



OPEN ACCESS

EDITED BY

Tinka Vidovic,
University of Zagreb, Croatia

REVIEWED BY

Panneerselvam Jayabal,
The University of Texas Health Science
Center at San Antonio, United States
Zongmei Gao,
Columbia University, United States
Dhiraj Kumar,
National Eye Institute (NIH), United States

*CORRESPONDENCE

Zhiyao Ren

✉ zhiyao.ren@ugent.be

Xin Wang

✉ xinwang.wang@ugent.be

[†]These authors have contributed
equally to this work and share
first authorship

[‡]These authors have contributed
equally to this work and share
last authorship

RECEIVED 14 June 2024

ACCEPTED 18 March 2025

PUBLISHED 04 April 2025

CITATION

Pang Y, Liang J, Deng Y, Chen W, Shen Y,
Li J, Wang X and Ren Z (2025) Identification
and validation of HOXC6 as a diagnostic
biomarker for Ewing sarcoma: insights
from machine learning algorithms
and *in vitro* experiments.
Front. Immunol. 16:1449355.
doi: 10.3389/fimmu.2025.1449355

COPYRIGHT

© 2025 Pang, Liang, Deng, Chen, Shen, Li,
Wang and Ren. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Identification and validation of HOXC6 as a diagnostic biomarker for Ewing sarcoma: insights from machine learning algorithms and *in vitro* experiments

Yonghua Pang^{1†}, Jiahui Liang^{2,3†}, Yakai Deng^{1†}, Weinan Chen¹,
Yunyan Shen¹, Jing Li⁴, Xin Wang^{5*†} and Zhiyao Ren^{5*†}

¹Department of Orthopedics, The 904th Hospital of the Joint Logistics Support Force, People's Liberation Army of China, Wuxi, Jiangsu, China, ²Department of Breast Surgery, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China, ³Department of General Surgery, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China, ⁴Department of Orthopedics, Linyi People's Hospital, Linyi, Shandong, China, ⁵Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium

Introduction: Early diagnosis of Ewing sarcoma (ES) is critical for improving patient prognosis. However, the accurate diagnosis of ES remains challenging, underscoring the need for novel diagnostic biomarkers to enhance diagnostic precision and reliability. This study aimed to identify potential gene expression-based biomarkers for the diagnosis of ES.

Methods: We selected the GSE17679, GSE45544, and GSE68776 datasets from the Gene Expression Omnibus (GEO) database. After correcting for batch effects, we combined ES and normal tissue samples from the GSE17679 and GSE45544 datasets to create a combined cohort. Two-thirds of both the tumor and normal samples from the combined cohort were randomly selected for the training cohort, while the remaining one-third served as the internal validation cohort. Additionally, the GSE68776 dataset was used for external validation. To identify key diagnostic genes, we applied three machine learning algorithms: least absolute shrinkage and selection operator (LASSO), support vector machine recursive feature elimination (SVM-RFE), and random forest (RF).

Results: HOXC6 was identified as a key diagnostic biomarker for ES. It demonstrated strong diagnostic performance across all cohorts, with area under the curve (AUC) values of 0.956 (95% CI: 0.909–0.990) in the training cohort, 0.995 (95% CI: 0.977–1.000) in the internal validation cohort, and 0.966 (95% CI: 0.910–0.999) in the external validation cohort. Functional validation through HOXC6 knockdown in the RD-ES cell line revealed that its suppression significantly inhibited cell proliferation and migration. Furthermore,

transcriptome sequencing suggested potential oncogenic mechanisms underlying HOXC6 function.

Discussion: These findings highlight HOXC6 as a promising diagnostic biomarker for ES, demonstrating robust performance across multiple datasets. Additionally, its functional role suggests potential as a therapeutic target.

KEYWORDS

Ewing sarcoma, bioinformatics, machine learning, HOXC6, diagnostic biomarker

1 Introduction

Ewing sarcoma (ES) is an invasive malignant tumor that primarily affects bone and soft tissue. ES is a rare malignant tumor with an annual incidence rate not exceeding 1%. ES is most common in children and adolescents, typically between the ages of 10 and 20. The majority of cases involve male patients (1–3). ES is a complex disease driven by the coordination of multiple signaling pathways, with EWSR1/FLI1 identified as a key contributor to its pathogenesis (4, 5). The absence of precursor lesions makes the diagnosis and treatment of ES a challenge (6). ES is characterized by rapid growth and a tendency to undergo metastasis (7). The early diagnosis and treatment of ES can effectively prevent its recurrence and metastasis, thereby improving its prognosis (8). Currently, diagnostic methods for ES primarily rely on clinical symptoms, imaging, and pathology. However, achieving an accurate diagnosis remains relatively challenging. ES patients typically present with mild symptoms in the early stages and ES is easily confused with trauma, sports injuries, or growth-related discomfort, increasing the likelihood of delayed diagnosis and treatment (9). In addition, imaging techniques have significant limitations in diagnosing ES because of the lack of characteristic features and a high rate of misdiagnosis (10). Moreover, the pathological diagnosis of ES is complex, requiring molecular pathology and multiple diagnostic approaches to ensure accuracy, particularly in challenging cases (6). Therefore, there is an urgent need for a precise and simple diagnostic approach for the early detection of ES.

In recent years, with the completion of the Human Genome Project, high-throughput sequencing technology has undergone groundbreaking advancements. These developments have enabled the acquisition of more accurate gene expression profiles, the identification of disease-related genes, and an analysis of the mechanisms underlying complex diseases, driving the advent of the precision medicine era (11). Additionally, the emergence of machine learning, which allows computer systems to automatically learn from data and algorithms to improve their performance, has shown great potential in omics research (12). Collectively, these advancements offer considerable potential for the discovery of novel diagnostic biomarkers. To date, numerous studies have reported the successful use of machine learning techniques to identify diagnostic biomarkers for various tumors, such as lung cancer (13), colorectal cancer (14), and breast

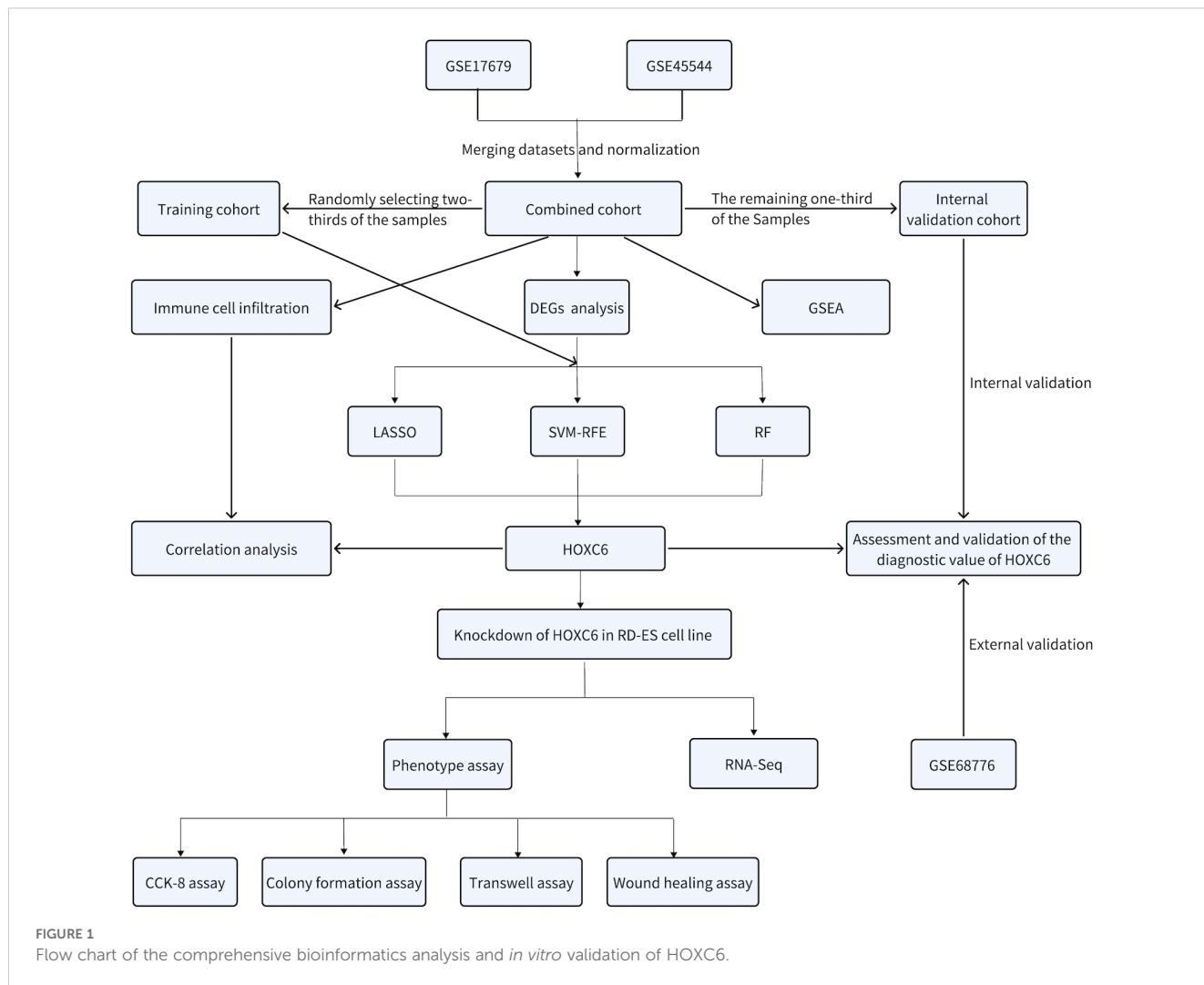
cancer (15). To our knowledge, no studies to date have focused on identifying diagnostic biomarkers for ES on the basis of transcriptome data. Therefore, this study aims to address this gap by identifying reliable diagnostic biomarkers derived from mRNA expression profiles.

According to the flowchart shown in Figure 1, we first merged the GSE17679 and GSE45544 datasets using batch effect correction to create a combined cohort. From this cohort, two-thirds of both the tumor and normal tissue samples were randomly selected to form the training cohort, while the remaining one-third were designated as the internal validation. Additionally, the GSE68776 dataset was utilized for external validation. We performed gene set enrichment analysis (GSEA) on the combined cohort to explore functional and pathway enrichment differences between tumor and normal tissues. Next, we analyzed the gene expression profiles of ES and normal tissues within the combined cohort to identify differentially expressed genes (DEGs). Furthermore, on the basis of the DEGs identified, we used the training cohort to apply least absolute shrinkage and selection operator (LASSO), support vector machine recursive feature elimination (SVM-RFE), and random forest (RF) machine learning algorithms to identify diagnostic biomarkers associated with the pathogenesis of ES. Among these, HOXC6 emerged as the sole diagnostic biomarker. We subsequently evaluated and validated the diagnostic performance of HOXC6 using receiver operating characteristic (ROC) curve analysis in the training, internal, and external validation cohorts. Additionally, we assessed immune cell infiltration via the CIBERSORT algorithm and investigated the relationship between infiltrating immune cells and HOXC6 expression. Finally, we performed *in vitro* experiments in which HOXC6 was knocked down in the RD-ES cell line to explore its functional role.

2 Methods

2.1 Data collection and processing

We selected the GSE17679, GSE45544, and GSE68776 datasets from the Gene Expression Omnibus (GEO) database (Table 1). After removing batch effects using the surrogate variable analysis (SVA) algorithm (16), the ES and normal tissue samples from the GSE17679 and GSE45544 datasets were combined. We randomly selected two-thirds of the tumor tissues and two-thirds of the normal tissues from



the combined cohort to construct the training cohort (68 tumor samples and 26 normal samples), while the remaining one-third was used for internal validation (34 tumor samples and 13 normal samples). We utilized the GSE68776 dataset for the purpose of external validation (32 tumor samples and 33 normal samples).

2.2 Functional and biological pathway enrichment analyses

GSEA was conducted to identify significantly altered biological functions and signaling pathways between tumor and normal tissues in

TABLE 1 Details of the datasets used in this study.

GEO series	Tissue sample		Cell line	
	ES	Normal	ES	Normal
GSE17679	88	18	0	0
GSE45544	14	21	7	1
GSE68776	32	33	0	0

the combined cohort. For the purpose of this study, the Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and hallmark gene sets were downloaded from the Molecular Signatures Database (<http://www.broadinstitute.org/msigdb>) (17).

2.3 Identification of DEGs

The R package “limma” (18) was employed, applying the criteria of an absolute value of $|\log_2(FC)| > 1$ and an FDR p value of < 0.05 to identify DEGs between ES and normal tissues in the combined cohort. The DEGs were subsequently visualized using heatmaps and volcano plots.

2.4 Screening diagnostic biomarkers via three machine learning methods

In this study, LASSO (19), SVM-RFE (20) and RF (21) machine learning algorithms were independently applied to the training cohort to further screen diagnostic genes from the DEGs. The

overlapping genes across the three sets of results were considered candidate diagnostic biomarkers for ES. LASSO logistic regression analysis was conducted using the “glmnet” package in R software. The SVM-RFE algorithm was applied using the “e1071” package in R software. The RF algorithm was implemented using the “randomForest” package in R software.

2.5 Assessment and validation of the diagnostic value of biomarkers in ES

To further assess the diagnostic value of the identified biomarkers in ES, ROC curves were generated, and the area under the curve (AUC) was calculated to evaluate and validate their predictive performance across the training, internal validation, and external validation cohorts. This analysis was performed using the R package “pROC”.

2.6 Evaluation of immune cell infiltration and correlation analysis between biomarkers and infiltrating immune cells

The quantification of different cell types involved in immune cell infiltration within ES gene expression profiles was conducted using the CIBERSORT algorithm (22). The correlations of infiltrating immune cells were visualized and analyzed using the R package “corrplot”. Additionally, a violin plot was generated using the R package “vioplot” to visualize the differences in infiltrating immune cells between ES and normal tissues. The correlation between diagnostic genes and immune cells was analyzed using the R packages “immuneCor” and “lollipop”. A lollipop chart was subsequently created to visualize the correlation between diagnostic gene levels and immune cell levels.

2.7 Cell culture

Human ES cells (RD-ES) were obtained from Qin Qi Biotechnology Development Co., Ltd., Shanghai, China. The cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% fetal bovine serum (FBS) under standard conditions in a humidified incubator with 5% CO₂ at 37°C.

2.8 Cell transfection

To knockdown HOXC6 expression, specific short hairpin RNAs (shRNAs) were designed and transfected into RD-ES cells. The sequences of the shRNAs used were as follows: sh-HOXC6-1: 5'-TGCTGTTGACAGTGAGCGCGGAGACAGAAATAAATATTAATAGTGAAGCCACAGATGTATTAATATTTATTTCTGTCTCCATGCCTACTGCCTCGGA-3'; sh-HOXC6-2:

5'-TGCTGTTGACAGTGAGCGACAGTAGGAGAAAATAAATAAATAGTGAAGCCACAGATGTATTTATTTATTTCTCCTACTGGTGCCTACTGCCTCGGA-3'. The knockdown efficiency was evaluated using real-time quantitative PCR (RT-qPCR) after 48 hours of transfection.

2.9 Real-time quantitative PCR

The sequences of the primers used in the experiment were as follows. For the HOXC6 gene, the primers used were as follows: forward primer, CCGTCAGTGTTCCTATCCAATTTTC; reverse primer, ATATTCGAGAACGGACCCAGAG. For ACTB, the primers used for the housekeeping gene were as follows: forward primer, CATGTACGTTGCTATCCAGGC; reverse primer, CTCCTTAATGTACGACGAT. After HOXC6 was knocked down in the RE-DS cell line, total mRNA was extracted from the cells with TRIzol reagent (TaKaRa, Japan). The concentration and purity were subsequently evaluated with a NanoDrop 2000 (Thermo Fisher, USA). The extracted RNA was then reverse transcribed into cDNA using the PrimeScript RT kit (TaKaRa, Japan) following the manufacturer’s instructions. RT-qPCR was subsequently performed using the SYBR Premix Ex Taq™ kit (TaKaRa, Japan) on an ABI StepOne Plus RT-qPCR system to detect SYBR Green fluorescence signals after each amplification cycle. Data processing was performed using GraphPad Prism 10.0.0, and a t test was conducted to compare the values for the experimental group with those of the control group.

2.10 Cell proliferation assay

Proliferation assays were performed over five consecutive days on cells seeded in a 96-well plate using the Cell Counting Kit-8 (CCK-8) reagent (Beyotime, China). A total of 2000 cells were plated per well and incubated at 37°C. The absorbance at 450 nm was measured daily for five days using a microplate reader.

2.11 Colony formation assays

Approximately 2000 cells per well were seeded into a 6-well culture plate and incubated at 37°C for two weeks. After being washed with PBS twice, the cells were fixed with 4% paraformaldehyde for 15 min and then stained with crystal violet. Each experiment was repeated three times. ImageJ was used for image analysis to convert images into cellular count data (23). The acquired counts were normalized by dividing them by the corresponding cell count in the control group, yielding percentage data. Data and image processing were performed using GraphPad Prism 10.0.0 and ImageJ. The statistical analysis consisted of a t test conducted on three replicate datasets to compare the values between the experimental and control groups.

2.12 Migration assays

A 24-well Transwell plate (Costar) was used for the cell migration assays. First, 50 μ l of serum-free Ham's F-12K medium was added to the upper chamber of the Transwell plate without the addition of a matrix. The plate was incubated at 37°C for 30 minutes. The knockdown cells or negative control (NC) cells were prepared as described previously. After a 20-minute incubation, 100 μ l of the shRNA mixture was mixed with 100 μ l of serum-free Ham's F-12K containing 1×10^5 cells. The mixture was transferred to the upper chamber of the Transwell system. In the lower chamber, 500 μ l of Ham's F-12K medium supplemented with 10% FBS was added. The mixture was incubated at 37°C for 24 hours. A cotton swab was used to remove nonmigrated cells from the lower chamber. The upper chamber was removed, and the cells were washed with PBS. The cells were fixed and stained with Giemsa. Five fields were randomly selected under an optical microscope, and the migrated cells were counted.

2.13 Wound healing assay

Wound healing assays were performed following previously described protocols (24). Briefly, cells were seeded in 6-well plates and incubated at 37°C until reaching full confluence. A scratch was then made across the middle of each well to create a wound, and the medium was replaced with serum-free medium. After 48 hours, the wound area was measured.

2.14 RNA-seq

RD-ES cells were subjected to RNA sequencing after HOXC6 knockdown. Approximately 2 μ g of total RNA was extracted from each sample and pretreated with the Epicenter Ribo-zeroTM rRNA Removal Kit. An RNA library was then constructed following the manufacturer's protocol for the NEBNext[®] UltraTM Directional RNA Library Prep Kit (NEB, USA). The procedure was as follows: RNA was first fragmented into small pieces by treatment with NEBNext First Strand Synthesis Reaction Buffer at high temperature, and first-strand cDNA was synthesized using random hexamer primers and M-MuLV reverse transcriptase. Next, second-strand cDNA was synthesized, and the fragment ends were repaired to blunt ends using exonuclease or polymerase. The 3' ends of the cDNA fragments were then adenylated and ligated to NEBNext adapters with a hairpin structure. After purification with the AMPure XP system (Beckman Coulter, Beverly, USA), 150–200 bp DNA fragments were selected and sequenced using the HiSeq 2500 platform (Illumina, CA, USA).

2.15 RNA-seq data processing and analysis

The FastQC program (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to assess the sequencing quality of all the sample data, which were trimmed using the FASTX-Toolkit. The

sequencing reads were mapped to the human reference genome assembly GRCh37 using TopHat (v2.0.9). Differential expression analysis was conducted on the basis of the gene expression matrix in count format. The R package edgeR was used to compare differential gene expression between shHOXC6-expressing cells and the NC group using a predetermined threshold (q value > 0.05). Volcano plots were generated to visualize the differentially expressed genes. Intersection analyses were performed separately for the upregulated and downregulated genes to identify the genes that were consistently differentially expressed across both replicates. GO and KEGG enrichment analyses were subsequently performed on the upregulated and downregulated genes, with a significance threshold of a q value < 0.05.

3 Results

3.1 Functional and biological pathway enrichment analyses

After correction, we merged the tissue samples from the GSE17679 and GSE45544 datasets, and the merging effect was satisfactory, allowing us to construct a combined cohort (Supplementary Figures 1A, B). To explore the biological pathways associated with ES, we performed GSEA using the cancer GO, KEGG, and hallmark gene sets. The results demonstrated that ES tissues were enriched in pathways associated with chromosomal abnormalities, the cell cycle, ribosomes, and the epithelial–mesenchymal transition (EMT) (Supplementary Figure 2, Supplementary Tables 1–3).

3.2 Identification of DEGs and diagnostic biomarker screening using three machine learning methods

DEGs were identified in the combined cohort. We identified a total of 1832 DEGs, comprising 1077 upregulated genes and 755 downregulated genes (Supplementary Table 4). DEGs were visualized using heatmaps (Figure 2A) and volcano plots (Figure 2B).

On the basis of the DEGs, we further employed three machine learning algorithms—LASSO, SVM-RFE, and RF—to screen for diagnostic biomarkers of ES in the training group, identifying 25, 8, and 20 potential diagnostic genes, respectively (Figures 3A–C). HOXC6 was identified as the only overlapping gene across the three machine learning analyses (Figure 3D). HOXC6 expression is significantly higher in tumor tissues than in normal tissues (Supplementary Figure 3). Interestingly, we also found that HOXC6 expression levels were higher in tumor cell lines than in the normal cell line on the basis of the GSE45544 dataset (Supplementary Figure 4). Furthermore, HOXC6 expression was elevated in multiple tumor types, including stomach adenocarcinoma, invasive breast carcinoma, and esophageal carcinoma (Supplementary Figure 5), highlighting its critical role in various cancers.

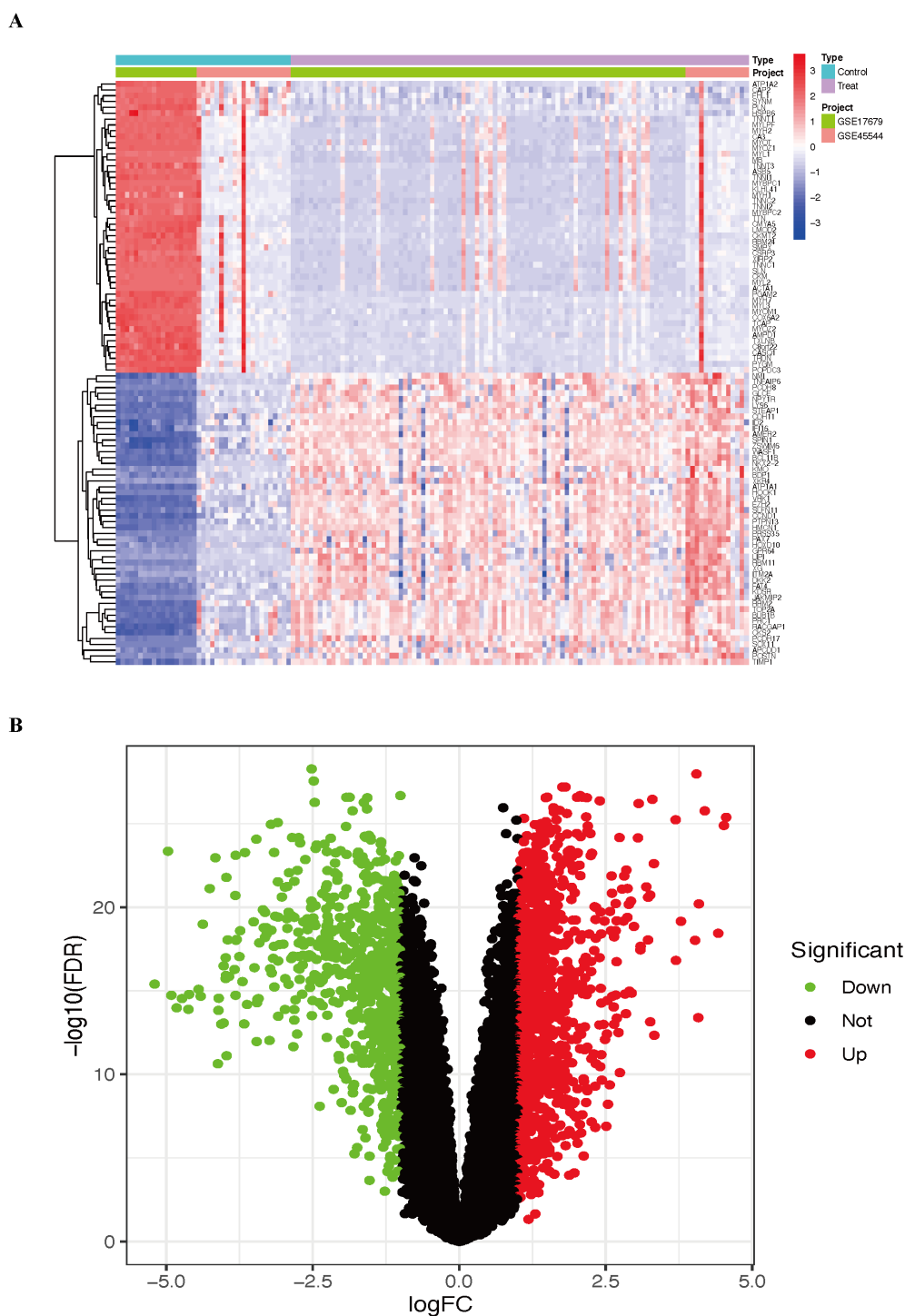
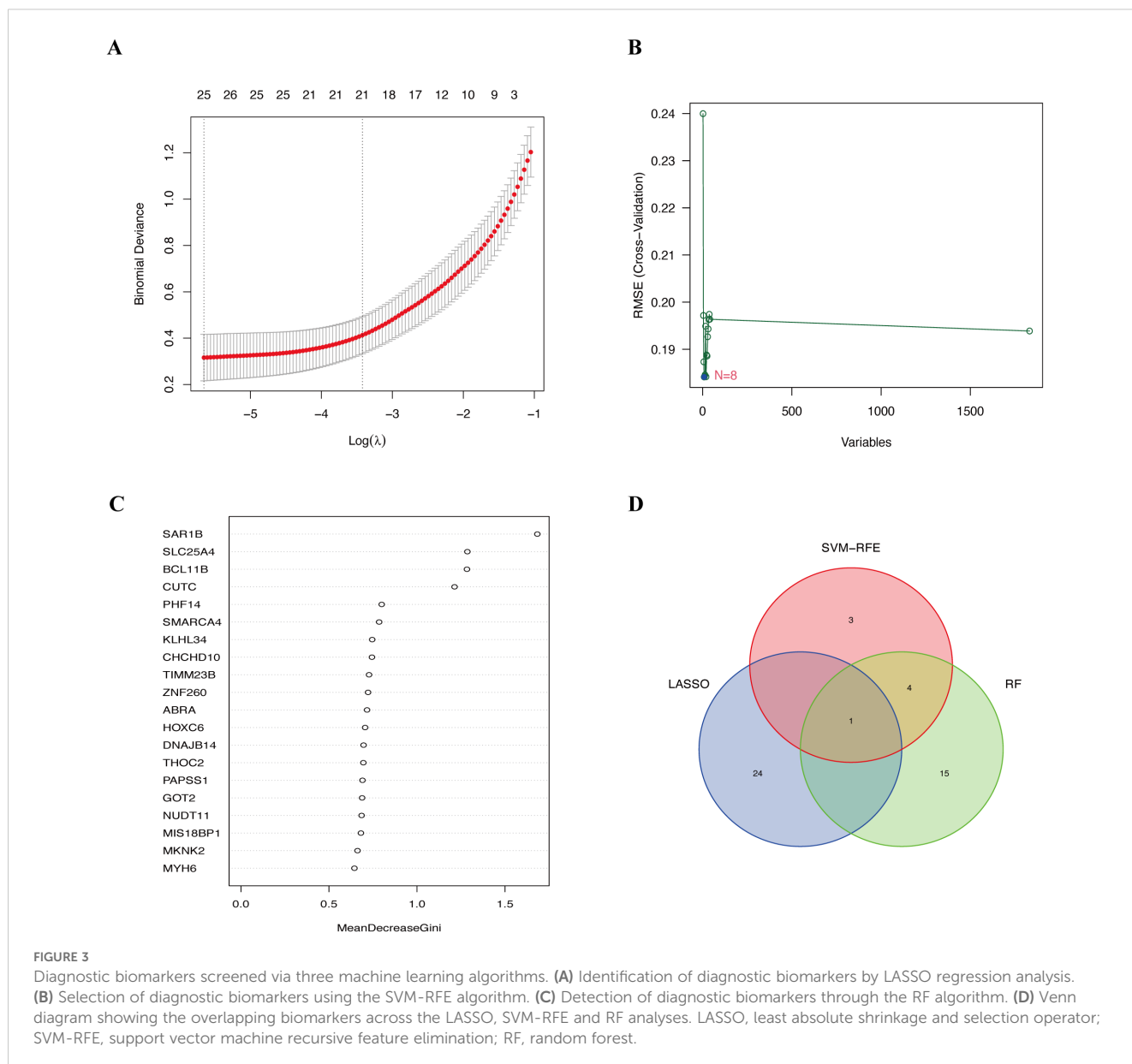


FIGURE 2
The results include the top 50 upregulated and downregulated DEGs identified in the combined cohort. **(A)** Heatmap of the DEGs. **(B)** Volcano plot of the DEGs. ES, Ewing sarcoma; DEGs, differentially expressed genes.

3.3 Assessment and validation of the diagnostic value of HOXC6 in ES

To further assess and validate the diagnostic value of HOXC6 in ES, ROC analysis was conducted for HOXC6 across the training, internal, and external validation cohorts. The results demonstrated

that HOXC6 exhibited strong diagnostic performance in the training cohort, with an AUC of 0.956 (95% CI: 0.909–0.990) (Figure 4A). Similarly, in the internal validation cohort, HOXC6 showed excellent diagnostic accuracy, with an AUC of 0.995 (95% CI: 0.977–1.000) (Figure 4B). Furthermore, HOXC6 was highly expressed in tumor tissues in the external validation cohort



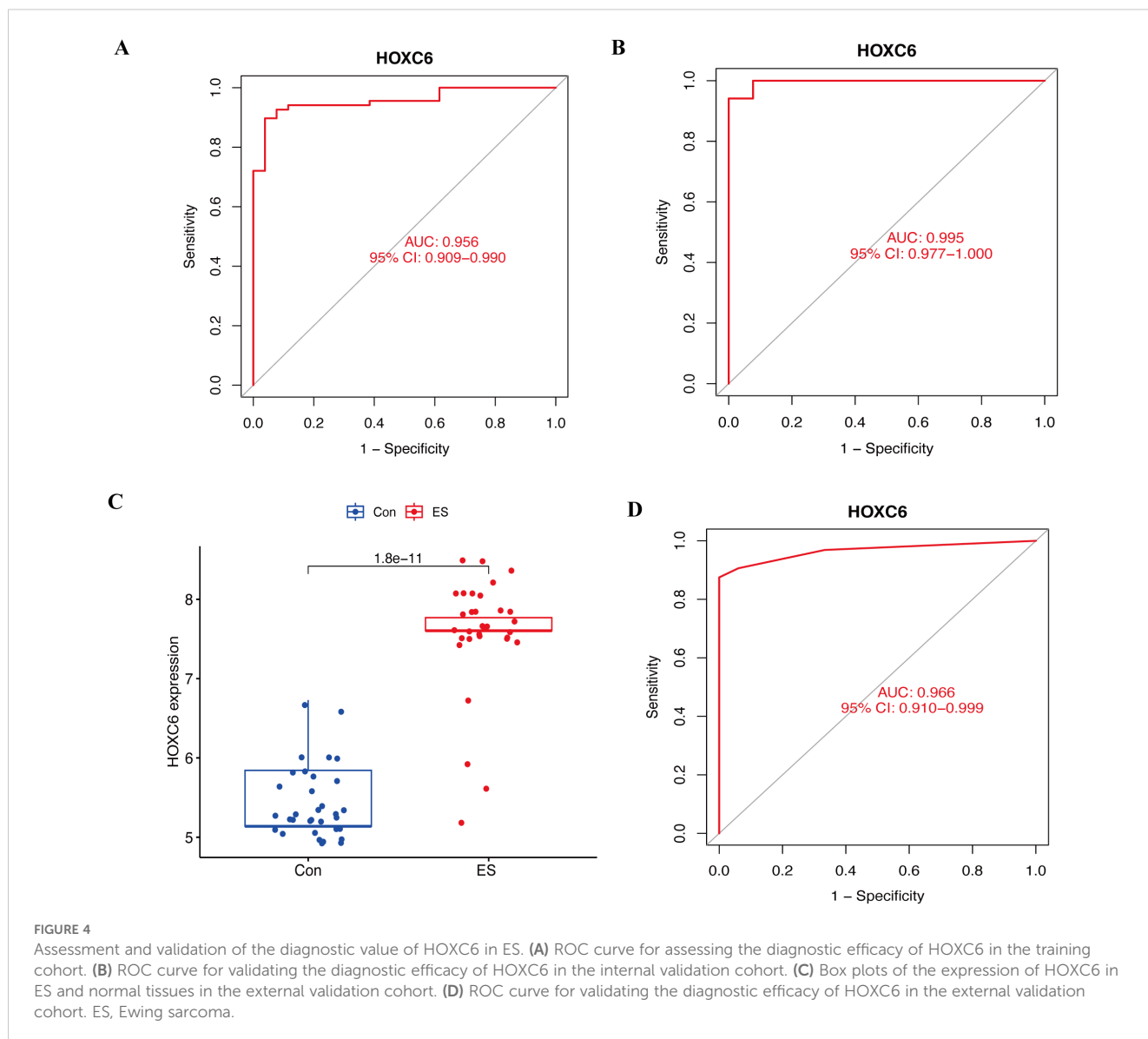
(Figure 4C) and maintained strong diagnostic efficacy, with an AUC of 0.966 (95% CI: 0.910–0.999) (Figure 4D). These results indicate that HOXC6 is a reliable diagnostic biomarker for ES.

3.4 Evaluation of immune cell infiltration and correlation analysis between HOXC6 and infiltrating immune cells

Initially, we assessed the infiltration of immune cells in the combined cohort using the CIBERSORT algorithm (Figure 5A). Compared with that in normal tissues, the degree of M0 macrophage infiltration in ES tissues was significantly greater

($p=0.016$). Conversely, the degree of infiltration of memory B cells ($p=0.009$) and CD8+ T cells ($p=0.009$) was greater in normal tissues than in ES tissues (Figure 5B). Furthermore, we calculated the correlations between the 22 types of infiltrating immune cells (Figure 5C).

HOXC6 expression was significantly positively correlated with the number of M0 macrophages ($p = 0.009$) and follicular helper T cells ($p = 0.021$) and significantly negatively correlated with the number of CD4 memory-activated T cells ($p = 0.022$), activated dendritic cells ($p = 0.018$), memory B cells ($p = 0.010$), CD8+ T cells ($p = 0.004$), and monocytes ($p < 0.001$) (Figure 5D). These findings suggest that HOXC6 may play a role in the tumor immune microenvironment.



3.5 Knockdown of HOXC6 inhibited ES cell proliferation and migration

We established two RD-ES shHOXC6 cell lines through lentiviral transduction. As shown in [Figure 6A](#), the HOXC6 knockdown efficiency in RD-ES cells was satisfactory.

We further investigated the impact of HOXC6 knockdown on the proliferation of ES cells through CCK-8 and colony formation assays. As shown in [Figure 6B](#), the CCK-8 assay revealed that HOXC6 knockdown suppressed RD-ES cell proliferation. Similarly, the colony formation assay confirmed a significant reduction in proliferation ([Figure 6C](#)). Together, these results suggest that HOXC6 positively regulates ES cell proliferation.

We next examined the cell migration ability. Transwell assays demonstrated that HOXC6 knockdown significantly inhibited the migration of RD-ES cells ([Figure 6D](#)). Similarly, the wound healing assay results revealed that reduced HOXC6 expression impaired the wound closure rate of RD-ES cells ([Figure 6E](#)). These findings

suggest that HOXC6 functions as an oncogene in the progression of ES.

3.6 Transcriptome-based analysis of HOXC6-related pathways

We performed transcriptome sequencing on two RD-ES shHOXC6 cell lines and the NC group, analyzing all upregulated and downregulated genes between the two knockdown cell lines and the NC group ([Figures 7A, B](#), [Supplementary Tables 5, 6](#)). First, we identified the intersection of commonly downregulated genes ([Figure 7C](#), [Supplementary Table 7](#)) and conducted GO and KEGG analyses ([Figures 7D, E](#), [Supplementary Tables 8, 9](#)). The results revealed associations with ribosomes, metabolism, and the cell cycle. Functional enrichment analysis was also performed on the intersecting upregulated genes ([Supplementary Figure 6](#), [Supplementary Tables 10–12](#)).

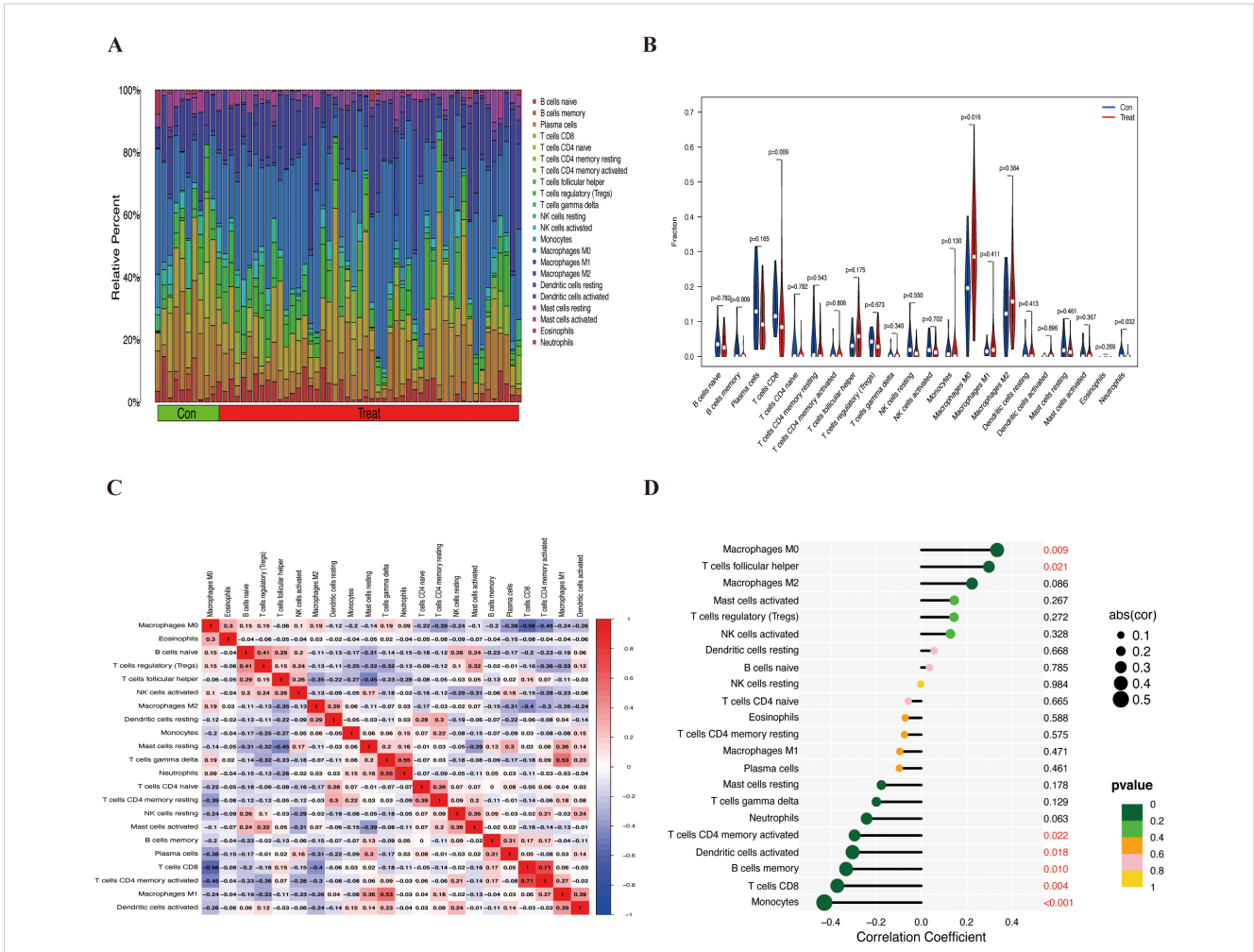


FIGURE 5
The infiltration of immune cells in the combined cohort and its correlation with HOXC6. **(A)** Bar chart of the proportions of 22 types of infiltrating immune cells. **(B)** Differential analysis of 22 types of infiltrating immune cells. **(C)** Heatmap showing the correlations between 22 immune cells. **(D)** Correlations between HOXC6 and 22 types of immune cells.

4 Discussion

ES ranks as the second most prevalent malignant bone tumor among children and adolescents (25). Although significant progress has been made in the early diagnosis and treatment of ES in recent years, the absence of specific diagnostic biomarkers still presents significant challenges for achieving accurate early diagnosis of ES. Reports in the literature indicate that up to 25% of patients present with metastasis at the time of diagnosis, contributing to a persistently poor clinical prognosis (26). Thus, the identification of diagnostic biomarkers for ES is crucial for improving the diagnostic accuracy and prognosis of ES. To our knowledge, this study is among the first to explore diagnostic biomarkers for ES on the basis of transcriptome data.

The GSEA comparing ES and normal tissues in this study revealed that chromosomal abnormalities, the cell cycle, ribosomes, and the EMT may play critical roles in the development and progression of ES. Chromosomal translocation involving EWSR1/FLI1 is a key driver in the pathogenesis of ES (27). While the EMT is

essential for normal embryonic development and tissue regeneration, its aberrant reactivation is associated with tumor malignancy, contributing to cancer progression and metastasis (28).

As genomic technology advances, there is a growing trend toward the use of bioinformatics analysis methods to explore the molecular mechanisms underlying gene expression profiles (29). This approach holds great promise for identifying specific molecular diagnostic markers or therapeutic targets for various diseases (30). LASSO analysis is a regression method known for its ability to efficiently process large datasets, effectively perform parameter shrinkage and variable selection, prevent overfitting, and more accurately screen variables (31). In omics research, the screening of differentially expressed biomarkers from tissue data is pivotal. SVM-RFE has emerged as an efficient technique for feature selection and has promising application prospects in the analysis of metabolomics data (20). RF is a flexible and powerful machine learning algorithm that offers advantages such as high accuracy, resistance to overfitting, the ability to handle missing data, and built-in feature selection. It is particularly well suited for complex

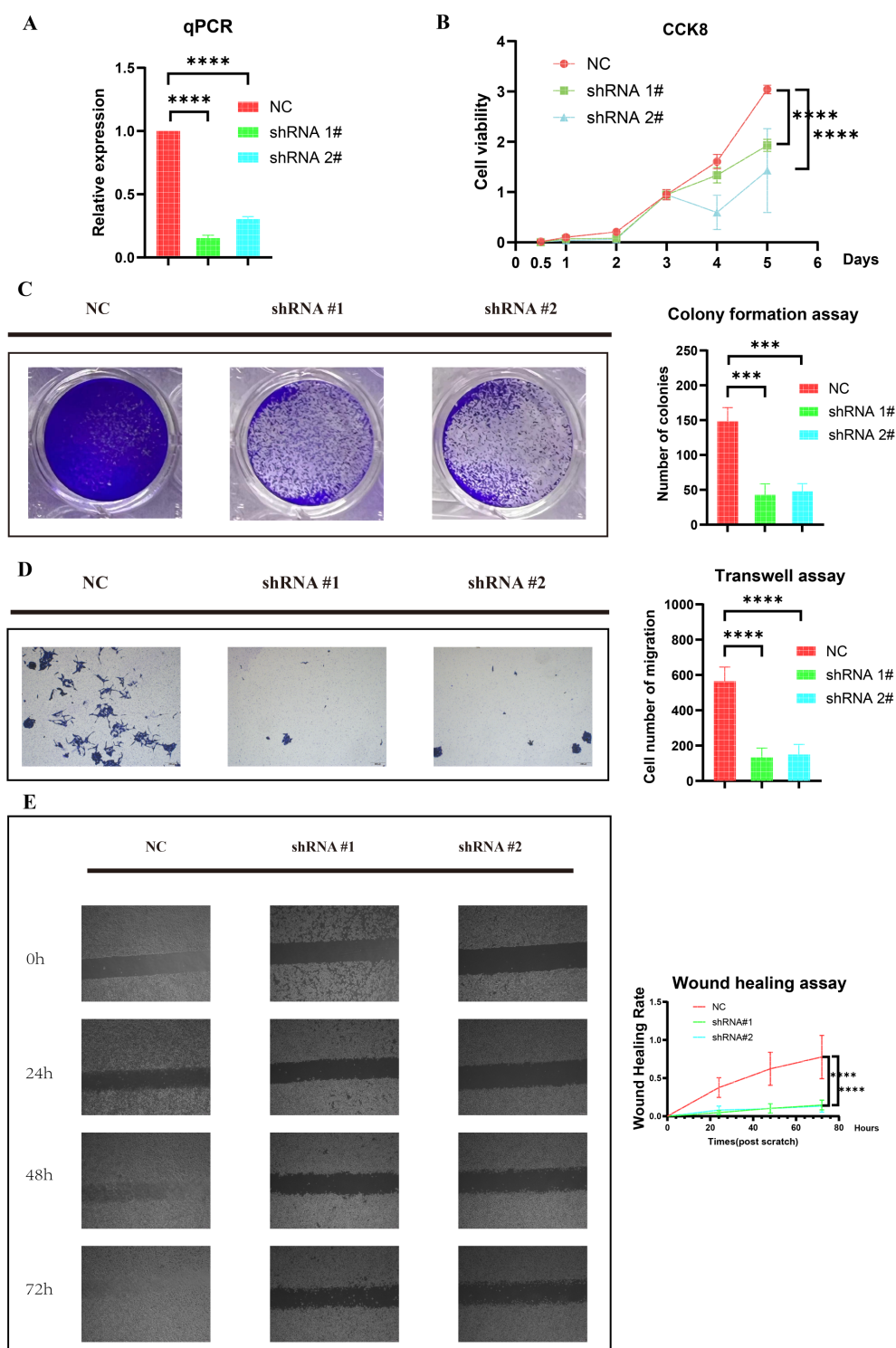
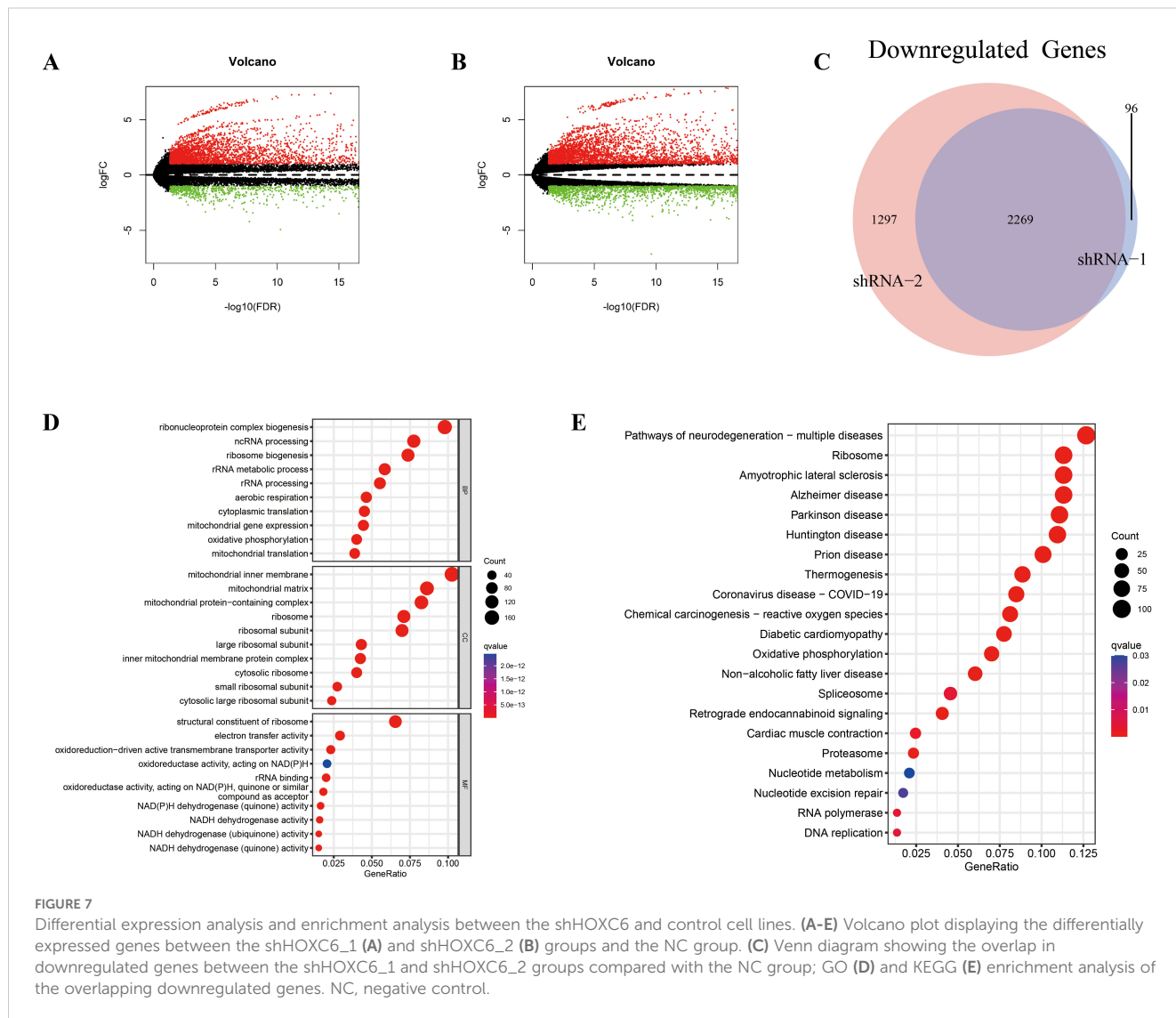


FIGURE 6 Knockdown of HOXC6 inhibited ES cell proliferation and migration. **(A)** Knockdown efficiency of HOXC6 in the RD-ES shHOXC6 cell line. **(B)** CCK-8 assay. **(C)** Colony formation assay. **(D)** Transwell assay. **(E)** Wound healing assay. ES, Ewing sarcoma. The data is presented as the mean from at least three independent experiments. (***) $p < 0.001$; **** $p < 0.0001$.

datasets and high-dimensional problems (21). To identify reliable diagnostic biomarkers, we first performed differential gene expression analysis in the combined cohort. On the basis of the identified DEGs, we applied three machine learning methods in the

training cohort and identified HOXC6 as the only diagnostic biomarker through intersection analysis. ROC curves demonstrated the excellent predictive performance of HOXC6 for ES. Internal validation is crucial for estimating the generalizability



of a model (32). In this study, the strong results from internal validation provide robust evidence supporting the reliability of the diagnostic biomarker HOXC6. Furthermore, external validation is essential for assessing the reproducibility of the model and its applicability to independent samples (33). The GSE68776 dataset was used for external validation in this study. The AUCs obtained in this study were satisfactory, indicating the reproducibility of HOXC6 use for ES diagnosis in practical applications.

HOXC6 plays a pivotal role in regulating embryonic development, cell differentiation, and organ formation (34). As one of the 39 HOX genes in humans (35), HOXC6 is overexpressed in several cancers, including osteosarcoma (36), lung adenocarcinoma (37), and prostate cancer (38). It is critically involved in tumor cell proliferation, growth, and metastasis by regulating various proteins, such as bone morphogenetic protein 7 (BMP7) (39, 40). Liu et al. (41) reported that HOXC6 is involved in various processes, including immune cell infiltration, immune-related genes, chemotherapy sensitivity, signaling pathways, and

transcriptional regulatory networks. Moreover, it may function as a radiosensitivity-related gene affecting the prognosis of rectal cancer patients and could serve as a potential target for radiotherapy. Huang et al. (42) noted that HOXC6 may play a significant role in promoting tumor development and glioma progression by regulating the EMT signaling pathway. Additionally, it may serve as a novel immunotherapeutic target for glioma treatment. Wang et al. (43) reported that HOXC6 overexpression enhances BCL2-mediated antiapoptotic effects, thereby promoting cervical cancer cell cycle progression and proliferation. These studies indicate that HOXC6 plays an important role in tumor diseases.

The tumor microenvironment plays a crucial role in all stages of cancer progression (44). In this study, we observed that CD8+ T-cell infiltration was significantly lower in ES tissues than in normal tissues and was negatively correlated with HOXC6 expression. CD8+ T cells, also known as cytotoxic T cells, can recognize and directly kill tumor cells. However, in the tumor immune microenvironment, CD8+ T cells often experience exhaustion (45). Previous studies

have indicated that HOXC6 can regulate the tumor immune microenvironment (42). Hence, our results indicate that HOXC6 may promote ES progression by mediating the exhaustion of CD8+ T cells.

To investigate the mechanisms by which HOXC6 promotes tumor progression, we performed transcriptome sequencing. Differential expression analysis between the knockdown and NC groups, with a focus on downregulated genes, revealed significant enrichment of ribosome-related pathways in both the GO and KEGG analyses. Interestingly, ribosome-related pathways were also enriched in the GSEA between ES and normal tissues. The ribosome, a complex molecular machine responsible for protein synthesis, has been shown to play critical roles in tumor proliferation, growth, and metastasis (46, 47). To date, no studies have reported that HOXC6 mediates tumor progression by regulating ribosomes. Therefore, this study may reveal a novel oncogenic mechanism of HOXC6 and highlight its potential as a therapeutic target.

In the future, the collection of clinical samples will be essential for validating the diagnostic accuracy of our findings. Additionally, further investigations utilizing a wider array of *in vitro* models, including diverse cell lines and patient-derived primary cells, as well as *in vivo* studies, will offer valuable insights into the role of HOXC6 in the pathogenesis of ES. Moreover, transcriptomic data can be leveraged to elucidate the downstream mechanisms regulated by HOXC6. Further experiments are needed to elucidate the downstream mechanisms of HOXC6. Overall, this study provides a simple and efficient diagnostic strategy for ES based on the expression of a single gene, HOXC6, which holds great potential for clinical application. Additionally, our findings establish HOXC6 as an oncogene, highlighting its promise as a therapeutic target pending further comprehensive research.

5 Conclusion

In this study, a simple and efficient diagnostic strategy for ES was developed on the basis of the expression of a single gene, HOXC6, which holds great potential for clinical application. Additionally, *in vitro* experiments revealed HOXC6 as an oncogene in ES tumorigenesis, highlighting its promise as a therapeutic target pending further comprehensive research.

Data availability statement

The data presented in this study are deposited in the GEO database repository, accession number GSE263430.

Ethics statement

Ethical approval was not required for the studies on humans in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used.

Author contributions

YP: Data curation, Formal Analysis, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. JL: Data curation, Formal Analysis, Investigation, Validation, Writing – original draft, Writing – review & editing. YD: Formal Analysis, Investigation, Validation, Writing – original draft, Writing – review & editing. WC: Formal Analysis, Investigation, Validation, Visualization, Writing – original draft, Writing – review & editing. YS: Formal Analysis, Software, Validation, Writing – original draft, Writing – review & editing. JL: Formal analysis, Investigation, Validation, Writing – review & editing. XW: Data curation, Formal Analysis, Investigation, Validation, Writing – original draft, Writing – review & editing. ZR: Data curation, Formal Analysis, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Acknowledgments

We acknowledge the GEO database for providing their platform and the contributors for uploading their meaningful datasets.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2025.1449355/full#supplementary-material>

SUPPLEMENTARY TABLE 1
GSEA of ES and normal tissues based on GO biological processes. GSEA, gene set enrichment analysis; ES, Ewing sarcoma; GO, Gene Ontology.

SUPPLEMENTARY TABLE 2

GSEA of ES and normal tissues based on the KEGG pathway analysis. GSEA, gene set enrichment analysis; ES, Ewing sarcoma; KEGG, Kyoto Encyclopedia of Genes and Genomes.

SUPPLEMENTARY TABLE 3

GSEA of ES and normal tissues on the basis of hallmark pathways. GSEA, gene set enrichment analysis; ES, Ewing sarcoma.

SUPPLEMENTARY TABLE 4

Differentially expressed genes between ES and normal tissues in the combined cohort. ES, Ewing sarcoma.

SUPPLEMENTARY TABLE 5

Differentially expressed genes between the shHOXC6_1 and NC groups.

SUPPLEMENTARY TABLE 6

Differentially expressed genes between the shHOXC6_2 and NC groups.

SUPPLEMENTARY TABLE 7

Downregulated gene sets.

SUPPLEMENTARY TABLE 8

GO enrichment analysis of the downregulated gene sets. GO, Gene Ontology.

SUPPLEMENTARY TABLE 9

KEGG enrichment analysis of the downregulated gene sets. KEGG, Kyoto Encyclopedia of Genes and Genomes.

SUPPLEMENTARY TABLE 10

Upregulated gene sets.

SUPPLEMENTARY TABLE 11

GO analysis of the upregulated gene sets. GO, Gene Ontology.

SUPPLEMENTARY TABLE 12

KEGG pathway analysis of the upregulated gene sets. KEGG, Kyoto Encyclopedia of Genes and Genomes.

References

- Liu D, Liu X, Li X, Liu Y, Yu J. Primary endobronchial multifocal ewing's sarcoma: A rare case report. *Front Oncol.* (2024) 14:1431950. doi: 10.3389/fonc.2024.1431950
- El Beaino M, Liu J, Wasylishen AR, Pourebrahim R, Migut A, Bessellieu BJ, et al. Loss of stag2 cooperates with *ews-flil1* to transform murine mesenchymal stem cells. *BMC Cancer.* (2020) 20:3. doi: 10.1186/s12885-019-6465-8
- Longhi A, Marrari A, Tetta C, Parmeggiani A, Parise O, Ferrari C, et al. The critical role of stereotactic body radiation therapy in multimodal treatment of lung metastasis from bone and soft tissue sarcomas. *Cancers (Basel).* (2024) 16(1):3593. doi: 10.3390/cancers16213593
- Mishra MN, Sharma R, Chandavarkar V, Premalatha BR. Pathogenesis of ewing sarcoma: existing and emerging trends. *Adv Cancer Biol - Metastasis.* (2021) 2:100008. doi: 10.1016/j.adcanc.2021.100008
- Lessnick SL, Ladanyi M. Molecular pathogenesis of ewing sarcoma: new therapeutic and transcriptional targets. *Annu Rev Pathol.* (2012) 7:145–59. doi: 10.1146/annurev-pathol-011110-130237
- Zollner SK, Amatruda JF, Bauer S, Collaud S, de Alava E, DuBois SG, et al. Ewing sarcoma-diagnosis, treatment, clinical challenges and future perspectives. *J Clin Med.* (2021) 10(8):1685. doi: 10.3390/jcm10081685
- El Harras Y, Choayb S, Laasri K, Allali N, Chat L, El Haddad S. Extra-skeletal ewing's sarcoma of the leg with multiple skeletal and pulmonary metastases: A rare pediatric case report. *SAGE Open Med Case Rep.* (2023) 11:2050313X231194815. doi: 10.1177/2050313X231194815
- Aran V, Devalle S, Meohas W, Heringer M, Cunha Caruso A, Pinheiro Aguiar D, et al. Osteosarcoma, chondrosarcoma and ewing sarcoma: clinical aspects, biomarker discovery and liquid biopsy. *Crit Rev Oncol Hematol.* (2021) 162:103340. doi: 10.1016/j.critrevonc.2021.103340
- Zarghooni K, Bratke G, Landgraf P, Simon T, Maintz D, Eysel P. The diagnosis and treatment of osteosarcoma and ewing's sarcoma in children and adolescents. *Dtsch Arztebl Int.* (2023) 120:405–12. doi: 10.3238/arztebl.m2023.0079
- Ribeiro GJ, Gillet R, Hossu G, Trinh JM, Euxibie E, Sirveaux F, et al. Solitary bone tumor imaging reporting and data system (Bti-rads): initial assessment of a systematic imaging evaluation and comprehensive reporting method. *Eur Radiol.* (2021) 31:7637–52. doi: 10.1007/s00330-021-07745-9
- Morganti S, Tarantino P, Ferraro E, D'Amico P, Viale G, Trapani D, et al. Complexity of genome sequencing and reporting: next generation sequencing (Ngs) technologies and implementation of precision medicine in real life. *Crit Rev Oncol Hematol.* (2019) 133:171–82. doi: 10.1016/j.critrevonc.2018.11.008
- Gong Y, Ding W, Wang P, Wu Q, Yao X, Yang Q. Evaluating machine learning methods of analyzing multiclass metabolomics. *J Chem Inf Model.* (2023) 63:7628–41. doi: 10.1021/acs.jcim.3c01525
- Guan S, Xu Z, Yang T, Zhang Y, Zheng Y, Chen T, et al. Identifying potential targets for preventing cancer progression through the pla2g1b recombinant protein using bioinformatics and machine learning methods. *Int J Biol Macromol.* (2024) 276:133918. doi: 10.1016/j.ijbiomac.2024.133918
- Wei W, Li Y, Huang T. Using machine learning methods to study colorectal cancer tumor micro-environment and its biomarkers. *Int J Mol Sci.* (2023) 24(13):11133. doi: 10.3390/ijms241311133
- Anh NK, Lee A, Phat NK, Yen NTH, Thu NQ, Tien NTN, et al. Combining metabolomics and machine learning to discover biomarkers for early-stage breast cancer diagnosis. *PLoS One.* (2024) 19:e0311810. doi: 10.1371/journal.pone.0311810
- Parker HS, Leek JT, Favorov AV, Considine M, Xia X, Chavan S, et al. Preserving biological heterogeneity with a permuted surrogate variable analysis for genomics batch correction. *Bioinformatics.* (2014) 30:2757–63. doi: 10.1093/bioinformatics/btu375
- Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (Msigdb) 3.0. *Bioinformatics.* (2011) 27:1739–40. doi: 10.1093/bioinformatics/btr260
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res.* (2015) 43:e47. doi: 10.1093/nar/gkv007
- Zhao E, Xie H, Zhang Y. Predicting diagnostic gene biomarkers associated with immune infiltration in patients with acute myocardial infarction. *Front Cardiovasc Med.* (2020) 7:586871. doi: 10.3389/fcvm.2020.586871
- Lin X, Yang F, Zhou L, Yin P, Kong H, Xing W, et al. A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information. *J Chromatogr B Analyt Technol BioMed Life Sci.* (2012) 910:149–55. doi: 10.1016/j.jchromb.2012.05.020
- Feng CH, Disis ML, Cheng C, Zhang L. Multimetric feature selection for analyzing multicategory outcomes of colorectal cancer: random forest and multinomial logistic regression models. *Lab Invest.* (2022) 102:236–44. doi: 10.1038/s41374-021-00662-x
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* (2015) 12:453–7. doi: 10.1038/nmeth.3337
- Kurnia KA, Sampurna BP, Audira G, Juniardi S, Vasquez RD, Roldan MJM, et al. Performance comparison of five methods for tetrahymena number counting on the imagej platform: assessing the built-in tool and machine-learning-based extension. *Int J Mol Sci.* (2022) 23(11):6009. doi: 10.3390/ijms23116009
- Zhu Y, Xiao B, Liu M, Chen M, Xia N, Guo H, et al. N6-methyladenosine-modified oncofetal lincrna mir4435-2hg contributed to stemness features of hepatocellular carcinoma cells by regulating rna 2'-O methylation. *Cell Mol Biol Lett.* (2023) 28:89. doi: 10.1186/s11658-023-00493-2
- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA Cancer J Clin.* (2021) 71:7–33. doi: 10.3322/caac.21654
- Spector LG, Hubbard AK, Diessner BJ, Machiela MJ, Webber BR, Schiffman JD. Comparative international incidence of ewing sarcoma 1988 to 2012. *Int J Cancer.* (2021) 149:1054–66. doi: 10.1002/ijc.33674
- Grünewald TGP, Cidre-Aranaz F, Surdez D, Tomazou EM, de Álava E, Kovar H, et al. Ewing sarcoma. *Nat Rev Dis Primers.* (2018) 4:5. doi: 10.1038/s41572-018-0003-x
- Huang Y, Hong W, Wei X. The molecular mechanisms and therapeutic strategies of emt in tumor progression and metastasis. *J Hematol Oncol.* (2022) 15:129. doi: 10.1186/s13045-022-01347-8
- Jiang Y, Li M, Fan Y, Di Z. Characterizing dissimilarity of weighted networks. *Sci Rep.* (2021) 11:5768. doi: 10.1038/s41598-021-85175-9
- Zhang P, Pei S, Gong Z, Feng Y, Zhang X, Yang F, et al. By integrating single-cell rna-seq and bulk rna-seq in sphingolipid metabolism, cacybp was identified as a potential therapeutic target in lung adenocarcinoma. *Front Immunol.* (2023) 14:1115272. doi: 10.3389/fimmu.2023.1115272
- Lemieux W, Fleischer D, Yang AY, Niemann M, Ouakacha K, Klement W, et al. Dissecting the impact of molecular T-cell hla mismatches in kidney transplant failure:

A retrospective cohort study. *Front Immunol.* (2022) 13:1067075. doi: 10.3389/fimmu.2022.1067075

32. Reps JM, Ryan P, Rijnbeek PR. Investigating the impact of development and internal validation design when training prognostic models using a retrospective cohort in big us observational healthcare data. *BMJ Open.* (2021) 11:e050146. doi: 10.1136/bmjopen-2021-050146
33. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J.* (2021) 14:49–58. doi: 10.1093/ckj/sfaa188
34. Chariot A, Gielen J. Molecules in focus the hoxc6 homeodomain-containing proteins. *Int J Biochem Cell Biol.* (1998) 30:651–5. doi: 10.1016/S1357-2725(97)00136-2
35. Hussain I, Bhan A, Ansari KI, Deb P, Bobzean SA, Perrotti LI, et al. Bisphenol-a induces expression of hoxc6, an estrogen-regulated homeobox-containing gene associated with breast cancer. *Biochim Biophys Acta.* (2015) 1849:697–708. doi: 10.1016/j.bbagr.2015.02.003
36. Bodey B, Bodey B, Siegel S, Luck J, Kaiser H. Homeobox B3, B4, and C6 gene product expression in osteosarcomas as detected by immunocytochemistry. *Anticancer Res.* (2000) 20:2717–21.
37. Xin M, Peng H, Zhang L. Exploring the prognosis value, immune correlation, and drug responsiveness prediction of homeobox C6 (Hoxc6) in lung adenocarcinoma. *Discovery Oncol.* (2024) 15:393. doi: 10.1007/s12672-024-01273-w
38. Zhou J, Yang X, Song P, Wang H, Wang X. Hoxc6 in the prognosis of prostate cancer. *Artif Cells Nanomed Biotechnol.* (2019) 47:2715–20. doi: 10.1080/21691401.2019.1635136
39. McCabe C, Spyropoulos D, Martin D, Moreno C. Genome-wide analysis of the homeobox C6 transcriptional network in prostate cancer. *Cancer Res.* (2008) 68:1988–96. doi: 10.1158/0008-5472.Can-07-5843
40. Moon S, Kim S, Yoon J, Ahn S. Hoxc6 is deregulated in human head and neck squamous cell carcinoma and modulates bcl-2 expression. *J Biol Chem.* (2012) 287:35678–88. doi: 10.1074/jbc.M112.361675
41. Liu Y, Yang Y, Ni F, Tai G, Yu C, Jiang X, et al. Research on radiotherapy related genes and prognostic target identification of rectal cancer based on multi-omics. *J Transl Med.* (2023) 21:856. doi: 10.1186/s12967-023-04753-9
42. Huang H, Huo Z, Jiao J, Ji W, Huang J, Bian Z, et al. Hoxc6 impacts epithelial-mesenchymal transition and the immune microenvironment through gene transcription in gliomas. *Cancer Cell Int.* (2022) 22:170. doi: 10.1186/s12935-022-02589-9
43. Wang Y, Wang C, Liu N, Hou J, Xiao W, Wang H. Hoxc6 promotes cervical cancer progression via regulation of bcl-2. *FASEB J.* (2019) 33:3901–11. doi: 10.1096/fj.201801099RR
44. de Visser KE, Joyce JA. The evolving tumor microenvironment: from cancer initiation to metastatic outgrowth. *Cancer Cell.* (2023) 41:374–403. doi: 10.1016/j.ccell.2023.02.016
45. Wang Q, Qin Y, Li B. Cd8+ T cell exhaustion and cancer immunotherapy. *Cancer Lett.* (2023) 559:216043. doi: 10.1016/j.canlet.2022.216043
46. Pelletier J, Thomas G, Volarević S. Ribosome biogenesis in cancer: new players and therapeutic avenues. *Nat Rev Cancer.* (2018) 18:51–63. doi: 10.1038/nrc.2017.104
47. Elhamamsy AR, Metge BJ, Alsheikh HA, Shevde LA, Samant RS. Ribosome biogenesis: A central player in cancer metastasis and therapeutic resistance. *Cancer Res.* (2022) 82:2344–53. doi: 10.1158/0008-5472.CAN-21-4087