



Construction of a novel mRNAsi-related risk model for predicting prognosis and immunotherapy response in osteosarcoma

Zhe Li, Chi Jin, Xinchang Lu, Yan Zhang, Yi Zhang, Jia Wen, Yongkui Liu, Xiaoting Liu, Jiazhen Li

Department of Orthopaedics, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Contributions: (I) Conception and design: Z Li, J Li; (II) Administrative support: C Jin, X Lu, Yan Zhang; (III) Provision of study materials or patients: Yi Zhang, J Wen; (IV) Collection and assembly of data: Y Liu, X Liu; (V) Data analysis and interpretation: J Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Jiazhen Li. Department of Orthopaedics, The First Affiliated Hospital of Zhengzhou University, No. 1, Jianshe East Road, Erqi District, Zhengzhou 410100, China. Email: jzhli6411@163.com.

Background: Targeting cancer stem cells (CSC) may represent a future therapeutic direction for osteosarcoma (OS), which mainly relies on the identification of CSC markers. This study aimed to classify OS based on messenger ribonucleic acid (mRNA) stemness indices (mRNAsi) and construct a mRNAsi-related risk model to predict the prognosis of OS.

Methods: The one-class logistic regression (OCLR) algorithm was applied to the RNA-sequencing (seq) data of human embryonic stem cells (hESC) and induced pluripotent stem cell (iPSC) lines to calculate mRNAsi. Weighted gene co-expression network analysis (WGCNA) was performed on data obtained from the TARGET database to screen the mRNAsi-related genes. Univariate Cox regression analysis was implemented to screen mRNAsi-related genes with prognostic significance for consensus clustering of OS. The least absolute shrinkage and selection operator (LASSO) and COX regression analysis were conducted to construct a risk model based on mRNAsi-related genes.

Results: Six gene modules were identified in the TARGET database. The yellow module showed the strongest negative correlation with mRNAsi and the strongest significant positive correlation with the immune score and stromal score. OS was divided into three molecular subtypes with significant survival differences based on 73 mRNAsi-related genes with prognostic value for OS. The survival rate was ranked as C3 < C1 < C2 from low to high. The levels of immune components in C2 was significantly higher than those in C1 and C3. HSD11B2, GBP1, RNF130, APBB1IP, and NPC2 in the yellow module were used as variables for building the mRNAsi-related risk model. The survival rate of the high-risk group (as defined by this model) was significantly higher than that of the low-risk group, and it had significant survival prediction ability in 28 types of cancer. In addition, the mRNAsi-related risk model was superior to the Tumor Immune Dysfunction and Exclusion (TIDE) model in predicting the prognosis and immunotherapy response in all three immunotherapy cohorts.

Conclusions: This study classified OS and constructed a mRNAsi-related risk model based on mRNAsi-related genes, which provides a potential tool for more accurate risk stratification of OS and prediction of immunotherapy response.

Keywords: mRNA stemness indices; osteosarcoma (OS); tumor microenvironment; prognosis; immunotherapy response

Submitted Nov 15, 2022. Accepted for publication Jan 06, 2023. Published online Jan 31, 2023.

doi: 10.21037/atm-22-6011

View this article at: <https://dx.doi.org/10.21037/atm-22-6011>

Introduction

Osteosarcoma (OS) is a primary bone malignancy in which malignant mesenchymal cells differentiate into osteoblasts to produce malignant osteoid matrix (1), and is the fifth most common primary malignancy in adolescence (2). OS is most common in the metaphysis of the long bone and originates from the intramedullary cavity, and is accompanied by changes in bone texture and structural integrity. It is characterized by the up-regulation of osteoclast activity, resulting in increased bone resorption and repair and compensatory deposition of osteoid extracellular matrix by reactive osteoblasts (3). With complete surgical resection and multidrug chemotherapy, about 70% of patients with OS can become long-term survivors (4). The prognosis of patients with primary metastasis or recurrent disease is poor, and the 5-year survival rate has dropped sharply to 20% (4,5). The sum of observed side effects, coupled with many drug resistance phenomena, indicates the limitations of this treatment strategy and the need to develop new treatments (6).

The theory of tumor stem cells posits that tumors are maintained by stem cells, and the incomplete eradication of tumor stem cells is the cause of tumor drug resistance and recurrence (7). Targeting cancer stem cells (CSCs) may represent the future treatment direction of OS (8). Osteosarcoma CSCs have been identified as the main cause of recurrence and metastasis (9,10). The development of therapeutic strategies targeting tumor stem cells mainly depends on the identification of cell surface markers (11). Several CSC surface markers have been identified, such as cellular myelocytomatosis oncogene (c-Myc) (12), cluster

of differentiation (CD)133 (13), aldehyde dehydrogenase (ALDH) 1A (14), and so on. Clinically, the use of identified CSC markers is very limited. Also, stem cells from different tissues are not identical; for example, the differences in location, self-renewal, and differentiation are usually reflected in specific combinations of phenotypic markers (15). Most marker-labeled heterogeneous stem cell populations indicate that their characterization and isolation must be based on a combination of markers using multiple surface markers, and extracellular as well as intracellular markers (16). The expression profile analysis of stem cells is a method to identify the combination of CSC markers. mRNAsi is a cancer stem cell index that describes the degree of similarity between tumors and stem cells and can be considered as a quantification of cancer stem cells (17). Previous studies reported that messenger ribonucleic acid (mRNA) stemness indices (mRNAsi)-related genes signature could effectively predict prognosis for breast cancer (18), Colon Adenocarcinoma (19) and Gastric Cancer (20). However, it hadn't reported in OS.

In this study, expression profiles of stem cell human embryonic stem cells (hESC) and induced pluripotent stem cell (iPSC) lines were used to calculate messenger ribonucleic acid (mRNA) stemness indices (mRNAsi) defining stem cell properties via the one-class logistic regression (OCLR) machine-learning algorithm. Patients with OS were classified based on mRNAsi-related genes and a mRNAsi-related gene signature was developed, which provides some new evidence for targeted CSC therapy. We present the following article in accordance with the TRIPOD reporting checklist (available at <https://atm.amegroups.com/article/view/10.21037/atm-22-6011/rc>).

Highlight box

Key findings

- Classifier of osteosarcoma and mRNAsi-related risk model based on mRNAsi-related genes may be useful for accurately prognostic prediction in osteosarcoma.

What is known and what is new?

- CSCs have been confirmed as a subpopulation of tumor cells that can drive tumor initiation and can cause relapses.
- The impact of the CSC-related classifier and prognostic model for osteosarcoma has only been rarely studied.

What is the implication, and what should change now?

- Classifier of osteosarcoma and mRNAsi-related risk model may be applied clinically in osteosarcoma to help clinicians develop personalized treatments.

Methods

Data download and collation

OS samples (n=84) of primary solid tumors with both gene expression profiling and survival data were obtained from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET, <https://ocg.cancer.gov/programs/target>) public database, and Ensembl ID was converted into gene symbols. The GSE21257 and GSE39058 datasets were selected by logging into the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo>) database, and samples with both gene expression profile and survival data were selected, with 53 and 41 OS samples selected in the two datasets, respectively. We also

downloaded RNA-sequencing (seq) data of the hESC and iPSC lines from the Progenitor Cell Biology Consortium (PCBC) synapse portal, including a total of 78 cases. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Calculation of the mRNA stemness indices (mRNAsi)

The average value was used to centralize each OS sample, and through the ‘gelnet’ package in R, the OCLR machine learning algorithm was applied to hESC and iPSC in the PCBC data to calculate the weight vector of each gene. iPSC and hESC are true stem cells, which have strong differentiation and regeneration ability and can differentiate into various cell types needed by various organs and tissues of the human body. The correlation coefficients (ρ) of gene expression and gene weight vector of each sample in the TARGET database were calculated by Spearman correlation analysis, and the mRNAsi for each sample was obtained by linear transformation

Weighted gene co-expression network analysis (WGCNA)

WGCNA of the sample data in the TARGET dataset was carried out by using the ‘WGCNA’ package in R. Firstly, the Pearson correlation coefficient between paired genes was calculated and a computational critical matrix was constructed. Distinct values of soft thresholding power (β) were evaluated for the network topology analysis, in which the minimum β , which accords with the distribution of scale-free networks, was used to construct the co-expression networks. The weighted adjacency matrix was transformed into a topological overlap metric (TOM) matrix to measure the network connectivity of genes. According to the average linkage hierarchical clustering based on the difference measurement of the TOM, genes with a similar co-expression were divided into modules, and genes lacking similar co-expression with other genes in the network were assigned to the grey module. The key modules between the modules and mRNAsi and the immune score were selected by Pearson correlation analysis.

Mining of mRNAsi-related genes and identification of the mRNAsi-related molecular subtypes

After selecting the key modules that were significantly related to mRNAsi, we performed univariate Cox regression analysis on the genes in the modules to screen the genes

related to OS prognosis, and their transcriptional profiles in TARGET were imported into the ‘ConsensusClusterPlus’ package for consensus clustering. The optimal number of clusters was determined according to the cumulative distribution function (CDF) and the relative change in the area under the increase of the CDF.

Difference analysis and functional enrichment of mRNAsi-related molecular subtypes

The differential expression of every two molecular subtypes was analyzed in the ‘Limma’ package, and genes with significant differences were evaluated by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis using the ‘WebGestaltR’ package to determine the functional differences among molecular subtypes at the molecular level. In addition, the Hallmark pathway was downloaded and Gene Set Enrichment Analysis (GSEA) was performed between every two clusters to evaluate the significant differences in function between every two molecular subtypes by generating the normalized enrichment score (NES).

Analysis of the immune infiltration characteristics of OS

The stromal, immune, and ESTIMATE scores were measured for each molecular subtype by importing the RNA-seq data from OS into ‘ESTIMATE’ (21), which is a tool that assesses stromal and immune cell infiltration of major non-tumor components in tumor samples. The Microenvironment Cell Populations-counter (MCP-counter) method (22) quantified the abundance of eight immune cell groups and two stromal cell groups in different molecular subtypes according to transcriptome data. We employed the ‘GSVA’ R package (23) to perform single sample Gene Set Enrichment Analysis (ssGSEA) based on 29 immune-associated gene sets (24) and quantified immune cell infiltration by calculating the ssGSEA score.

Prediction of the immunotherapy response and drug sensitivity analysis

We obtained the information on 20 immune checkpoints from the HisgAtlas database (25), which has collected 995 human immunosuppressive genes. According to the data of the OS samples in TARGET, the expression levels of 20 immune checkpoints in each molecular subtype were analyzed. Tumor Immune Dysfunction and Exclusion

(TIDE) integrates the expression signatures of T-cell dysfunction and exclusion signatures in tumors, as well as that of three cell types reported to restrict T-cell infiltration in tumors, including cancer-associated fibroblasts (CAFs), myeloid-derived suppressor cells (MDSCs), and the M2 subtype of tumor-associated macrophages (TAMs) (26). We calculated the TIDE scores using the TIDE algorithm. Ridge regression was performed using the 'pRRophetic' algorithm, to predict the chemotherapy response of OS. Internal cross-validation was performed using 10-fold cross-validation based on the Genomics of Drug Sensitivity in Cancer (GDSC).

Establishment of the mRNAsi-related risk model

According to the 1:1 random allocation principle, the samples in TARGET were divided into training and verification cohorts. We conducted univariate Cox regression analysis on the training cohort of TARGET for genes in key mRNAsi-related modules screened by WGCNA. Genes with $P < 0.01$ were screened and included in the 'glmnet' package of R for conducting the least absolute shrinkage and selection operator (LASSO) Cox regression analysis. Multivariate Cox regression analysis was performed using the 'MASS' package to reduce the number of genes screened by LASSO Cox and to calculate the predictive value of simulated genes. The product of the gene expression level and multivariate Cox regression coefficient were added to form a risk model. After calculating the risk score, the samples in all of the cohorts were classified into high- and low-risk groups using the 'survminer' package. The 'survival' and 'timeROC' packages were adopted to analyze survival and draw the time-dependent receiver operating characteristic (ROC) curve.

Evaluation of the mRNAsi-related risk model in predicting the prognosis of pan-cancer

The expression profiles and survival information of 33 cancer types were downloaded from The Cancer Genome Atlas (TCGA). The mRNAsi-related risk model was applied to the samples of each cancer type to calculate the risk scores. The samples were grouped using the same method as in the TARGET training set, and the survival differences between the risk groups were calculated by the Kaplan-Meier survival curve.

Statistical analysis

All data in this study were statistically analyzed by the R programming language (27). The statistical significance of normally distributed variables between two groups was assessed using the unpaired *t*-test, and a comparison of the non-normally distributed variables was performed using the Wilcoxon rank-sum test. If the P value representing the statistical difference was not specified, then the default $P < 0.05$ was considered statistically significant.

Results

Clinical and immunological correlation of mRNAsi

In the TARGET dataset, we calculated the mRNAsi of the samples and sorted them in ascending order. The change in mRNAsi was also accompanied by changes in the stromal, immune, and ESTIMATE scores; higher mRNAsi corresponded to lower stromal, immune, and ESTIMATE scores (*Figure 1A*). There was no significant correlation between mRNAsi and age, gender, and the degree of metastasis (*Figure 1B*). The Pearson correlation test showed that mRNAsi was negatively correlated with the stromal, immune, and ESTIMATE scores (*Figure 1C*). mRNAsi was also significantly negatively correlated with several kinds of immune cells; for instance, central memory cluster of differentiation (CD)4 T cells, central memory CD8 T cells, effector memory CD8 T cells, regulatory T cells, T follicular helper cells, activated dendritic cells, CD56 dim natural killer cells, and plasmacytoid dendritic cells (*Figure S1*).

Identification of the mRNAsi-related gene module

WGCNA was performed on TARGET data to identify the gene modules related to mRNAsi. We found that the minimum β in accordance with the distribution of the scale-free networks was six (*Figure 2A*). A total of six gene modules were detected in the dynamic cutting tree (*Figure 2B*). In the eigengene dendrogram and eigengene adjacency heatmap, we observed that the yellow module had the strongest statistically significant correlation with mRNAsi and the immune, stromal, and ESTIMATE scores, with the highest correlation coefficient absolute value. This module was significantly negatively correlated with mRNAsi and positively correlated with immune, stromal, and ESTIMATE scores (*Figure 2C*). By analyzing the signaling pathways and

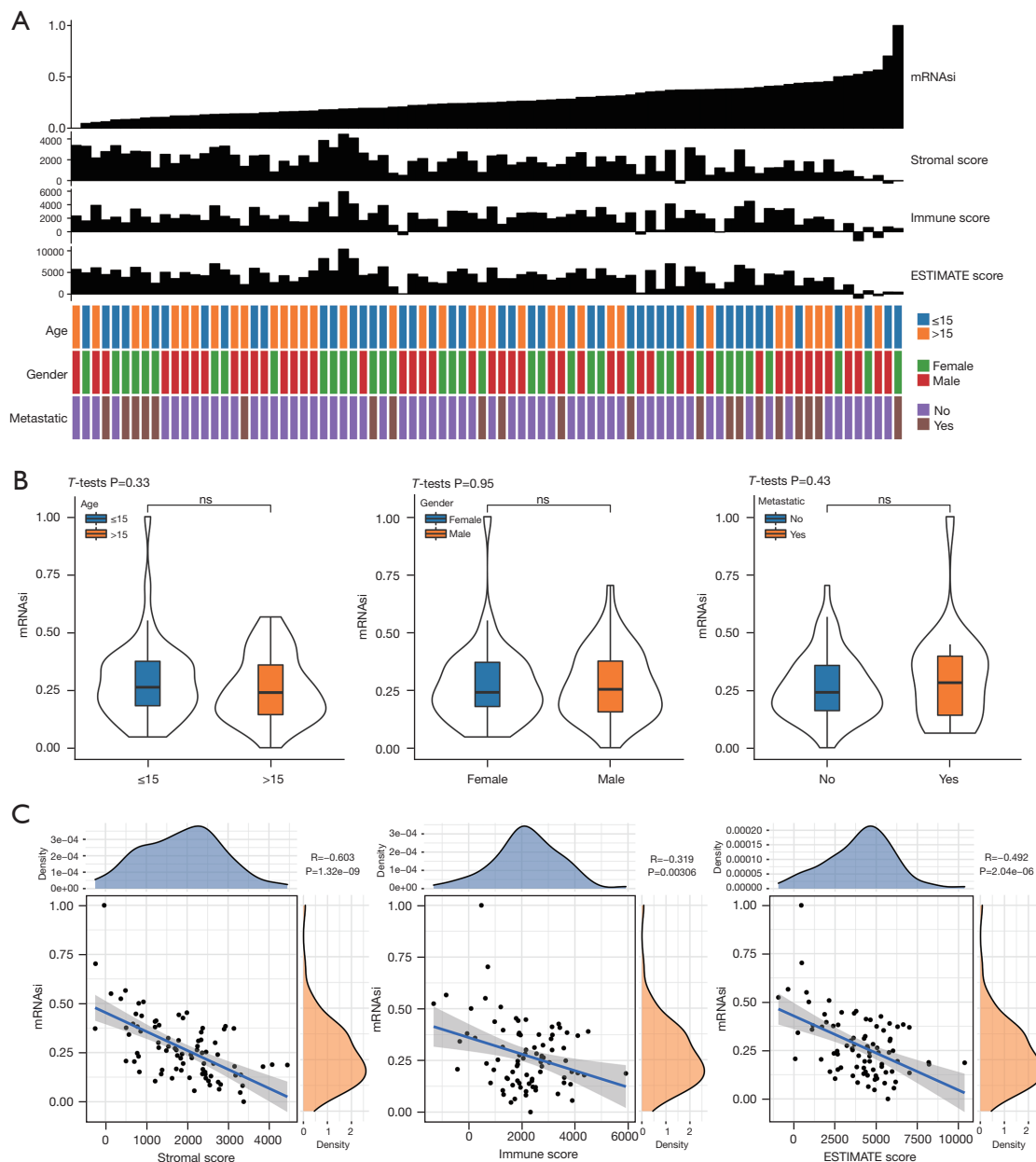


Figure 1 Clinical and immunological correlation of mRNAsi. (A) The corresponding stromal, immune, ESTIMATE scores and clinical features in ascending order of mRNAsi in TARGET; (B) the level of mRNAsi between the samples grouped according to age, gender, and metastatic status; (C) the Pearson correlations analysis between mRNAsi and the stromal, immune, and ESTIMATE scores. ns, no significance; mRNAsi, mRNA stemness indices.

biological processes enriched by the genes in the yellow module, we learned that they were mainly significantly related to many KEGG pathways, molecular functions (MF), cellular components (CC), and biological processes (BP) that dominate the immune response (Figure 2D).

Identification of three molecular subtypes of OS based on the mRNAsi-related genes

A total of 1,313 genes were gathered in the yellow module. Seventy-three genes that were significantly associated with OS prognosis were obtained by univariate Cox regression

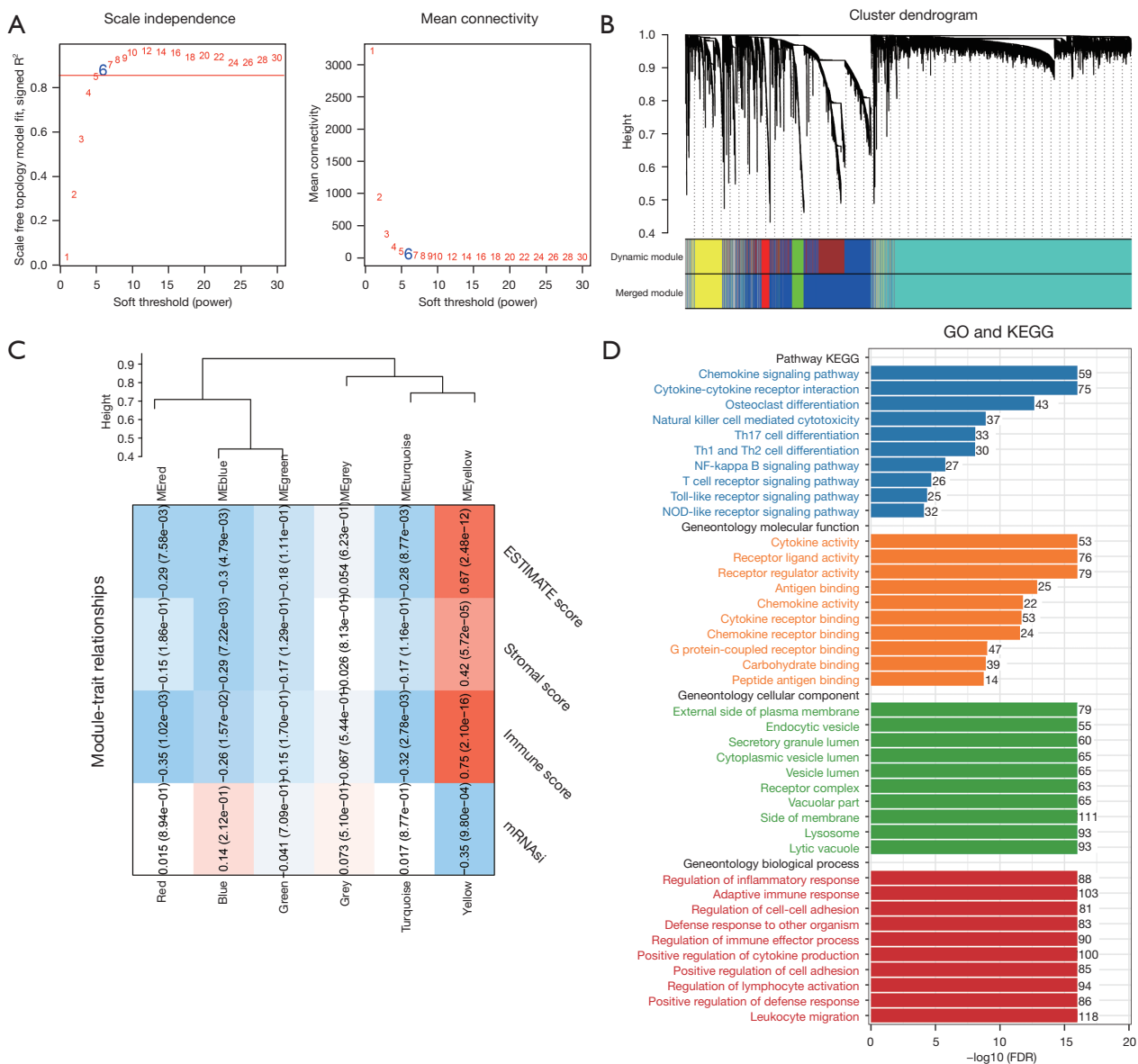


Figure 2 Identification of the mRNAsi-related gene module. (A) The scale-free topology model fit index (left) and the mean connectivity (right) for each soft-thresholding power; (B) the clustering dendrogram of the weighted gene co-expression network; (C) the module-trait relationship; (D) KEGG pathway and GO terms enriched by genes in the yellow module. mRNAsi, mRNA stemness indices; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology.

analysis based on the expression of these genes in TARGET, including seven genes with a hazard ratio (HR) >1 and 66 genes with a HR <1 (Figure 3A). Pearson correlation analysis showed that most of the 73 genes were positively correlated with each other (Figure 3B). During consensus clustering, we clustered and ranked the samples by applying the clustering variable (k) from 2 to 10. According to CDF

and the CDF Delta area curve, when k=4, the area under the curve increases the least. Since the CDF Delta area curve showed the relative change of the area under the curve between k and K-1, 3 was the most appropriate k value (Figure 3C,3D). Also, the clustering heatmap showed that when k=3, both the inter-group separation and intra-group aggregation exhibited obvious trends (Figure 3E).

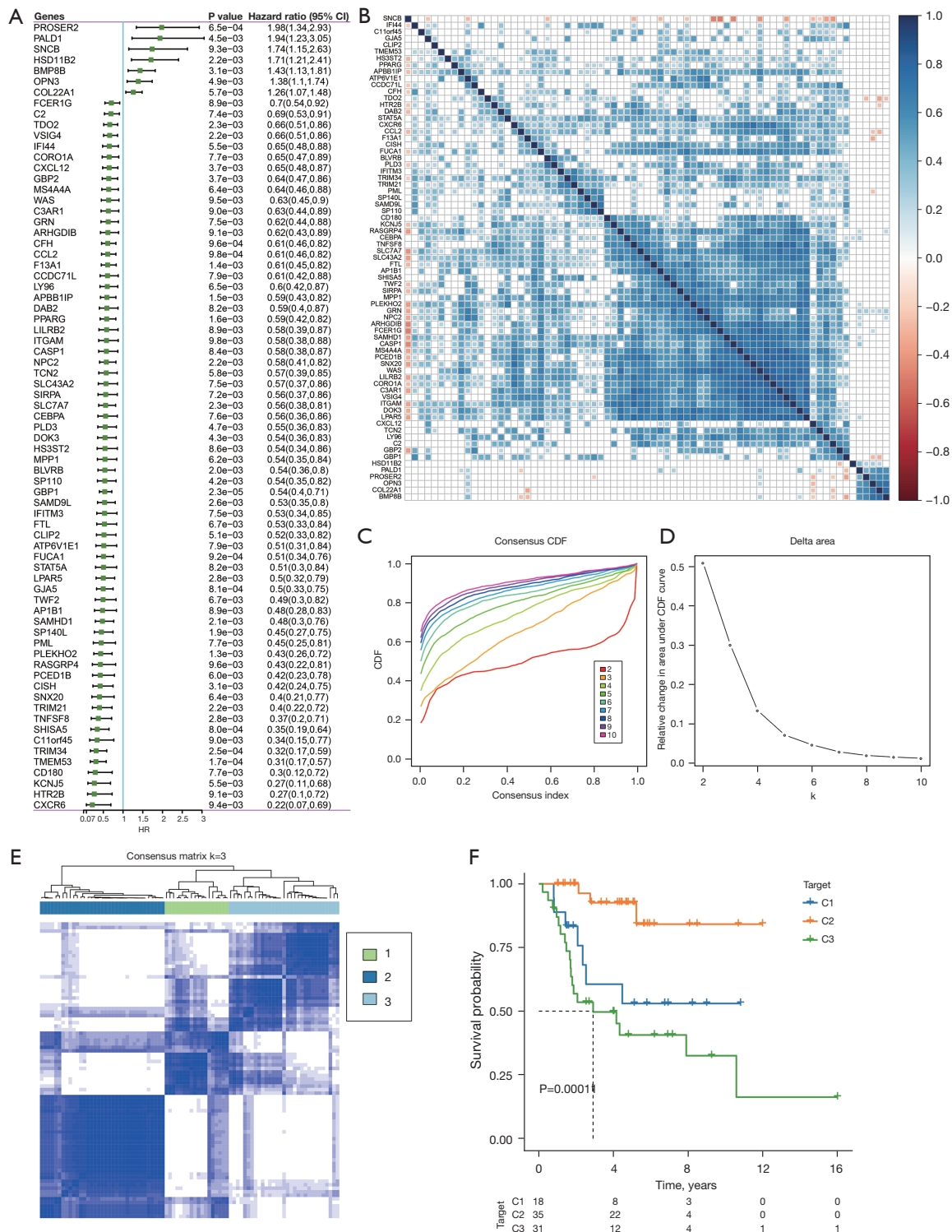


Figure 3 Identification of three molecular subtypes of OS based on mRNAsi-related genes. (A) The HRs and P values of 73 genes that were significantly related to OS prognosis; (B) the heatmap displays the correlation between 73 genes; (C) CDF curves for k from 2 to 10; (D) the relative change in area under the CDF curve for k from 2 to 10; (E) heatmap of clustering at k=3; (F) the Kaplan-Meier curve shows the difference in survival among the three clusters. OS, osteosarcoma; mRNAsi, mRNA stemness indices; HR, hazards ratio; CDF, cumulative distribution function.

Therefore, the OS sample was divided into three clusters. The survival differences among the three clusters were analyzed, and a significant difference in survival rate among clusters was detected, which was ranked as C3 < C1 < C2 according to the survival rate from low to high (Figure 3F).

Differences in the regulatory pathways among the three clusters

By analyzing the differences between every two molecular subtypes, 807 differentially expressed genes (DEGs) were identified between C1 and C2, 552 DEGs were identified between C3 and C2, and 1,089 DEGs were identified between C3 and C1 (Figure S2A-S2C). The biological function enrichment of each type of DEG was then analyzed. The top three KEGG pathways related to DEGs between C1 and C2 were *Staphylococcus aureus* infection, phagosome, tuberculosis, and these DEGs were also significantly correlated with antigen binding, peptide antigen binding, immunoglobulin binding, granulocyte activation, leukocyte degranulation, myeloid cell activation involved in the immune response, and other GO terms (Figure 4A). The DEGs between C3 and C2 were significantly enriched in *Staphylococcus aureus* infection, phagosome, hematopoietic cell lineage KEGG pathways, receptor-ligand activity, receptor regulator activity, cytokine activity, T-cell activation, regulation of cell-cell adhesion, activation of the immune response, and other GO terms (Figure 4B). The DEGs-enriched KEGG pathway and GO terms between C3 and C1 coincided with the above pathways and biological processes (Figure 4C).

The molecular pathways of differential enrichment between subtypes were analyzed by GSEA. The results showed that compared with C2, the pathways controlling the cell cycle and growth in C1 were significantly activated, while epithelial-mesenchymal transition (EMT) and apoptosis and immunoreactive pathways were significantly inhibited (Figure 4D). The pathways that were significantly activated by C3 relative to C2 included Early 2 factor (E2F) targets, oxidative phosphorylation, myelocytomatosis oncogene (MYC) targets, G2M checkpoint, and so on. Most of the pathways that were significantly activated in C1 compared with C2 were also significantly activated in C3, such as interferon alpha/gamma response, Interleukin (IL)6-Janus kinases (JAK)-signal transducer and activator of transcription (STAT) signaling, EMT, apoptosis, allograft rejection, Kirsten rat sarcoma virus (KRAS) signaling, apical junction, complement, and coagulation (Figure 4E).

The activities of C3 oxidative phosphorylation, interferon alpha/gamma response, coagulation, allograft rejection, and complement were significantly improved, while G2M checkpoint, E2F targets and mitotic spindle, and EMT were significantly inhibited compared with C1 (Figure 4F).

Tumor microenvironment (TME) status of the three mRNAsi-related clusters

The above results indicated that mRNAsi was significantly associated with the stromal and immune scores, and the genes belonging to the yellow module, which were divided OS into three clusters, were also significantly associated with these scores. We further explored the TME status differences in the three clusters. SsGSEA quantified the scores of 29 immune cell groups, compared these immune cell scores between the clusters, and detected a significant difference in 26 immune cell scores among the three clusters. The abundance of most tumor suppressor immune cells, such as activated B cells, activated CD8⁺ T cells, central memory CD4⁺ T cells, central memory CD8⁺ T cells, immature B cells, macrophages, natural killer cells, and so on, was the highest in C2 (Figure 5A).

MCP-counter quantified the levels of eight immune cell groups and two stromal cell groups in the three mRNAsi-related clusters. The T-cell, monocytic lymphocyte, and fibroblast scores in C2 were significantly higher than those in C1 and C3 (Figure 5B). Moreover, there were significant differences in the stromal, immune, and ESTIMATE scores among the three mRNAsi-related clusters, which were always the highest in C2 (Figure 5C).

Sensitivity of the three mRNAsi-related clusters to immunotherapy and tumor inhibitors

We analyzed whether immunotherapy and tumor inhibitors were effective in the three mRNAsi-related clusters. At present, immunotherapy mainly includes immune checkpoint blockade (ICB) and chimeric antigen receptor (CAR) T-cell therapies. Considering the significant challenges faced by CAR-T therapy in solid tumors, no CAR-T therapy has yet been approved for these tumors (28). Therefore, we mainly explored the response of the three mRNAsi-related clusters to ICB.

Firstly, we compared the expression differences in the immune checkpoints among the three mRNAsi-related clusters. Among the 20 immune checkpoints analyzed, more than half showed significant expression differences among

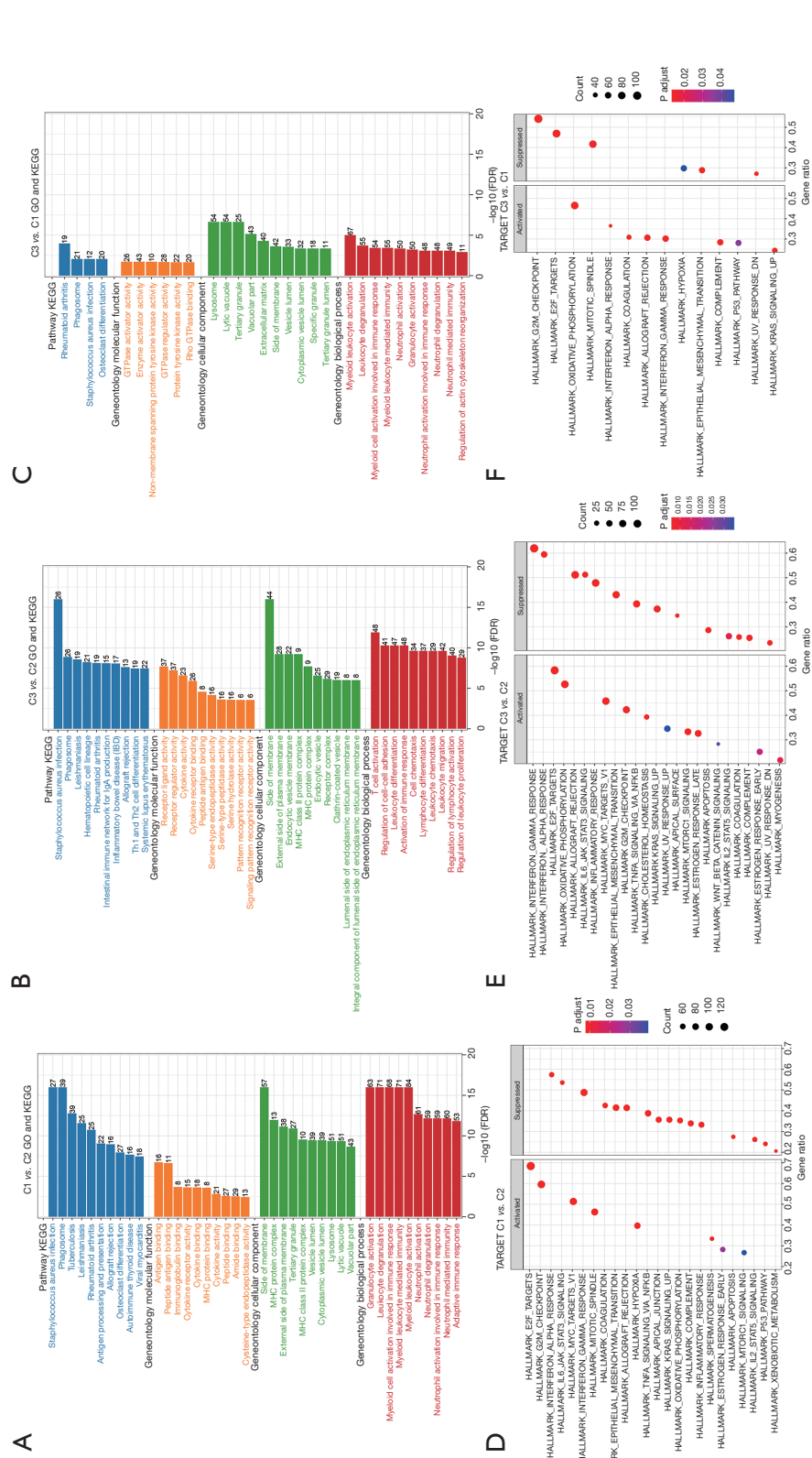


Figure 4 Differences in the regulatory pathways among the three clusters. (A) Bar plot showing KEGG pathways and GO terms with significantly enriched DEGs between C1 and C2; (B) KEGG pathways and GO terms annotated by the DEGs between C3 and C2; (C) KEGG pathways and GO terms that were significantly related to DEGs between C3 and C1; (D) the bubble plot showing the pathways in which C1 was significantly up-regulated and down-regulated compared with C2; (E) significantly activated and inhibited pathways in C3 compared with C2; (F) the signaling pathway activity in C3 versus C1. The left frame bubble chart represents the significantly activated pathway, and the right frame bubble map refers to the significantly suppressed pathway. KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology; DEGs, differentially expressed genes.

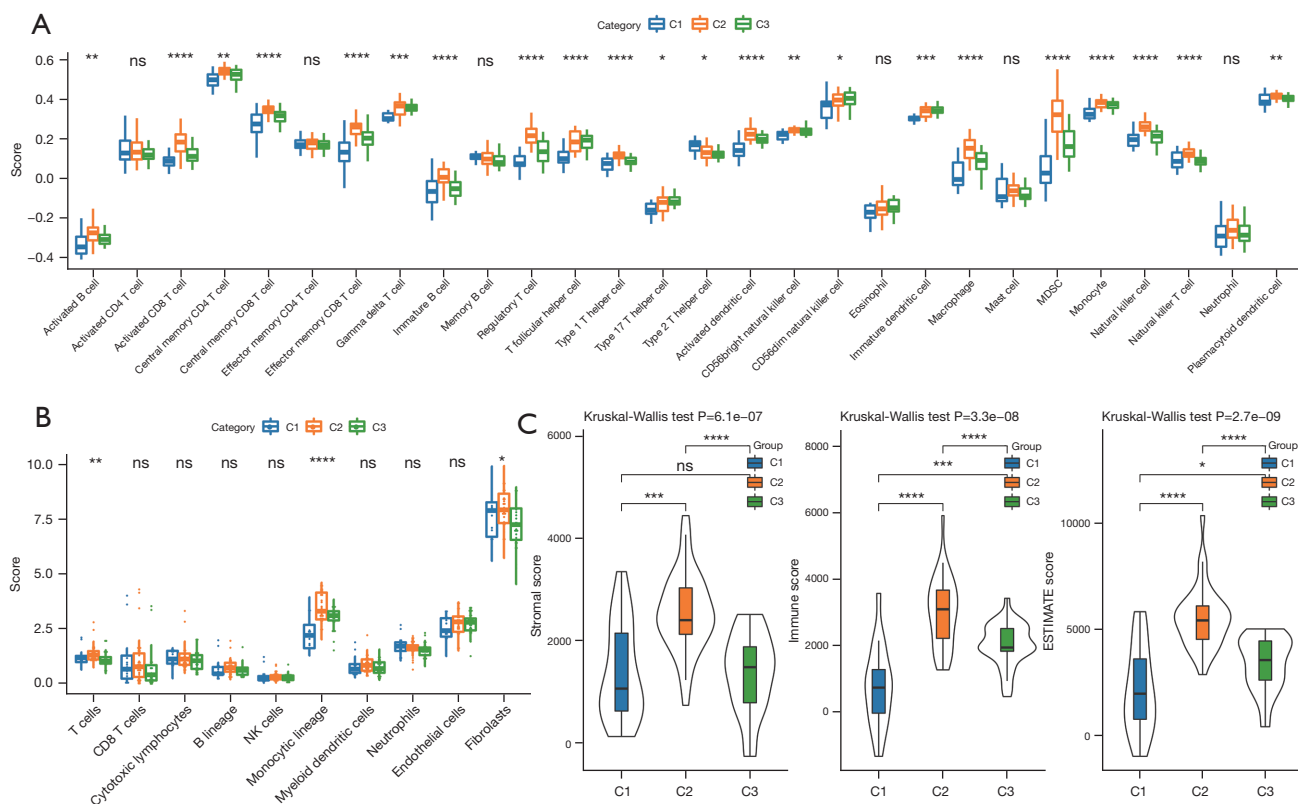


Figure 5 TME status of the three mRNAsi-related clusters. (A) Differences in the scores of 29 immune cell populations between the three mRNAsi-related clusters; (B) differences in the levels of eight immune cell populations and two stromal cell populations among the three mRNAsi-associated clusters; (C) the stromal, immune, and ESTIMATE score differences among the three mRNAsi-related clusters. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, no significance. TME, tumor microenvironment; mRNAsi, mRNA stemness indices.

the three subtypes (Figure 6A). The T-cell dysfunction, T-cell exclusion, TIDE, as well as MDSC and M2 TAM scores calculated by TIDE exhibited significant differences among the three subgroups, indicating that the responses of the three clusters to immunotherapy were different (Figure 6B).

The sensitivity of tumor inhibitors erlotinib, rapamycin, pyrimethamine, bortezomib, roscovitine, and midostaurin in three mRNAsi-related subtypes was evaluated based on the estimated inhibitory concentration (IC₅₀) value. Our comparison showed that Erlotinib, Rapamycin, Pyrimethamine, Bortezomib, and Roscovitine were all significantly different in terms of the estimated IC₅₀ values among the three mRNAsi-related subtypes. Erlotinib, Rapamycin, Pyrimethamine, and Roscovitine showed higher sensitivity in C3, while Bortezomib had greater resistance in C3 (Figure 6C).

Development and validation of a mRNAsi-related risk model

In addition to screening genes from the yellow module to cluster OS, we also used this module to screen the most valuable genes to build a risk model. The risk model was constructed using half of the samples in the TARGET database. Univariate Cox regression analysis of the genes in the yellow module in the TARGET training cohort showed that 20 genes were significantly associated with OS prognosis. Seven genes were screened at $\lambda=5$ by LASSO regression (Figure S3A). Following multivariate Cox regression, only five of these genes remained as variables for constructing a risk model (Figure S3B).

The risk model was calculated as follows: risk score = $-0.683 \times \text{Ring finger protein 130 (RNF130)} - 0.607 \times$

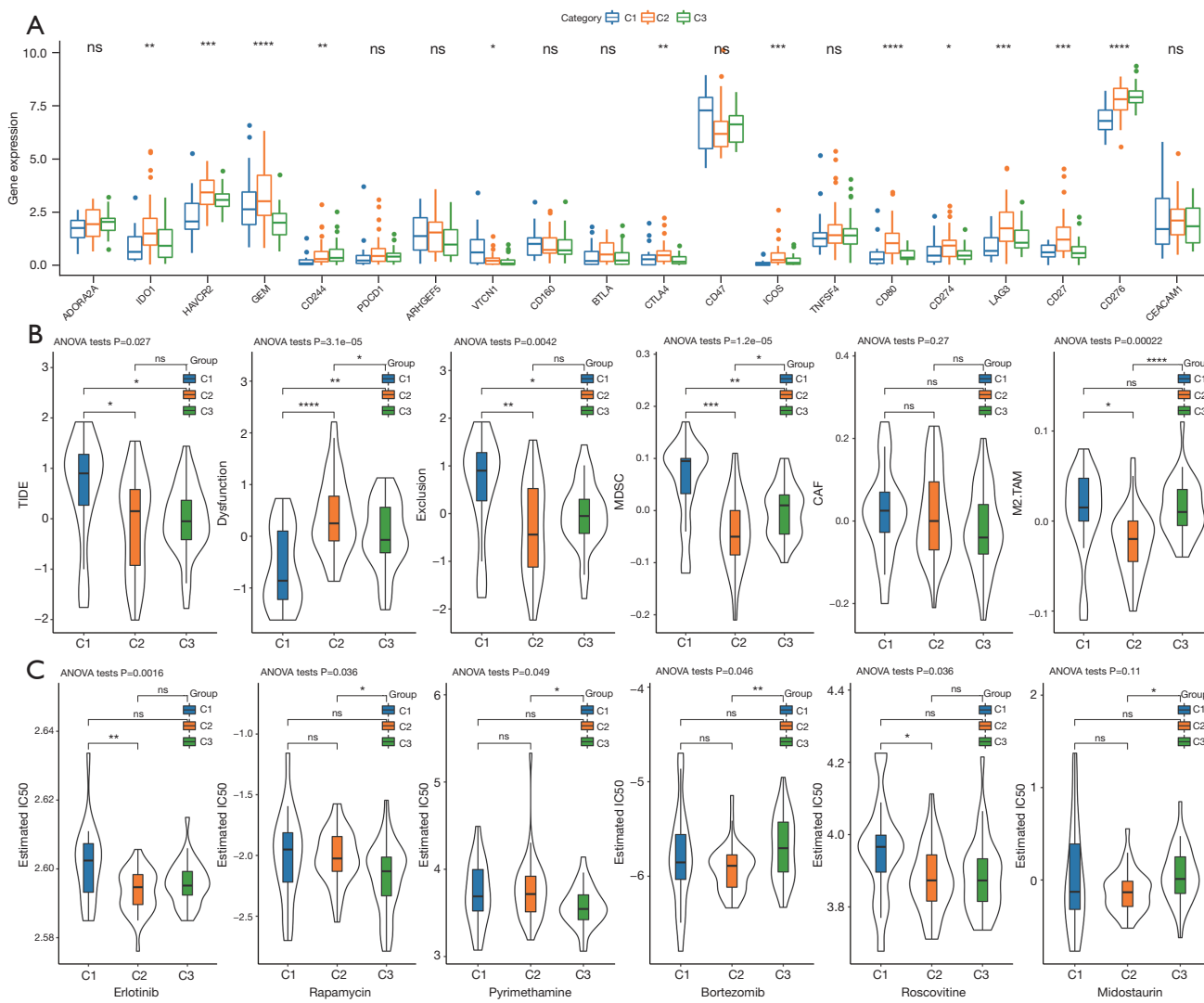


Figure 6 Sensitivity of the three mRNAsi-related clusters to immunotherapy and tumor inhibitors. (A) Differences in the expressions of the three immune checkpoints among the three mRNAsi-related clusters; (B) differences in the T-cell dysfunction, T-cell exclusion, TIDE, MDSC, CAF, and M2.TAM scores among the three mRNAsi-related clusters; (C) the estimated IC50 values of Erlotinib, Rapamycin, Pyrimethamine, Bortezomib, Roscovitine and Midostaurin in the three mRNAsi-related subtypes. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, no significance. mRNAsi, mRNA stemness indices; TIDE, Tumor Immune Dysfunction and Exclusion; MDSC, myeloid-derived suppressor cell; CAF, cancer-associated fibroblast; TAM, tumor-associated macrophage.

guanylate binding protein 1 (GBP1) – $0.808 \times$ amyloid beta (A4) precursor protein-binding, family B, member 1 interacting protein (APBB1IP) + $0.661 \times$ human 11beta-hydroxysteroid dehydrogenase type 2 (HSD11B2) – $0.814 \times$ Niemann-Pick C2 (NPC2). This formula was then introduced into two cohorts of the TARGET database, an unsplit TARGET cohort, and two external verification cohorts (GSE21257 and GSE39058) to calculate the risk score of the sample. The risk groups were divided using the

‘survminer’ package.

Significant differences between the two risk groups in terms of survival rates were detected in each cohort. Furthermore, the 5-year survival prediction results were satisfactory. The areas under the ROC curve (AUC) were ≥ 0.73 . The 5-year AUC of TARGET training, TARGET verification, unsplit TARGET, and GSE21257 and GSE39058 cohorts were 0.92, 0.84, 0.87, 0.73, and 0.76, respectively (Figure 7A-7E).

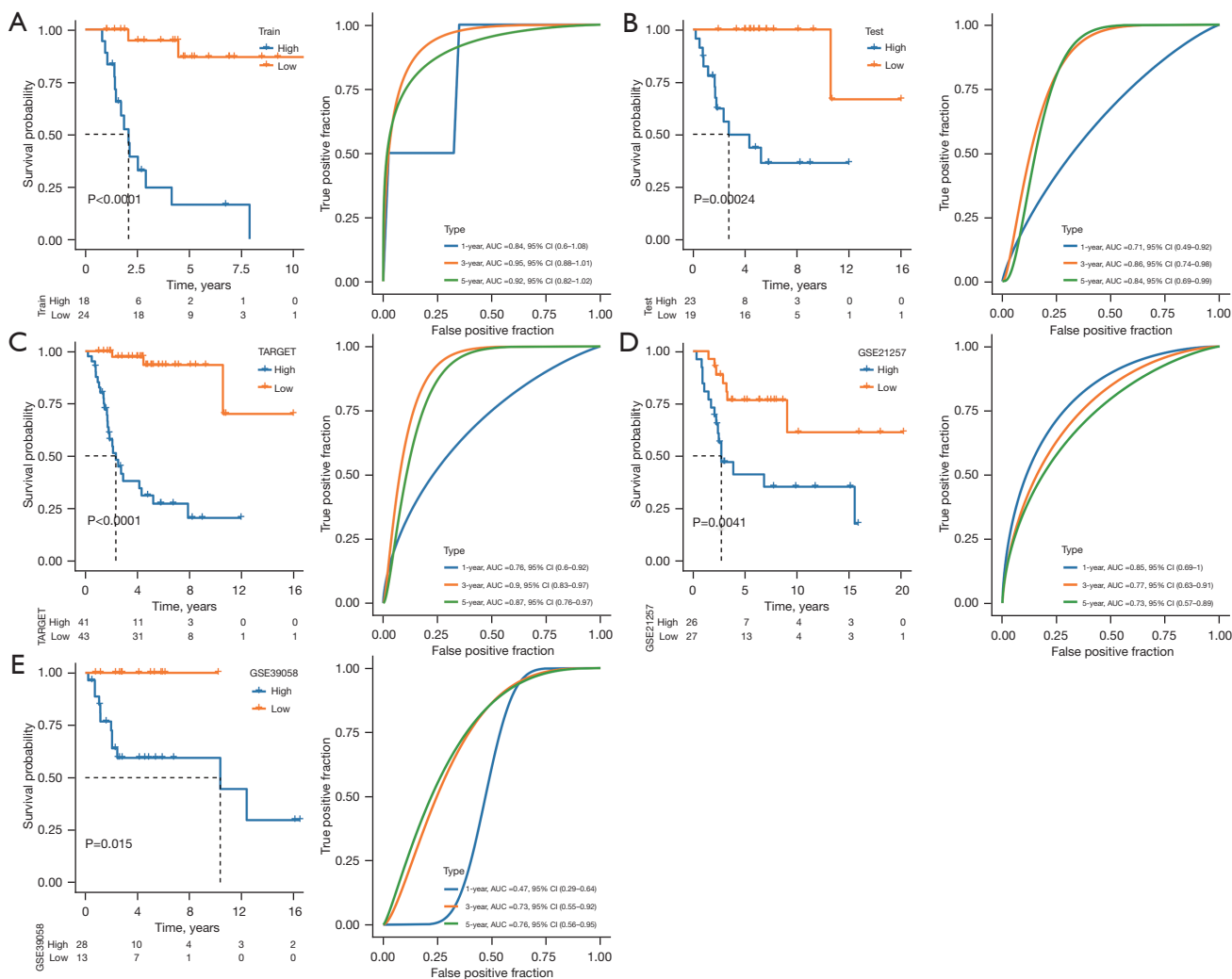


Figure 7 Development and validation of a mRNAsi-related risk model. (A-E) Risk score-based the survival and ROC curves in the TARGET training cohort (A), TARGET authentication cohort (B), unsplit TARGET cohort (C), and GSE21257 (D) and GSE39058 (E) cohorts. mRNAsi, mRNA stemness indices; ROC, receiver operating characteristic.

Prognostic effect of the mRNAsi-related risk model in pan-cancer

The risk score of the sample in each cancer type was calculated by introducing the mRNAsi-related risk model into 33 cancer types. At the same time, the ‘survminer’ package was used to split the risk groups of samples from each type of cancer dataset. Through survival analysis, we found that the mRNAsi-related risk model could significantly distinguish the survival rates of 28 types of cancer samples, and the survival time of the high-risk group was always markedly longer than that of the low-risk group (Figure 8).

Comparison of the mRNAsi-related risk and TIDE scores in predicting prognosis and immunotherapy response

Finally, we also investigated and compared the applicability of the mRNAsi-related risk and TIDE scores in predicting prognosis and immunotherapy response. Data from the IMvigort210, GSE91061, and GSE135222 immunotherapy cohorts were downloaded from the ‘IMvigort210CoreBiology’ R package and the GEO database, respectively. The signatures obtained by the mRNAsi-related risk model and TIDE were substituted into each immunotherapy cohort to calculate the risk and TIDE scores of the sample.

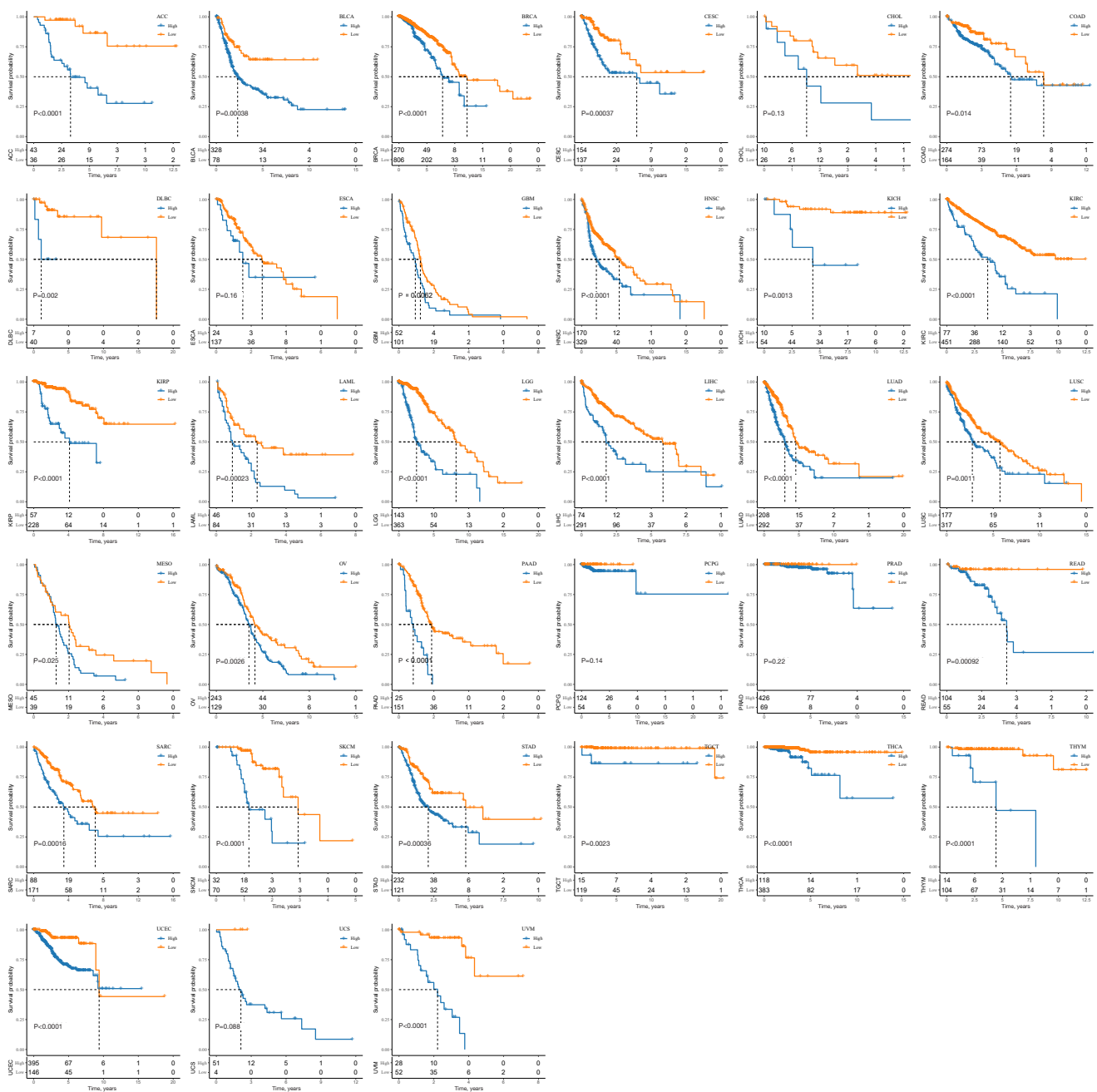


Figure 8 Prognostic effect of the mRNasi-related risk model in pan-cancer patients. mRNasi, mRNA stemness indices.

In the IMvigor210 cohort, both the risk and TIDE scores were significantly correlated with the prognosis of the sample. Also, the AUC of the risk model for predicting survival was higher than that of the TIDE model at 0.5, 1, and 1.5 years (Figure 9A,9B). Moreover, the AUC of the mRNasi-related risk model for evaluating the immunotherapy response of the samples in the IMvigor210

cohort was also significantly higher than that of the TIDE model (Figure 9C). In the GSE91061 cohort, the mRNasi-related risk model could also markedly distinguish the differences in sample survival, and the AUC that predicted 1-, 2-, 2.5-year survival was 0.6, 0.62 and 0.66 (Figure 9D). IDE model could not significantly distinguish the survival differences between two groups in GSE91061 dataset

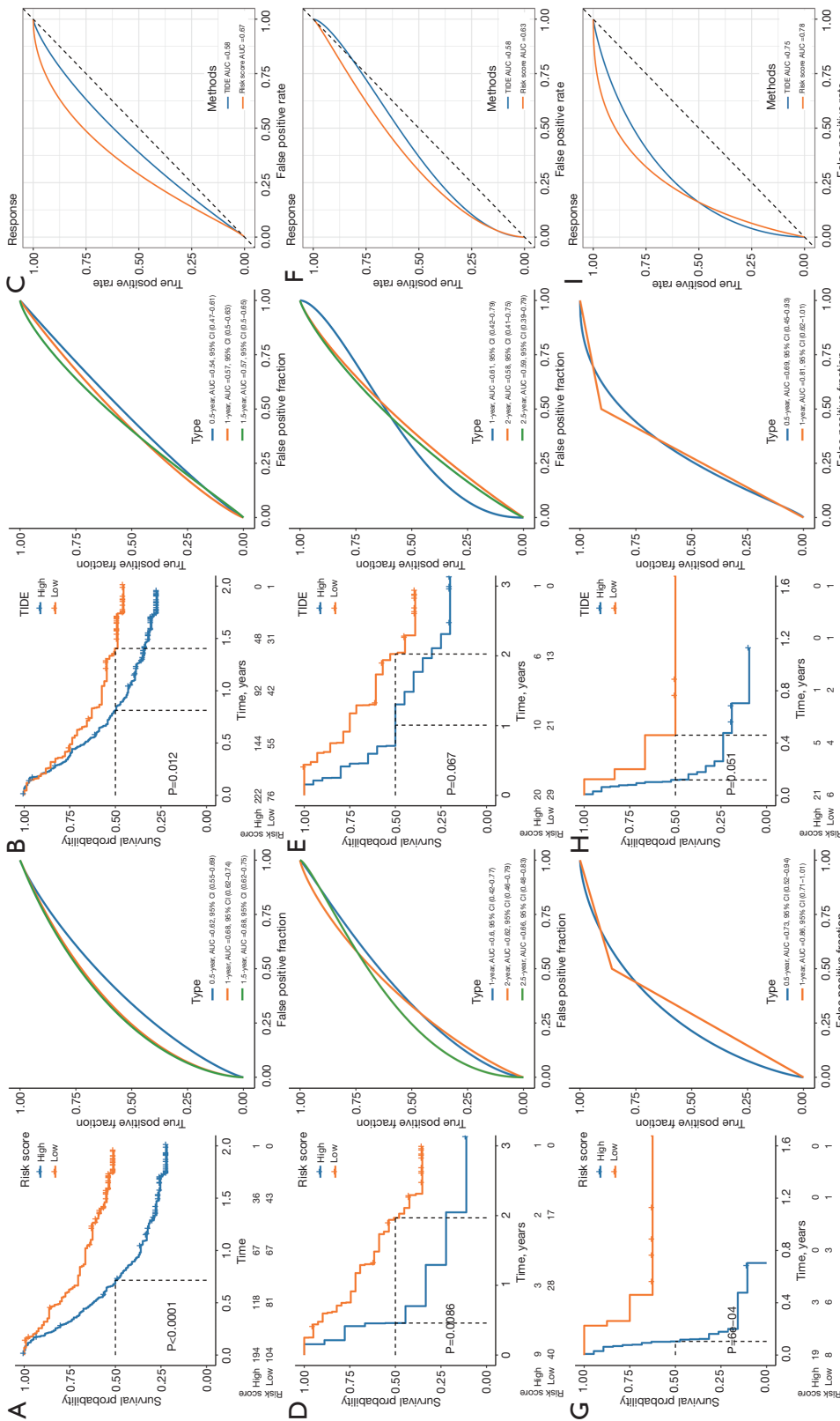


Figure 9 Comparison of the mRNAasi-related risk and TIDE scores in predicting prognosis and immunotherapy response. (A,B) Kaplan–Meier and ROC curves of the mRNAasi-related risk model (A) and TIDE model (B) for predicting survival in the IMvigor210 cohort. (C) The mRNAasi-related risk model and TIDE models evaluated the AUCs of immunotherapy response of samples in the IMvigor210 cohort. (D,E) Kaplan–Meier and ROC curves of the mRNAasi-related risk model (D) and TIDE model (E) for predicting survival in the GSE91061 cohort. (F) AUCs of the mRNAasi-related risk model and TIDE model to assess immunotherapy response in the GSE91061 cohort. (G–H) Kaplan–Meier and ROC curves of the mRNAasi-related risk model (G) and TIDE model (H) for predicting survival in the GSE135222 cohort. (I) AUCs of the mRNAasi-related risk and TIDE models to assess the immunotherapy response in the GSE135222 cohort. mRNAasi, mRNA stemness indices; TIDE, Tumor Immune Dysfunction and Exclusion; ROC, receiver operating characteristic; AUC, area under the curve.

(Figure 9E). The AUC of TIDE was 0.58, which lower than 0.63 of risk model (Figure 9F). In the GSE135222 cohort, the risk score could predict survival, and the 1-year AUC reaching 0.86 (Figure 9G). However, the TIDE model could not significantly distinguish the survival differences between samples in these two cohorts (Figure 9H). Regarding the prediction of immunotherapy response, the AUCs of the mRNAsi-related risk model in the GSE135222 cohorts were also higher than that of the TIDE model (Figure 9I).

Discussion

CSC surface markers are more difficult to identify in OS originating from mesenchymal cells than in tumors originating from other tissue types. The analysis of histopathological specimens also indicates that in some cases, tumors are organized in a stratified manner, and the leading CSC is the producer of phenotypic and functional heterogeneity (15). mRNAsi is a cancer stem cell index that describes the degree of similarity between tumors and stem cells and can be considered as a quantification of cancer stem cells (17). In the present study, we employed the OCLR algorithm to hESC and iPSC samples and defined stemness by calculating mRNAsi. There was a significant negative correlation between mRNAsi and immune activity in OS, which was consistent with the previously reported trend in non-small cell lung cancer (29). Then, we identified the module with the strongest negative correlation with mRNAsi and the highest positive correlation with the immune score by WGCNA and screened the mRNAsi-related genes that were significantly related to OS prognosis from the module. Based on the expressions of the mRNAsi-related genes, we identified the molecular subtypes and risk models in OS and indirectly indicated the importance of stemness in OS. There are three main differences between our study and others that have been published recently. Firstly, the analysis was conducted based on mRNAsi for the first time. We then propose that the unique treatment of osteosarcoma CSC can be addressed by focusing on CSC-related genes. Finally, the prognostic model was further constructed based on the differences of tumor stem cell subtypes.

Specifically, the yellow module (divided by WGCNA) had the strongest negative correlation with mRNAsi and the highest positive correlation with the immune score. From this module, we identified 73 genes that were significantly related to the prognosis of OS and divided OS into three subgroups according to their consensus

clustering expression. In terms of clinical and biological characteristics, the survival rate of C2 was the highest among the three subgroups, and this subgroup also exhibited a markedly active TME status compared with C1 and C3; for example, it had the highest abundance of activated B cells, activated CD8⁺ T cells, central memory CD4⁺ T cells, central memory CD8⁺ T cells, immature B cells, macrophages, and natural killer cells, as well as the highest stromal and immune scores. Available evidence suggests that a large number of these immune components are associated with anti-tumor responses and patient outcomes. Tumor-infiltrating B cells inhibit tumor progression by secreting immunoglobulin, promoting the T-cell response, and directly killing cancer cells, which is beneficial for the prognosis of tumor patients (30). Activated CD8⁺ T cells induce the apoptosis of mesenchymal tumor stromal cells and inhibit tumor invasion and metastasis by releasing extracellular vesicles (31). Central memory CD8⁺ T cells have been shown to exert strong antitumor activity and have been associated with a better prognosis in cancer patients (32). Central memory CD4⁺ T cells have also been proposed as an independent prognostic biomarker of oral squamous cell carcinoma (33). Natural killer cells not only exert antitumor effects via direct cytolytic activity but also indirectly exert antitumor effects through cytokine production (34). These findings provide evidence for the positive clinical outcomes of C2.

In the process of building the risk model, we took half of the samples in the TARGET as the training cohort and the other half as the verification cohort. Of course, the performance of the risk model was also tested on the unsplit TARGET-OS dataset and two other independent validation cohorts, GSE21257 and GSE39058. The risk model is composed of five mRNAsi, including HSD11B2, GBP1, RNF130, APBB1IP, and NPC2. It has been found that the silencing of HSD11B2 prevents the formation of adenoma as well as the growth and metastasis of colon cancer in mice (35), suggesting that HSD11B2 is a tumor-driving gene in colorectal cancer. GBP1 is a unique large guanosine triphosphate (GTP) enzyme that controls the cellular response to infection, inflammation, and environmental stressors. It is highly environmentally dependent and may act as a double-edged sword in cancer. In ovarian cancer, the above responses are hijacked by upstream carcinogenic events to induce cancer treatment resistance and tumor progression. In breast and colorectal cancers, it inhibits cancer cell proliferation (36). APBB1IP has prognostic significance in many human cancers, and its correlation

with prognosis varies with cancer types. For example, the high expression of APBB1IP is related to survival advantages in patients with endocervical adenocarcinoma (CESC), head and neck squamous cell carcinoma (HNSC), kidney renal papillary cell carcinoma (KIRP), skin cutaneous melanoma (SKCM), thymoma (THYM), and uterine corpus endometrial carcinoma (UCEC), and its low expression is related to better prognoses in esophageal cancer (ESCA), liver hepatocellular carcinoma (LIHC), stomach adenocarcinomas (STAD), and tenosynovial giant cell tumor (TGCT) (37). The down-regulation of NPC2 in hepatocellular carcinoma is significantly related to vascular infiltration and late pathological stages of the tumor. Meanwhile, the up-regulation of its expression weakens the proliferation and migration of cancer cells (38). In the present research, these genes were combined into a risk model to significantly distinguish not only the different death risks of OS samples but also the survival rates of 28 types of cancer samples in TCGA, and the survival rates of high-risk samples defined by the mRNAsi-related gene signature were significantly higher than those of low-risk samples.

TIDE is a common calculation method applied to predict the response to immunotherapy. Herein, we compared the sensitivity of the mRNAsi-related risk model and TIDE in predicting sample survival and immunotherapy response in the immunotherapy cohort and confirmed that the mRNAsi-related risk model was more sensitive than the TIDE model in predicting prognosis and immunotherapy response in the three immunotherapy cohorts. The limitations of this study cannot be ignored. Firstly, all data were downloaded from public databases, and the sample size and clinical information were limited. Second, although a risk score system consisting of mRNAsi-related gene has been created, the regulatory network and biological effects between these genes remain to be explored.

Conclusions

In summary, this study classified OS and constructed a mRNAsi-related risk model based on mRNAsi-related genes. We also identified three miRNA-related molecular subtypes of OS that showed differences in survival and TME. The developed mRNAsi-related risk model performed well in predicting the prognosis of OS and predicted the immunotherapy response of OS more reliably than TIDE. The results of this study may provide clues for the treatment of targeted CSC.

Acknowledgments

Funding: This work was supported by Key Research Program for University of Henan Province (No. 23A320047) and Application of 3D-printed polyether-ether-ketone (PEEK) composite with sustained release doxorubicin in the comprehensive treatment of malignant bone tumors (No. 222102310245).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://atm.amegroups.com/article/view/10.21037/atm-22-6011/rc>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://atm.amegroups.com/article/view/10.21037/atm-22-6011/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Ferguson JL, Turner SP. Bone Cancer: Diagnosis and Treatment Principles. *Am Fam Physician* 2018;98:205-13.
2. Kim HJ, Chalmers PN, Morris CD. Pediatric osteogenic sarcoma. *Curr Opin Pediatr* 2010;22:61-6.
3. Shoaib Z, Fan TM, Irudayaraj JMK. Osteosarcoma mechanobiology and therapeutic targets. *Br J Pharmacol* 2022;179:201-17.
4. Kager L, Tamamyan G, Bielack S. Novel insights and therapeutic interventions for pediatric osteosarcoma.

- Future Oncol 2017;13:357-68.
5. Marchandet L, Lallier M, Charrier C, et al. Mechanisms of Resistance to Conventional Therapies for Osteosarcoma. *Cancers (Basel)* 2021;13:683.
 6. Lallier M, Marchandet L, Moukengue B, et al. Molecular Chaperones in Osteosarcoma: Diagnosis and Therapeutic Issues. *Cells* 2021;10:754.
 7. Basu-Roy U, Basilico C, Mansukhani A. Perspectives on cancer stem cells in osteosarcoma. *Cancer Lett* 2013;338:158-67.
 8. Zhang W, Wei L, Weng J, et al. Advances in the research of osteosarcoma stem cells and its related genes. *Cell Biol Int* 2022;46:336-43.
 9. Wang JH, Gong C, Guo FJ, et al. Knockdown of STIP1 inhibits the invasion of CD133-positive cancer stem-like cells of the osteosarcoma MG63 cell line via the PI3K/Akt and ERK1/2 pathways. *Int J Mol Med* 2020;46:2251-9.
 10. Brown HK, Tellez-Gabriel M, Heymann D. Cancer stem cells in osteosarcoma. *Cancer Lett* 2017;386:189-95.
 11. Kim WT, Ryu CJ. Cancer stem cell surface markers on normal stem cells. *BMB Rep* 2017;50:285-98.
 12. Yoshida GJ. Correction to: Emerging roles of Myc in stem cell biology and novel tumor therapies. *J Exp Clin Cancer Res* 2018;37:285.
 13. Barzegar Behrooz A, Syahir A, Ahmad S. CD133: beyond a cancer stem cell biomarker. *J Drug Target* 2019;27:257-69.
 14. Belayneh R, Weiss K. The Role of ALDH in the Metastatic Potential of Osteosarcoma Cells and Potential ALDH Targets. *Adv Exp Med Biol* 2020;1258:157-66.
 15. Martins-Neves SR, Sampaio-Ribeiro G, Gomes CMF. Chemoresistance-Related Stem Cell Signaling in Osteosarcoma and Its Plausible Contribution to Poor Therapeutic Response: A Discussion That Still Matters. *Int J Mol Sci* 2022;23:11416.
 16. Walcher L, Kistenmacher AK, Suo H, et al. Cancer Stem Cells-Origins and Biomarkers: Perspectives for Targeted Personalized Therapies. *Front Immunol* 2020;11:1280.
 17. Malta TM, Sokolov A, Gentles AJ, et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* 2018;173:338-54.e15.
 18. Zhao X, Lin J. Construction and Validation of a Prognostic Model Based on mRNAsi-Related Genes in Breast Cancer. *Comput Math Methods Med* 2022;2022:6532591.
 19. Wang W, Xu C, Ren Y, et al. A Novel Cancer Stemness-Related Signature for Predicting Prognosis in Patients with Colon Adenocarcinoma. *Stem Cells Int* 2021;2021:7036059.
 20. Chen X, Zhang D, Jiang F, et al. Prognostic Prediction Using a Stemness Index-Related Signature in a Cohort of Gastric Cancer. *Front Mol Biosci* 2020;7:570702.
 21. Yoshihara K, Shahmoradgoli M, Martinez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013;4:2612.
 22. Becht E, Giraldo NA, Lacroix L, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* 2016;17:218.
 23. Hanzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013;14:7.
 24. He Y, Jiang Z, Chen C, et al. Classification of triple-negative breast cancers based on Immunogenomic profiling. *J Exp Clin Cancer Res* 2018;37:327.
 25. Liu Y, He M, Wang D, et al. HisgAtlas 1.0: a human immunosuppression gene database. *Database (Oxford)* 2017;2017:bax094.
 26. Jiang P, Gu S, Pan D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* 2018;24:1550-8.
 27. Chan BKC. Data Analysis Using R Programming. *Adv Exp Med Biol* 2018;1082:47-122.
 28. Pan K, Farrukh H, Chittepu V, et al. CAR race to cancer immunotherapy: from CAR T, CAR NK to CAR macrophage therapy. *J Exp Clin Cancer Res* 2022;41:119.
 29. Li N, Li Y, Zheng P, et al. Cancer Stemness-Based Prognostic Immune-Related Gene Signatures in Lung Adenocarcinoma and Lung Squamous Cell Carcinoma. *Front Endocrinol (Lausanne)* 2021;12:755805.
 30. Tokunaga R, Naseem M, Lo JH, et al. B cell and B cell-related pathways for novel cancer treatments. *Cancer Treat Rev* 2019;73:10-9.
 31. Seo N, Shirakura Y, Tahara Y, et al. Activated CD8(+) T cell extracellular vesicles prevent tumour progression by targeting of lesional mesenchymal cells. *Nat Commun* 2018;9:435.
 32. Han J, Khatwani N, Searles TG, et al. Memory CD8(+) T cell responses to cancer. *Semin Immunol* 2020;49:101435.
 33. Wu J, Zhang T, Xiong H, et al. Tumor-Infiltrating CD4(+) Central Memory T Cells Correlated with Favorable Prognosis in Oral Squamous Cell Carcinoma. *J Inflamm Res* 2022;15:141-52.
 34. Vivier E, Ugolini S, Blaise D, et al. Targeting natural killer cells and natural killer T cells in cancer. *Nat Rev Immunol* 2012;12:239-52.
 35. Zhang MZ, Xu J, Yao B, et al. Inhibition of 11beta-

- hydroxysteroid dehydrogenase type II selectively blocks the tumor COX-2 pathway and suppresses colon carcinogenesis in mice and humans. *J Clin Invest* 2009;119:876-85.
36. Honkala AT, Tailor D, Malhotra SV. Guanylate-Binding Protein 1: An Emerging Target in Inflammation and Cancer. *Front Immunol* 2019;10:3139.
37. Ge Q, Li G, Chen J, et al. Immunological Role and Prognostic Value of APBB1IP in Pan-Cancer Analysis. *J Cancer* 2021;12:595-610.
38. Liao YJ, Fang CC, Yen CH, et al. Niemann-Pick type C2 protein regulates liver cancer progression via modulating ERK1/2 pathway: Clinicopathological correlations and therapeutical implications. *Int J Cancer* 2015;137:1341-51.
- (English Language Editor: A. Kassem)

Cite this article as: Li Z, Jin C, Lu X, Zhang Y, Zhang Y, Wen J, Liu Y, Liu X, Li J. Construction of a novel mRNAsi-related risk model for predicting prognosis and immunotherapy response in osteosarcoma. *Ann Transl Med* 2023;11(2):61. doi: 10.21037/atm-22-6011