin complex. *Oncol. Rep.* **38** (5), 2597–2606 (2017).
77. Kosari-Monfared, M. et al. CTNNBIP1 downregulation is associated with tumor grade and viral infections in gastric adenocarcinoma. *J. Cell. Physiol.* **234** (3), 2895–2904 (2019).
78. Cassoni, P. et al. Ghrelin and cortistatin in lung cancer: expression of peptides and related receptors in human primary tumors and in vitro effect on the H345 small cell carcinoma cell line. *J. Endocrinol. Invest.* **29** (9), 781–790 (2006).
79. Cassoni, P. et al. Cortistatin-14 inhibits cell proliferation of human thyroid carcinoma cell lines of both follicular and parafollicular origin. *J. Endocrinol. Invest.* **25** (4), 362–368 (2002).
80. Wu, Z. L. et al. Development of a novel immune-related genes prognostic signature for osteosarcoma. *Sci. Rep.* **10** (1), 18402 (2020).
81. Yao, H., Li, C. & Tan, X. An age stratified analysis of the biomarkers in patients with colorectal cancer. *Sci. Rep.* **11** (1), 22464 (2021).
82. Guterres, A. N. & Villanueva, J. Targeting telomerase for cancer therapy. *Oncogene* **39** (36), 5811–5824 (2020).
83. Khermlina, G., Ikeda, S. & Kurzrock, R. The biology of hepatocellular carcinoma: implications for genomic and immune therapies. *Mol. Cancer.* **16** (1), 149 (2017).
84. Chafai, A., Fromm, M. F., König, J. & Maas, R. The prognostic biomarker L-homoarginine is a substrate of the cationic amino acid transporters CAT1, CAT2A and CAT2B. *Sci. Rep.* **7** (1), 4767 (2017).
85. Robles, E. F. et al. Homeobox NKX2-3 promotes marginal-zone lymphomagenesis by activating B-cell receptor signalling and shaping lymphocyte dynamics. *Nat. Commun.* **7**, 11889 (2016).
86. Li, X. Y. et al. Targeting CD39 in Cancer reveals an extracellular ATP- and Inflammasome-Driven tumor immunity. *Cancer Discov.* **9** (12), 1754–1773 (2019).
87. Flores, C. et al. Massive clonal expansion of medulloblastoma-specific T cells during adoptive cellular therapy. *Sci. Adv.* **5** (11), eaav9879 (2019).

Acknowledgements

We sincerely thank all authors for their significant contributions to this study. We would like to extend our special thanks to Suping Hu and Dedong Liu for their valuable contributions in preparing the figures and insightful suggestions. We also appreciate Elsevier Language Editing Services for their linguistic editing and proofreading of the manuscript.

Author contributions

Q.Z., W.H., Q.W., and T.Z. designed the study; Q.Z., W.H., X.X., and Y.X. collected and analyzed the data; X.W., X.Y., and J.C. prepared the figures, and drafted the manuscript. Q.Z., Q.W., T.Z., S.D., and T.Z. reviewed and revised the manuscript. All listed authors have actively participated in the research, contributed to the manuscript preparation, and have read and approved the final manuscript.

Funding

This research was approved by the National Natural Science Foundation of China (No. 82260422), the Science and Technology Program of Jiangxi Provincial Health Commission, China (No. 202310050), the Science and Technology Program of Ganzhou, China (No. 2023LNS36838), the Key Project of Jiangxi Provincial Natural Science Foundation (No. 20242BAB26154), the Science and Technology Plan of Health Commission of Jiangxi Province (No. 202311903 and No. 202310736), the Jiangxi Traditional Chinese Medicine Science and Technology Program (No. 2022A136 and No. SZYYB20231890), the City-Level Scientific Research Program of the Ganzhou Health Commission (GZWJW202402043), the General Program of the Jiangxi Administration of Traditional Chinese Medicine (2023A0217), and the “Technology+Medical” - Targeted Project of the First Affiliated Hospital of Gannan Medical University (No. 2023NS326774).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-00179-z>.

Correspondence and requests for materials should be addressed to T.Z., X.X. or Q.W.

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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025