Original Article

Development and validation of a novel stemness-related prognostic model for neuroblastoma using integrated machine learning and bioinformatics analyses

Yuren Xia[1,2#], Chaoyu Wang[1#], Xin Li[1,3#], Mingyou Gao[1], Henry David Jeffry Hogg[4], Thara Tunthanathip[5], Tim Hulsen[6], Xiangdong Tian[1], Qiang Zhao[1]

[1]National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute & Hospital, Tianjin, China; [2]Department of General Surgery, Tianjin Cancer Hospital Airport Hospital, Tianjin, China; [3]Department of Pathology, Tianjin Cancer Hospital Airport Hospital, Tianjin, China; [4]Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle Upon Tyne, UK; [5]Division of Neurosurgery, Department of Surgery, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla, Thailand; [6]Data Science & AI Engineering, Philips, Eindhoven, The Netherlands

Contributions: (I) Conception and design: Y Xia, C Wang, X Li; (II) Administrative support: Q Zhao; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: M Gao, X Tian; (V) Data analysis and interpretation: Y Xia, C Wang, X Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Qiang Zhao, PhD, MD. National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute & Hospital, West Huan-Hu Rd., Ti Yuan Bei, Hexi District, Tianjin 300060, China. Email: qiangzhao169@aliyun.com.

Background: Neuroblastoma (NB) is a common solid tumor in children, with a dismal prognosis in high-risk cases. Despite advancements in NB treatment, the clinical need for precise prognostic models remains critical, particularly to address the heterogeneity of cancer stemness which plays a pivotal role in tumor aggressiveness and patient outcomes. By utilizing machine learning (ML) techniques, we aimed to explore the cancer stemness features in NB and identify stemness-related hub genes for future investigation and potential targeted therapy.

Methods: The public dataset GSE49710 was employed as the training set for acquire gene expression data and NB sample information, including age, stage, and MYCN amplification status and survival. The messenger RNA (mRNA) expression-based stemness index (mRNAsi) was calculated and patients were grouped according to their mRNAsi value. Stemness-related hub genes were identified from the differentially expressed genes (DEGs) to construct a gene signature. This was followed by evaluating the relationship between cancer stemness and the NB immune microenvironment, and the development of a predictive nomogram. We assessed the prognostic outcomes including overall survival (OS) and event-free survival, employing machine learning methods to measure predictive accuracy through concordance indices and validation in an independent cohort E-MTAB-8248.

Results: Based on mRNAsi, we categorized NB patients into two groups to explore the association between varying levels of stemness and their clinical outcomes. High mRNAsi was linked to the advanced International Neuroblastoma Staging System (INSS) stage, amplified MYCN, and elder age. High mRNAsi patients had a significantly poorer prognosis than low mRNAsi cases. According to the multivariate Cox analysis, the mRNAsi was an independent risk factor of prognosis in NB patients. After least absolute shrinkage and selection operator (LASSO) regression analysis, four key genes (*ERCC6L*, *DUXAP10*, *NCAN*, *DIRAS3*) most related to mRNAsi scores were discovered and a risk model was built. Our model demonstrated a significant prognostic capacity with hazard ratios (HR) ranging from 18.96 to 41.20, P values below 0.0001, and area under the receiver operating characteristic curve (AUC) values of 0.918 in the training set, suggesting high predictive accuracy which was further confirmed by external verification. Individuals with a low four-gene signature score had a favorable outcome and better immune responses.

Finally, a nomogram for clinical practice was constructed by integrating the four-gene signature and INSS stage.

Conclusions: Our findings confirm the influence of CSC features in NB prognosis. The newly developed NB stemness-related four-gene signature prognostic signature could facilitate the prognostic prediction, and the identified hub genes may serve as promising targets for individualized treatments.

Keywords: Neuroblastoma (NB); cancer stemness; messenger RNA expression-based stemness index (mRNAsi); prognosis

Submitted Dec 04, 2023. Accepted for publication Jan 05, 2024. Published online Jan 12, 2024.
doi: 10.21037/tp-23-582
View this article at: https://dx.doi.org/10.21037/tp-23-582

Introduction

Cancers in children and young adults are relatively uncommon and have unique diagnostic features compared to those affecting older individuals (1). These malignancies often originate from immature cells that possess stem cell-like properties, potentially contributing to the observed clinical and biological variations (2). Neuroblastoma (NB) is currently the most prevalent extracranial solid tumor in childhood, with the vast majority of cases (90%) being identified in children under the age of 5 years (3). The clinical prognosis of NB is highly variable, with patients displaying a broad range of outcomes from complete remission to refractory disease that is challenging to treat (3). Despite intensive efforts to improve outcomes, current treatments still encounter high rates of resistance, recurrence, and progression in high-risk NB cases, with long-term survival still less than 40% (4). Previous studies on NB have identified several prognostic biomarkers, such as MYCN amplification, ALK mutations, and age at diagnosis, among others (5,6). While these markers are informative, their predictive power is often limited by NB's extensive biological diversity, which stems from its complex genetic and epigenetic landscape. Therefore, finding a robust index to predict and evaluate clinical prognosis and therapeutic efficacy has become an urgent issue, with the goal of achieving accurate clinical interventions.

Many investigations have shown that in various types of tumors, a minute fraction of undifferentiated cells, resembling stem cells, have the capability to initiate the development of cancer. These cells have been designated as cancer stem cells (CSCs) (7). CSCs are believed to play a critical role in the growth, progression, recurrence, and resistance to therapy of solid malignancies (8). These uncommon and difficult-to-detect cells have the ability to spread at an early stage and conceal themselves in specific niches located in remote organs. This makes them capable of causing disease relapse even after successful treatment of the primary tumor (9). It has been demonstrated that NB tumors with poor prognosis harbor an undifferentiated stem cell population that is responsible for their highly aggressive nature (10-12). Pandian *et al.* successfully isolated highly aggressive cellular clones from metastatic NB tumors, and these clones were found to retain their plasticity and adaptive stemness despite exposure to different culture conditions, indicating their potential as NB CSCs (13).

Highlight box

Key findings

- The newly developed neuroblastoma (NB) stemness-related prognostic signature in this study could facilitate the prognostic prediction and immune responses.

What is known and what is new?

- NB tumors with poor prognosis harbor an undifferentiated stem cell population that is responsible for their highly aggressive nature.
- By utilizing the stemness index and machine learning methods, our study delved into the features of cancer stem cells in NB, further providing evidence of their essential role in the development and progression of the disease. For the first time, we constructed a novel mRNAsi-associated signature for NB that exhibited significant associations with prognosis, clinical characteristics, and tumor immune response in NB.

What is the implication, and what should change now?

- The identification of hub genes in this study significantly advances our understanding of NB stem cell maintenance and suggests that they may serve as promising therapeutic targets for inhibiting NB stemness characteristics.

Drug resistance in treatment is a major challenge in the management of refractory cancers. It is widely accepted that CSCs play a significant role in the development of therapy resistance in tumors (14-16). During prolonged drug selection of NB cells, a subpopulation of cells with CSC characteristics becomes enriched (17). Despite the evidence that the CSC subpopulation in NB contributes to disease metastases and therapy resistance, the origin and characteristics of this cell subset are still not fully understood (13,18,19). The heterogeneity of CSC populations within NB tumors contributes to the variability in patient responses to treatment, necessitating the exploration of new biomarkers that can encompass this diversity. Currently, there are no established therapeutic modalities specifically targeting CSCs, scientists have adopted artificial intelligence (AI) to further examine the characteristics of CSCs within tumors, in order to obtain a more comprehensive understanding of these cells. In a study by Malta *et al.*, a novel messenger RNA (mRNA) expression-based stemness index (mRNAsi) was generated by machine learning algorithm to quantify tumor stemness (20). By quantifying tumor stemness, mRNAsi may offer a more comprehensive biomarker for NB prognosis. Stemness attributes, reflective of a tumor's ability to sustain growth, resist therapy, and metastasize, are particularly pertinent to the aggressive and often treatment-resistant nature of high-risk NB. Prognostic signatures based on mRNAsi have been successfully developed in various types of cancer, indicating the stemness of CSCs and predicting unfavorable outcomes (21-23). Despite this, few studies have sought to determine the predictive and prognostic value of mRNAsi-related genes in NB.

Our study proposes mRNAsi as a promising tool for prognostic assessment, with the potential to enhance the precision of clinical interventions in NB. We utilized the mRNAsi in a comprehensive bioinformatics analysis to identify genes associated with stemness, with the aim of discovering potential therapeutic targets and developing a prognostic signature for NB. The mRNAsi of NB patients was evaluated and a risk model was developed using the four hub genes that we identified. A comprehensive evaluation of prognosis, clinical characteristics, tumor microenvironment (TME), and response to immunotherapy was conducted across different subtypes. Our findings offer novel insights into CSC research in NB and have the potential to improve individualized treatments for NB patients. We present this article in accordance with the TRIPOD reporting checklist (available at https://tp.amegroups.com/article/view/10.21037/tp-23-582/rc).

Methods

Data acquisition

The microarray data were downloaded from the Gene Expression Omnibus (GEO) database (24). The gene expression profiles and clinical information of 498 NB patients were obtained from the dataset GSE49710, which is hosted on the GEO. Overall survival (OS) is defined as the duration from the date of diagnosis or initiation of treatment to either the date of death from any cause or the last follow-up. The probes were mapped to gene symbols using the platform GPL16876's gene mapping information. We obtained the E-MTAB-8248 dataset (N=223) from the ArrayExpress database (https://www.ebi.ac.uk/biostudies/arrayexpress) to externally validate our signature. Table S1 provides a detailed illustration of patient information derived from two datasets. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Calculation of stemness index

By using the one-class logistic regression (OCLR) machine learning (ML) algorithm, we calculated the mRNAsi of all 498 samples to represent their stemness values (20). To normalize the mRNAsi values to a range of 0 to 1, we employed a linear transformation technique, which involved subtracting the minimum value from each data point and then dividing the result by the range, defined as the maximum value minus the minimum value. Patients were stratified into two groups based on the median value of mRNAsi. The prognostic significance of mRNAsi was evaluated using Kaplan-Meier (KM) analysis for OS difference between the patient groups. Multivariate Cox regression was employed to validate the prognostic independence of mRNAsi.

Analysis of enriched gene sets

We employed gene set enrichment analysis (GSEA) to investigate the differences in functional pathways between the two mRNAsi groups. The dataset was obtained from the Molecular Signatures Database (MSigDB). The latest MSigDB consists of nine categories (C1–C8 and Hallmark) (25). The R package "clusterProfiler" was used for analysis (26).

Construction and verification of a four-gene signature

The 498 samples of GSE49710 were used for signature

development and 223 samples of E-MTAB-8248 for model validation. The R package "limma" was used to identify differentially expressed genes (DEGs) (27). Cox analysis and least absolute shrinkage and selection operator (LASSO) regression analysis were used to further screen DEGs. Ultimately, the risk score for the prognostic gene signature was derived by multiplying the mRNA expression level with the regression coefficient assigned to each gene. Then, two patient risk groups were built according to the median score. The assessment of the predictive performance of the model was conducted by KM analysis, decision curve analysis (DCA), and area under the curve (AUC) of the receiver operating characteristic (ROC) curve. We adopted AUC threshold values of >0.7 for acceptable, >0.8 for excellent, and >0.9 for outstanding predictive performance.

Analysis of immune microenvironment and prediction of therapeutic sensitivity

To evaluate the degree of immune infiltration, four algorithms were used, including Estimation of STromal and Immune cells in MAlignant Tumours using Expression data (ESTIMATE) (28), Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORT) (4), Microenvironment Cell Populations-counter (MCPcounter) (29), and xCell (30). The activity of 14 cancer-related pathways was analyzed by the "PROGENy" R package (31). In order to assess the potential immunotherapy response, we used the Tumor Immune Dysfunction and Exclusion (TIDE, http://tide.dfci.harvard.edu/) (32) and immunophenoscore (IPS, https://tcia.at/tools/toolsMain) (33) algorithms. A high IPS score and low TIDE score may suggest a more favorable response to immunotherapy. From the database of Genomics of Drug Sensitivity in Cancer (GDSC, https://www.cancerrxgene.org/) (34), we obtained the anti-cancer drugs information. To determine the IC50 of drugs, we employed the "pRRophetic" R package (35).

Predictive nomogram development and validation

A nomogram plot incorporating the risk score of the four-gene model and the International Neuroblastoma Staging System (INSS) stage was built using the "rms" R package. The predictive ability of the nomogram was assessed by analyzing the ROC curve and the calibration plot. The global and individual Schoenfeld tests were used to estimate the time-varying covariance of the Cox proportional risk regression analysis hypothesis. We assumed that Schoenfeld residuals were time independent. A P value <0.05 indicated rejection of this assumption.

Statistical methods

R software (version 4.2.1; The R Foundation for Statistical Computing, Vienna, Austria) was employed for statistical analysis and data visualization. We used two-tailed unpaired Student's *t*-test and Wilcoxon test to compare differences between two groups. The differences between rates were tested using either the chi-square test or Fisher's exact test. To generate images, we used Adobe Illustrator (version 22.0; Adobe, San Jose, CA, USA). In this study, a two-sided P values less than 0.05 were considered as statistically significant.

Results

The relationship between clinical features and mRNAsi in NB

Following the computation with the OCLR algorithm, the mRNAsi values of 498 samples were subsequently arranged in ascending order. The correlation between clinical characteristics and mRNAsi was assessed and is visualized in *Figure 1A*. The analysis revealed that patients with high mRNAsi tended to exhibit clinical characteristics indicative of more aggressive disease, such as higher INSS stage and poor OS status. Notably, MYCN amplification, a well-established marker of poor prognosis in NB, was also more prevalent in this group. Additionally, these patients were more likely to be over 18 months at diagnosis, another factor associated with a less favorable outcome in NB. Statistically significant differences in mRNAsi were observed among patients with different clinical characteristics (*Figure 1B-1D*). Patients with INSS stage 4 were observed to have significantly higher mRNAsi levels compared to the other stages (P<0.0001), and those with MYCN amplification displayed a markedly higher mRNAsi than patients lacking amplification (P<0.0001). Furthermore, the mRNAsi levels of patients aged over 18 months at diagnosis were markedly higher than those diagnosed before 18 months (P<0.0001). In addition, patients with lower mRNAsi had significantly better survival outcomes than those with high mRNAsi (P<0.0001) (*Figure 1E*). The Cox regression indicated that mRNAsi could independently predict the OS of NB, in addition to INSS stage, MYCN, and age (*Figure 1F*).

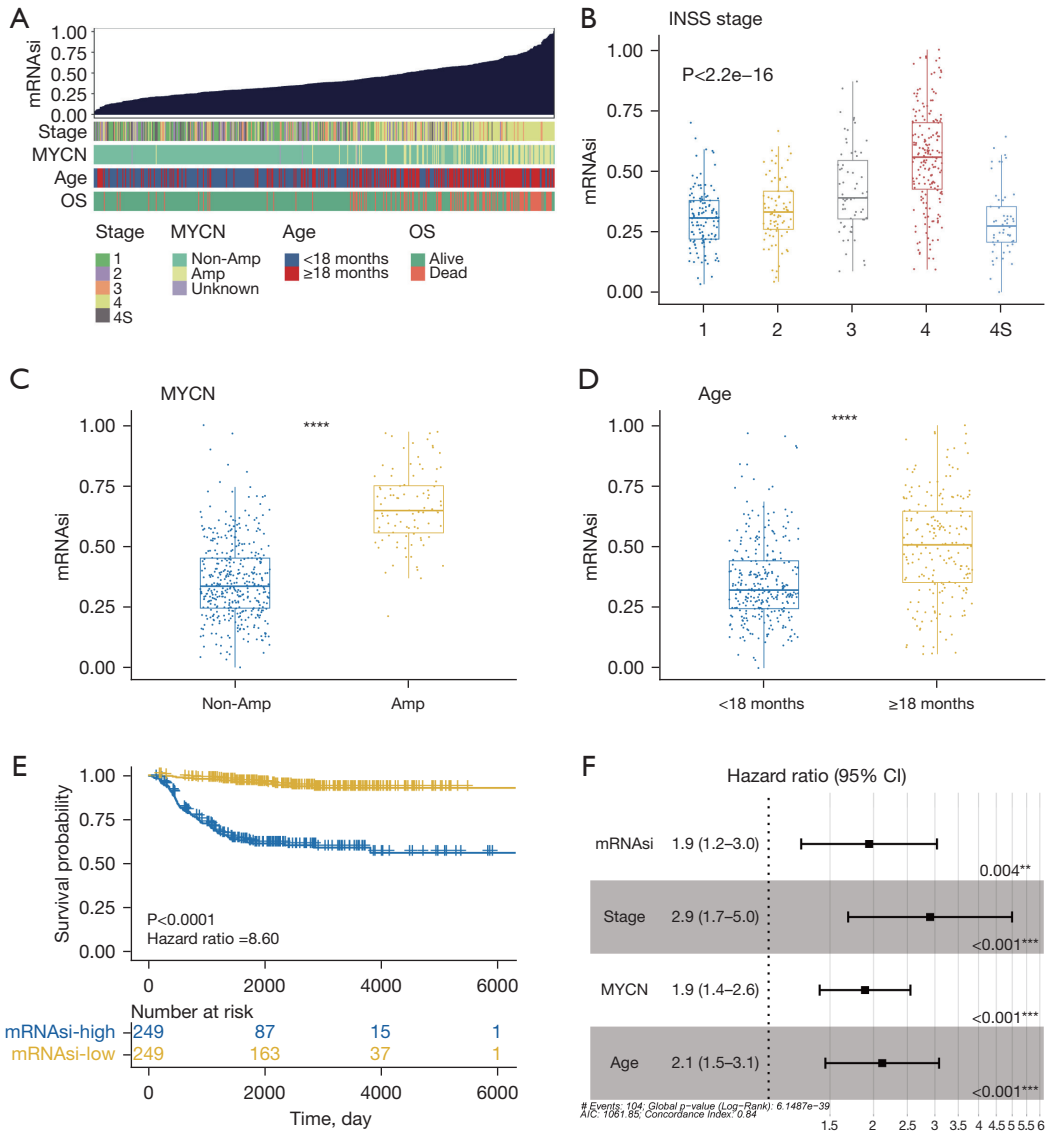

Figure 1 Correlation between mRNAsi and clinical characteristics of NB. (A) A summary of the association between mRNAsi and clinical characteristics of NB. (B-D) The discrepancies in mRNAsi of different clinical characteristics patients. (E) The KM curve showed differences between patients with high and low mRNAsi. (F) Forest diagram displaying the multivariate Cox proportional hazard regression model for mRNAsi and clinical parameters. (The KM curve = the Kaplan-Meier curve. **, P<0.01; ***, P<0.001; ****, P<0.0001). mRNAsi, messenger RNA expression-based stemness index; OS, overall survival; Amp, amplification; INSS, International Neuroblastoma Staging System; AIC, Akaike Information Criterion; NB, neuroblastoma.

Significant survival discrepancies between different mRNAsi groups

According to the respective mRNAsi value of each patient, two groups were created: low mRNAsi and high mRNAsi. As indicated by the survival analysis, high mRNAsi patients had significantly shorter OS, regardless of their INSS stages (*Figure 2A*,*2B*), MYCN statuses (*Figure 2C*,*2D*), or ages (*Figure 2E*,*2F*). Among the 131 NB patients aged over 18 months who were also INSS stage 4, those with high mRNAsi also had a worse prognosis (P=0.0072) (*Figure 2G*). Similarly, in the subset of 47 patients with MYCN amplification, those categorized in the high mRNAsi group demonstrated a poorer survival outcome

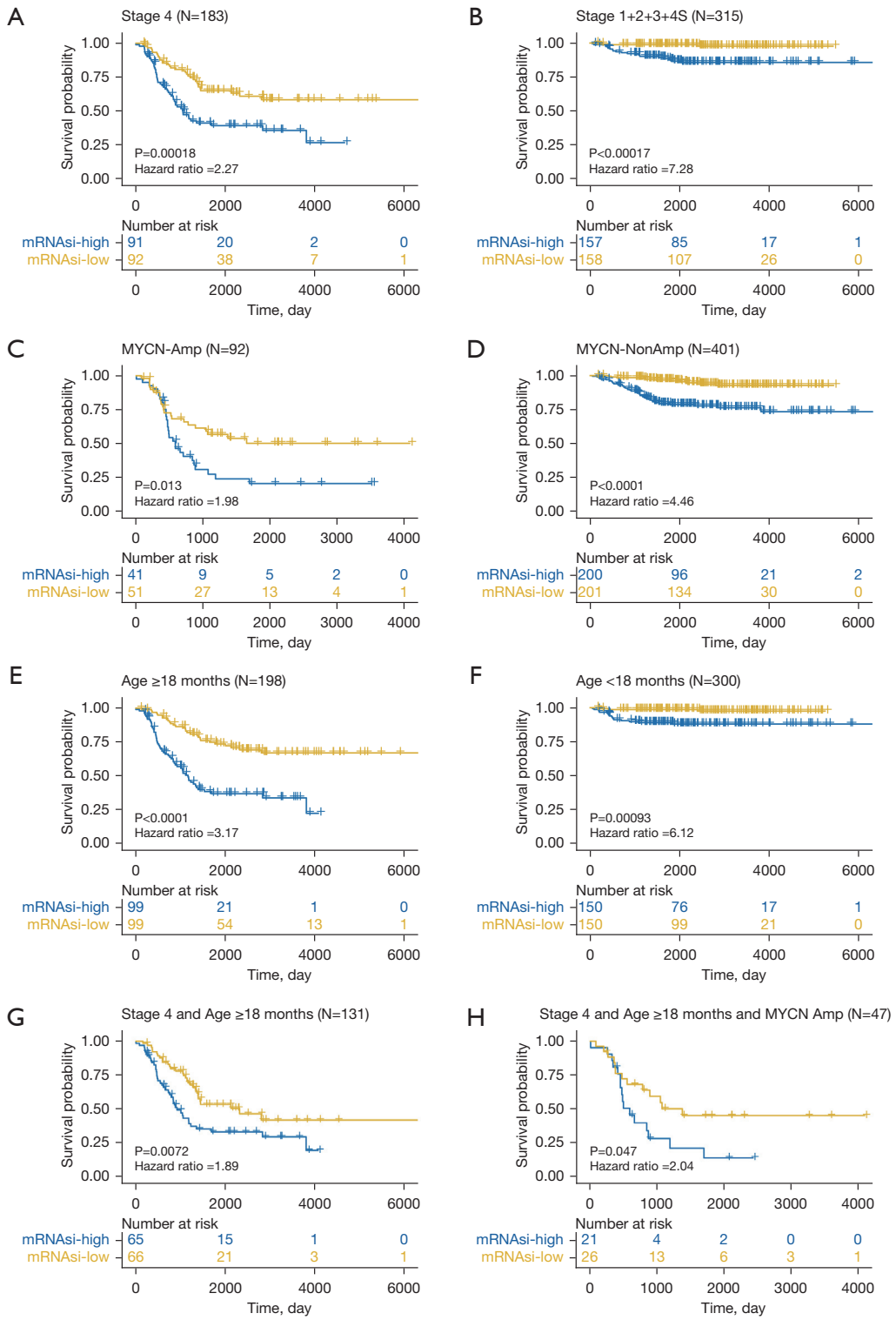

Figure 2 Significant survival discrepancies between different mRNAsi groups. (A-F) Significant survival difference between high and low mRNAsi patients in each clinicopathological subgroup. (G) Significant survival difference between high and low mRNAsi patients in patients with stage 4 and over 18 months. (H) Significant survival difference between high and low mRNAsi patients in patients with stage 4, over 18 months, and MYCN amplification. Amp, amplification; mRNAsi, messenger RNA expression-based stemness index.

(P=0.047) (*Figure 2H*).

GSEA of DEGs

To investigate the function of DEGs between patients with low and high mRNAsi, we used MSigDB to conduct an enrichment analysis. The top two enrichments with the lowest P values in each of the nine categories (C1–C8 and Hallmark) are presented in *Figure 3*. In Figure S1A,S1B, the top 10 enrichment with the lowest P values of C5 [Gene Ontology (GO) gene sets] and H (Hallmark gene sets) were presented, respectively. The gene set on the left represents normalized enrichment score (NES) <0. As depicted in *Figure 3*, genes located on chr1p36 were found to be more enriched in the low mRNAsi group, suggesting a potential higher frequency of 1p36 deletion in patients with high mRNAsi. Loss of tumor suppressor genes at the 1p36 locus is crucial in the oncogenesis and progression of NB (36,37). According to GO enrichment analysis, the low mRNAsi group showed enrichment of gene sets associated with positive regulation of cell adhesion, whereas the high mRNAsi group exhibited enrichment of gene sets related to chromosome separation and nuclear division (Figure S1A). The Hallmark enrichment results showed that the gene sets related to MYC targets, DNA repair, G2/M checkpoint, and E2F targets were enriched in the high mRNAsi group (Figure S1B). These pathways related to cell division, DNA damage repair, and cell cycle are critical for the survival of CSCs.

Construction of the four-gene signature

Analysis of gene expression differences revealed 20 DEGs, as shown in *Figure 4A*. Univariate Cox analysis demonstrated that all 20 genes were significantly associated with OS (*Figure 4B*). LASSO regression analysis identified seven variables as the most influential (*Figure 4C*,*4D*), and three collinear variables were excluded by multivariate Cox regression (*Figure 5A*). Finally, four genes (*ERCC6L*, *DUXAP10*, *NCAN*, *DIRAS3*) were included in the prediction signature, and the chordal graph displayed in *Figure 5B* demonstrates their close interrelation. Based on the coefficients of the four genes, the risk score was calculated by the following formula:

$$\begin{aligned}\text{Four-gene signature risk score} = &(0.602 \times \text{the expression of } \textit{ERCC6L}) \\ &+ (0.139 \times \text{the expression of } \textit{DUXAP10}) \\ &+ (0.154 \times \text{the expression of } \textit{NCAN}) \\ &- (0.101 \times \text{the expression of } \textit{DIRAS3})\end{aligned} \quad [1]$$

Internal and external verification of the four-gene signature

By using the median risk score (6.534) as a cutoff point, the 498 NB samples were categorized into two groups. The two groups showed significant differences in survival, with shorter OS in the high-risk score group (P<0.0001) (*Figure 5C*,*5D*). *KCAN*, *DUXAP10*, and *ERCC6L* were upregulated in the high-risk score group, whereas *DIRAS3* was overexpressed in the low-risk score group. The ROC curve demonstrates the risk score's predictive ability for OS, with high AUC values at different time points (*Figure 5E*).

The E-MTAB-8248 dataset was employed to evaluate the predictive ability of the risk score. The KM analysis revealed that, in line with the findings of the training set, the high-risk group demonstrated shorter OS (P<0.0001) (Figure S2A), highlighting significant survival differences between the two groups (Figure S2B).

Comparative analysis and correlation of the four-gene signature with other clinical indicators

The significant and positive association between high-risk scores and high mRNAsi indicated that the four-gene signature can effectively reflect the stemness of NB specimens (*Figure 6A*). The advanced INSS stage, MYCN amplification, and age above 18 months are established clinical factors that are highly linked with a dismal prognosis in NB (3). *Figure 6B* depicts the correlation between these features and the risk score for predicting the prognosis. High-risk scores were observed to be associated with clinical indicators such as INSS stage 4, MYCN amplification, age >18 months, and deceased OS status. Moreover, compared to the low mRNAsi group, all low-risk patients in the four-gene model were non-MYCN-amplified, indicating a higher accuracy of the gene signature, as illustrated in *Figure 6C*. It is worth mentioning that the predictive ability of the four-gene signature exceeded that of MYCN status, INSS stage, and mRNAsi. This was evident in the area under the ROC curve (*Figure 6D*) and the DCA curve for 5-year OS prediction (*Figure 6E*), both of which demonstrated the superior predictive ability of the model.

Immune infiltration and sensitivity to treatment

The low and high-risk score groups exhibited remarkable differences in immune infiltration, as calculated by the four-gene signature. *Figure 7A* and Table S2 provide an

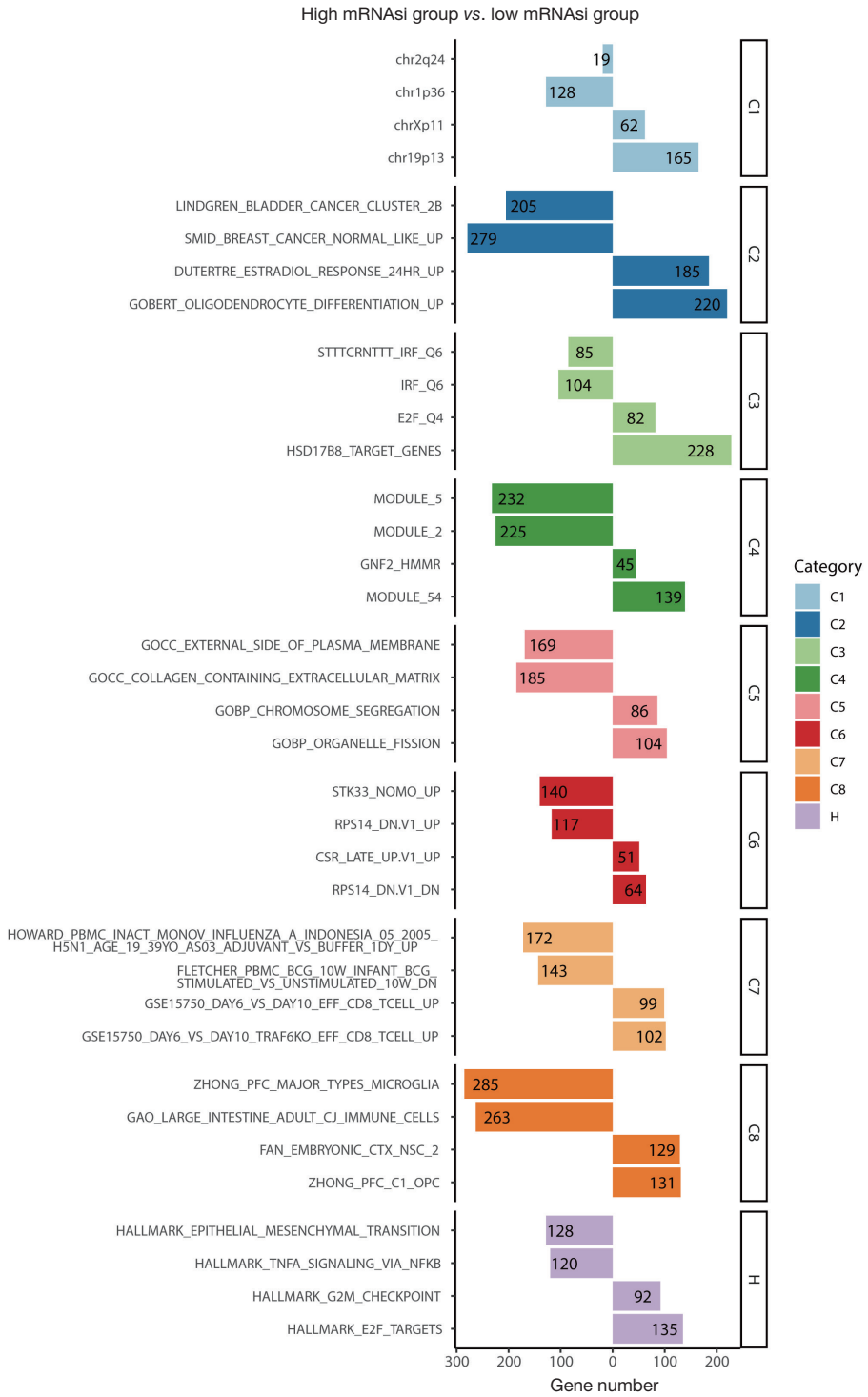

Figure 3 The GSEA of DEGs between high and low mRNAsi. The enrichment analysis of DEGs, the top two enrichments of up and down-regulated were displayed in each category, respectively. The gene sets with NES <0 were listed on the left, and the gene sets with NES >0 were listed on the right. C, category; H, Hallmark; GSEA, gene set enrichment analysis; DEGs, differentially expressed genes; mRNAsi, messenger RNA expression-based stemness index; NES, normalized enrichment score.

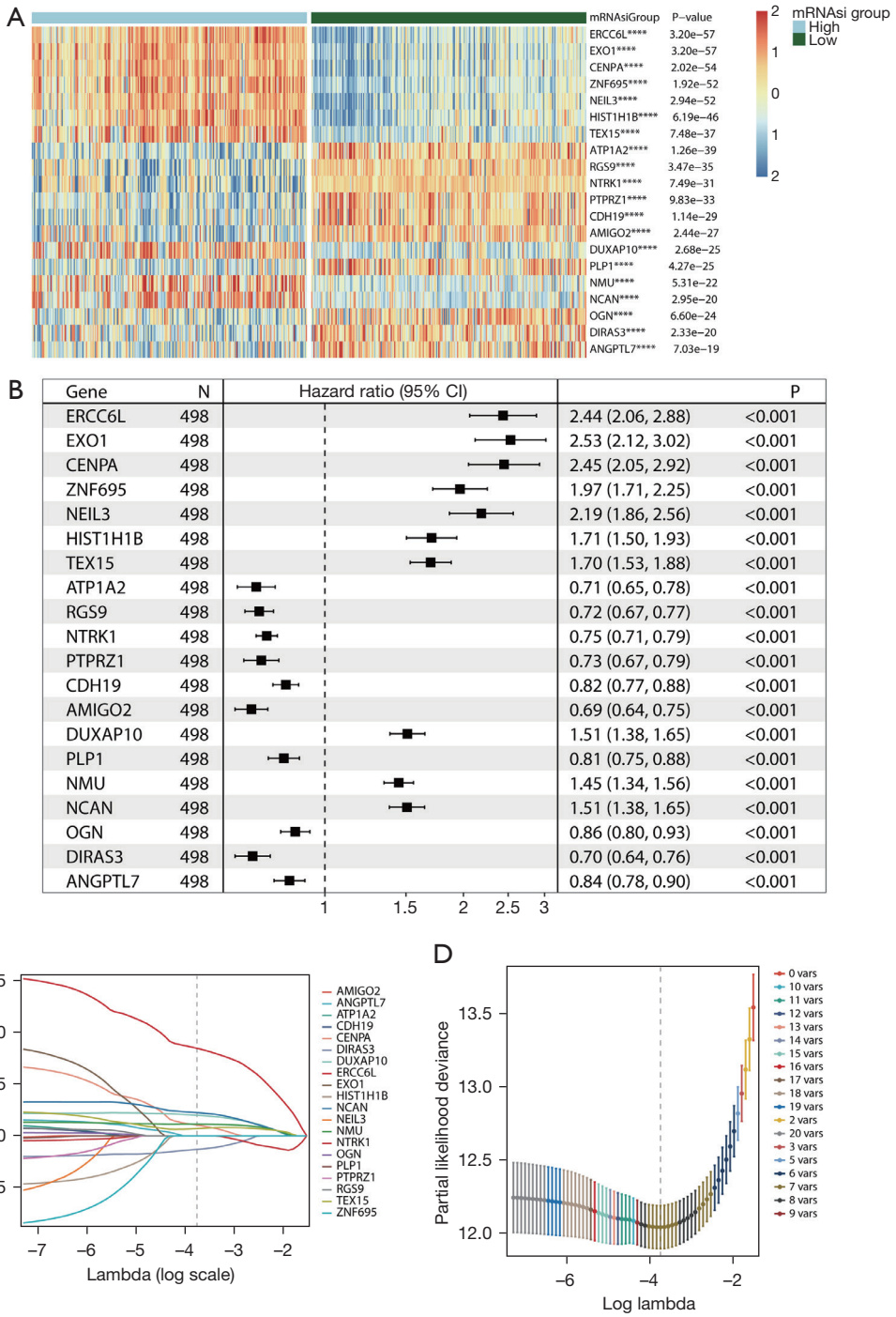

Figure 4 Construction of the four-gene signature. (A) Heat map of 20 DEGs between high and low mRNAsi NB patients. (B) The univariate COX analysis for selecting genes significantly associated with OS. (C,D) The LASSO regression analysis for filtering genes. ****, P<0.0001. CI, confidence interval; vars, variables; DEGs, differentially expressed genes; mRNAsi, messenger RNA expression-based stemness index; NB, neuroblastoma; OS, overall survival; LASSO, least absolute shrinkage and selection operator.

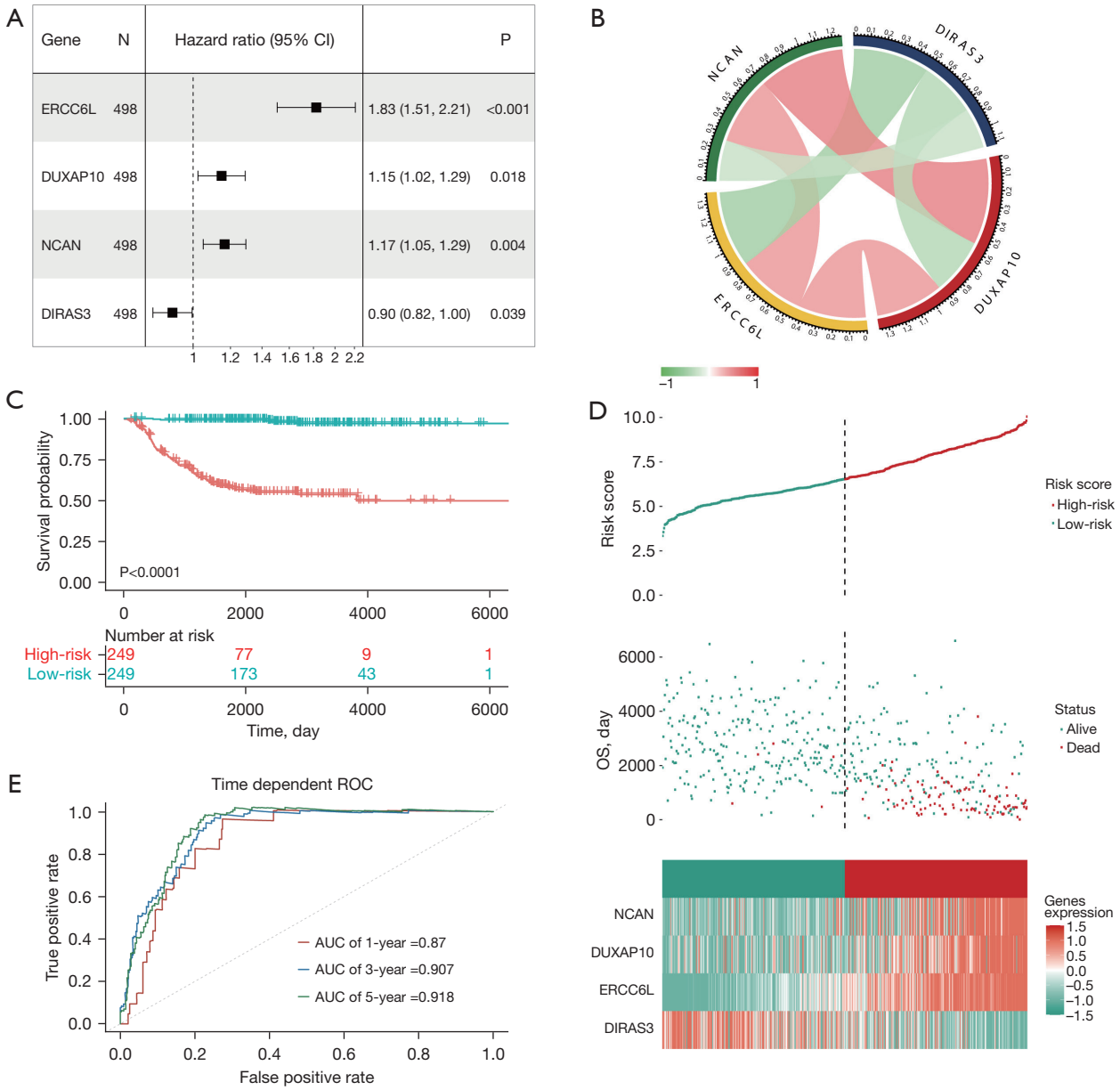

Figure 5 Verification of the four-gene signature. (A) Multivariate Cox regression analysis for removing collinearity genes. (B) Circle plot of the correlation between the four genes. (C) KM curve for prognostic prediction of OS using the four-gene signature. (D) The distribution of risk scores, the association of risk scores and OS, and the four-gene mRNA expression. (E) The ROC curve for verification of the prediction performance of four-gene signature. CI, confidence interval; OS, overall survival; ROC, receiver operating characteristic; AUC, area under the curve; KM, Kaplan-Meier; mRNA, messenger RNA.

overview of the immune infiltration differences, as assessed by ESTIMATE, CIBERSORT, MCPcounter, and xCell. Patients identified as low risk by the ESTIMATE algorithm tend to have higher ImmuneScore, StromalScore, and ESTIMATEScore (P<0.0001). The results obtained from CIBERSORT showed that the low-risk score group had a significantly higher infiltration abundance of M2 macrophages compared to the high-risk score group (P<0.0001). The significant distinctions in the abundance of naïve B cells (P<0.01) and memory B cells (P<0.0001)

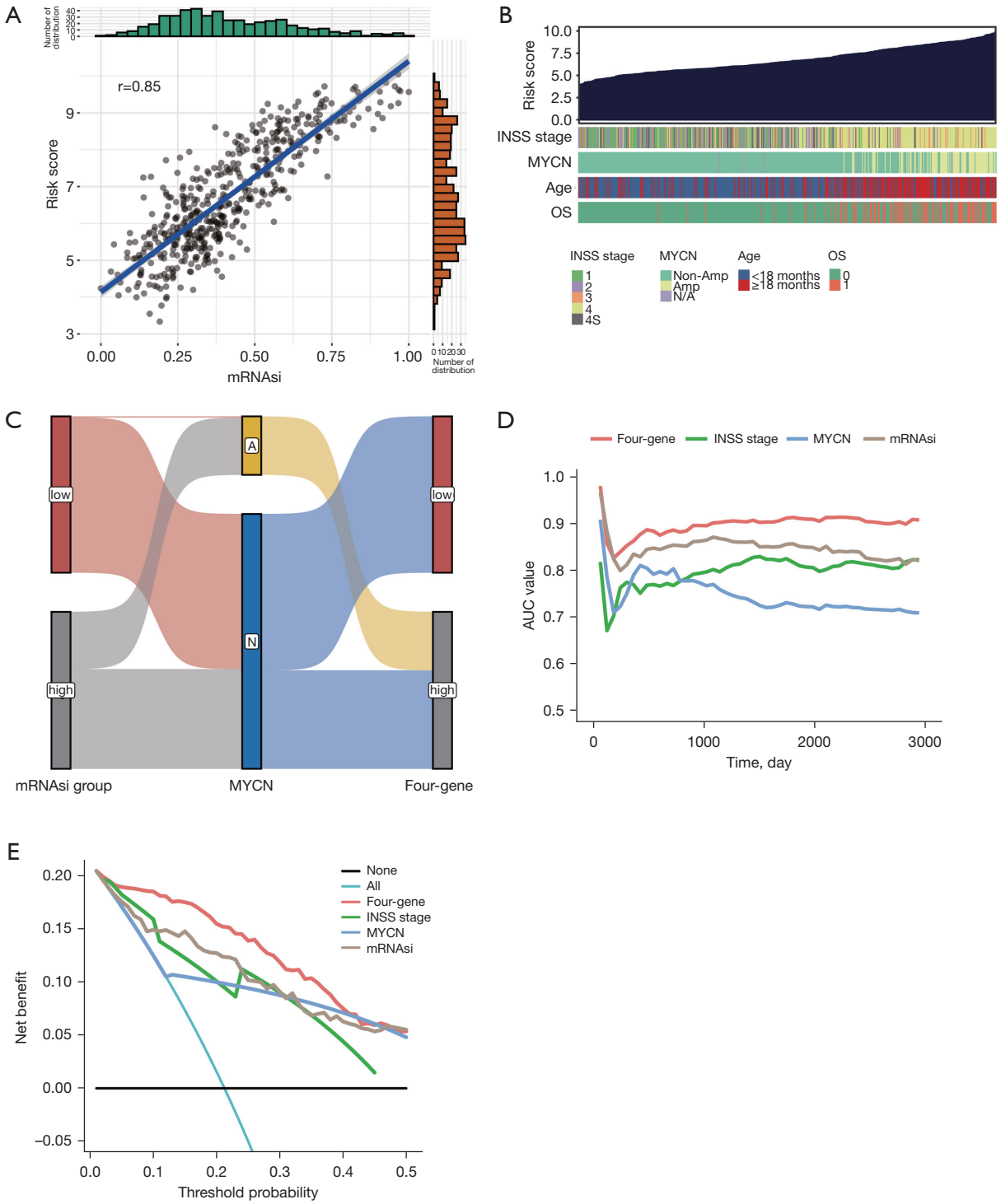

Figure 6 Comparison and association of four-gene signature with other clinical indicators. (A) Correlation between four-gene risk scores and mRNAsi. (B) Distribution of INSS stage, MYCN status, age at diagnosis, OS status, and four-gene signature risk scores in mRNAsi. (C) Sankey diagram for mRNAsi, MYCN status, and four-gene risk scores. (D) The AUC of the four-gene signature, INSS stage, MYCN status, and mRNAsi. (E) The DCA curves of the four-gene signature, INSS stage, MYCN status, and mRNAsi at 5 years. mRNAsi, messenger RNA expression-based stemness index; INSS, International Neuroblastoma Staging System; OS, overall survival; Amp, amplification; N/A, not applicable; AUC, area under the curve; DCA, decision curve analysis.

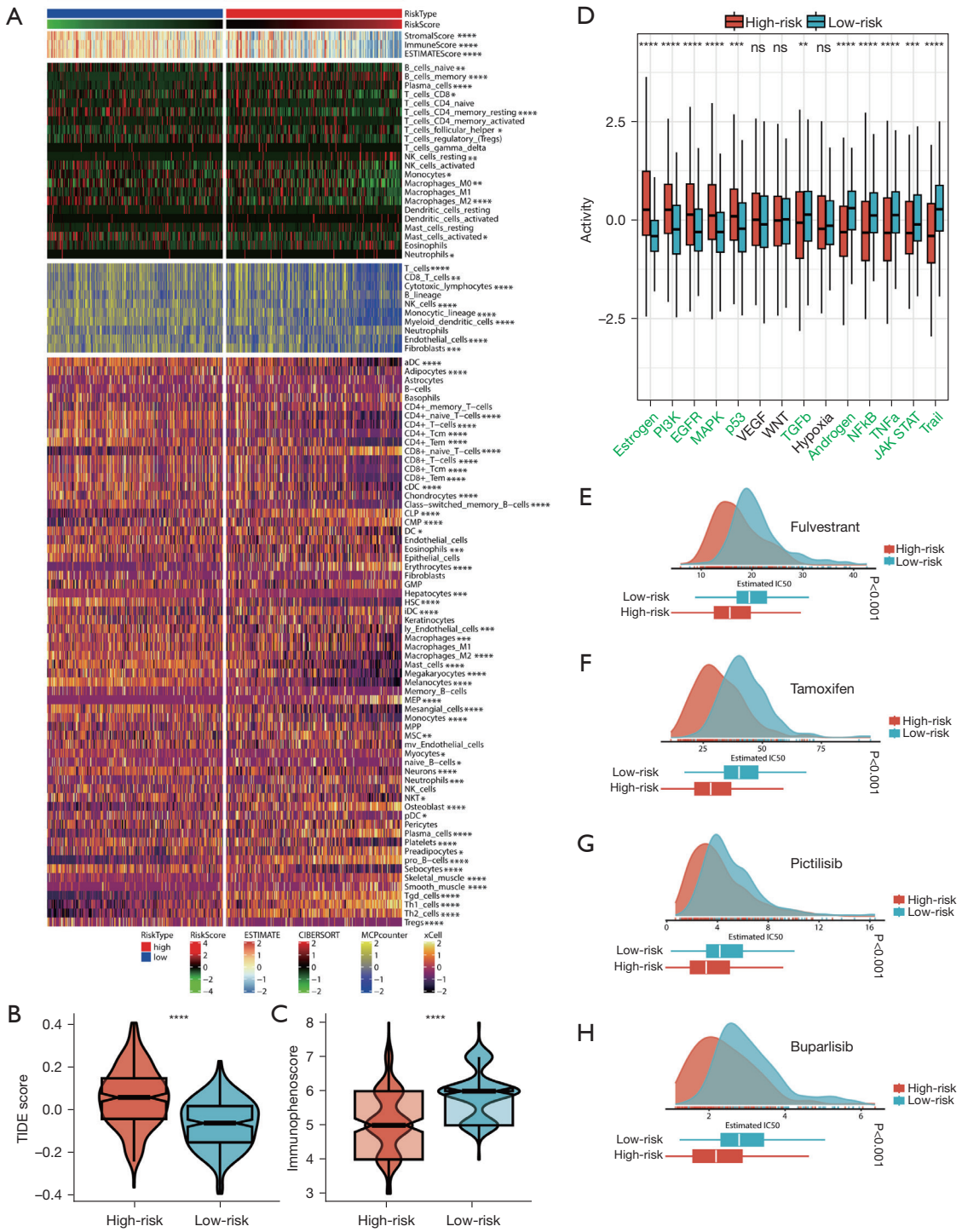

Figure 7 Immune infiltration and sensitivity to chemotherapy. (A) Heat map of immune infiltration calculated by ESTIMATE, CIBERSORT, MCPcounter and xCell. (B,C) Violin plots of TIDE score and IPS. (D) Boxplot of tumor-related pathways activity. (E-H) Ridge plot for IC50 of four anti-cancer drugs (fulvestrant, tamoxifen, pictilisib, buparlisib). *, P<0.05; **, P<0.01; ***, P<0.001; ****, P<0.0001. ESTIMATE, Estimation of STromal and Immune cells in MAlignant Tumours using Expression data; CIBERSORT, Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts; MCPcounter, Microenvironment Cell Populations-counter; TIDE, Tumor Immune Dysfunction and Exclusion; ns, not significant; IPS, immunophenoscore.

also warrant attention. Naïve B cells have been reported to secrete cytokines that hinder the proliferation of cancer cells (38). Memory B cells have a significant impact on the tumor immune microenvironment by recognizing and binding to tumor-associated antigens. They initiate the production of antibodies and stimulate the immune system to attack cancer cells (38). The application of MCPcounter revealed substantial differences in 8 out of 10 cell types between the two groups (P<0.05). Among them, the significantly higher abundance of natural killer (NK) cells and cytotoxic lymphocytes in patients with low-risk score attracted our attention in particular (P<0.0001). Additionally, as per xCell, the low-risk group showed significantly higher infiltration abundance of $CD4^+$ T cells, $CD8^+$ T cells, and M2 macrophages (P<0.0001). Furthermore, both the TIDE score (*Figure 7B*) and the IPS (*Figure 7C*) were significantly different between the high and low-risk score groups, with the former being higher and the latter being lower in the high-risk group. These findings suggested a poor immunotherapy response in the high-risk score group. Specifically, low-risk score patients (65%) were more responsive to immunotherapy than high-risk patients (29%), as revealed by TIDE algorithms (P<0.001) (Figure S3).

As depicted in *Figure 7D*, between the high- and low-risk score groups, significant differences in activity were observed in 11 out of the 14 tumor-related pathways. Specifically, the high-risk score group showed significantly higher activity in the estrogen, PI3K, EGFR, MAPK, and p53 pathways, compared to the low-risk score group. Conversely, the TGF-β, androgen, NF-κB, TNF-α, JAK-STAT, and Trail pathways exhibited higher activity in the low-risk group. Furthermore, a significant reduction in the IC50 of four anti-cancer drugs related to the estrogen and PI3K pathways was observed in the high-risk score group: fulvestrant (P<0.001) (*Figure 7E*), tamoxifen (P<0.001) (*Figure 7F*), pictilisib (P<0.001) (*Figure 7G*), and buparlisib (P<0.001) (*Figure 7H*). This indicates that these drugs are more effective for the high-risk group.

Nomogram development and validation

The nomogram was constructed to make the clinical application of the four-gene signature more feasible for prognosis prediction. The Schoenfeld tests indicated that the INSS stage and risk score did not violate the proportional risk hypothesis (P>0.05) (Figure S4). Hence, the nomogram was constructed by integrating the risk score and INSS stage (*Figure 8A*). The ROC curve (*Figure 8B*) and calibration curve (*Figure 8C*) demonstrate the nomogram's predictive value. The AUC for 1-, 3-, and 5-year OS prediction in the ROC curves were 0.865, 0.915, and 0.926, respectively.

Discussion

By employing the OCLR machine-learning algorithm, we computed the mRNAsi for NB patients, which enabled us to uncover novel biological mechanisms associated with CSCs. We observed that higher mRNAsi scores were correlated with worse clinicopathological factors, such as advanced INSS stage, MYCN amplification, and older age. Our study revealed that high mRNAsi patients tended to have a worse prognosis, irrespective of their INSS stages, MYCN status, and age. Furthermore, mRNAsi was found to be an independent risk factor with strong predictive value for the survival of NB patients. These findings suggest that mRNAsi has potential as a promising prognostic indicator for NB.

Through GSEA, we observed that the low mRNAsi group had enriched genes associated with cell adhesion, whereas gene sets representing active cell division, such as chromosome segregation and nuclear division, were found to be more enriched in the high mRNAsi group.

CSCs differ from normal stem cells in their inability to regulate their mode of cell division, which results in perpetual cell division and consequent unchecked growth of tumors (39). Moreover, some CSCs have acquired specific genetic alterations that confer them with migratory abilities, enabling them to detach from adhesion with other tumor cells and further disperse and metastasize (40). Gene sets related to DNA repair and cell cycle were predominantly enriched in the high mRNAsi group. It is known that DNA repair pathways have a crucial role in maintaining genome integrity, but CSCs activate these pathways inappropriately to evade anticancer therapies (41). Stem cells are able to maintain their longevity by exiting the cell cycle for extended periods, which allows them to preserve their replicative potential and reduce DNA damage (42). Consequently, slow-cycling dormant CSCs can survive cytotoxic drugs and differentiate into highly proliferative cells that have the ability to regenerate tumors (43). Our study also demonstrated that high mRNAsi NB patients were more likely to have 1p36 deletion, which is a common segmental chromosomal loss in NB and is strongly linked to poor OS (44). In another study, it was shown that miR-200b, located at chromosome 1p36, is downregulated in

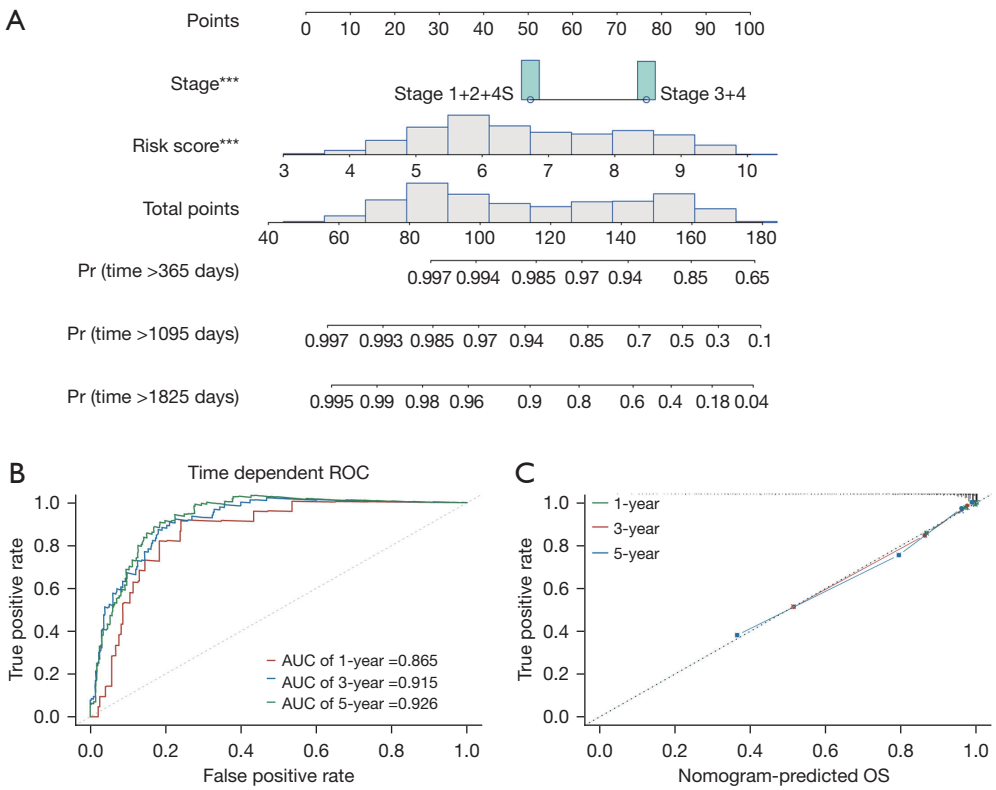

Figure 8 Construction and validation of the nomogram. (A) The nomogram plot constructed based on four-gene signature risk score and INSS stage. (B) ROC curves validating the predictive performance of nomogram. (C) Calibration plot validating the predictive performance of nomogram. ***, P<0.001. Pr, probability; ROC, receiver operating characteristic; AUC, area under the curve; OS, overall survival; INSS, International Neuroblastoma Staging System.

hepatocellular carcinoma (HCC) and is correlated with the stemness of CSCs. Through the inhibition of ZEB1 expression, miR-200b can suppress the stemness of HCC CSCs (45).

Various prognostic signatures for NB have been reported in previous studies (12,46,47). However, to our knowledge, no previous study in NB has reported measuring stemness for prognostication. In our study, we first utilized bioinformatics methods to construct a gene prediction model for NB that is related to cancer stemness. By employing an ML approach, we identified potential hub biomarkers from mRNAsi phenotype-based DEGs. Among them, four genes (*ERCC6L*, *DUXAP10*, *NCAN*, and *DIRAS3*) were identified as promising cancer stemness-related biomarkers for NB. The four-gene signature remained an effective reflection of the NB stemness value while outperforming the mRNAsi subgrouping in terms of prognostic accuracy. These stemness-related hub genes are very likely to be crucial in NB progression and may hold promise as potential targets for treatment.

Drug resistance in treatment is a major challenge in the management of refractory cancers. It is widely accepted that CSCs play a significant role in the development of therapy resistance in tumors (14-16). During prolonged drug selection of NB cells, a subpopulation of cells with CSC characteristics becomes enriched (17). A recent study has shown that drug resistance induced by CSCs may be closely associated with immune evasion and the formation of a distinct tumor immune microenvironment (48). Therefore, targeting CSCs and enhancing the immune microenvironment represent promising strategies to overcome drug resistance and improve treatment efficacy. In recent research, mRNAsi was found to be higher in tumors with lower anti-tumor immunity, suggesting its strong association with the TME (22,23). We demonstrated through multiple algorithms that, in our stemness predictor gene model, low-risk score NB patients responded better to immunotherapy than did high-risk patients. Moreover,

the distribution of immune cells between high- and low-risk score NB patients is significantly different. Typically, there is a higher infiltration abundance of $CD4^+$ and $CD8^+$ T-cells in the low-risk group, reflecting a better immune activity in low-risk score patients. Interestingly, we observed higher levels of M2 type macrophages in the low-risk score group, corroborating earlier research reports. In NB, M2 type macrophage has been reported to be linked with a favorable outcome, and more M2-like macrophages have been found in stage 1/2 than in stage 4S (49,50). By conducting single-cell analysis, Liu *et al.* observed that M2-like macrophages were the predominant immune cell type within the NB tumor immune microenvironment, with their abundance decreasing progressively from favorable stages to advanced stages (50). This finding suggests that M2-like macrophages are likely to have a critical function in inhibiting tumor growth in NB, which is somewhat contradictory to their tumor-promoting characteristics observed in other cancers (51,52). Additional investigations are required to elucidate the precise subtypes and phenotypes of M2-like macrophages present in NB. In sum, our four-gene signatures established based on the NB mRNAsi confirmed the close relationship between CSCs and TME, and demonstrated its potential to predict response to immunotherapy in NB patients. The identified four hub genes may serve as potential targets to aid in the immunotherapy of NB.

Despite advances in the selection of CSCs in NB, targeting these cells effectively remains challenging (53). Our study examined the activity of signaling pathways associated with mRNAsi-related genes, and identified several drugs targeted at certain pathways that suggested better efficacy in the high-risk score group.

Among the four pivotal genes, *ERCC6L*, also known as PICH, is a DNA helicase that has recently been identified. It has been shown to be involved in embryonic development, suggesting its important role in growth regulation. *ERCC6L* dysfunction can lead to DNA damage, affecting the cell cycle and division, and promoting cancer development (54,55). Its significant role in tumorigenesis and progression has been demonstrated in many studies (56-59). The long non-coding RNA (lncRNA) *DUXAP10* is a pseudogene located on chromosome 14q11.2 and has been found to be upregulated in several types of cancer (60-62). Studies have demonstrated that upregulation of *DUXAP10* is linked to advanced disease features and poor outcome, indicating its potential as a prognostic biomarker for predicting cancer outcomes (63-65). Belonging to the lectican family, neurocan (*NCAN*) is a secreted chondroitin sulfate proteoglycan that inhibits axon regeneration after nervous injury (66). Increased *NCAN* expression levels have been associated with poor OS in NB patients (67). Additionally, *NCAN* overexpression in NB cells has been shown to stimulate sphere formation and enhance malignancy. Knockdown of *NCAN* led to the suppression of potential stemness markers. Based on these results, Su *et al.* proposed that *NCAN* may serve as a constituent of the extracellular matrix, creating a specific microenvironment that promotes the growth of CSCs (67). *DIRAS3*, also known as ARHI, is commonly downregulated or lost in various types of cancer and acts as a tumor suppressor (68-70). Our research has indicated that *DIRAS3* also exhibits a protective effect. The restoration of *DIRAS3* expression results in inhibited malignant biological behavior in various types of cancer (71-73). Nevertheless, it remains unclear whether *DIRAS3* can also suppress tumor growth in NB. To summarize, the mechanistic plausibility of our findings is supported by these genes' influence on the progression and prognosis of many cancers, including NB. However, some of these genes have not been thoroughly investigated in the context of NB and may represent potential targets for suppressing NB stemness features.

Finally, we developed a prognostic nomogram that provides a user-friendly scoring system. In the clinical setting, it is customary to forecast the prognosis of NB patients by evaluating their clinicopathological features. However, the accuracy of these predictions is often suboptimal. Our nomogram demonstrated prediction accuracy superior to that held by prognostic indicators currently used in clinical practice (stage, MYCN status, age at diagnosis), as indicated by the calibration curves. By combining the four-gene signatures with staging system, this nomogram has the potential to serve as a precise and efficient tool to help evaluate the outcome of NB patients. Such a tool could prove invaluable in facilitating medical decision-making.

Several limitations should be considered in this study. Firstly, the retrospective nature of the analysis and the reliance on bioinformatics approaches are significant drawbacks. Thus, it would be advantageous to obtain prospective sequencing data for future investigations. Secondly, while the stemness index-related signature and the developed prognostic nomogram have shown impressive predictive capabilities for NB survival, it is essential to validate these findings through expanded research involving larger sample sizes, comprehensive data from a broader

range of centers, and in-depth prospective studies. Thirdly, to ensure the model aligns with the needs and expectations of key stakeholders, including healthcare professionals and patients, comprehensive qualitative research is imperative. This will involve gathering perspectives on the prognostic model's utility, challenges, and potential integration into clinical practice, particularly in the context of its deployment as an AI medical device. The nomogram is a two-dimensional graphic scoring method that might be difficult to assign a value to in continuous variables. Alternative methods for simplifying the use of the prediction tool in common practice include web-based applications and scoring tables. As computerized clinical decision support systems (CCDSS), the online application will simplify the predictive model for use in general practice or external validation by other hospitals. A prior systematic review discovered that CCDSS improved care processes such as screening and treatment while having a challenging influence on patient outcomes, healthcare costs, and patient safety (74).

Conclusions

By utilizing the stemness index and ML methods, our study evaluated the prognostic value of CSC features in NB, confirming their role in the development and progression of the disease. For the first time, we constructed a novel mRNAsi-associated signature for NB that exhibited significant associations with prognosis, clinical characteristics, and tumor immune response in NB. The identification of four influential genes for NB stemness in this study improves on prior approaches to NB prognostication and identifies promising therapeutic targets for inhibiting NB stemness characteristics.

Acknowledgments

We express our gratitude to the GEO and ArrayExpress networks for generously sharing the gene expression data utilized in this study.
Funding: This research was supported by the Tianjin Health Science and Technology Project (No. TJWJ2022QN105 and No. TJWJ2022QN106), National Natural Science Foundation of China (No. 81903055), National Key R&D Program of China (No. 2018YFC1313000), and Tianjin Key Medical Discipline (Specialty) Construction Project (No. TJYXZDXK-009A).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://tp.amegroups.com/article/view/10.21037/tp-23-582/rc

Peer Review File: Available at https://tp.amegroups.com/article/view/10.21037/tp-23-582/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tp.amegroups.com/article/view/10.21037/tp-23-582/coif). H.D.J.H. has been supported in reviewing this manuscript through the National Institute for Health Research (NIHR) doctoral fellowship award, which had no role in the design or delivery of this study. T.H. is an employee of Philips and an editor for several scientific journals, outside the submitted work. The other 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).

References

1. Kattner P, Strobel H, Khoshnevis N, et al. Compare and contrast: pediatric cancer versus adult malignancies. Cancer Metastasis Rev 2019;38:673-82.
2. Marshall GM, Carter DR, Cheung BB, et al. The prenatal origins of cancer. Nat Rev Cancer 2014;14:277-89.
3. Chung C, Boterberg T, Lucas J, et al. Neuroblastoma. Pediatr Blood Cancer 2021;68 Suppl 2:e28473.
4. Newman EA, Abdessalam S, Aldrink JH, et al. Update on

neuroblastoma. J Pediatr Surg 2019;54:383-9.
5. Tang J, Lu H, Yang Z, et al. Associations between WTAP gene polymorphisms and neuroblastoma susceptibility in Chinese children. Transl Pediatr 2021;10:146-52.
6. Qiu B, Matthay KK. Advancing therapy for neuroblastoma. Nat Rev Clin Oncol 2022;19:515-33.
7. Reya T, Morrison SJ, Clarke MF, et al. Stem cells, cancer, and cancer stem cells. Nature 2001;414:105-11.
8. Ayob AZ, Ramasamy TS. Cancer stem cells as key drivers of tumour progression. J Biomed Sci 2018;25:20.
9. Phan TG, Croucher PI. The dormant cancer cell life cycle. Nat Rev Cancer 2020;20:398-411.
10. Singh SK, Clarke ID, Terasaki M, et al. Identification of a cancer stem cell in human brain tumors. Cancer Res 2003;63:5821-8.
11. Buhagiar A, Ayers D. Chemoresistance, cancer stem cells, and miRNA influences: the case for neuroblastoma. Anal Cell Pathol (Amst) 2015;2015:150634.
12. Tian X, Cao F, Li X, et al. Tumor dormancy is closely related to prognosis prediction and tumor immunity in neuroblastoma. Transl Pediatr 2023;12:445-61.
13. Pandian V, Ramraj S, Khan FH, et al. Metastatic neuroblastoma cancer stem cells exhibit flexible plasticity and adaptive stemness signaling. Stem Cell Res Ther 2015;6:2.
14. Anjomshoaa A, Nasri S, Humar B, et al. Slow proliferation as a biological feature of colorectal cancer metastasis. Br J Cancer 2009;101:822-8.
15. Moore N, Lyle S. Quiescent, slow-cycling stem cell populations in cancer: a review of the evidence and discussion of significance. J Oncol 2011;2011:396076.
16. Damen MPF, van Rheenen J, Scheele CLGJ. Targeting dormant tumor cells to prevent cancer recurrence. FEBS J 2021;288:6286-303.
17. Zheng X, Naiditch J, Czurylo M, et al. Differential effect of long-term drug selection with doxorubicin and vorinostat on neuroblastoma cells with cancer stem cell characteristics. Cell Death Dis 2013;4:e740.
18. Walton JD, Kattan DR, Thomas SK, et al. Characteristics of stem cells from human neuroblastoma cell lines and in tumors. Neoplasia 2004;6:838-45.
19. Veschi V, Verona F, Thiele CJ. Cancer Stem Cells and Neuroblastoma: Characteristics and Therapeutic Targeting Options. Front Endocrinol (Lausanne) 2019;10:782.
20. Malta TM, Sokolov A, Gentles AJ, et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell 2018;173:338-354.e15.
21. 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.
22. Chen D, Liu J, Zang L, et al. Integrated Machine Learning and Bioinformatic Analyses Constructed a Novel Stemness-Related Classifier to Predict Prognosis and Immunotherapy Responses for Hepatocellular Carcinoma Patients. Int J Biol Sci 2022;18:360-73.
23. 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.
24. Clough E, Barrett T. The Gene Expression Omnibus Database. Methods Mol Biol 2016;1418:93-110.
25. Liberzon A, Birger C, Thorvaldsdóttir H, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 2015;1:417-25.
26. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284-7.
27. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
28. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013;4:2612.
29. 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.
30. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol 2017;18:220.
31. Schubert M, Klinger B, Klünemann M, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. Nat Commun 2018;9:20.
32. 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.
33. Charoentong P, Finotello F, Angelova M, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep 2017;18:248-62.
34. Yang W, Soares J, Greninger P, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res 2013;41:D955-61.

35. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. Genome Biol 2014;15:R47.
36. Laut AK, Dorneburg C, Fürstberger A, et al. CHD5 inhibits metastasis of neuroblastoma. Oncogene 2022;41:622-33.
37. Liu Z, Zhang X, Xu M, et al. Loss of CASZ1 tumor suppressor linked to oncogenic subversion of neuroblastoma core regulatory circuitry. Cell Death Dis 2022;13:871.
38. Downs-Canner SM, Meier J, Vincent BG, et al. B Cell Function in the Tumor Microenvironment. Annu Rev Immunol 2022;40:169-93.
39. Pine SR, Liu W. Asymmetric cell division and template DNA co-segregation in cancer stem cells. Front Oncol 2014;4:226.
40. Brabletz T, Jung A, Spaderna S, et al. Opinion: migrating cancer stem cells - an integrated concept of malignant tumour progression. Nat Rev Cancer 2005;5:744-9.
41. Maugeri-Saccà M, Vigneri P, De Maria R. Cancer stem cells and chemosensitivity. Clin Cancer Res 2011;17:4942-7.
42. Wilson A, Laurenti E, Oser G, et al. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell 2008;135:1118-29.
43. Chen J, Li Y, Yu TS, et al. A restricted cell population propagates glioblastoma growth after chemotherapy. Nature 2012;488:522-6.
44. Yue ZX, Xing TY, Zhao W, et al. MYCN amplification plus 1p36 loss of heterozygosity predicts ultra high risk in bone marrow metastatic neuroblastoma. Cancer Med 2022;11:1837-49.
45. Tsai SC, Lin CC, Shih TC, et al. The miR-200b-ZEB1 circuit regulates diverse stemness of human hepatocellular carcinoma. Mol Carcinog 2017;56:2035-47.
46. Xia Y, Li X, Tian X, et al. Identification of a Five-Gene Signature Derived From MYCN Amplification and Establishment of a Nomogram for Predicting the Prognosis of Neuroblastoma. Front Mol Biosci 2021;8:769661.
47. Gupta M, Kannappan S, Jain M, et al. Development and validation of a 21-gene prognostic signature in neuroblastoma. Sci Rep 2023;13:12526.
48. Nallasamy P, Nimmakayala RK, Parte S, et al. Tumor microenvironment enriches the stemness features: the architectural event of therapy resistance and metastasis. Mol Cancer 2022;21:225.
49. Stahl D, Knoll R, Gentles AJ, et al. Prognostic Gene Expression, Stemness and Immune Microenvironment in Pediatric Tumors. Cancers (Basel) 2021;13:854.
50. Liu Q, Wang Z, Jiang Y, et al. Single-cell landscape analysis reveals distinct regression trajectories and novel prognostic biomarkers in primary neuroblastoma. Genes Dis 2022;9:1624-38.
51. Chen Y, Zhang S, Wang Q, et al. Tumor-recruited M2 macrophages promote gastric and breast cancer metastasis via M2 macrophage-secreted CHI3L1 protein. J Hematol Oncol 2017;10:36.
52. Yamaguchi T, Fushida S, Yamamoto Y, et al. Tumor-associated macrophages of the M2 phenotype contribute to progression in gastric cancer with peritoneal dissemination. Gastric Cancer 2016;19:1052-65.
53. Farina AR, Cappabianca LA, Zelli V, et al. Mechanisms involved in selecting and maintaining neuroblastoma cancer stem cell populations, and perspectives for therapeutic targeting. World J Stem Cells 2021;13:685-736.
54. Abbasi R, Ramroth H, Becher H, et al. Laryngeal cancer risk associated with smoking and alcohol consumption is modified by genetic polymorphisms in ERCC5, ERCC6 and RAD23B but not by polymorphisms in five other nucleotide excision repair genes. Int J Cancer 2009;125:1431-9.
55. Hübner NC, Wang LH, Kaulich M, et al. Re-examination of siRNA specificity questions role of PICH and Tao1 in the spindle checkpoint and identifies Mad2 as a sensitive target for small RNAs. Chromosoma 2010;119:149-65.
56. Zhang G, Ma J, Xiong J, et al. Upregulation of Excision Repair Cross-Complementation Group 6-Like (ERCC6L) Promotes Tumor Growth in Hepatocellular Carcinoma. Dig Dis Sci 2021;66:1097-109.
57. Chen D, Liu Q, Cao G. ERCC6L promotes cell growth and metastasis in gastric cancer through activating NF-κB signaling. Aging (Albany NY) 2021;13:20218-28.
58. Xie Y, Yu J, Wang F, et al. ERCC6L promotes cell growth and invasion in human colorectal cancer. Oncol Lett 2019;18:237-46.
59. Huang X, Jiang L, Lu S, et al. Overexpression of ERCC6L correlates with poor prognosis and confers malignant phenotypes of lung adenocarcinoma. Oncol Rep 2022;48:131.
60. Zhu Q, Liu J, Tang J, et al. Overexpression of long non-coding RNAs DUXAP9 and DUXAP10 is associated with prognosis in patients with hepatocellular carcinoma after hepatectomy. Int J Clin Exp Pathol 2018;11:1407-14.
61. Wu B, Yang C, Fang Y, et al. Long noncoding RNA DUXAP10 promotes the stemness of glioma cells by

recruiting HuR to enhance Sox12 mRNA stability. Environ Toxicol 2021;36:840-9.
62. Lv XY, Ma L, Chen JF, et al. Knockdown of DUXAP10 inhibits proliferation and promotes apoptosis in bladder cancer cells via PI3K/Akt/mTOR signaling pathway. Int J Oncol 2018;52:288-94.
63. Chen J, Wang XF, Qin YC, et al. Downregulation of long non-coding RNA DUXAP10 inhibits proliferation, migration, and invasion of renal cell carcinoma. Eur Rev Med Pharmacol Sci 2020;24:11041-51.
64. Lian Y, Xu Y, Xiao C, et al. The pseudogene derived from long non-coding RNA DUXAP10 promotes colorectal cancer cell growth through epigenetically silencing of p21 and PTEN. Sci Rep 2017;7:7312.
65. Lin HP, Wang Z, Yang C. LncRNA DUXAP10 Upregulation and the Hedgehog Pathway Activation Are Critically Involved in Chronic Cadmium Exposure-Induced Cancer Stem Cell-Like Property. Toxicol Sci 2021;184:33-45.
66. Asher RA, Morgenstern DA, Fidler PS, et al. Neurocan is upregulated in injured brain and in cytokine-treated astrocytes. J Neurosci 2000;20:2427-38.
67. Su Z, Kishida S, Tsubota S, et al. Neurocan, an extracellular chondroitin sulfate proteoglycan, stimulates neuroblastoma cells to promote malignant phenotypes. Oncotarget 2017;8:106296-310.
68. Zhang S, Feng XL, Shi L, et al. Genome-wide analysis of DNA methylation in tongue squamous cell carcinoma. Oncol Rep 2013;29:1819-26.
69. Zou CF, Jia L, Jin H, et al. Re-expression of ARHI (DIRAS3) induces autophagy in breast cancer cells and enhances the inhibitory effect of paclitaxel. BMC Cancer 2011;11:22.
70. Li J, Cui G, Sun L, et al. ARHI overexpression induces epithelial ovarian cancer cell apoptosis and excessive autophagy. Int J Gynecol Cancer 2014;24:437-43.
71. Sutton MN, Lu Z, Li YC, et al. DIRAS3 (ARHI) Blocks RAS/MAPK Signaling by Binding Directly to RAS and Disrupting RAS Clusters. Cell Rep 2019;29:3448-3459.e6.
72. Zhao X, Li J, Zhuo J, et al. Reexpression of ARHI inhibits tumor growth and angiogenesis and impairs the mTOR/VEGF pathway in hepatocellular carcinoma. Biochem Biophys Res Commun 2010;403:417-21.
73. Badgwell DB, Lu Z, Le K, et al. The tumor-suppressor gene ARHI (DIRAS3) suppresses ovarian cancer cell migration through inhibition of the Stat3 and FAK/Rho signaling pathways. Oncogene 2012;31:68-79.
74. Souza NM, Sebaldt RJ, Mackay JA, et al. Computerized clinical decision support systems for primary preventive care: a decision-maker-researcher partnership systematic review of effects on process of care and patient outcomes. Implement Sci 2011;6:87.

Cite this article as: Xia Y, Wang C, Li X, Gao M, Hogg HDJ, Tunthanathip T, Hulsen T, Tian X, Zhao Q. Development and validation of a novel stemness-related prognostic model for neuroblastoma using integrated machine learning and bioinformatics analyses. Transl Pediatr 2024;13(1):91-109. doi: 10.21037/tp-23-582

Supplementary

Table S1 Patients Characteristics of the GSE49710 and E-MTAB-8248 dataset

Characteristic	Subtypes	GSE49716	E-MTAB-8248
		Overall, mean ± SD or N (%)	Overall, mean ± SD or N (%)
N		498	223
Age (days)		758±1031	1040 ±1320
Age_group	<18 months	300 (60.2)	103 (46.2)
	≥18 months	198 (39.8)	120 (53.8)
Status	Live	393 (78.9)	181 (81.2)
	Death	105 (21.1)	42 (18.8)
Overall survival (days)		2185.42±1370.89	2230±1450
INSS_STAGE	1	121 (24.3)	29 (13.0)
	2	78 (15.7)	39 (17.5)
	3	63 (12.7)	36 (16.1)
	4	183 (36.7)	89 (39.9)
	4S	53 (10.6)	30 (13.5)
MYCN status	Non-amplification	401 (80.5)	176 (78.9)
	Amplification	92 (18.5)	46 (20.6)
	No data	5 (1.0)	1 (0.4)

https://dx.doi.org/10.21037/tp-23-582

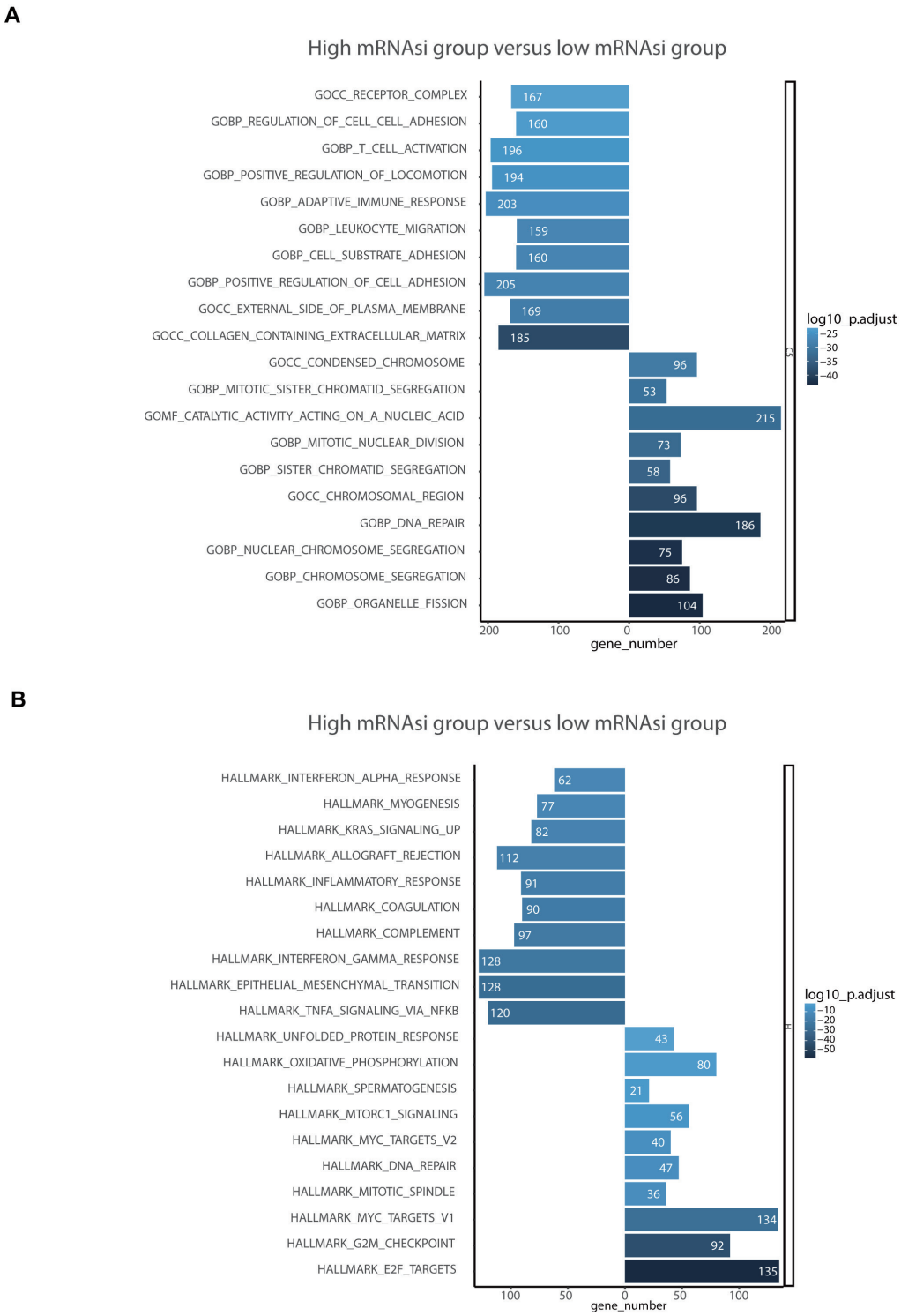

Figure S1 The GSEA of DEGs between high and low mRNAsi. (A) The top 10 GO gene sets enrichment (C5). (B) The top 10 hallmark gene sets enrichment. The gene sets with NES <0 are listed on the left, and the gene sets with NES >0 are listed on the right. GO, Gene Ontology; NES, normalized enrichment score; GSEA, gene set enrichment analysis; DEGs, differentially expressed genes; mRNAsi, messenger RNA expression-based stemness index.

https://dx.doi.org/10.21037/tp-23-582

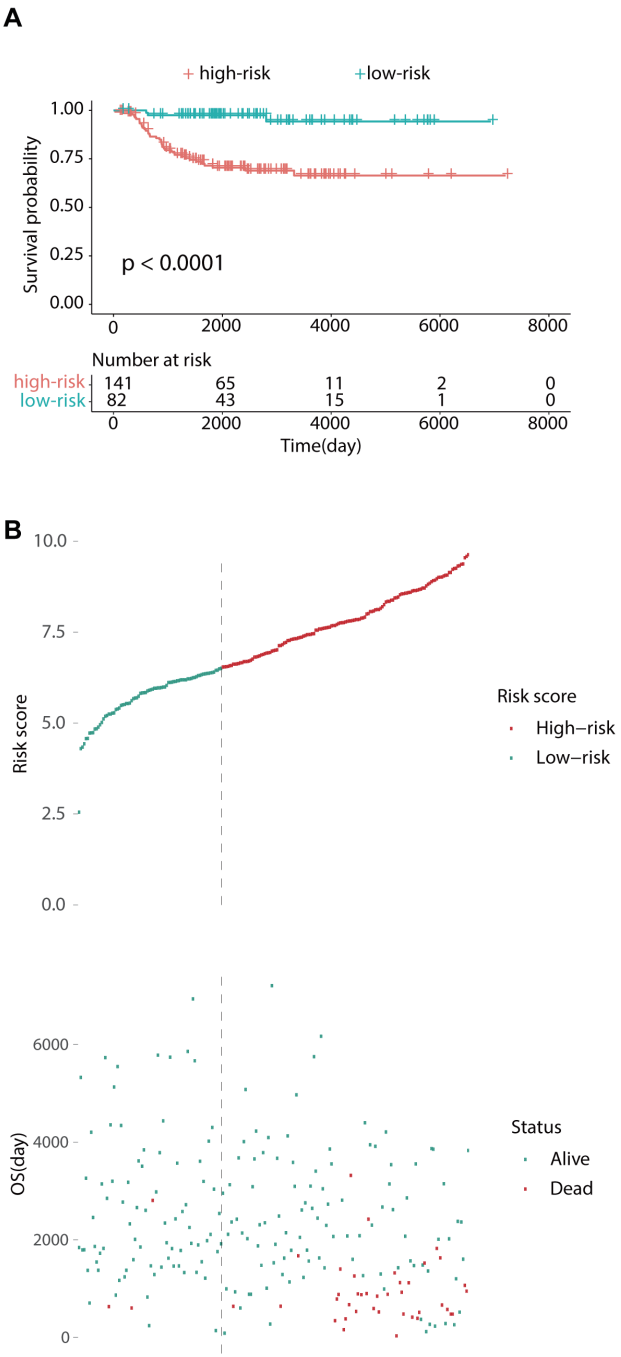

Figure S2 The external verification of the four-gene signature. (A) The KM curve for OS predicting of the four-gene signature in E-MTAB-8248. (B) The distribution of risk scores, the association of risk scores and OS in E-MTAB-8248. KM, Kaplan-Meier curve; OS, overall survival.

https://dx.doi.org/10.21037/tp-23-582

Table S2 Comparison of immune infiltration-related genes as assessed by ESTIMATE, CIBERSORT, MCPcounter, and xCell

Sig_names	P.value	High	Low	Statistic	P.adj	Log10pvalue	Significance
pro_B-cells_xCell	7.59E-52	0.624503	-0.65648	1.28098	7.51E-50	51.11974	****
Melanocytes_xCell	2.57E-47	-0.60958	0.588222	-1.1978	1.27E-45	46.58964	****
Tgd_cells_xCell	1.52E-44	0.602928	-0.61795	1.220878	5.03E-43	43.81732	****
Th2_cells_xCell	8.38E-33	0.511512	-0.5229	1.03441	2.07E-31	32.07679	****
Th1_cells_xCell	5.91E-30	0.49735	-0.50403	1.001382	1.17E-28	29.22868	****
Osteoblast_xCell	1.50E-26	0.432907	-0.46269	0.895594	2.47E-25	25.82401	****
HSC_xCell	3.00E-25	-0.45143	0.415564	-0.86699	4.24E-24	24.52357	****
CLP_xCell	2.24E-23	0.412387	-0.42497	0.837353	2.77E-22	22.64973	****
Erythrocytes_xCell	8.44E-23	0.361926	-0.43594	0.797871	9.28E-22	22.07385	****
CD4+_naive_T-cells_xCell	6.26E-22	-0.40959	0.365033	-0.77462	6.20E-21	21.20322	****
CD8+_naive_T-cells_xCell	8.41E-22	0.408336	-0.42712	0.835461	7.57E-21	21.07499	****
Mast_cells_xCell	1.19E-20	-0.41393	0.408401	-0.82233	9.82E-20	19.92451	****
Macrophages_M2_xCell	1.77E-17	-0.36736	0.319039	-0.6864	1.35E-16	16.75229	****
CD4+_Tem_xCell	4.49E-17	-0.35314	0.322876	-0.67602	3.18E-16	16.3475	****
Mesangial_cells_xCell	1.16E-16	-0.373	0.365348	-0.73835	7.63E-16	15.9373	****
Sebocytes_xCell	2.21E-15	-0.35344	0.324693	-0.67813	1.37E-14	14.65609	****
Monocytic_lineage_MCPcounter	1.31E-14	-0.31702	0.331742	-0.64876	7.65E-14	13.88132	****
Plasma_cells_xCell	7.90E-14	0.292339	-0.34219	0.634534	4.35E-13	13.10226	****
aDC_xCell	2.04E-13	-0.31209	0.344778	-0.65686	1.06E-12	12.69127	****
Myeloid_dendritic_cells_MCPcounter	4.65E-13	-0.30445	0.309747	-0.6142	2.30E-12	12.33209	****
CD8+_Tem_xCell	5.25E-13	-0.28343	0.243849	-0.52728	2.48E-12	12.2797	****
ImmuneScore_estimate	2.00E-12	-0.30277	0.28725	-0.59002	9.00E-12	11.69908	****
ESTIMATEScore_estimate	2.19E-12	-0.29812	0.296791	-0.59491	9.41E-12	11.66027	****
Class-switched_memory_B-cells_xCell	7.58E-12	-0.2531	0.165702	-0.4188	3.13E-11	11.12053	****
CD8+_T-cells_xCell	1.31E-11	-0.28693	0.238596	-0.52552	5.20E-11	10.88188	****
CD4+_T-cells_xCell	5.61E-11	-0.28669	0.250334	-0.53703	2.14E-10	10.25117	****
Neurons_xCell	7.02E-11	-0.20129	0.282099	-0.48339	2.57E-10	10.15364	****
MEP_xCell	5.33E-10	0.227634	-0.38165	0.609285	1.89E-09	9.273112	****
NK_cells_MCPcounter	1.97E-09	-0.25785	0.244688	-0.50254	6.71E-09	8.706543	****
Megakaryocytes_xCell	2.07E-09	-0.26074	0.246052	-0.50679	6.84E-09	8.683333	****
StromalScore_estimate	3.09E-09	-0.25252	0.255219	-0.50774	9.88E-09	8.509465	****
CD4+_Tcm_xCell	3.46E-09	-0.28923	0.222837	-0.51207	1.07E-08	8.460616	****
Tregs_xCell	6.80E-09	-0.24958	0.126033	-0.37561	2.04E-08	8.167262	****
Macrophages_M2_CIBERSORT	9.37E-09	-0.25036	0.19064	-0.441	2.73E-08	8.028062	****
Endothelial_cells_MCPcounter	2.19E-08	-0.21682	0.221515	-0.43834	6.21E-08	7.658569	****
T_cells_MCPcounter	3.51E-08	-0.24834	0.219247	-0.46759	9.66E-08	7.454143	****
Platelets_xCell	1.13E-07	-0.23725	0.219874	-0.45712	3.01E-07	6.9486	****
B_cells_memory_CIBERSORT	1.40E-07	0.117725	-0.23613	0.35385	3.66E-07	6.852464	****
cDC_xCell	1.78E-07	-0.23876	0.215837	-0.4546	4.51E-07	6.75046	****
Monocytes_xCell	2.48E-07	-0.20594	0.147721	-0.35366	6.13E-07	6.606303	****
iDC_xCell	1.94E-06	0.200012	-0.22561	0.425625	4.68E-06	5.712619	****
Plasma_cells_CIBERSORT	2.89E-06	0.063779	-0.20606	0.269841	6.81E-06	5.539192	****
Smooth_muscle_xCell	3.09E-06	0.1174	-0.27333	0.390731	7.11E-06	5.51046	****
Cytotoxic_lymphocytes_MCPcounter	5.61E-06	-0.20408	0.19616	-0.40024	1.26E-05	5.250774	****
CMP_xCell	6.58E-06	0.146851	-0.18558	0.332435	1.45E-05	5.182073	****
T_cells_CD4_memory_resting_CIBERSORT	7.84E-06	-0.21436	0.129451	-0.34381	1.69E-05	5.105585	****
Chondrocytes_xCell	3.80E-05	-0.18345	0.142869	-0.32632	8.01E-05	4.419815	****
Adipocytes_xCell	4.13E-05	-0.1941	0.158121	-0.35223	8.53E-05	4.38364	****
Skeletal_muscle_xCell	5.40E-05	0.035792	-0.17556	0.211354	1.09E-04	4.267362	****
CD8+_Tcm_xCell	7.08E-05	-0.18586	0.122644	-0.30851	1.40E-04	4.150135	****
Fibroblasts_MCPcounter	1.50E-04	-0.157	0.19706	-0.35406	2.91E-04	3.823753	***
Eosinophils_xCell	1.84E-04	-0.17355	0.154897	-0.32845	3.47E-04	3.735455	***
Neutrophils_xCell	1.86E-04	0.10236	-0.18902	0.291377	3.47E-04	3.730593	***
Hepatocytes_xCell	2.51E-04	-0.12878	-0.05211	-0.07667	4.60E-04	3.600515	***
Macrophages_xCell	3.20E-04	-0.17599	0.141214	-0.3172	5.75E-04	3.494816	***
ly_Endothelial_cells_xCell	3.25E-04	-0.1372	0.129039	-0.26624	5.75E-04	3.487546	***
B_cells_naive_CIBERSORT	0.001635	-0.13531	0.029551	-0.16486	0.00284	2.786407	**
MSC_xCell	0.002266	0.131946	-0.12956	0.261509	0.003868	2.644725	**
NK_cells_resting_CIBERSORT	0.002739	-0.01073	-0.17912	0.168389	0.004537	2.562474	**
CD8_T_cells_MCPcounter	0.00275	-0.12974	0.114663	-0.2444	0.004537	2.560687	**
Macrophages_M0_CIBERSORT	0.004947	-0.12929	0.096471	-0.22576	0.008028	2.305689	**
Myocytes_xCell	0.011185	-0.10466	-0.03279	-0.07187	0.017859	1.951379	*
Mast_cells_activated_CIBERSORT	0.013301	-0.17647	0.082774	-0.25925	0.020902	1.876115	*
NKT_xCell	0.014771	-0.13352	0.097664	-0.23119	0.022849	1.830596	*
DC_xCell	0.016906	-0.10336	0.108747	-0.2121	0.025749	1.771962	*
Preadipocytes_xCell	0.018384	0.092771	-0.12549	0.218264	0.027576	1.735564	*
Monocytes_CIBERSORT	0.025282	-0.10741	0.001993	-0.1094	0.037357	1.597192	*
pDC_xCell	0.027544	-0.11099	0.030046	-0.14104	0.040101	1.559968	*
Neutrophils_CIBERSORT	0.032906	-0.0526	-0.10597	0.053378	0.047213	1.482727	*
naive_B-cells_xCell	0.036533	-0.09526	-0.02248	-0.07278	0.051669	1.43731	*
T_cells_follicular_helper_CIBERSORT	0.041315	0.005431	-0.13411	0.139544	0.057608	1.383891	*
T_cells_CD8_CIBERSORT	0.048048	-0.12291	0.041053	-0.16396	0.066067	1.318321	*
MPP_xCell	0.060971	0.062202	-0.12066	0.182866	0.082687	1.214877	+
Memory_B-cells_xCell	0.066906	-0.09823	-0.04316	-0.05506	0.089509	1.174534	+
Macrophages_M1_CIBERSORT	0.084148	-0.06063	-0.04395	-0.01668	0.111075	1.074956	+
Pericytes_xCell	0.097515	0.06508	-0.07266	0.137739	0.127026	1.010927	+
Macrophages_M1_xCell	0.118875	-0.06599	0.068703	-0.13469	0.15284	0.924908	+
T_cells_CD4_naive_CIBERSORT	0.136547	-0.11527	-0.04897	-0.06629	0.172592	0.864716	+
NK_cells_activated_CIBERSORT	0.137725	-0.10295	0.033783	-0.13673	0.172592	0.860987	+
T_cells_CD4_memory_activated_CIBERSORT	0.143946	-0.11947	-0.05846	-0.06101	0.178133	0.841801	+
Basophils_xCell	0.153977	0.034953	-0.10776	0.142709	0.188194	0.812544	+
mv_Endothelial_cells_xCell	0.223623	-0.04889	0.026174	-0.07507	0.269984	0.650484	+
Dendritic_cells_activated_CIBERSORT	0.238855	-0.04326	-0.08503	0.041776	0.284899	0.621866	+
T_cells_gamma_delta_CIBERSORT	0.27677	-0.06823	-0.10448	0.036254	0.326194	0.55788	+
Mast_cells_resting_CIBERSORT	0.308185	-0.04197	-0.08704	0.045073	0.358945	0.511188	+
Endothelial_cells_xCell	0.318709	-0.04281	0.03971	-0.08252	0.366886	0.496605	+
Eosinophils_CIBERSORT	0.362759	-0.02664	-0.05918	0.032536	0.412795	0.440381	+
B_lineage_MCPcounter	0.395245	-0.05244	-0.00442	-0.04802	0.444651	0.403133	+
CD4+_memory_T-cells_xCell	0.425404	0.001076	-0.05692	0.057995	0.473202	0.371198	+
Fibroblasts_xCell	0.445453	-0.10574	-0.00843	-0.09731	0.488115	0.351198	+
T_cells_regulatory_(Tregs)_CIBERSORT	0.448672	-0.05767	-0.02854	-0.02914	0.488115	0.348071	+
Epithelial_cells_xCell	0.459906	-0.05085	9.07E-04	-0.05176	0.494899	0.337331	+
Neutrophils_MCPcounter	0.487413	-0.02897	-0.06849	0.039523	0.518859	0.312103	+
NK_cells_xCell	0.529771	-0.04409	-0.02236	-0.02172	0.557951	0.275911	+
B-cells_xCell	0.537145	-0.0503	-0.06909	0.018789	0.559762	0.269908	+
Astrocytes_xCell	0.668341	-0.05423	-0.04964	-0.00459	0.689227	0.175002	+
Keratinocytes_xCell	0.704456	-0.01324	-0.00849	-0.00474	0.718981	0.152146	+
GMP_xCell	0.941392	0.002616	-0.01869	0.021307	0.950998	0.026229	+
Dendritic_cells_resting_CIBERSORT	0.988882	-0.08526	-0.07567	-0.00959	0.988882	0.004856	+

*, P<0.05; **, P<0.01; ***, P<0.001; ****, P<0.0001; +, not significant.

https://dx.doi.org/10.21037/tp-23-582

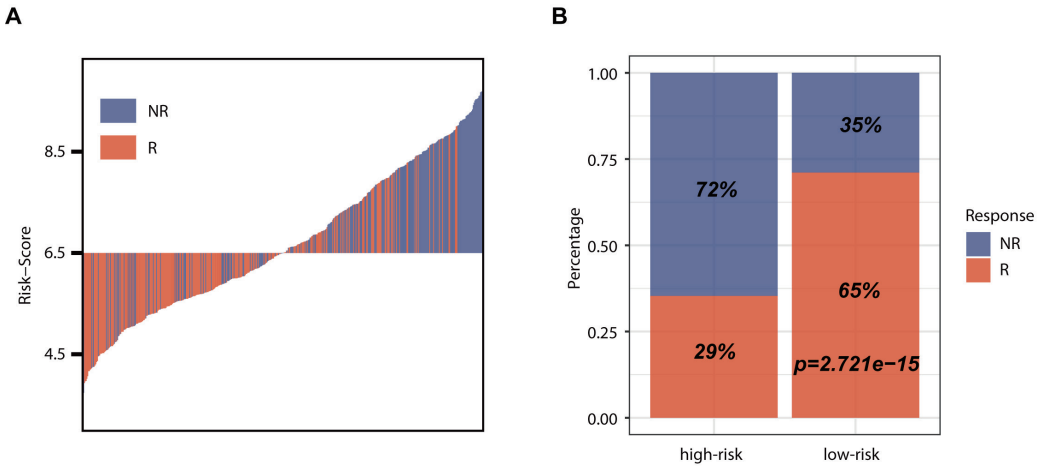

Figure S3 Differences in immunotherapy estimated response between high- and low-risk score groups revealed by TIDE algorithm. (A) Risk score stratification reveals diverse immune response patterns in patients. (B) Proportion of patients responding to immunotherapy in each risk group. R, response; NR, no response

https://dx.doi.org/10.21037/tp-23-582

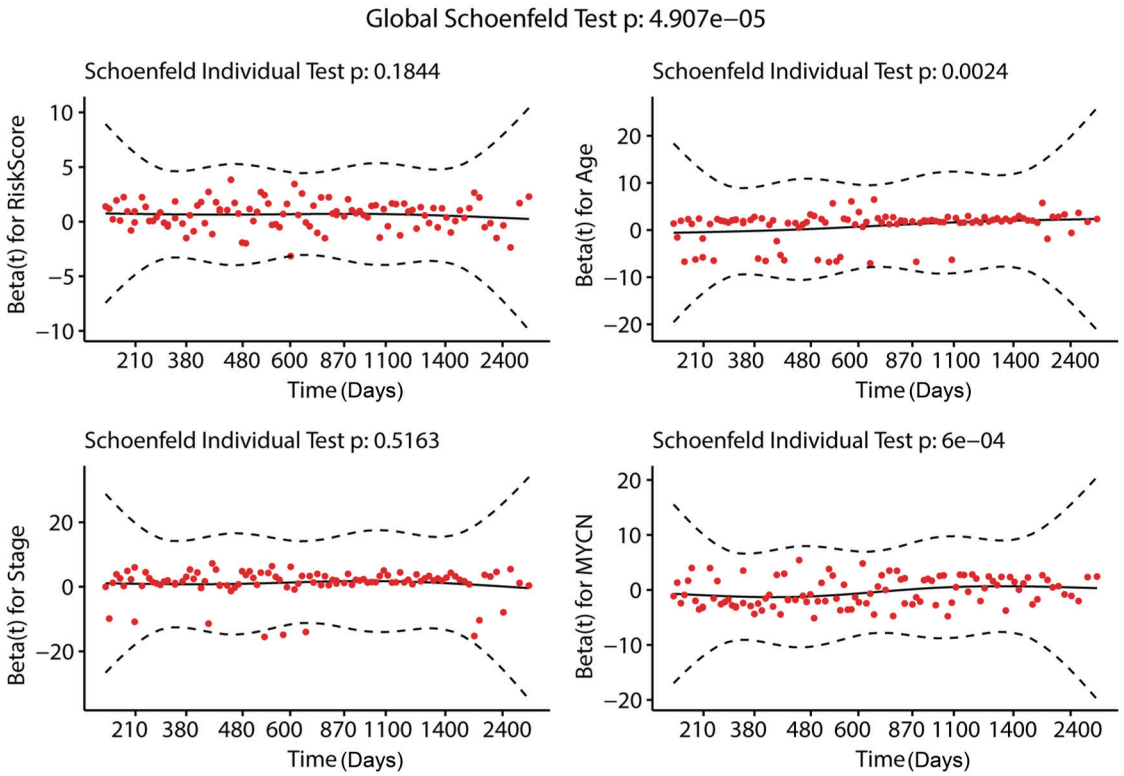

Figure S4 Schoenfeld test on proportional hazard assumption for the nomogram model. The Schoenfeld test for estimating time-varying covariance of the nomogram model.

https://dx.doi.org/10.21037/tp-23-582