scientific reports

Check for updates

OPEN

The homeobox family gene signature predicts the prognosis of osteosarcoma and correlates with immune invasion

Wenda Liu[1,2], Kezhou Xia[1,2], Xinghan Huang[1], Zhun Wei[1], Zicheng Wei[1] & Weichun Guo[1]✉

Osteosarcoma (OS) is a prevalent invasive bone cancer, with numerous homeobox family genes implicated in tumor progression. This study aimed to develop a prognostic model using HOX family genes to assess osteosarcoma patient outcomes. Data from osteosarcoma patients in The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) cohorts were collected. LASSO regression and multivariate COX regression analyses were employed to create and validate a risk-prognosis model in a validation cohort. Four genes (HOXA1, HOXA5, HOXA6, HOXA13) were identified to construct the risk-prognosis model. Patients were categorized into high-risk and low-risk groups, with significantly better prognosis observed in the low-risk group. A nomogram was developed to predict patients' overall survival. Variances in gene function were primarily concentrated in immune-related pathways. ssGSEA indicated that immune cell content and function were relatively deficient in the high-risk group. Notably, HOXA1 overexpression suppressed osteosarcoma cell proliferation, migration, invasion, and tumor growth. The model exhibited high accuracy and versatility, enhancing early diagnosis rates and aiding clinicians in decision-making and personalized treatment.

Keywords Osteosarcoma, Homeobox family genes, Risk-prognostic model, Immune-related pathways, Individualized treatment

Osteosarcoma, a highly aggressive primary bone malignancy commonly affecting long bones, particularly in the extremities[1], is the predominant primary bone cancer in children and adolescents, constituting approximately 20% of all pediatric bone tumors[2]. Osteosarcoma exhibits a bimodal age distribution, with the first peak occurring around the age of 15 during the growth spurt period, and a smaller peak in older adults[3,4]. Osteosarcoma has a high propensity for metastasis to distant sites, most commonly the lungs[1,5,6]. Treatment for osteosarcoma is usually surgical resection and chemotherapy[7]. In recent years, molecular profiling studies have identified genetic alterations and signaling pathways involved in osteosarcoma development, providing potential targets for novel therapies[8]. Immunotherapies have shown promise in early clinical trials and may play a role in future treatment strategies[9–11]. Although there have been significant advances in the treatment of osteosarcoma, long-term survival remains unsatisfactory[12]. Accurate and timely diagnosis remains a major barrier, with most patients diagnosed with metastasis or recurrence, and 5-year survival rates of less than 20%[13]. Therefore, we need to further explore the molecular profile of osteosarcoma and identify reliable biomarkers to improve early detection, personalized treatment, and long-term survival.

The homeobox genes (***HOX***) are a group of important transcriptional regulatory factors, which mainly combine DNA to act on target genes and are highly conserved in evolution[14–16]. They can participate in embryonic development, cell identification, cell differentiation, cell metabolism, apoptosis, autophagy, and other processes[17,18]. A total of 39 ***HOX*** genes have been found in mammals, located in clusters of four different chromosomes (7p15, 17q21.2, 12q13, 2q31)[19]. In recent years, studies have found that ***HOX*** gene family members are closely related to tumors[20,21]. For example, in osteosarcoma, Hsa_circ_0007031 promotes the proliferation and migration of osteosarcoma cells by sponging miR-196a-5p to regulate the HOXB6[22]. HOXB5 promotes the progression and metastasis of osteosarcoma cells by activating the JAK2/STAT3 signaling pathway[23]. Knockdown of ***HOXB8*** efficiently prevented the activation of the Wnt/β-catenin signaling pathway and dramatically repressed the migration and invasion of OS cells[24]. ***HOXA9*** can also regulate OS cell proliferation, invasion, and

[1]Department of Orthopaedics, Renmin Hospital of Wuhan University, 238 Jiefang Road, Wuhan 430060, Hubei Province, China. [2]Wenda Liu and Kezhou Xia have contributed equally to this work and share the first authorship. ✉email: guoweichun@aliyun.com

migration[25]. ***HOX-C6***, ***-B3***, and ***-B4*** were also shown to be more abundant in osteosarcoma cells than in normal cells[26]. However, the role of HOX family genes in the development of osteosarcoma remains unclear. Therefore, it is promising to investigate how HOX family genes regulate the progression of osteosarcoma and the tumor immune microenvironment (TIME).

Based on the roles and characteristics of the ***HOX*** gene family, we used online databases The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) to analyze and study their roles in osteosarcoma. We constructed a risk prognostic model composed of four ***HOX*** family genes, which can predict the prognosis of osteosarcoma patients. We verified the accuracy and applicability of this prognostic model in internal and external cohorts. We also built a nomogram to verify whether it can be used for clinical decision-making. Finally, we analyzed the relationship between this model and immune cell function and immune microenvironment in osteosarcoma patients. Our objective is to develop a HOX family prognostic model to predict prognosis in osteosarcoma patients and to evaluate the tumor microenvironment and new targets for immunotherapy.

Materials and methods

Data acquisition

RNA sequencing data and clinical information for 88 osteosarcoma patients were obtained from The Cancer Genome Atlas (https://portal.gdc.cancer.gov/) in FPKM format. Additionally, gene expression and clinical data from the GSE39055 cohort in the Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo/) were collected to validate the prognostic model. After data curation, 123 patients were included in the study. Batch effect was mitigated during data processing.

Identification and expression analysis of HOX family genes

We got the ***HOX*** gene family members from previous research and the GeneCards database (https://www.genecards.org/). Then we analyzed the expression correlation according to The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) web-based database (string-interaction.org)[27]. Heatmaps visually represented the expression patterns and correlations of HOX genes in osteosarcoma patients.

Construction and validation of risk prognosis model based on HOX family

Osteosarcoma patients in the TCGA dataset were randomly divided into training and testing cohorts. LASSO regression and multivariate Cox regression analyses were performed to identify four HOX genes and their coefficients for constructing the prognostic model. We calculated the risk score for each osteosarcoma patient in all cohorts using the following formula: Risk score $=\sum_{k=1}^{n}(Coefficient\,(i) * Expr(i))$

, among them, coefficient is the regression coefficient we calculated , Expr is the expression of ***HOXs***. Based on the median risk score, we divided patients in each cohort into high- and low-risk groups. Survival analysis and receiver operating characteristic (ROC) curves validated the model's predictive accuracy across cohorts.

Nomogram construction

A nomogram incorporating patient risk scores and clinical characteristics was simulated to predict overall survival in osteosarcoma patients. The nomogram's predictive efficacy was assessed using TCGA and GSE39055 cohorts[28].

Differential genes between high- and low-risk groups

We divided patients into high- and low-risk groups in the TCGA entire cohort based on the median risk score and then analyzed and screened for the differentially expressed genes (DEGs) in osteosarcoma patients in the two groups, and the screening standard was $|\log2FC| > 0.5$ and p-value < 0.05. We then analyzed the expression of these significant differential genes in each osteosarcoma patient and drew a heatmap. Finally, we studied the expression correlation of these genes and identified the hub genes.

Functional enrichment analysis

Through functional enrichment analysis, we can understand which biological functions or pathways may affect the survival and prognosis of osteosarcoma patients. we performed the Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. To validate the results of the above functional analysis and to more intuitively see which pathways and functions affect the survival and prognosis of osteosarcoma patients, we performed a GSEA enrichment analysis between the two groups. We used the standard genetic set " c2.cp.kegg.v2023.1.Hs.symbols.gmt " to analyze.

Difference analysis of immune infiltration

To assess the prognostic risk score's effectiveness in the tumor microenvironment (TME), we evaluated immunological scores, estimate scores, stromal scores, and tumor purity in the immune microenvironment of osteosarcoma patients in both high and low-risk groups. We utilized the CIBERSORT algorithm to analyze each sample's expression data and determine the abundance of 22 different immune cell types[29,30]. We conducted a single-sample gene set enrichment analysis (ssGSEA) and acquired differences in expression and immune function scores of 22 types of infiltrated cells between high- and low-risk groups.

Cell line culture and gene transfection

We acquired the human osteosarcoma cell lines 143B and U2OS from the China Center for Type Culture Collection (Wuhan, China). RPMI-1640 medium (Invitrogen, USA) supplemented with 10% fetal bovine serum (FBS; Tian Hang, China) and 1% Penicillin–Streptomycin were used to culture the cells, which were incubated at 37 °C with 5% CO2. We procured the ***HOXA1*** expression plasmids from Sangon Biotech Co.,

Ltd (Shanghai, China) and cultured osteosarcoma cells in 6-well and 96-well plates for further experiments. Plasmid transfections were performed following the manufacturer's instructions, utilizing HighGene Plus. We conducted several assays using ***HOXA1***-overexpressing cells, including western blot analysis, CCK-8 assay, transwell invasion assay, and wound healing assay.

Western blot analysis

We used RIPA buffer (Servicebio Technology, Wuhan, China) to extract total protein according to the instructions. Subsequently, the obtained proteins were subjected to 10% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and transferred onto polyvinylidene fluoride (PVDF) membranes. Sealed with skim milk powder for 1 h, then washed with tris-buffered saline with tween (TBST) 3 times, each time for 10 min, we cut the membrane and then sealed it with the diluent of primary antibodies at 4℃ overnight, washed with TBST three times the next day, incubated with secondary antibody for 1 h and washed[31]. Finally, we visualized these bands with the ECL kit (Thermo Fisher Scientific).

CCK-8 assay

Osteosarcoma cells in optimal growth state were seeded onto 96-well plates with three wells in each group. The complete medium containing CCK-8 reagent was replaced at 0, 24, 48, 72, 96, and 120 h after plating, and incubated for 2 h each time. The viability of each cell line was measured by recording the absorbance at OD 450 nm (optical density) using a microplate reader (Bio-Rad Laboratories, Inc.).

Transwell invasion assay

Matrigel (Corning, USA) was mixed with 8 times the volume of medium to create a high-concentration mixture. 70 μL of this mixture was added to the upper layer of each chamber, and 200 μL of medium containing 1×10^5 cells was added after coagulation. Additionally, 600 μL of complete medium (containing 10% FBS) was added to the lower layer of each chamber. After 48 h of cell culture, the cells were fixed with 4% paraformaldehyde, stained with 1% crystal violet, and imaged using an inverted microscope (Olympus, Japan)[32]. The number of cells was then counted.

Wound healing assay

Osteosarcoma cell lines, 143B and U2OS were seeded onto 6-well plates with 2 mL of the complete medium in each well. When the cell density reached 95%, the bottom of each well was marked using a 100μL gun tip, and the culture medium was replaced with a serum-free medium. The intermediate scratch area was observed and imaged after 36 h of culture using an inverted microscope (Olympus, Japan). The wound area was measured using ImageJ software, and the wound healing percentage was calculated using the following formula: Wound healing percentage = (area of 0 h—area of 36 h) / (area of 0 h) × 100.

Animal studies

Experiments were authorized by the Renmin Hospital of Wuhan University Ethics Committee. Male BALB/c nude mice, aged 4–6 weeks, were obtained from Beijing HFK Experiment Animal Center (Beijing, China) and randomly divided into two groups of six mice each (NC, ***HOXA1*** OE). Stably transfected 143B osteosarcoma cells were subcutaneously injected into the mic. Tumor size and volume were monitored weekly, and the tumor volume was calculated as $V = L \times W^2/2$, where L and W are the longest and shortest diameters (mm). After four weeks, all mice were euthanized with 2% pentobarbital sodium (150 mg/kg), and tumors were excised and weighed. We either preserved the tumor in liquid nitrogen or immobilized it in 4% paraformaldehyde. See "Statistical analysis" Section for specific statistical methods. The care of the laboratory animals and animal experiments were performed according to the animal ethics guidelines and approved protocols of Renmin Hospital of Wuhan University.

Immunohistochemistry (IHC) staining

The tumor tissue was fixed with 4% paraformaldehyde for 24 h and sliced into 4-μm sections. Sections were blocked with 1% bovine serum albumin at room temperature for 1 h. Next, primary antibodies were applied and incubated overnight at 4 ℃, followed by a 1-h incubation with secondary antibodies at room temperature. Chromogenic detection was performed using a DAB kit (CST, USA), and the sections were observed using an inverted microscope (Olympus).

Statistical analysis

We use R software (version 4.3.1) and Prism 8.0 for plotting and data processing. The KM survival curve and log-rank test were used for statistical survival analysis. The data were tested for normal distribution. The two-tailed Student's t-test was used to statistically compare the two groups conforming to the normal distribution. One-way analysis of variance (ANOVA) is used for statistical comparison between groups conforming to a normal distribution. The Wilcox nonparametric test is used for data that does not conform to a normal distribution. $p < 0.05$ was considered statistically significant.

Results

Identification and expression analysis of HOX family genes

We got the ***HOX*** gene family members from previous research and the GeneCards database (https://www.genecards.org/). Then we analyzed the expression correlation of all genes in the ***HOX*** family and the hub genes among them according to The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) web-based database (string-interaction.org) (Fig. 1A, B). Then, we visually and quantitatively demonstrated the expression

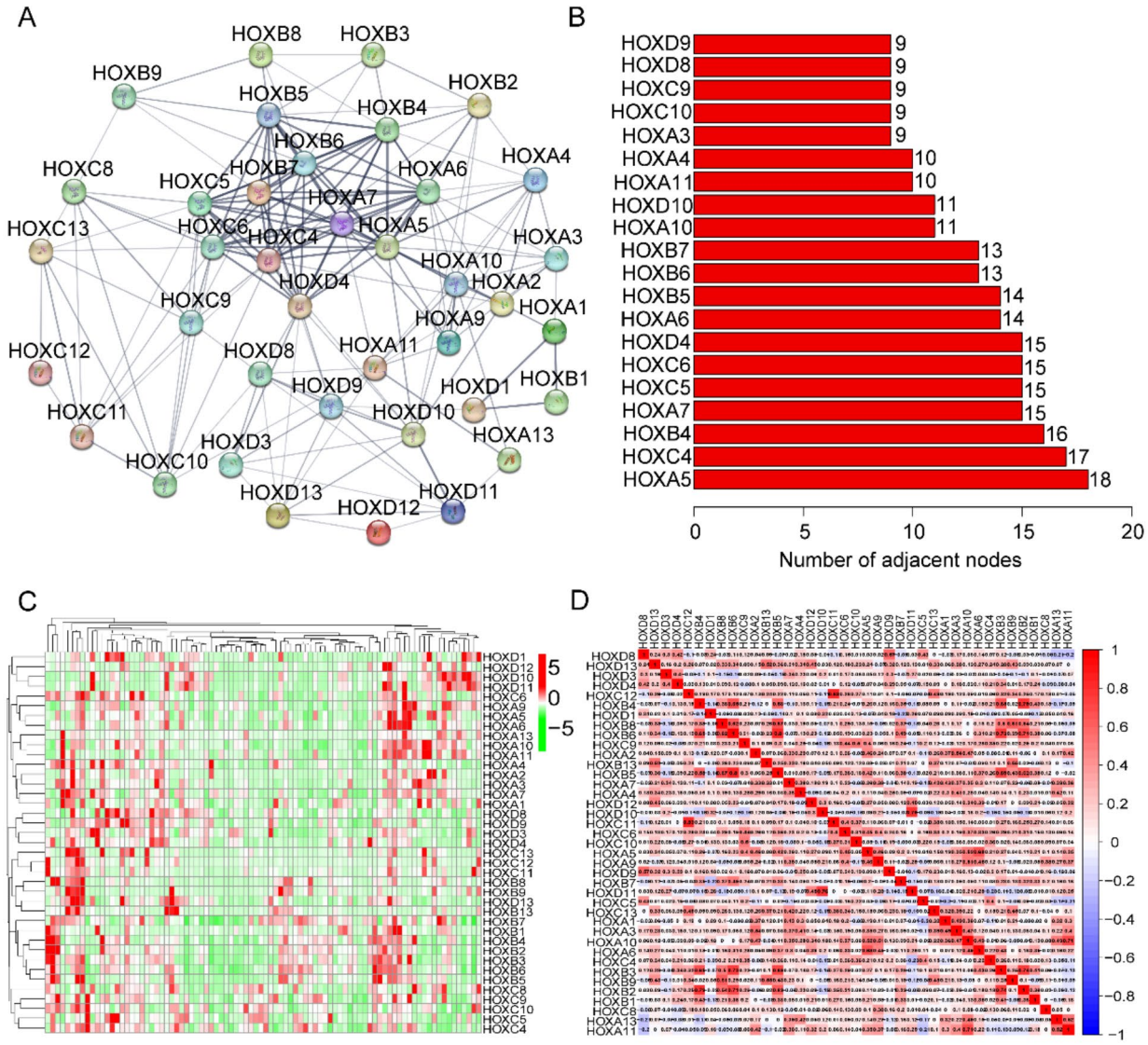

Fig. 1. Homeobox family analysis in osteosarcoma based on TCGA database. (**A**) Protein–protein interaction network of all the ***HOX*** family genes. (**B**) The hub genes in the PPI network. (**C**) Heatmap of expression patterns of all the ***HOX*** family genes in the TCGA dataset. (**D**) Correlation analysis of expression of all the ***HOX*** family genes.

of each gene in each osteosarcoma patient and the expression correlation between each other through heatmaps (Fig. 1C, D).

Construction and validation of risk prognosis model based on HOX family

Osteosarcoma patients in the TCGA dataset were randomly divided into a training cohort and a testing cohort. We then studied osteosarcoma patients in the training cohort and performed the Least Absolute Shrinkage and Selection Operator (LASSO) regression and multivariate Cox regression analysis for ***HOX*** family genes (Fig. 2A, B). According to the minimum partial probability of deviance, hub overall survival-related genes remained[33]. After analysis, we identified four genes in the ***HOX*** family and their corresponding regression coefficients to construct a prognostic risk model (Fig. 2C). We used the "glmnet" package in R to optimize our prognosis model. We calculated the risk score for each osteosarcoma patient in all cohorts using the following formula: Risk score$=\sum_{k=1}^{n}(Coefficient\,(i) * Expr(i))$. We calculated the risk score of osteosarcoma patients in the training cohort and based on the median risk score, we divided patients into high- and low-risk groups. The heatmap showed the expression patterns of the four genes in high-risk and low-risk groups (Fig. 2D). The expression of ***HOXA1***, ***HOXA5,*** and ***HOXA6*** was significantly different between the two groups. The risk score plot shows that as the risk score increases, the overall survival of patients with OS decreases (Fig. 2E, F). The Kaplan–Meier survival analysis showed that the survival of the high-risk group was significantly lower than

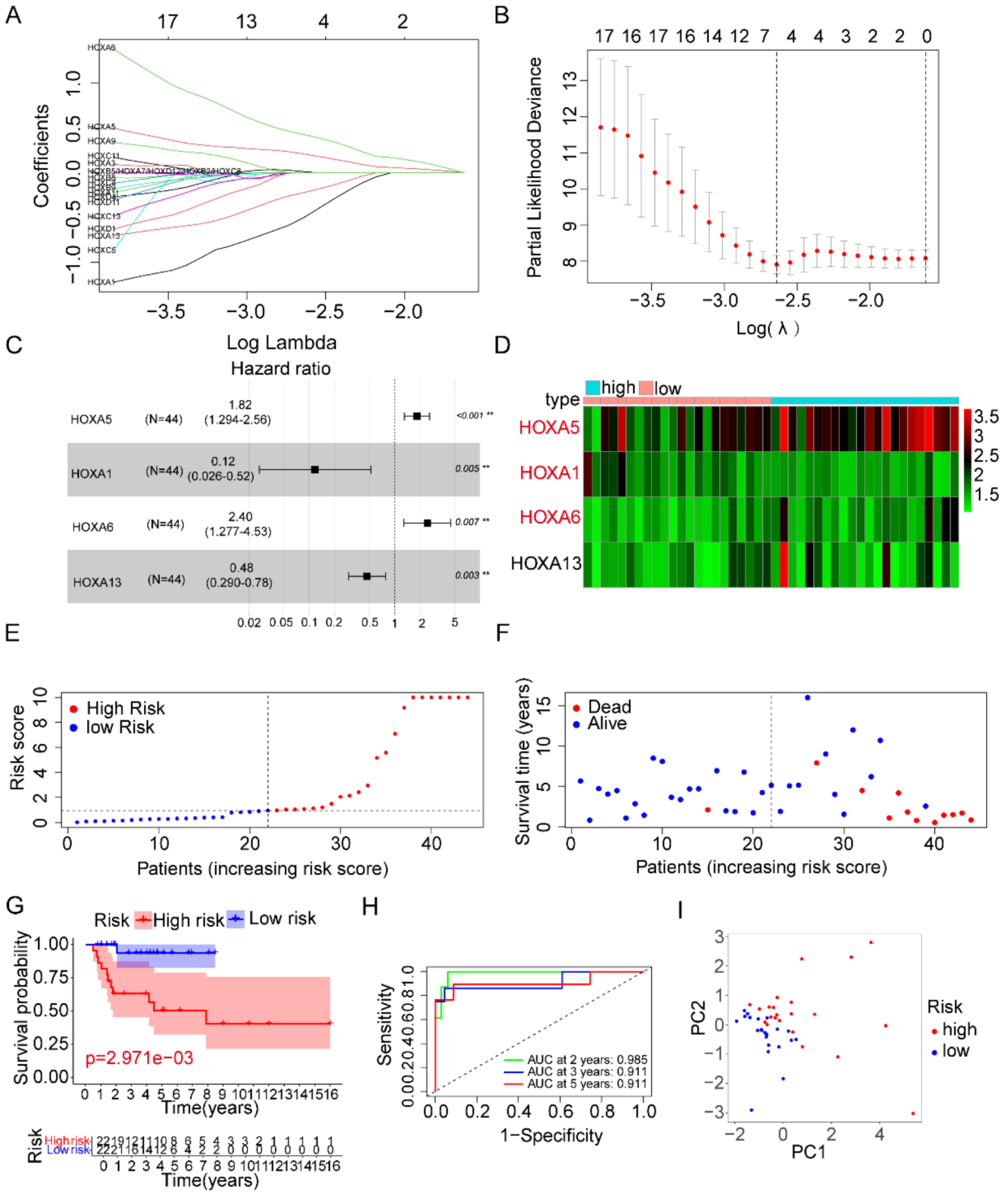

Fig. 2. Establishment of prognostic risk model based on the ***HOX*** family genes. (**A**, **B**) LASSO regression with Multivariate Cox analysis. (**C**) Forest map of four genes (***HOXA5***, ***HOXA1***, ***HOXA6***, ***HOXA13***) in the prognostic model. (**D**) Heatmap of the four genes in the training cohorts. (**E, F**) Distribution of survival time and risk scores of osteosarcoma patients in the training cohort. (**G**) Kaplan–Meier survival analysis of the two risk groups in the training cohort. (**H**) Time-dependent ROC curve analysis of the training cohort. (**I**) PCA based on the confirmed four genes in the training cohort.

that of the low-risk group (Fig. 2G). The ROC curve can well predict the survival time of patients, with a 2-year Area Under Curve (AUC) of 0.985, 3-year AUC of 0.911, and 5-year AUC of 0.911 (Fig. 2H). The PCA revealed different distribution patterns for the two groups of patients (Fig. 2I).

Validation of a risk-prognosis model in the internal cohort and the external cohort

We validated the prognostic model with three cohorts to test its accuracy and repeatability. We see the distribution of patients with the increasing risk score (Fig. 3A–C). Patients with lower risk scores had better outcomes (Fig. 3D–F). We built a heatmap that visually showed the amount of expression of each gene in each cohort (Fig. 3G–I). The results showed that in the three cohorts, the expression of the ***HOXA5*** gene was significantly lower in the low-risk group than in the high-risk group (Supplementary Fig. S1A–C), while the expression of the ***HOXA1*** gene was significantly higher (Supplementary Figure S1 G-I). However, ***HOXA13*** gene expression was elevated only in the low-risk group in the testing cohort (Supplementary Fig. S1D–F), while ***HOXA6*** gene expression was significantly lower in the low-risk group in both the testing cohort and the entire cohort (Supplementary Fig. S1J–L). The Kaplan–Meier survival analysis showed significant differences in overall survival between high- and low-risk groups (Fig. 3J–L), the low-risk group has a longer overall survival. In addition, the ROC curve showed the same results, with a 2-year AUC of 0.686, 3-year AUC of 0.619, and 5-year AUC of 0.737 in the testing cohort (Fig. 3M), and 2-year AUC of 0.807, 3-year AUC of 0.691 and 5-year AUC of 0.758 in the entire cohort (Fig. 3N) and 2-year AUC of 0.877, 3-year AUC of 0.836, and 5-year AUC of 0.850 in the GSE39055 dataset (Fig. 3O). The PCA revealed different distribution patterns for the two groups of patients in the testing cohort, the entire cohort, and the external cohort (Fig. 3P-R). This is consistent with the conclusion we got above.

Nomogram construction

We simulated a nomogram and incorporated the patient's risk score and clinical characteristics such as gender and age to predict overall survival (Fig. 4A). We diagnosed the constructed nomogram using PH assumptions (Fig. 4B). It can be seen from the above results that the P values of the three variables are all greater than 0.05, indicating that each variable meets the PH test, while the P value of the overall test of the model is 0.1006, indicating that the overall model meets the PH test. The graphical DCA indicated the expected net benefit per patient (Fig. 4C). Intervention decisions based on predictive models are beneficial in the 0% to 30% threshold risk range. We then used the TCGA and GSE39055 cohorts to predict the osteosarcoma patient's 1-year, 3-year, and 5-year survival (Fig. 4D–E), to verify that the nomogram we constructed can accurately predict the overall survival of patients in different cohorts.

Differential genes between high- and low-risk groups

We divided patients into high- and low-risk groups in the TCGA entire cohort based on the median risk score and then analyzed and screened for the differentially expressed genes (DEGs) in osteosarcoma patients in the two groups, and the screening standard was $|\log2FC| > 0.5$ and p-value < 0.05 (Fig. 5A). We then analyzed the expression of these significant differential genes in each osteosarcoma patient and drew a heatmap (Fig. 5B). Finally, we studied the expression correlation of these genes and identified the hub genes (Fig. 5C, D).

Functional enrichment analysis

We performed the Gene Ontology (GO) enrichment analysis (Fig. 6A–C) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses (Fig. 6D–F). GO enrichment analysis showed that differential genes mainly acted on extracellular matrix organization and were also related to immune function, such as immunoglobulin complex collagen-containing and antigen binding. KEGG enrichment analysis showed that differential genes were more involved in the extracellular matrix and protein metabolism, such as Protein digestion and absorption and ECM-receptor interaction. We then performed a GSEA enrichment analysis between the two groups (Fig. 7A). We used the standard genetic set " c2.cp.kegg.v2023.1.Hs.symbols.gmt " to analyze. The two groups differ in immune function pathways, such as complement and coagulation cascades (Fig. 7B), cytokine receptor interaction (Fig. 7C), nicotinate and nicotinamide metabolism (Fig. 7D), JAK-STAT signaling pathway (Fig. 7E). This suggests that the prognostic model we constructed plays an important role in the immune function of osteosarcoma.

Difference analysis of immune infiltration

To assess the prognostic risk score's effectiveness in the tumor microenvironment (TME), we evaluated estimate scores (Fig. 8A), immunological scores (Fig. 8B), stromal scores (Fig. 8C), and tumor purity (Fig. 8D) in the immune microenvironment of osteosarcoma patients in both high and low-risk groups. Moreover, estimate scores, stromal scores, and tumor purity were significantly different between the two groups, suggesting that the prognostic model we constructed has a significant impact on the immune microenvironment of osteosarcoma. The relationship between immunity, estimation, stromal score, tumor purity, and survival in patients with osteosarcoma has previously been investigated[34]. We conducted a single-sample gene set enrichment analysis (ssGSEA) and acquired differences in expression (Fig. 9A, B) and immune function scores (Fig. 9C, D) of 22 types of infiltrated cells between high- and low-risk groups in the TCGA and GSE39055 cohorts. In the TCGA dataset, there were many differences in immune cell content between the two risk groups, such as Th1 cells, Th2 cells, and Til. Both risk groups are also enriched in many immune-related functions such as CCR, Check-point, Cytolytic activity, Inflammation-promoting, and T-cell co-inhibition. This suggests that our model is associated with immune dysfunction in osteosarcoma.

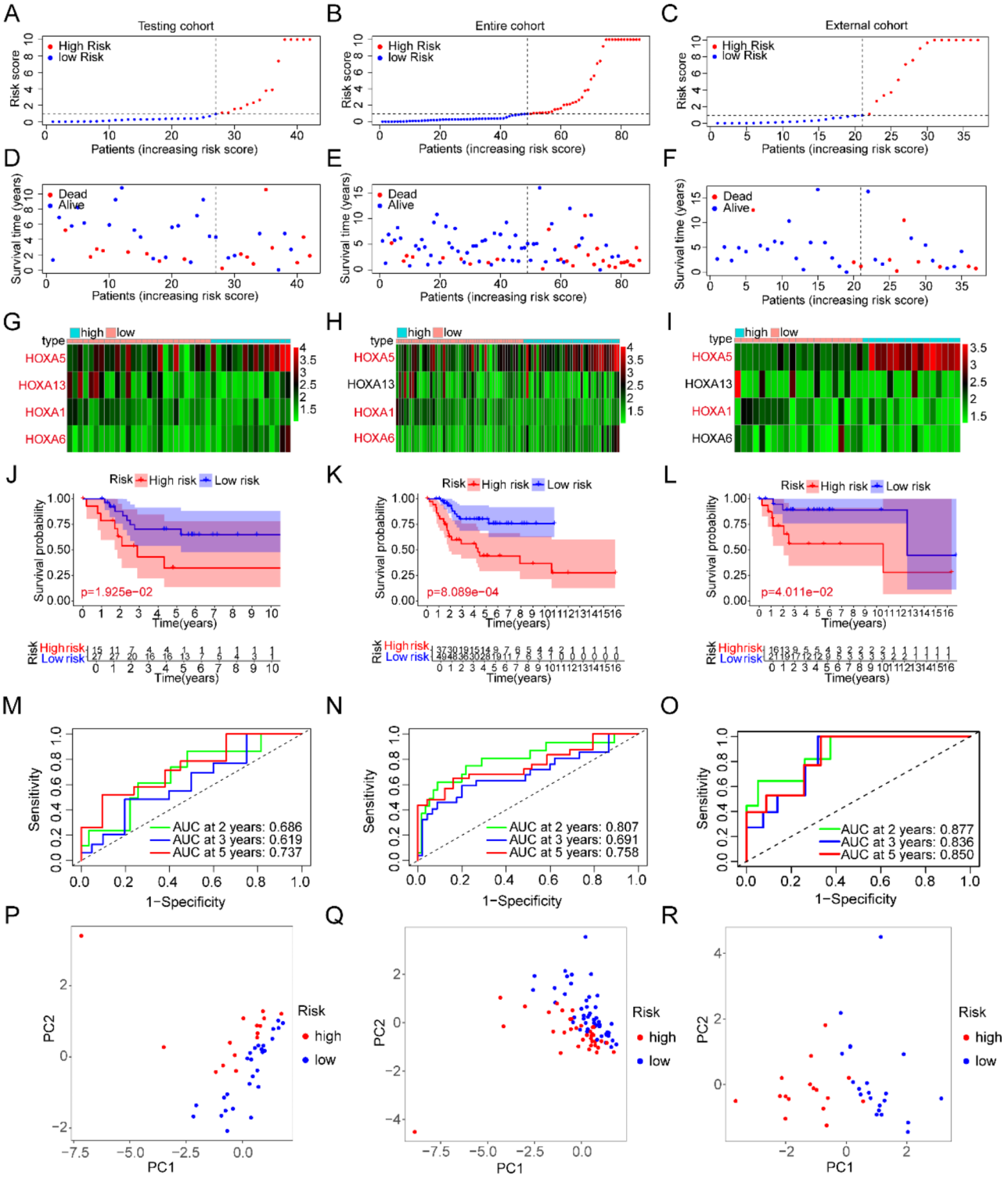

Fig. 3. Validation of the prognostic model in the testing cohort, the entire cohort, and the GSE39055 cohort. (**A**–**C**) Distribution of risk scores. (**D**–**F**) Distribution of survival time and status. (**G**–**I**) Heatmap of the four genes. (**J**–**L**) Kaplan–Meier survival analysis of the two risk groups. (**M**–**O**) Time-dependent ROC curve analysis. (**P**–**R**) PCA based on the four genes.

Overexpression of HOXA1 inhibits the proliferation, migration, and invasion of osteosarcoma cell lines

To prove that the prognostic model we constructed was more accurate, we selected ***HOXA1*** used to construct the prognostic model for subsequent experiments, and conducted a western blot experiment, CCK-8 experiment,

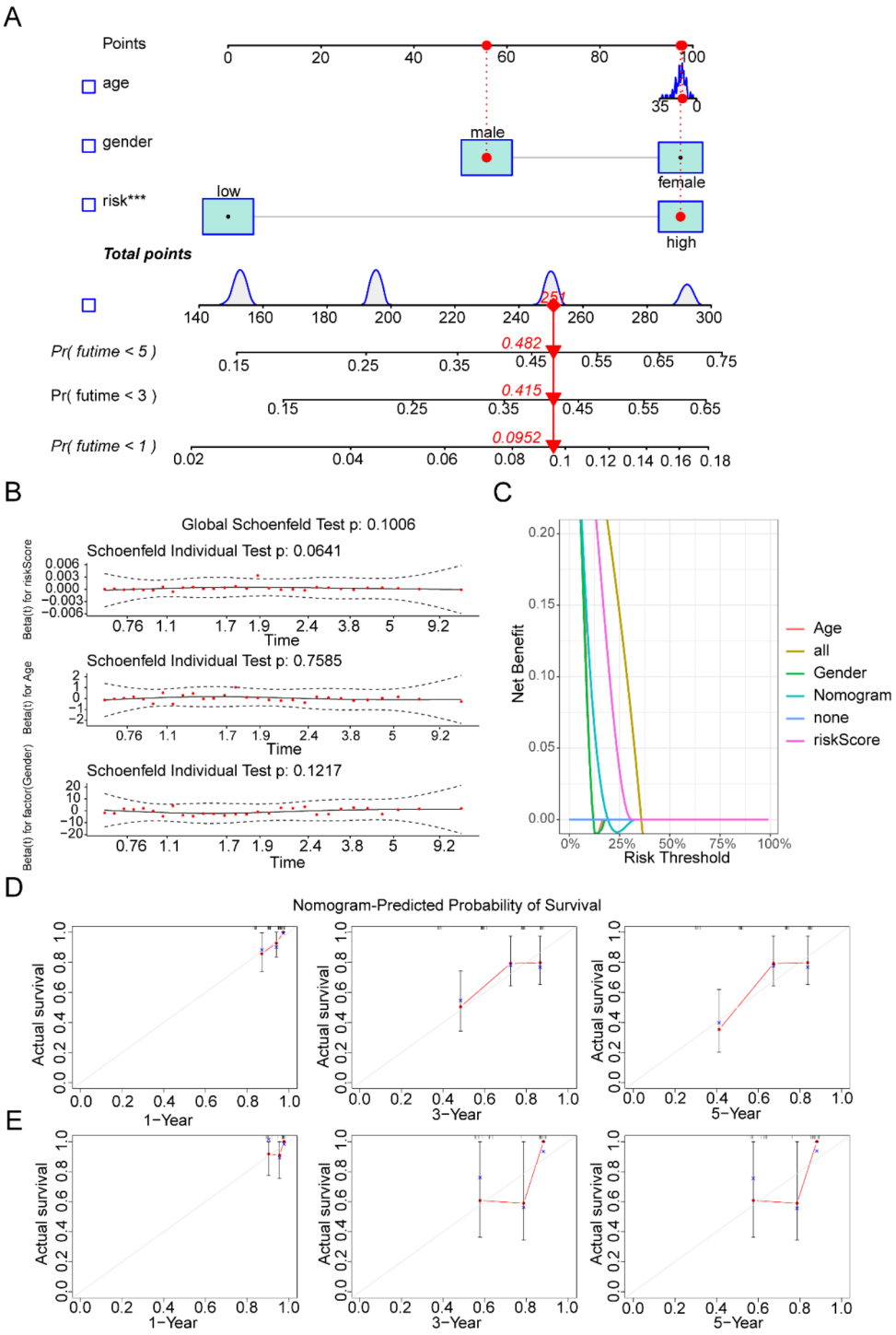

Fig. 4. A nomogram was drawn to predict the overall survival of osteosarcoma patients in TCGA and GSE39055 cohorts. (**A**) Nomogram to evaluate overall survival in osteosarcoma patients. (**B**) The PH assumption of the nomogram. (**C**) The Decision Curve Analysis. (**D**, **E**) The calibration curves for predicting 1-, 3-, and 5-year overall survival in GSE39055 and TCGA entire cohort. The X-axis represents expected survival and the Y-axis represents actual survival. The nomogram can predict overall survival accurately and reliably. "***"Represented "$p < 0.001$".

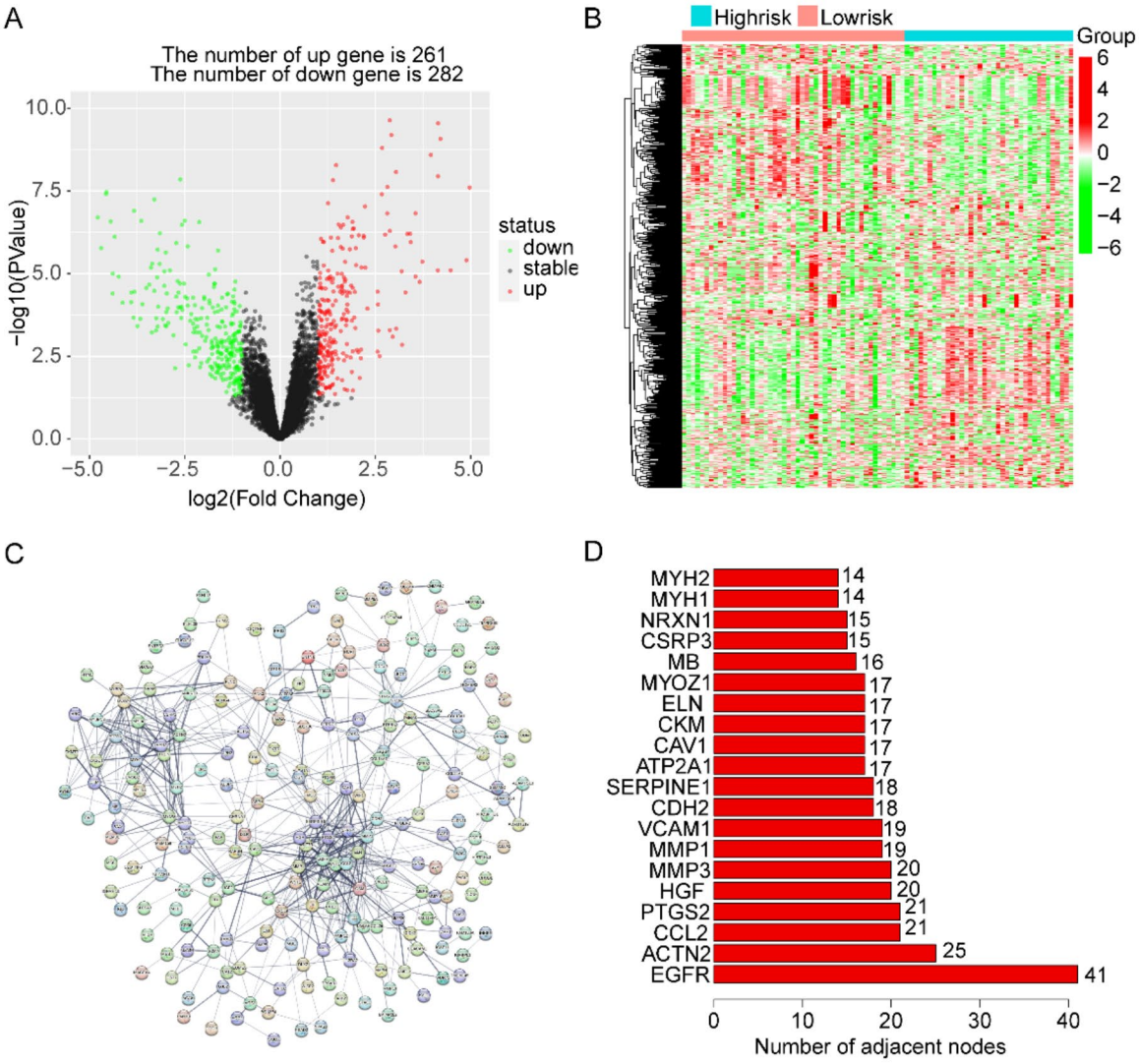

Fig. 5. Analysis of gene difference and PPI network. (**A**) The volcano plot shows the differentially expressed genes between high- and low-risk groups. (**B**) Heatmap of differential gene expression between the two groups. (**C**) PPI network of differential genes in the two groups. (**D**) The hub genes of the PPI network.

wound healing experiment, and transwell experiment. Western blot results showed low ***HOXA1*** expression in osteosarcoma cell lines compared with osteoblasts, consistent with our results above, and we successfully overexpressed ***HOXA1*** in osteosarcoma cells (143B and U2OS) (Fig. 10A–D). Subsequently, we found that when osteosarcoma cells overexpress ***HOXA1***, cell activity is inhibited (Fig. 10E, F). The results of the wound healing experiment were consistent with the above results, and the migration ability of osteosarcoma cell lines overexpressing ***HOXA1*** was much lower than that of the NC group (Fig. 10G, H). In addition, the results of the transwell experiment also confirmed our analysis that the invasion ability of ***HOXA1*** overexpressed osteosarcoma cell lines was greatly inhibited (Fig. 10I, J). These experimental results show that the proliferation, migration, and invasion ability of osteosarcoma cell lines with ***HOXA1*** overexpression are reduced, which can provide help for the diagnosis and evaluation of osteosarcoma patients, and also provide ideas for the treatment of osteosarcoma patients.

Increased expression of HOXA1 inhibits tumor growth in vivo

We injected osteosarcoma cells 143B and NC cells with stable overexpression of ***HOXA1*** into the skin of mice to construct a subcutaneous tumor model. The western blot results show that we have built successfully ($p < 0.05$) (Fig. 11D, E). We measured the size and weight of the subcutaneous tumors in the mice every 7 days, and the results showed that the tumor volume ($p < 0.0001$) and weight ($p < 0.001$) in the control group was significantly larger than that in the ***HOXA1*** overexpressed group (Fig. 11A–C), so overexpression of ***HOXA1*** can inhibit the

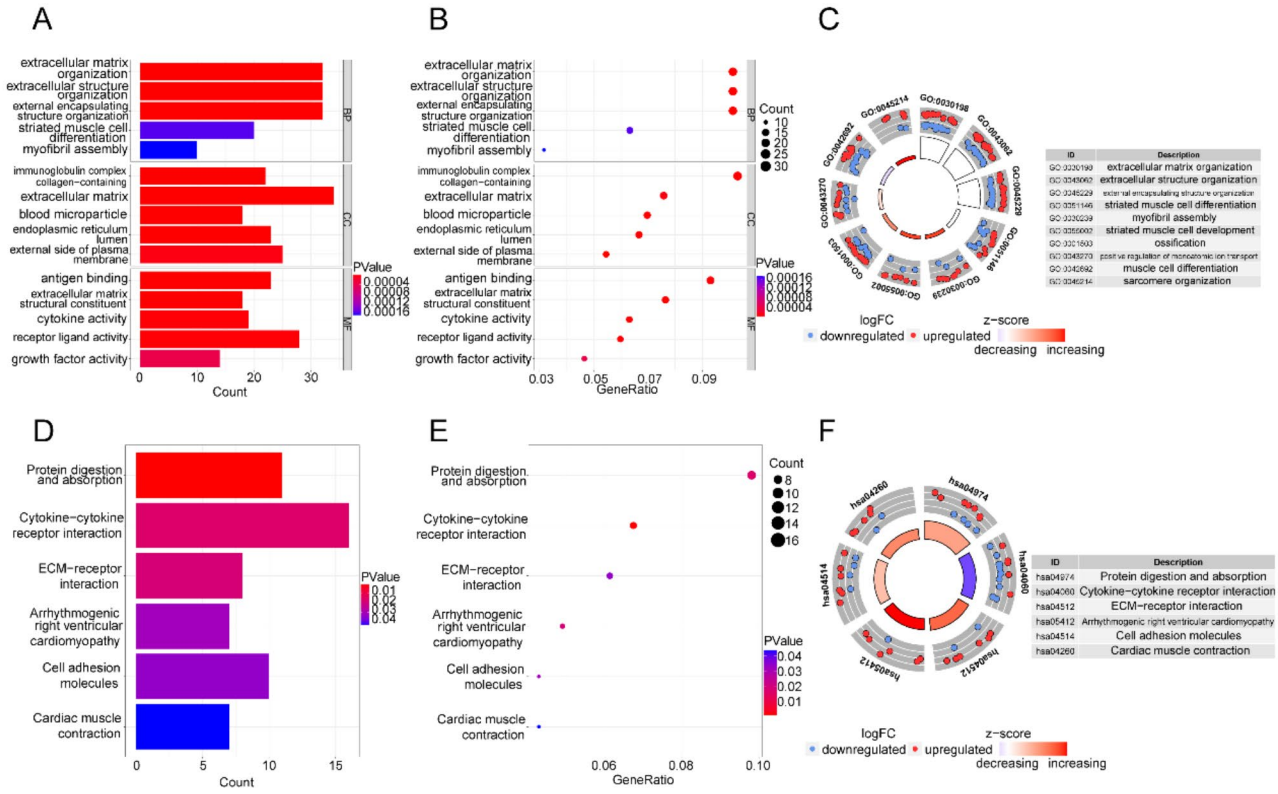

Fig. 6. GO and KEGG functional analysis. (**A**–**C**) The GO enrichment analysis of the differential genes between the two groups. (**D**–**F**) The KEGG enrichment analysis of the differential genes between the two groups. There are a variety of immune-related functional pathways enriched, such as immunoglobulin receptor binding, chemokine activity, and Cytokine-cytokine receptor interaction, etc.

growth of osteosarcoma in vivo. In addition, immunohistochemical results showed that ***HOXA1*** expression was increased in the HOXA1 OE group ($p < 0.001$), while Ki-67 level was significantly decreased ($p < 0.001$) (Fig. 11F, G), suggesting that overexpression of ***HOXA1*** may inhibit the growth of osteosarcoma. In summary, our results in vivo that ***HOXA1*** overexpression in osteosarcoma can inhibit tumor growth are consistent with the above.

Discussion

Osteosarcoma is a rare form of bone cancer that primarily affects children and adolescents[1,35,36]. With a reexamination rate of less than 10%, accurate diagnosis, and effective treatment become challenging. Due to the development of medical technology, the prognosis of patients with osteosarcoma has improved significantly, with the five-year survival rate for local cases exceeding 60–70%[37,38]. However, for patients with metastatic or recurrent disease, the prognosis is still not well[34]. With the development of high-throughput sequencing technology, we can improve the early diagnosis rate, improve treatment strategies, and improve the prognosis of osteosarcoma patients by constructing prognostic models.

Studies have found that in the process of tumor development, the ***HOX*** gene can play the role of transcriptional activator or suppressor[39–42]. For example, overexpression of ***HOXB5*** can activate ***FGFR4*** and ***CXCL1*** in hepatocellular carcinoma, thus accelerating the metastasis of HCC and affecting the prognosis of patients[43]. ***HOXB7*** may increase the expression of ***p-PI3K*** and ***p-Akt***, thus promoting the proliferation of osteosarcoma[44]. Therefore, ***HOX*** protein may be a diagnostic marker and therapeutic target for tumors. However, the clinical significance of ***HOX*** family genes in patients with osteosarcoma remains to be further explored.

Since the ***HOX*** gene family can affect the occurrence and development of osteosarcoma, we systematically analyzed the ***HOX*** gene family using the information of osteosarcoma patients in an online database to construct a risk model that can accurately predict the prognosis of patients. We conducted LASSO regression analysis and multivariate Cox analysis for ***HOX*** family genes and selected four genes to construct a risk prognosis model. Validation is then performed on internal and external cohorts. Patients in all cohorts were divided into high- and low-risk groups, and both prognostic risk score maps and Kaplan–Meier survival curve showed that the survival time of high-risk patients was significantly lower than that of low-risk patients, indicating that our risk model had good universality and accuracy in predicting patient survival. Moreover, the ROC curve also proved the accuracy and universality of the model used to predict the prognosis of patients. In summary, we

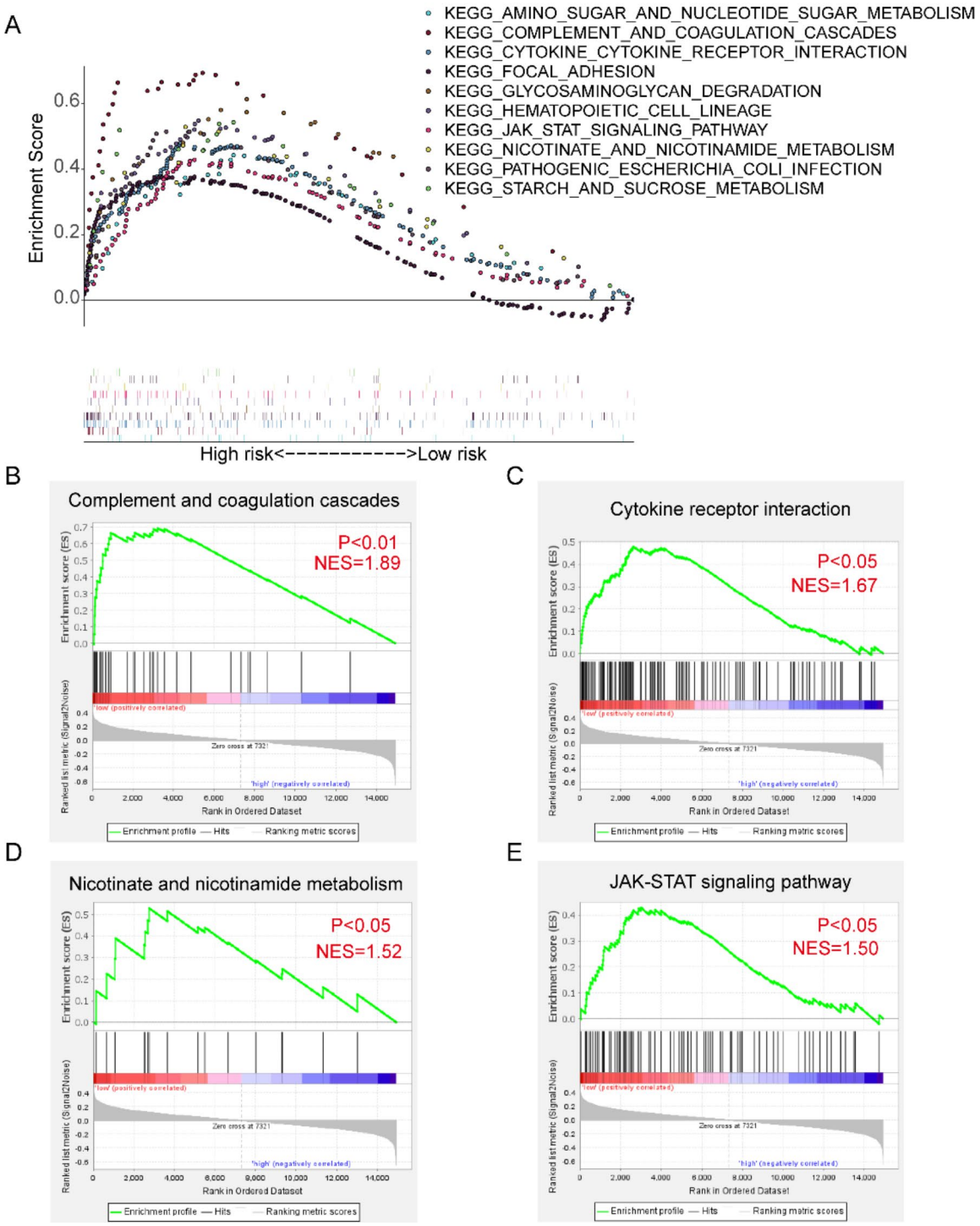

Fig. 7. GSEA analysis between high- and low- osteosarcoma groups. (**A**) The top 10 changed pathways were screened according to KEGG functional enrichment analysis. (**B**) Complement and coagulation cascades. (**C**) Cytokine receptor interaction. (**D**) Nicotinate and nicotinamide metabolism. (**E**) JAK-STAT signaling pathway.

constructed a risk-prognostic model that can well predict the survival time of patients and may be used as a biomarker for clinical decision-making. We then combined the patient's risk score and clinical characteristics such as gender and age to create a nomogram and this nomogram can predict the prognosis of the patient. Then we identified differential genes in the high- and low-risk groups and performed functional enrichment analyses. Functional enrichment analysis showed that the four genes were mainly involved in immune-related

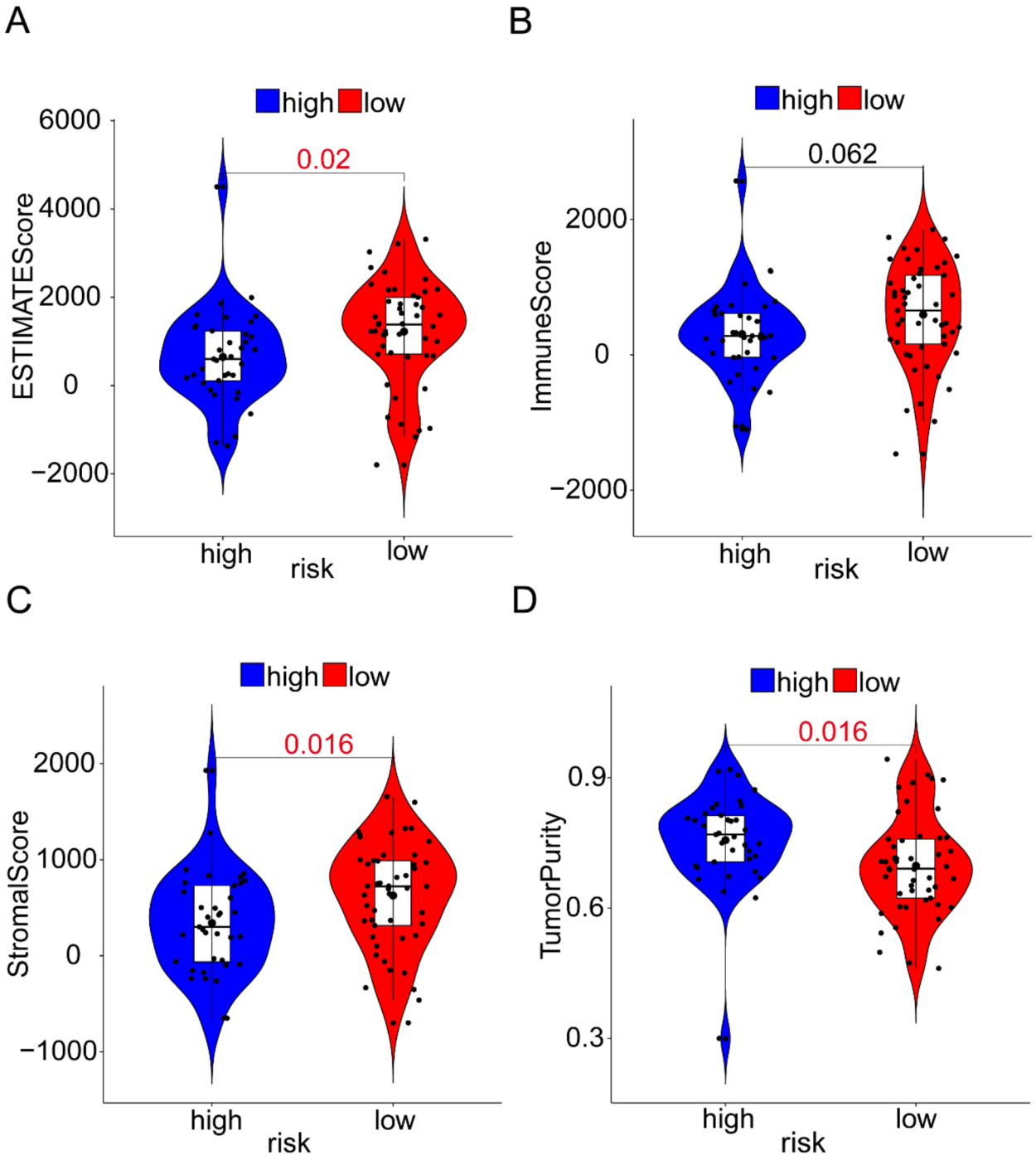

Fig. 8. Analysis of immune-related scores for the two groups in the TCGA cohort. (**A**) The estimated score. (**B**) The immune score. (**C**) The stromal score. (**D**) The tumor purity.

pathways and functions. A subsequent GSEA analysis showed that functional differences between high- and low-risk groups were mainly concentrated in immune-related signaling pathways. We used the ESTIMATE algorithm and ssGSEA to verify the relationship between the risk-prognosis model and the osteosarcoma immune microenvironment. The results showed that patients in the high-risk group had lower ESTIMATE scores, immune scores, stromal scores, and higher tumor purity. These results indicate that our risk prognostic model is closely related to the immune microenvironment of osteosarcoma patients, and can be used for clinical decision-making and treatment evaluation.

We constructed a risk prognosis model for osteosarcoma patients, in which ***HOXA1*** expression was significantly higher in the low-risk group. Conversely, ***HOXA5*** and ***HOXA6*** expression in the high-risk group

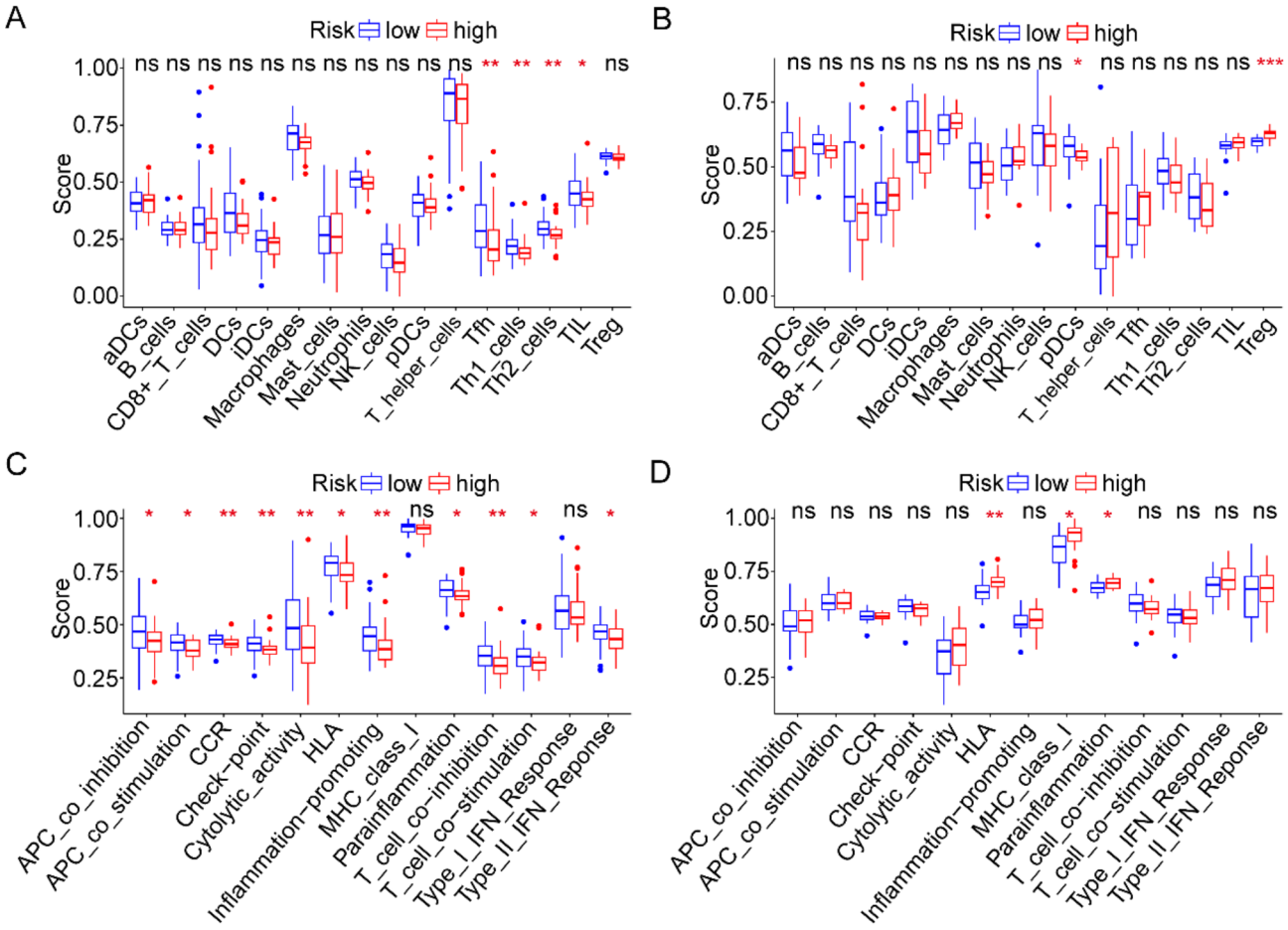

Fig. 9. The four ***HOX*** gene signatures were correlated with immune cells and immune function based on ssGSEA. (**A**, **B**) Box charts showed the scores of various immunocytes in the two groups in the TCGA and GSE39055 cohort. (**C**, **D**) Box charts showed scores of the related immune function between the two groups in TCGA and GSE39055 cohorts. "*" represented "$p < 0.05$", "**"Represented "$p < 0.01$", "***"Represented "$p < 0.001$".

was notably higher. ***HOXA1*** is a key regulator of intestinal formation during embryogenesis[45]. In bladder cancer (Bca), overexpression of ***HOXA1*** can promote the transcription of ***SMAD3*** and thus promote the proliferation and lymphatic metastasis of cancer cells[46]. However, these findings are inconsistent with those obtained in osteosarcoma, where ***HOXA1*** expression is higher in the low-risk group. So, the specific role and mechanism of ***HOXA1*** in osteosarcoma need to be verified in subsequent experiments. ***HOXA5*** can promote the development of human bones and organ formation[47], and the most common functional abnormality controlling ***HOXA5*** expression is CpG island hypermethylation[48]. In osteosarcoma, ***HOXA5*** may promote apoptosis of osteosarcoma by regulating ***p53*** and ***p38α MAPK*** pathways[49,50]. The results showed that although ***HOXA5*** expression was increased in high-risk patients, it acted as a tumor suppressor gene. The expression of ***HOXA6*** in the high-risk group was higher than that in the low-risk group, suggesting that ***HOXA6*** acts as an oncogene in osteosarcoma. In gastric cancer and colorectal cancer (CRC), ***HOXA6*** can inhibit the apoptosis of tumor cells by binding to other genes or acting on other pathways to promote the occurrence and progression of tumors[51,52]. However, no studies have reported the specific role of ***HOXA6*** in osteosarcoma, so we will explore the role of ***HOXA6*** in the progression of osteosarcoma in the future. In the prognostic model we constructed, there was no significant difference in the expression of ***HOXA13*** between the high- and low-risk groups. However, studies have shown that the expression of ***HOXA13*** in osteosarcoma is significantly higher than that in adjacent tissues, and overexpression of ***HOXA13*** can promote the malignancy of osteosarcoma cells[53]. In summary, ***HOX*** family genes may regulate the progression of osteosarcoma through their unique mechanisms, which provides a new direction for the clinically targeted therapy of osteosarcoma patients.

We enhanced our prognostic model's accuracy by validating the role of ***HOXA1*** through in vitro and in vivo experiments. We discovered that ***HOXA1*** overexpression inhibited osteosarcoma cell proliferation, migration, and invasion. In vivo experiments demonstrated that ***HOXA1*** overexpression suppressed tumor growth in mice, and immunohistochemical analysis revealed a decrease in ***Ki-67*** level in the ***HOXA1*** overexpression group, indicating that it may inhibit the proliferation of tumor cells. When analyzing the experimental data, we used the two-sample t test to compare the results of the experimental group and the control group. $P < 0.05$ indicated

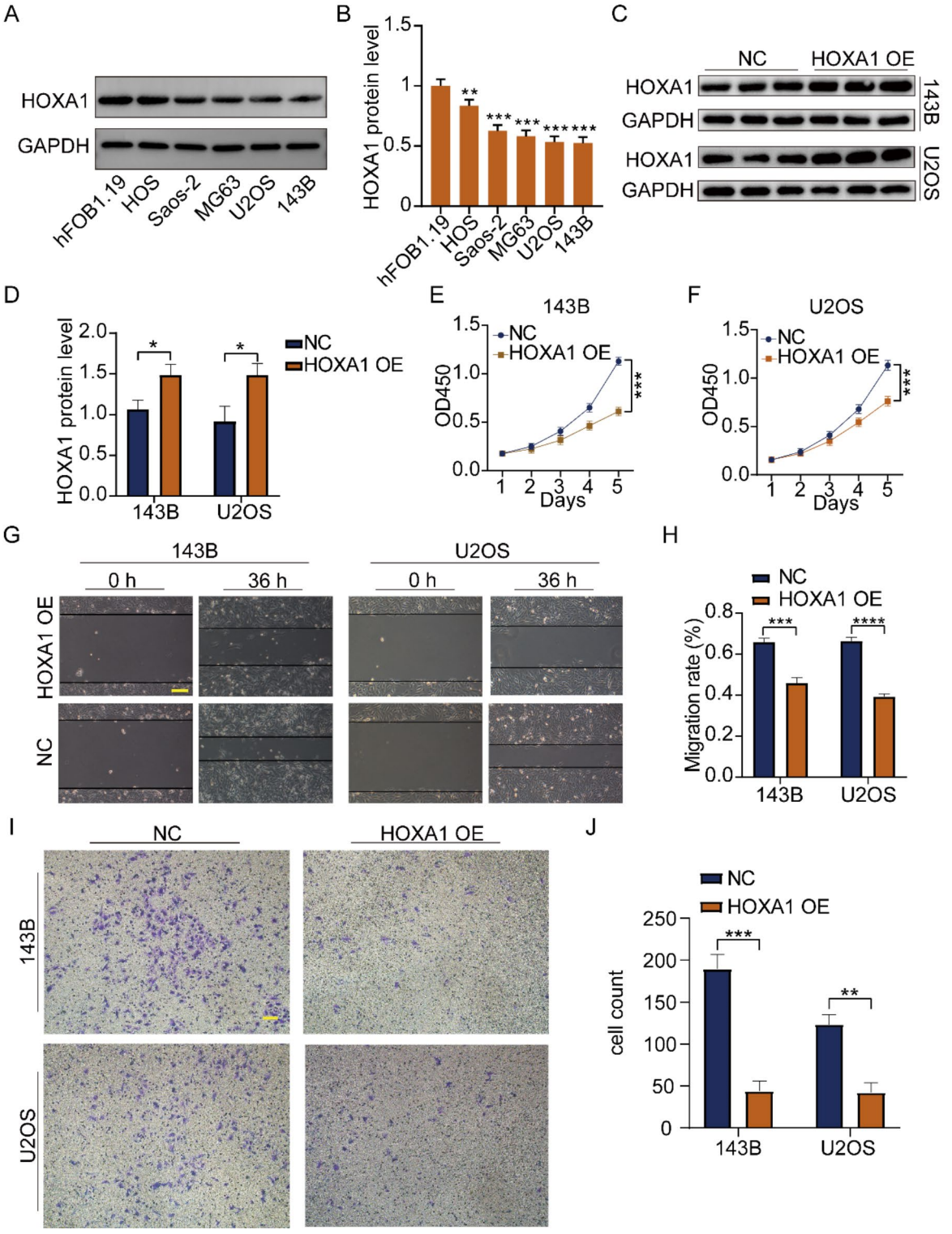

Fig. 10. Overexpress ***HOXA1*** inhibits the activity of OS cell lines. (**A**, **B**) The protein level of ***HOXA1*** in human osteoblastic cell and osteosarcoma cell lines. (**C**, **D**) The protein content of ***HOXA1***. (**E**, **F**) CCK-8 assay was applied to measure the viability of osteosarcoma cells. (**G**) Photos of 0 h and 36 h wound healing assay. (**H**) Statistical analysis of wound healing assay. (**I**, **J**) The transwell assay was applied to measure the invasion ability of osteosarcoma cells. All data are from three independent experiments and are shown as mean ± SD. "**" represented "$p < 0.01$", "***"Represented "$p < 0.001$", "****"Represented "$p < 0.0001$".

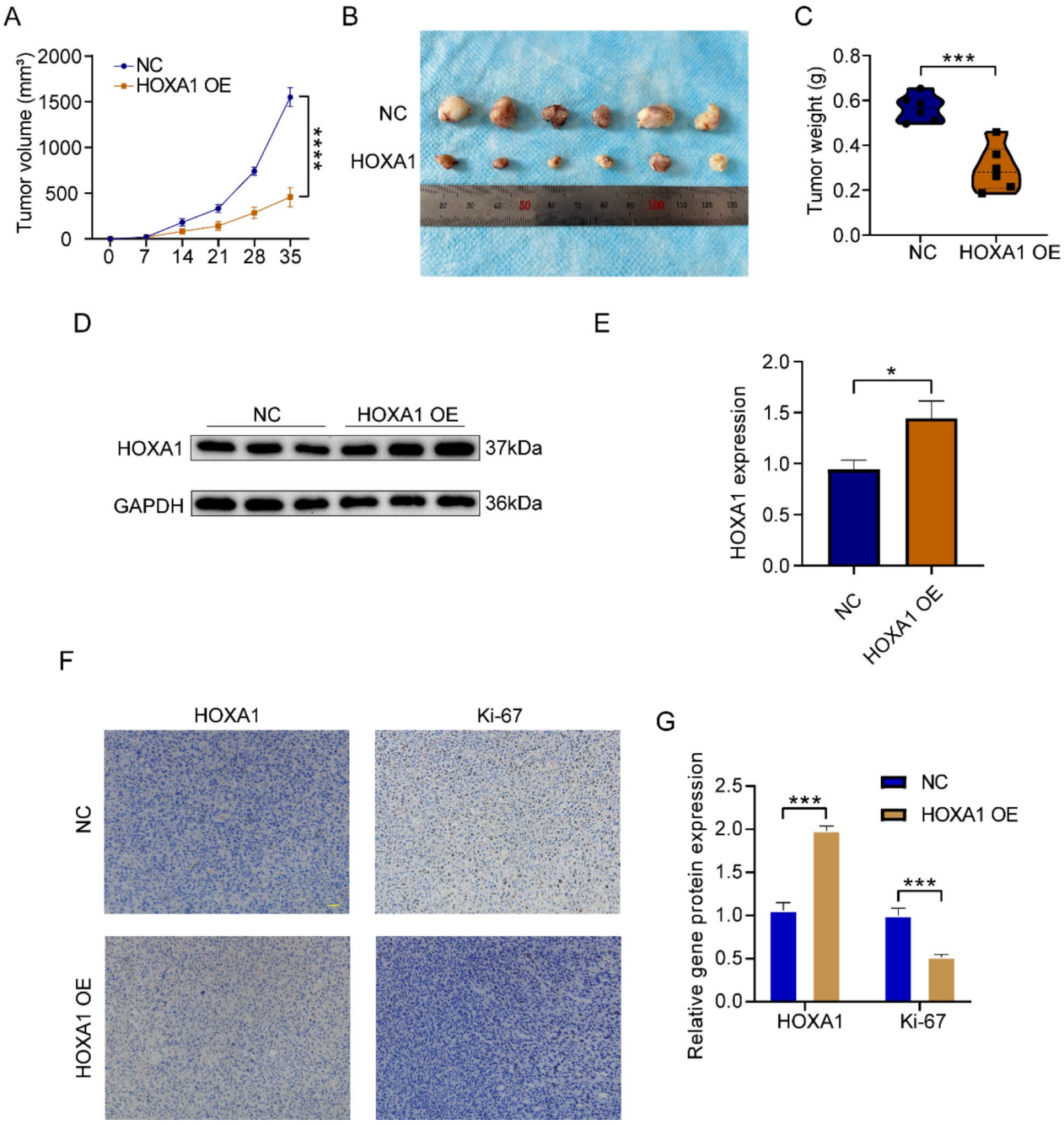

Fig. 11. Increased expression of ***HOXA1*** inhibits tumor growth in vivo. (**A**) Tumor growth was assessed by monitoring its volume in vivo. (**B**) Subcutaneous tumor formation was performed with NC/***HOXA1*** OE osteosarcoma cells, and the size of the transplanted tumor was observed in nude mice. (**C**) The weight of tumors. (**D**, **E**) Western blot of subcutaneous tumor. (**F**, **G**) The subcutaneous tumor tissue was detected by immunohistochemistry. Scale bar: 200 μm. All data are from three independent experiments and are shown as mean ± SD. "**"Represented "$p < 0.01$", "***"Represented "$p < 0.001$", "****"Represented "$p < 0.0001$".

that the two groups were statistically significant. We noticed a certain degree of variability in the experimental results. This variability may be due in part to technical differences in experimental operations as well as inherent variability in the biological samples themselves. To control for variability, we plan to increase the sample size to improve the statistical power of the experiment and optimize the experimental design to exclude factors that may introduce variability. We will also delve into other potential factors in the course of the experiment to more fully explain the statistical significance and variability of the results. In conclusion, ***HOXA1*** likely plays a positive role in inhibiting osteosarcoma growth and progression.

Of course, there are inevitably some shortcomings in our research. First of all, we only collected two datasets. Although we randomly divided the TCGA dataset into two for research, the sample size was relatively small,

which led to the possibility that some genes could be significantly different in the high-low-risk group, this difference could not be detected, which also made it difficult to generalize our results to the whole population, limiting the external validity of the study. Second, the data we collected on osteosarcoma patients lacked clinical information such as tumor stage and grade, which may make it difficult to popularize the research results to actual clinical practice, and the relationship between the model and clinical information cannot be verified, which may lead to the inaccuracy of the analysis results and make the conclusion unconvincing. In addition, we selected a specific ***HOX*** gene family to construct the prognostic model, which may be related to a specific subpopulation of patients, so that the prediction results of the model cannot be generalized to the entire tumor patient population. In addition, overfitting training data of the model will reduce the generalization ability of the model, thus affecting the prediction accuracy of the model. Finally, we were unable to collect samples from osteosarcoma patients for verification, which means that the experimental results may be biased and do not fully reflect the real clinical situation. Therefore, the specific role of ***HOXA6*** in osteosarcoma remains to be further explored.

In summary, we created a risk-prognostic model based on four ***HOX*** family genes that accurately predicted survival time and outcomes for patients with osteosarcoma. Moreover, this model is closely related to the proportion of infiltration and immune function of immune cells in the immune microenvironment of osteosarcoma.

Data availability

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 10 April 2024; Accepted: 30 December 2024
Published online: 06 January 2025

References

1. Dong, Z. et al. Advances in the biological functions and mechanisms of miRNAs in the development of osteosarcoma. *Technol. Cancer Res. Treat.* **21**, 15330338221117386 (2022).
2. Lindsey, B. A., Markel, J. E. & Kleinerman, E. S. Osteosarcoma overview. *Rheumatol. Ther.* **4**(1), 25–43 (2017).
3. Beird, H. C. et al. Osteosarcoma. *Nat. Rev. Dis. Primers* **8**(1), 77 (2022).
4. Meltzer, P. S. & Helman, L. J. New horizons in the treatment of osteosarcoma. *N. Engl. J. Med.* **385**(22), 2066–2076 (2021).
5. Tian, H., Wu, R., Feng, N., Zhang, J. & Zuo, J. Recent advances in hydrogels-based osteosarcoma therapy. *Front. Bioeng. Biotechnol.* **10**, 1042625 (2022).
6. Xie, D. et al. Targeted delivery of chemotherapeutic agents for osteosarcoma treatment. *Front. Oncol.* **12**, 843345 (2022).
7. Smrke, A., Anderson, P. M., Gulia, A., Gennatas, S., Huang, P. H. & Jones, R. L. Future directions in the treatment of osteosarcoma. *Cells* 2021, **10**(1).
8. Jiang, Z. Y., Liu, J. B., Wang, X. F., Ma, Y. S. & Fu, D. Current Status and prospects of clinical treatment of osteosarcoma. *Technol. Cancer Res. Treat.* **21**, 15330338221124696 (2022).
9. Cersosimo, F., Lonardi ,S., Bernardini, G., Telfer, B., Mandelli, G. E., Santucci, A., Vermi, W. & Giurisato, E. Tumor-associated macrophages in osteosarcoma: From mechanisms to therapy. *Int. J. Mol. Sci.* **21**(15) (2020).
10. Shaikh, A. B. et al. Present advances and future perspectives of molecular targeted therapy for osteosarcoma. *Int. J. Mol. Sci.* **17**(4), 506 (2016).
11. Tarhini, A. A. & Kirkwood, J. M. How much of a good thing? What duration for interferon alfa-2b adjuvant therapy?. *J. Clin. Oncol.* **30**(31), 3773–3776 (2012).
12. Kempf-Bielack, B. et al. Osteosarcoma relapse after combined modality therapy: An analysis of unselected patients in the Cooperative Osteosarcoma Study Group (COSS). *J. Clin. Oncol.* **23**(3), 559–568 (2005).
13. 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).
14. Mallo, M. Reassessing the role of hox genes during vertebrate development and evolution. *Trends Genet.* **34**(3), 209–217 (2018).
15. Cillo, C., Cantile, M., Faiella, A. & Boncinelli, E. Homeobox genes in normal and malignant cells. *J. Cell. Physiol.* **188**(2), 161–169 (2001).
16. Soubeyran, P. et al. Homeobox gene Cdx1 regulates Ras, Rho and PI3 kinase pathways leading to transformation and tumorigenesis of intestinal epithelial cells. *Oncogene* **20**(31), 4180–4187 (2001).
17. Li, P. D. et al. HOXC6 predicts invasion and poor survival in hepatocellular carcinoma by driving epithelial-mesenchymal transition. *Aging (Albany NY)* **10**(1), 115–130 (2018).
18. Tan, Z. et al. Overexpression of HOXC10 promotes angiogenesis in human glioma via interaction with PRMT5 and upregulation of VEGFA expression. *Theranostics* **8**(18), 5143–5158 (2018).
19. Feng, Y. et al. Homeobox genes in cancers: From carcinogenesis to recent therapeutic intervention. *Front. Oncol.* **11**, 770428 (2021).
20. Bach, C. et al. Leukemogenic transformation by HOXA cluster genes. *Blood* **115**(14), 2910–2918 (2010).
21. Kwon, S. M. et al. Recurrent glioblastomas reveal molecular subtypes associated with mechanistic implications of drug-resistance. *PLoS ONE* **10**(10), e0140528 (2015).
22. Huang, H. et al. Hsa_circ_0007031 promotes the proliferation and migration of osteosarcoma cells by sponging miR-196a-5p to regulate the HOXB6. *Biochem. Pharmacol.* **214**, 115667 (2023).
23. Ma, Q. et al. HOXB5 promotes the progression and metastasis of osteosarcoma cells by activating the JAK2/STAT3 signalling pathway. *Heliyon* **10**(9), e30445 (2024).
24. Guo, J., Zhang, T. & Dou, D. Knockdown of HOXB8 inhibits tumor growth and metastasis by the inactivation of Wnt/beta-catenin signaling pathway in osteosarcoma. *Eur. J. Pharmacol.* **854**, 22–27 (2019).
25. Zhang, N., Meng, X., Mei, L., Zhao, C. & Chen, W. LncRNA DLX6-AS1 promotes tumor proliferation and metastasis in osteosarcoma through modulating miR-641/HOXA9 signaling pathway. *J. Cell. Biochem.* **120**(7), 11478–11489 (2019).
26. Bodey, B., Bodey, B. Jr., Siegel, S. E., Luck, J. V. & Kaiser, H. E. Homeobox B3, B4, and C6 gene product expression in osteosarcomas as detected by immunocytochemistry. *Anticancer Res.* **20**(4), 2717–2721 (2000).
27. Luo, Z. et al. Cytokine-induced apoptosis inhibitor 1: A comprehensive analysis of potential diagnostic, prognosis, and immune biomarkers in invasive breast cancer. *Transl. Cancer Res.* **12**(7), 1765–1786 (2023).
28. Wang, Y. et al. Tubulin alpha-1b chain was identified as a prognosis and immune biomarker in pan-cancer combing with experimental validation in breast cancer. *Sci. Rep.* **14**(1), 8201 (2024).

29. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**(5), 453–457 (2015).
30. Pan, Q., Wang, L., Chai, S., Zhang, H. & Li, B. The immune infiltration in clear cell renal cell carcinoma and their clinical implications: A study based on TCGA and GEO databases. *J. Cancer* **11**(11), 3207–3215 (2020).
31. Szebenyi, K. et al. Effective targeting of breast cancer by the inhibition of P-glycoprotein mediated removal of toxic lipid peroxidation byproducts from drug tolerant persister cells. *Drug Resist. Updat.* **71**, 101007 (2023).
32. Ji, X. et al. Intermittent F-actin perturbations by magnetic fields inhibit breast cancer metastasis. *Research (Wash D C)* **6**, 0080 (2023).
33. Mao, Y. et al. Identification of a prognostic model based on costimulatory molecule-related subtypes and characterization of tumor microenvironment infiltration in acute myeloid leukemia. *Front. Genet.* **13**, 973319 (2022).
34. Zheng, D., Wei, Z. & Guo, W. Identification of a solute carrier family-based signature for predicting overall survival in osteosarcoma. *Front. Genet.* **13**, 849789 (2022).
35. Mann, K. et al. HIAYA CHAT study protocol: A randomized controlled trial of a health insurance education intervention for newly diagnosed adolescent and young adult cancer patients. *Trials* **23**(1), 682 (2022).
36. Peng, Z. et al. Self-assembling imageable silk hydrogels for the focal treatment of osteosarcoma. *Front. Cell Dev. Biol.* **10**, 698282 (2022).
37. Shen, S. et al. CircECE1 activates energy metabolism in osteosarcoma by stabilizing c-Myc. *Mol. Cancer* **19**(1), 151 (2020).
38. Zalacain, M. et al. Local administration of IL-12 with an HC vector results in local and metastatic tumor control in pediatric osteosarcoma. *Mol. Ther. Oncolytics* **20**, 23–33 (2021).
39. Chen, M., Qu, Y., Yue, P. & Yan, X. The prognostic value and function of HOXB5 in acute myeloid leukemia. *Front. Genet.* **12**, 678368 (2021).
40. Tan, X. et al. miR-138-5p-mediated HOXD11 promotes cell invasion and metastasis by activating the FN1/MMP2/MMP9 pathway and predicts poor prognosis in penile squamous cell carcinoma. *Cell Death Dis.* **13**(9), 816 (2022).
41. Xu, F. et al. HOXD13 suppresses prostate cancer metastasis and BMP4-induced epithelial-mesenchymal transition by inhibiting SMAD1. *Int. J. Cancer* **148**(12), 3060–3070 (2021).
42. Yin, J. & Guo, Y. HOXD13 promotes the malignant progression of colon cancer by upregulating PTPRN2. *Cancer Med.* **10**(16), 5524–5533 (2021).
43. Li, W. et al. miR-200a-3p- and miR-181-5p-mediated HOXB5 upregulation promotes HCC progression by transcriptional activation of EGFR. *Front. Oncol.* **12**, 822760 (2022).
44. Yang, L., Xie, F. & Li, S. Downregulation of homeobox B7 Inhibits the tumorigenesis and progression of osteosarcoma. *Oncol. Res.* **25**(7), 1089–1095 (2017).
45. Essner, J. J., Johnson, R. G. & Hackett, P. B. Jr. Overexpression of thyroid hormone receptor alpha 1 during zebrafish embryogenesis disrupts hindbrain patterning and implicates retinoic acid receptors in the control of hox gene expression. *Differentiation* **65**(1), 1–11 (1999).
46. Chen, S. et al. HOXA1 promotes proliferation and metastasis of bladder cancer by enhancing SMAD3 transcription. *Pathol. Res. Pract.* **239**, 154141 (2022).
47. Jeannotte, L., Lemieux, M., Charron, J., Poirier, F. & Robertson, E. J. Specification of axial identity in the mouse: Role of the Hoxa-5 (Hox1.3) gene. *Genes Dev.* **7**(11), 2085–2096 (1993).
48. Fan, F. et al. HOXA5: A crucial transcriptional factor in cancer and a potential therapeutic target. *Biomed. Pharmacother.* **155**, 113800 (2022).
49. Chen, Y. Q. et al. HOXA5 overexpression promotes osteosarcoma cell apoptosis through the p53 and p38alpha MAPK pathway. *Gene* **689**, 18–23 (2019).
50. Yu, T., Chen, D., Zhang, L. & Wan, D. microRNA-26a-5p Promotes proliferation and migration of osteosarcoma cells by targeting HOXA5 in vitro and in vivo. *OncoTargets Ther.* **12**, 11555–11565 (2019).
51. Lin, J. et al. Coexpression of HOXA6 and PBX2 promotes metastasis in gastric cancer. *Aging (Albany NY)* **13**(5), 6606–6624 (2021).
52. Wu, S., Wu, F. & Jiang, Z. Effect of HOXA6 on the proliferation, apoptosis, migration and invasion of colorectal cancer cells. *Int. J. Oncol.* **52**(6), 2093–2100 (2018).
53. Ji, S., Wang, S., Zhao, X. & Lv, L. Long noncoding RNA NEAT1 regulates the development of osteosarcoma through sponging miR-34a-5p to mediate HOXA13 expression as a competitive endogenous RNA. *Mol. Genet. Genomic Med.* **7**(6), e673 (2019).

Acknowledgments

We sincerely thank the TCGA (https://portal.gdc.cancer.gov/) andGEO database (https://www.ncbi.nlm.nih.gov/geo/) for the use of theirdata.

Author contributions

WG designed the study. WL and KX conducted bioinformatic analysis, wrote the manuscript, and was responsible for language revisions. XH, ZW, and ZW helped with some of the experiments. All authors reviewed the manuscript.

Funding

This study was supported by funds from The Open Project of Hubei Key Laboratory (2023KFZZ022).

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/s41598-024-84924-w.

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