frontiers | Research Topics

Advances in genetic and epigenetic mechanisms of therapeutic resistance in cancer

Edited by
Yilin Zhang and Peter Hart

Published in
Frontiers in Genetics

ISSN 1664-8714
ISBN 978-2-8325-2741-2
DOI 10.3389/978-2-8325-2741-2

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Advances in genetic and epigenetic mechanisms of therapeutic resistance in cancer

Topic editors
Yilin Zhang — The University of Chicago, United States
Peter Hart — Roosevelt University College of Pharmacy, United States

Citation
Zhang, Y., Hart, P., eds. (2023). *Advances in genetic and epigenetic mechanisms of therapeutic resistance in cancer*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-2741-2

Table of contents

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 05 August 2022
DOI 10.3389/fgene.2022.947000

OPEN ACCESS

EDITED BY
Peter Hart,
Roosevelt University College of Pharmacy, United States

REVIEWED BY
Chunlin Zhuang,
Second Military Medical University, China
Hongyan Guo,
Louisiana State University Health Science Center Shreveport, United States

*CORRESPONDENCE
Jing Chu,
chujing198603@163.com

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 18 May 2022
ACCEPTED 30 June 2022
PUBLISHED 05 August 2022

CITATION
Jing Chu (2022), Development of a prognostic model for children with neuroblastoma based on necroptosis-related genes.
Front. Genet. 13:947000.
doi: 10.3389/fgene.2022.947000

Development of a prognostic model for children with neuroblastoma based on necroptosis-related genes

Jing Chu*

Department of Pathology, Anhui Provincial Children's Hospital, Hefei, China

Background: Neuroblastoma (NBL) is a rare malignant tumor of the peripheral sympathetic nervous system in children with a low overall survival rate. Recent studies have revealed the important role of necroptosis in the occurrence and development of many kinds of tumors. In this study, a prognostic model based on necroptosis-related genes was constructed for NBL.

Methods: Expression profiles and clinical information for patients with NBL were downloaded from TARGET. Data for necroptosis-related genes were extracted for Cox regression and lasso regression analyses to evaluate factors associated with prognosis and to construct a prognostic model. Data from the GEO datasets GSE62564 and GSE85047 were used for external verification. Associations between risk scores were calculated, and immune infiltration, drug sensitivity, and mutation analyses were conducted. Functional enrichment analyses of genes in the prognostic model were performed.

Results: Six necroptosis-related genes (i.e., *CYLD, JAK1, APC, ERH, CNBP*, and *BAX*) were selected to construct a prognostic risk model. The risk score was highly correlated with levels of infiltration of multiple immune cells and sensitivity to common antineoplastic drugs. In addition, the risk score was identified as an independent prognostic factor for patients with NBL.

Conclusion: We constructed and validated a prognostic model based on necroptosis-related genes, providing insights into the development and progression of NBL and a basis for improved management. In addition to providing a tool for clinical decision-making, these findings support the importance of necroptosis in NBL and may guide the development of therapeutic strategies targeting this process.

KEYWORDS

neuroblastoma, tumor microenvironment, necroptosis, risk score, immunotherapy

Abbreviations: GN, Ganglioneuroma; GO, Gene ontology; GSEA, Gene set enrichment analysis; GSVA, Gene set variation analysis; IHC, Immunohistochemistry; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, Least absolute shrinkage and selection operator; NBL, Neuroblastoma; ROC, Receiver operating characteristic.

Introduction

Neuroblastoma (NBL) is a malignant tumor of the peripheral nervous system originating from primitive neural crest cells, accounting for 8%–10% of all malignant tumors and approximately 15% of tumor-related deaths in children. The morbidity is slightly higher in boys than in girls (ratio 1.2:1). The incidence peaks at 0–4 years of age, with a median age of 23 months (Park et al., 2010). NBL shows clinical and biological heterogeneity, and the disease spectrum ranges from spontaneous regression under no medical intervention or differentiation to an aggressive state with treatment resistance and tumor metastasis, despite intensive treatment. Therapy based on risk stratification by clinicopathological (diagnostic age, clinical staging, and histopathology) and genetic factors (MYCN amplification) significantly improves prognosis in low- and medium-risk patients, with 5-years survival rates ranging from 70% to 98%. However, about 50% of patients have high-risk characteristics with a 5-years survival rate after diagnosis of less than 40% (Fusco et al., 2018). Therefore, a comprehensive understanding of the pathogenesis of NBL, biomarker identification, and the development of an effective prognostic model are of great significance for improving outcomes in NBL.

Programmed cell death is a natural barrier to the occurrence and development of cancer and can be classified as apoptotic and non-apoptotic, including ferroptosis, pyroptosis, autophagy, and necroptosis (Dai et al., 2020). Evasion and resistance to programmed cell death are acquired by cancer cells (Hanahan and Weinberg, 2011). Resistance to apoptosis is an important cause of chemotherapeutic drug resistance in patients with cancer (Johnstone et al., 2002). Therefore, it is imperative to develop methods to induce non-apoptotic forms of programmed cell death as alternative therapeutic approaches. Necroptosis is a recently caspase-independent mechanism of cell death. It is mainly mediated by receptor-interacting protein kinase-1 (RIPK1) and -3 (RIPK3) and their target, mixed lineage kinase domain-like (MLKL). It is related to a variety of human diseases, including ischemia-reperfusion injury, inflammation, allograft rejection, neurodegenerative diseases, autoimmune diseases, and cancer (Negroni et al., 2020). Necroptosis plays dual roles in cancer development. On the one hand, adaptation to necroptosis in the tumor microenvironment promotes metastasis, suggesting that the inhibition of necroptosis is an anti-metastasis strategy. On the other hand, the expression levels of key mediators of necroptosis in some cancers are downregulated, suggesting that necroptosis has anticancer effects (Najafov et al., 2017).

However, the prognostic value of necroptosis-related genes in children with NB has not been evaluated. In this study, the association between necroptosis-related genes and prognosis in NB was evaluated and a prognostic model was constructed. These findings provide insight into the prognostic value of genes related to necroptosis and preliminarily uncover the complex biological functions and immunoregulatory effects of these genes and their regulatory networks.

Materials and methods

Data acquisition

The TARGET database (https://portal.gdc.cancer.gov/), as the largest cancer gene information database, stores data for gene expression, miRNA expression, copy number variation, DNA methylation, single nucleotide polymorphisms, and so on. Processed raw mRNA expression data for NBL were downloaded, including data for 158 NBL samples. The Series Matrix File for GSE62564 was downloaded from NCBI GEO and the annotation platform was GPL11154. Data for 495 patients with NBL with complete expression profiles and survival information were extracted. The Series Matrix File for GSE85047 whose annotation platform was GPL5175 was obtained. Data for 275 patients with NBL with complete expression profiles and survival information were retrieved. A total of 604 gene sets including necroptosis-related genes were obtained from the GeneCards database (https://www.genecards.org/).

Gene ontology and encyclopedia of genes and genomes functional annotation

Prognostic genes were annotated using clusterProfiler (R3.6) to thoroughly explore their functions. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed; terms and pathways with *P*- and *Q*-values of less than 0.05 were considered significant.

Metascape

Metascape (http://metascape.org/) is a powerful analytical tool for functional annotation of genes and proteins that allows users to apply current popular bioinformatics methods to batch gene and protein analyses to achieve an understanding of gene or protein function. GO/KEGG functional annotation of genes was performed using the Metascape database. Minimum overlap ≥3 and $p \leq 0.05$ were considered to be significant.

Model construction and prognosis

Necroptosis-related genes were selected, and a univariate Cox proportional hazards regression model was applied. Necroptosis genes with $p < 0.05$ were considered statistically significant and

included in the subsequent analysis. Lasso penalized Cox regression analysis was performed using 10-fold cross-validation based on the "glmnet" package in R to further reduce the number of necroptosis genes with the best predictive performance in the selected panels. After including the expression values for each specific gene, a risk score formula for each patient was constructed and weighted by its estimated regression coefficients in the lasso regression analysis. According to the risk score formula, the patients were divided into a low-risk group and a high-risk group with the median risk score value as the cut-off point. Differences in survival between the two groups were assessed by Kaplan-Meier analysis and compared using the log-rank statistical method. Lasso regression and stratified analyses were performed to examine the role of risk scores in the prediction of patient outcomes. The "survivalROC" package was used to derive receiver operator curves (ROCs) to investigate the accuracy of the model predictions. Univariate and multivariate Cox analyses including age, sex, tumor stage, and necroptosis score were performed to identify independent prognostic factors.

Drug sensitivity analysis

Based on the largest pharmacogenomics database (GDSC Cancer Drug sensitivity Genomics Database, https://www.cancerrxgene.org/), the R package "pRRophetic" was used to predict the chemosensitivity of each tumor sample. The estimated IC50 values for each specific chemotherapeutic drug were obtained by regression, and the prediction accuracy was measured by 10-fold cross-validation with the GDSC training set. Default values were selected for all parameters, using "combat" to remove the batch effect and the average value of repeated gene expression estimates.

Analysis of immune cell infiltration

The RNA-seq data for patients with NBL in different subgroups were analyzed by the CIBERSORT algorithm to infer the relative proportions of 22 kinds of immune-infiltrating cells. A Spearman correlation analysis was used to analyze the risk score and levels of infiltrating immune cells. Results were considered statistically significant at $p < 0.05$.

Gene set variation analysis

A gene set variation analysis (GSVA) is a non-parametric and unsupervised method to evaluate gene enrichment. By comprehensively scoring the set of genes of interest, GSVA converts gene level changes into pathway level changes to gain insight into biological functions. In this study, gene sets were downloaded from Molecular Signatures Database (v7.0), and each gene set was scored by the GSVA algorithm to evaluate differences in biological functions between samples.

Gene set enrichment analysis

A gene set enrichment analysis (GSEA) was performed with predefined gene sets to rank genes according to the degree of differential expression between two groups of samples and to determine whether a predefined gene set was enriched. GO terms and KEGG signaling pathways were obtained for differentially expressed genes between the high-risk group and low-risk group by GSEA; the number of replacements was set to 1,000 and the type of replacement was set to phenotype.

Regulatory network analysis of prognostic genes

Cistrome DB is a comprehensive database for ChIP-seq and DNase-seq analyses, containing data for transcription factors, histone modifications, and chromatin accessibility of 30,451 human and 26,013 mouse samples. The regulatory relationships between transcription factors and genes in the prognostic model were evaluated using Cistrome DB, in which the genome file was set to hg38 and the transcription initiation site was set to 10 kb. The results were visualized using Cytoscape.

Immunohistochemical staining analysis

To verify the protein expression of the necroptosis-related genes, ganglioneuroma samples were chosen as a control group, and IHC staining was used to evaluate the expression of these genes in paraffin-embedded tissues in the NBL group and control group. The paraffin-embedded serial tissue sections were cut at a thickness of 4 μm, and IHC was used to detect CYLD, JAK1, APC, ERH, CNBP, and BAX. The SP method was used to conduct IHC, and the primary antibodies against CYLD, JAK1, APC, ERH, CNBP, and BAX were all purchased from Abcam (Cambridge, United Kingdom). All experiments were carried out at least three times independently. Pannoramic SCAN (3DHISTECH, Budapest, Hungary) was used for observations and to obtain images. Image Pro Plus 6.0 was used to analyze the IHC results.

Statistical analysis

Survival curves were generated by the Kaplan-Meier method and compared by the log-rank test. A Cox proportional risk model was used for multivariate analyses. All statistical analyses were performed using R (version 3.6). All statistical tests were two-sided, and results were considered statistically significant at $p < 0.05$.

TABLE 1 Expression of necroptosis-related genes in NBL.

Gene	HR	z	*p*-value	Lower	Upper
BAX	1.98371619284125	5.48987193560628	4.02225271441093e-08	1.55337039660584	2.53328500552022
ERH	1.85233482253478	4.92782242546772	8.31511352013574e-07	1.44956691399528	2.36701338975658
CYLD	0.517434644838194	−4.34424835249258	1.3975334136114e-05	0.384377274471897	0.696551615978544
CPSF3	1.61988040244062	4.14992974066646	3.32577342461729e-05	1.28987185896012	2.03432030862867
JAK1	0.578925385579156	−3.77867763652999	0.00015766336556394	0.436010753947224	0.76868425614232
EMD	1.60016941917157	3.72777940585821	0.000193174345858099	1.24974414578669	2.04885310220044
EIF4EBP1	1.44851990579632	3.64660523798732	0.000265727677537716	1.24974414578669	2.04885310220044
ATAD3A	1.54022966941195	3.57903960381036	0.000344859209857309	1.21579227879252	1.95124403725686
HNRNPF	1.52677719453705	3.48401010645772	0.000493960835195072	1.20334696141417	1.93713756423083
ADRM1	1.49234932554181	3.48123248071431	0.00049911208043814	1.19118688309065	1.86965331893737
FUS	1.53804846136142	3.47146778172561	0.000517621441105841	1.20616906533564	1.96124501737073
APC	0.591730613045381	−3.47129043850349	0.000517963442372491	0.440007876948232	0.795770114034248
CCT5	1.56674940484423	3.45305994008363	1.21427595692257	1.21427595692257	2.0215369361351
CNBP	1.58200627442629	3.42348400836818	1.21663520141189	1.21663520141189	2.05710294213068

Results

Expression of necroptosis-related genes in NBL

The processed raw mRNA expression data for NBL in the TARGET database (FPKM) were downloaded, and necroptosis-related gene sets were obtained using the GeneCards database. We used clinical information for patients with NBL for a Cox univariate regression analysis to screen for necroptosis-related genes associated with prognosis in NBL. The following 14 prognosis-related genes were filtered ($p < 0.001$) by Cox univariate regression (in decreasing order based on significance): *BAX, ERH, CYLD, CPSF3, JAK1, EMD, EIF4EBP1, ATAD3A, HNRNPF, ADRM1, FUS APC, CCT5,* and *CNBP* (Table 1).

Functional enrichment of prognosis-related genes and construction of protein interaction networks

GO and KEGG pathway enrichment analyses revealed that the prognostic genes were significantly enriched in a large number of pathways. Enriched GO terms included cytoplasmic microtubule and translation regulator activity (Figure 1A). Enriched KEGG pathways included Human papillomavirus infection and Basal cell carcinoma (Figure 1B). A functional analysis using Metascape revealed that these prognostic genes were also highly enriched in several related pathways (Figure 1C).

Prognostic model construction and internal validation of necroptosis-related genes

We randomly divided patients from the TARGET database into training and validation sets at a ratio of 4:1. We obtained the risk score value for each sample based on the model obtained by lasso regression (Risk Score = *CYLD* × (−0.105492134) + *JAK1* × (−0.059871886) + *APC* × (−0.0109496) + *ERH* × 0.014856456 + *CNBP* × 0.059824071 + *BAX* × 0.39721508) (Figures 2A–C). Patients were divided into high-risk and low-risk groups using the median risk score as a threshold for analyses by Kaplan-Meier curves. Overall survival (OS) was significantly lower in the high-risk group than in the low-risk group in both the training set and the test set (Figures 2D,E). In addition, an ROC curve analysis showed that the AUC values in the training set and the test set for periods of 1, 3, and 5 years were all greater than 0.70 (Figures 2F,G), suggesting that the model was effective.

We integrated the clinical information as well as risk scores for patients in the high- and low-risk groups for regression analyses. A logistic regression analysis showed that in all of our samples, the distribution of values for multiple clinical indicators and risk scores for pediatric NBL contributed to multiple scoring processes. Age, gender, stage, and risk scores were evaluated with respect to 3-years and 5-years OS (Figure 2H). We also corrected the predicted OS in NBL for two periods of 3 and 5 years (Figure 2I). The risk score was identified as an independent prognostic factor for NBL by univariate and multivariate analyses (Figures 2J,K).

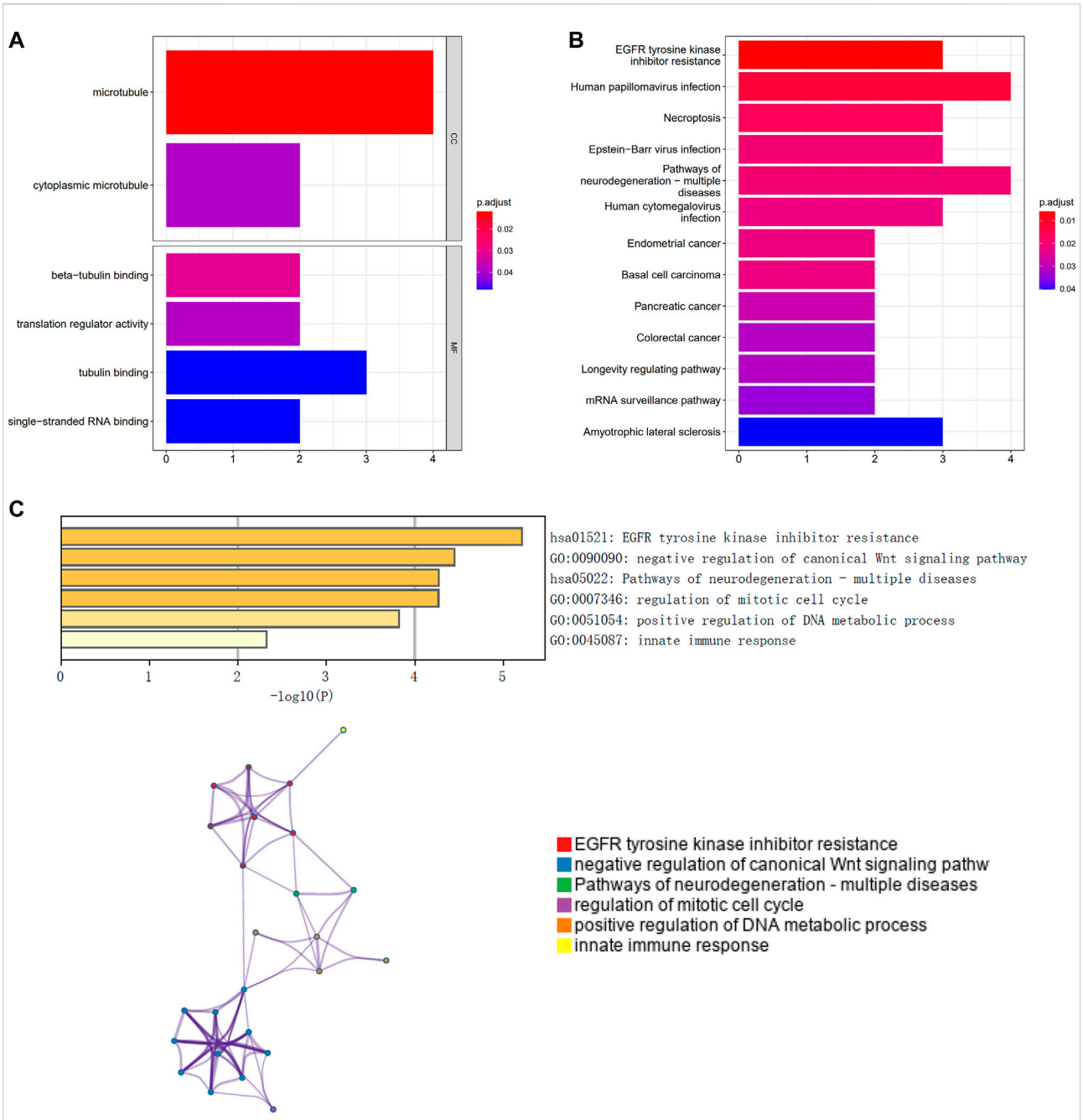


FIGURE 1
Functional enrichment of prognosis-related genes and construction of protein interaction networks. **(A)** GO analysis; **(B)** KEGG pathway analysis; **(C)** Metascape functional analysis.

Multi-omics analysis of the clinical predictive value of the model

The tumor microenvironment is mainly composed of tumor-associated fibroblasts, immune cells, extracellular matrix, various growth factors, inflammatory factors, specific physical and chemical characteristics, and cancer cells. The tumor microenvironment significantly affects tumor diagnosis, survival outcomes, and sensitivity to therapies. By analyzing the relationship between the risk score and tumor immune cell infiltration, we further explored the molecular mechanisms by which the risk score affects NBL progression. The distribution of levels of infiltration of different immune cell types in the samples differed between groups (Figure 3A). Correlations were detected

between the risk score and multiple cell types in the tumor microenvironment (Figure 3B). Additionally, levels of plasma cells were significantly lower in the low-risk group than in the high-risk group (Figure 3C). The risk score was significantly correlated with plasma cells and CD4 memory resting T cells (Figure 3D). Drug sensitivity data were obtained from the GDSC database, and the R package "pRRophetic" was used to predict the sensitivity of each tumor sample to chemotherapy. The risk score was significantly associated with sensitivity to various drugs, including AS601245, AZD.0530, AZD6244, AZD6482, CHIR.99021, and CCT007093 (Figure 3E). We further explored the mutation profiles of patients in the high- and low-risk groups. The mutation frequency in genes, such as *ALK*, was significantly higher in the high-risk group than in the low-risk group (Figure 3F).

Exploration of specific signaling mechanisms associated with prognostic models

We next investigated the specific signaling pathways related to a high and low risk to explore the molecular mechanisms by which risk scores influence tumor progression. The results of GSVA showed that the enriched differential pathways between the two groups were mainly TGF BETA SIGNALING, UV RESPONSE DN, MTORC1 SIGNALING, ALLOGRAFT REJECTION, and OXIDATIVE PHOSPHORYLATION (Figure 4), suggesting that perturbations in these signaling pathways in patients in the high- and low-risk groups affected prognosis in NB.

Validation of the robustness of the prognostic model using an external dataset

Expression data and survival information for patients with NBL were downloaded from the GEO database (GSE62564 and GSE85047) to predict the clinical stage based on the model and for a Kaplan-Meier analysis of the survival difference between groups. In the two GEO external validation sets, OS was significantly lower in the high-risk group than in the low-risk group (Figures 5A,B). As determined by ROC curve analyses, the model had strong predictive power for prognosis (AUC values for 1 year, 3 years, and 5 years with the GEO verification data set were all greater than 0.70) (Figures 5C,D).

Signaling mechanisms associated with the prognostic model

We next investigated the specific signaling pathways differentiating samples in the high- and low-risk groups based on the genes in the prognostic models to evaluate the factors contributing to tumor progression. By a GSEA, we found significant enrichment for many related pathways, including the

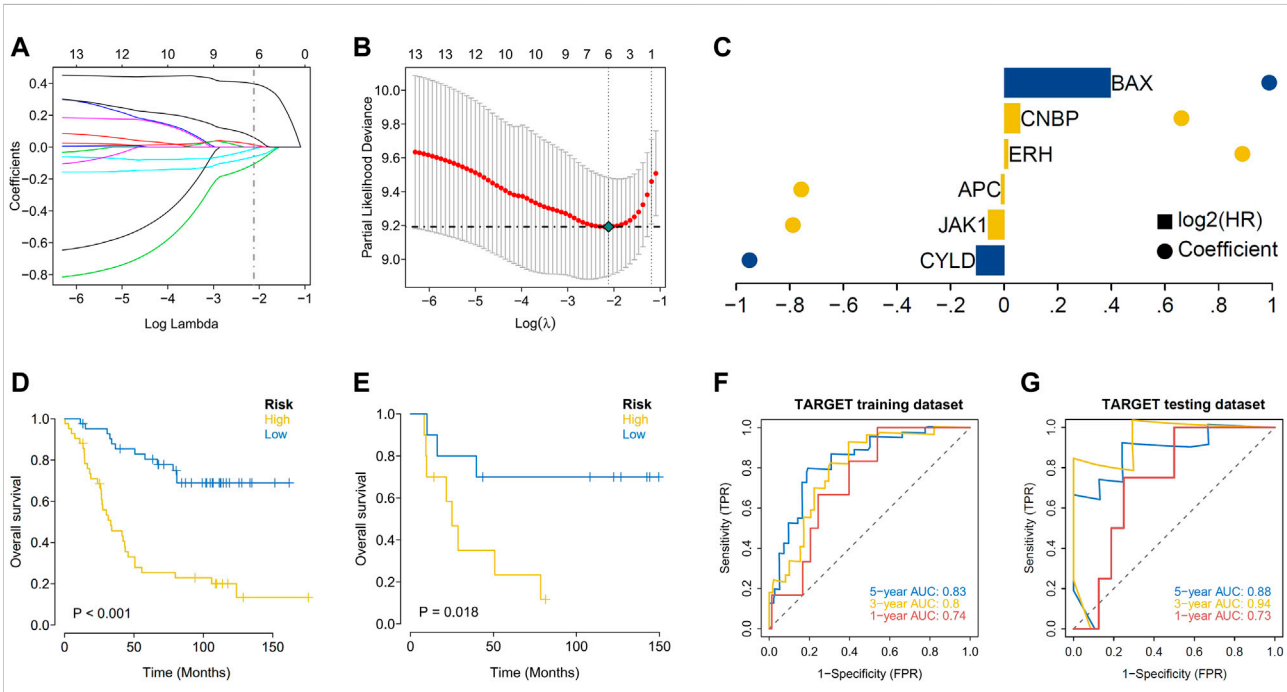


FIGURE 2
(continued)

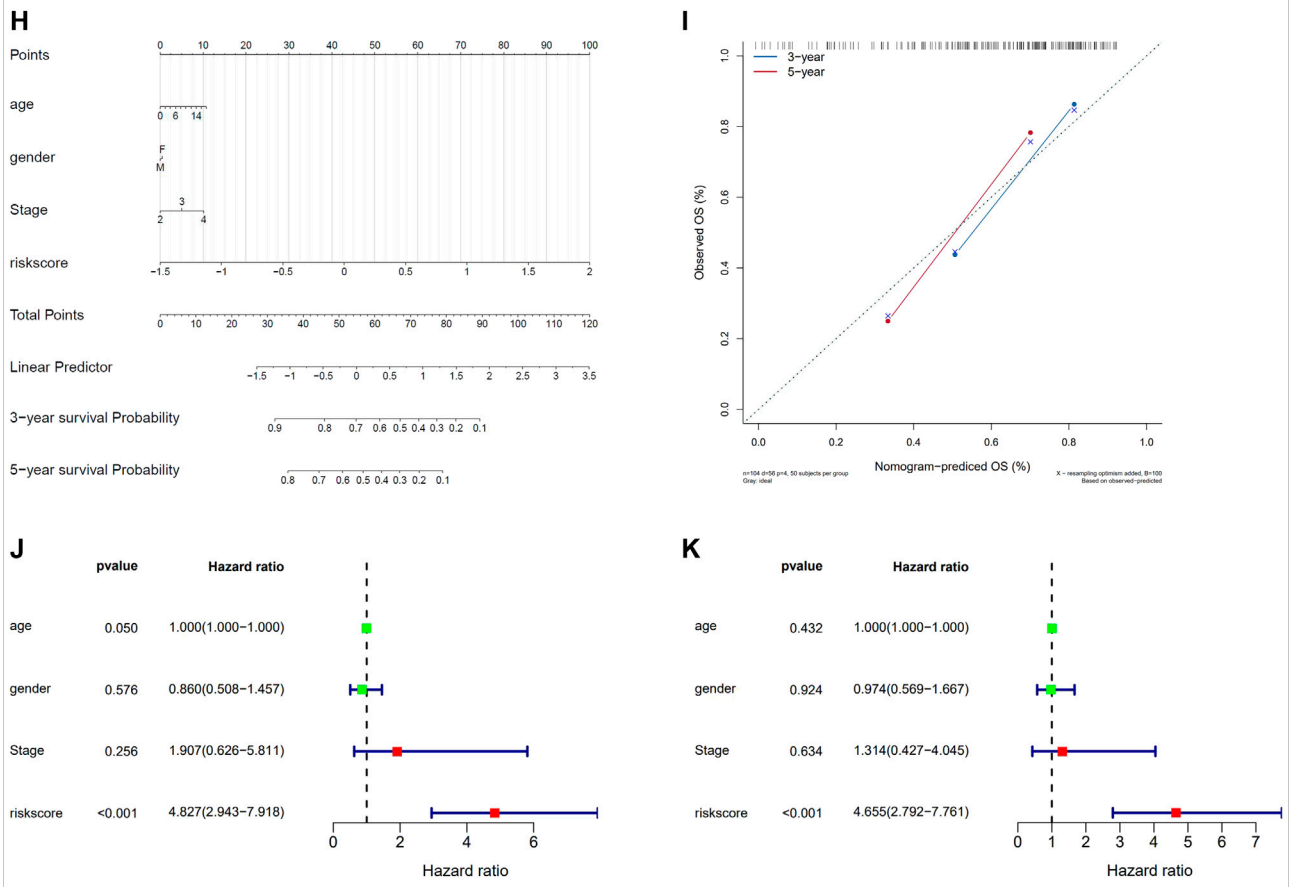


FIGURE 2
Lasso analysis and model construction. **(A)** LASSO coefficient profiles of the 13 IRGs in TARGET-NBL; **(B)** A coefficient profile plot was generated against the log (lambda) sequence. Selection of the optimal parameter (lambda) in the LASSO model for TARGET; **(C)** Lasso Coefficient HR; **(D)** Kaplan-Meier survival curve analysis in the high-risk and low-risk NBL patients in the training subset; **(E)** Kaplan-Meier survival curve analysis in the high-risk and low-risk NBL patients in the testing subset; **(F)** time-dependent ROC curve for 1-year, 3-years, and 5-years prediction (training subset); **(G)** time-dependent ROC curve for 1-year, 3-years, and 5-years prediction (testing subset); **(H)** the nomogram for predicting the 3- and 5-years OS of NBL patients; **(I)** the calibration curve of the nomogram for predicting 3- and 5-years OS of NBL patients; **(J)** Univariate Cox regression analyses in the TARGET cohort NBL patients; **(K)** multivariate Cox regression analyses in the TARGET cohort NBL patients.

GO terms AXON EXTENSION, DENDRITE MORPHOGENESIS, and REGULATION OF VIRAL TRANSCRIPTION and the KEGG pathways BASE EXCISION REPAIR, LONG TERM POTENTIATION, and AXON GUIDANCE (Figures 6A,B), suggesting that the disturbance of these signaling pathways in high and low risk groups affected prognosis in NBL.

Relationships between expression levels of genes in the prognostic model and immune cell infiltration

Several genes in the prognostic model were highly correlated with levels of infiltrating immune cells. For example, *BAX* was positively correlated with regulatory T cells (Tregs) and plasma cells and negatively correlated with CD4 memory resting T-cells and resting mast cells. *ERH* was positively correlated with plasma cells and neutrophils and negatively correlated with CD4 memory activated T cells and monocytes. *CYLD* was positively correlated with CD4 memory activated T cells and activated dendritic cells and negatively correlated with plasma cells and M0 macrophages. *JAK1* was positively correlated with CD4 memory resting T cells and activated dendritic cells and negatively correlated with follicular helper T cells and plasma cells. *APC* was positively correlated with memory B cells and CD4 memory resting T cells and negatively correlated with M0 macrophages and activated mast cells. *CNBP* was positively correlated with plasma cells and activated mast cells and negatively correlated with CD4 memory activated T cells and Tregs (Figure 7A). We further evaluated the correlations between genes in the prognostic model and immune factors, including immunomodulators, chemokines, and cellular receptors, using TISIDB (Figure 7B). These analyses confirmed that the prognostic genes are closely related to

levels of immune cell infiltration and play important roles in the immune microenvironment.

Regulatory network analysis

We evaluated the transcriptional regulatory network of the six genes in the prognostic model. Using Cistrome DB, 91 transcription factors were related to *BAX*, 98 were related to *ERH*, 78 were related to *CYLD*, 85 were related to *JAK1*, 10 were related to *APC*, and 68 were related to *CNBP*. The results were visualized using Cytoscape to obtain a comprehensive transcriptional regulatory network involving genes in the prognostic model (Figure 8).

ceRNA network analysis

The six genes in the prognostic model were analyzed using the miRWalk and ENCORI databases to predict interacting miRNAs and lncRNAs, respectively. Interacting mRNA–miRNA pairs associated with these six key mRNAs were first extracted using the miRWalk database, and only 605 mRNA–miRNA pairs with a TargetScan score of one or miRDB score of one were retained, involving 5 mRNAs and 131 miRNAs. Then, the interacting lncRNAs were predicted based on these miRNAs. A total of 18,244 pairs of interactions were predicted (involving 42 miRNAs and 3,868 lncRNAs). Finally, we constructed the ceRNA network using Cytoscape (v3.7) (Figure 9).

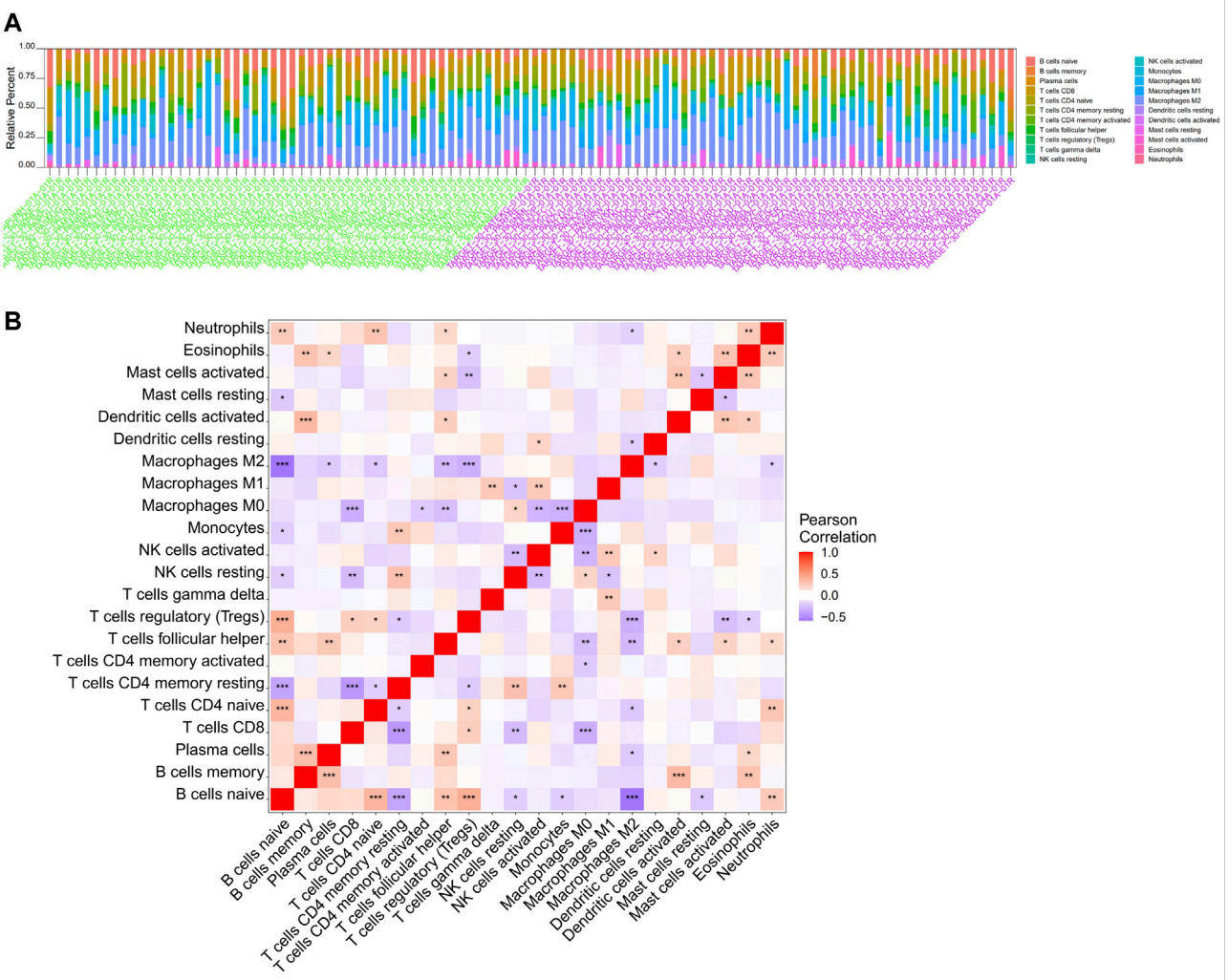


FIGURE 3
(continued)

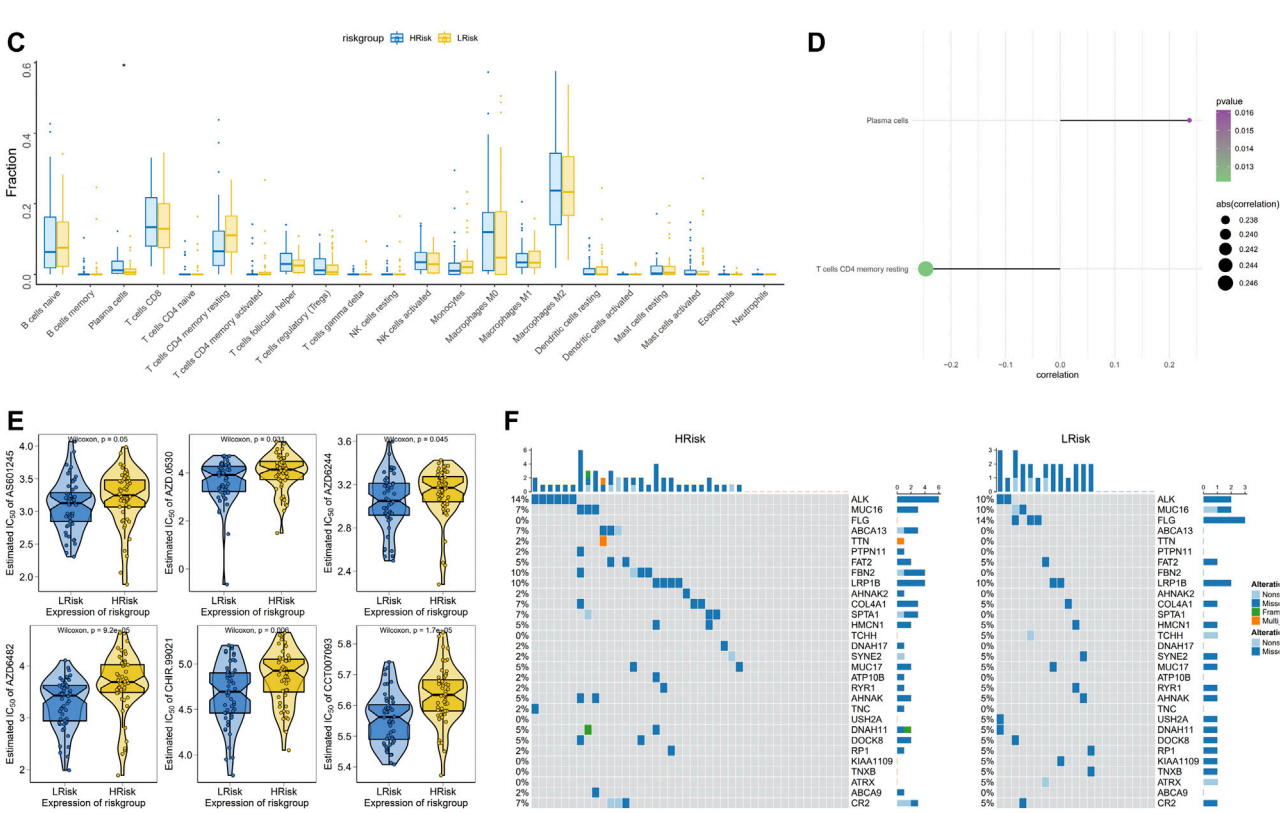


FIGURE 3
Multi-omics analysis of the clinical predictive value of the model. **(A)** Stacked bar chart of the distribution of 22 immune cells in each NBL sample of the TARGET cohort; **(B)** Pearson correlation between immune cells, red for positive correlation, purple for negative correlation; **(C)**Differences in immune cell counts between the high-risk and low-risk groups; **(D)**The correlation between the risk score and the immune cells, the circle size indicates the strength of the correlation, and the color indicates the *p*-value; **(E)** The difference on the therapeutic sensitivities of six chemotherapy drugs; **(F)** Mutation profiles of patients in the high and low risk groups, yellow for high risk group and blue for low risk group.

Immunohistochemical staining analysis

IHC assays were performed to verify the expression levels of proteins encoded by these necroptosis-related genes in the NBL and ganglioneuroma (control) groups. As shown in Figure 10, the expression levels of BAX ($p = 0.0017$), ERH ($p = 0.0067$), APC ($p = 0.0416$), and CNBP ($p = 0.0244$) were significantly higher in tumor tissues of the NBL group than in the control group. Other indexes did not differ significantly between groups.

Disscussion

Necroptosis has conflicting roles in malignant tumors, with both tumor-promoting and inhibitory effects in different types of adult cancers. In particular, necroptosis can inhibit tumor progression but can also trigger inflammatory responses and vascular endothelial cell necrosis, in turn promoting tumor cell extravasation and cancer metastasis (Gong et al., 2019). The delayed activation or disruption of normal apoptotic pathways may be an important cause of chemotherapeutic drug resistance in patients with NBL; therefore, the induction of necroptosis in NBL may be an alternative therapeutic approach to eliminate anti-apoptotic tumor cells and improve the anti-tumor immune microenvironment.

In this study, we constructed the first prognostic model for NBL based on necroptosis-related genes. We first obtained information on necroptosis-related genes from the GeneCards database and used univariate Cox regression and LASSO regression analyses to screen for necroptosis-related genes associated with prognosis, revealing six genes, i.e., *CYLD*, *JAK1*, *APC*, *ERH*, *CNBP*, and *BAX*, which were used to construct a prognostic risk model. The conserved cylindromatosis (CYLD) is a deubiquitinating (DUB) enzyme with an important regulatory role in a variety of cellular processes, including the immune response, inflammation, and necrosis. Small ubiquitin-related modifier (SUMO) can post-translationally modify CYLD to impair its DUB function. After 8 days of treatment with all-trans-retinoic acid (ATRA) on the NB SK-N-BE 2) C cell line, the SUMOization of CYLD decreased, while its expression increased, which blocked the NF-κB signal transduction pathway and promoted cell death (Kobayashi et al.,

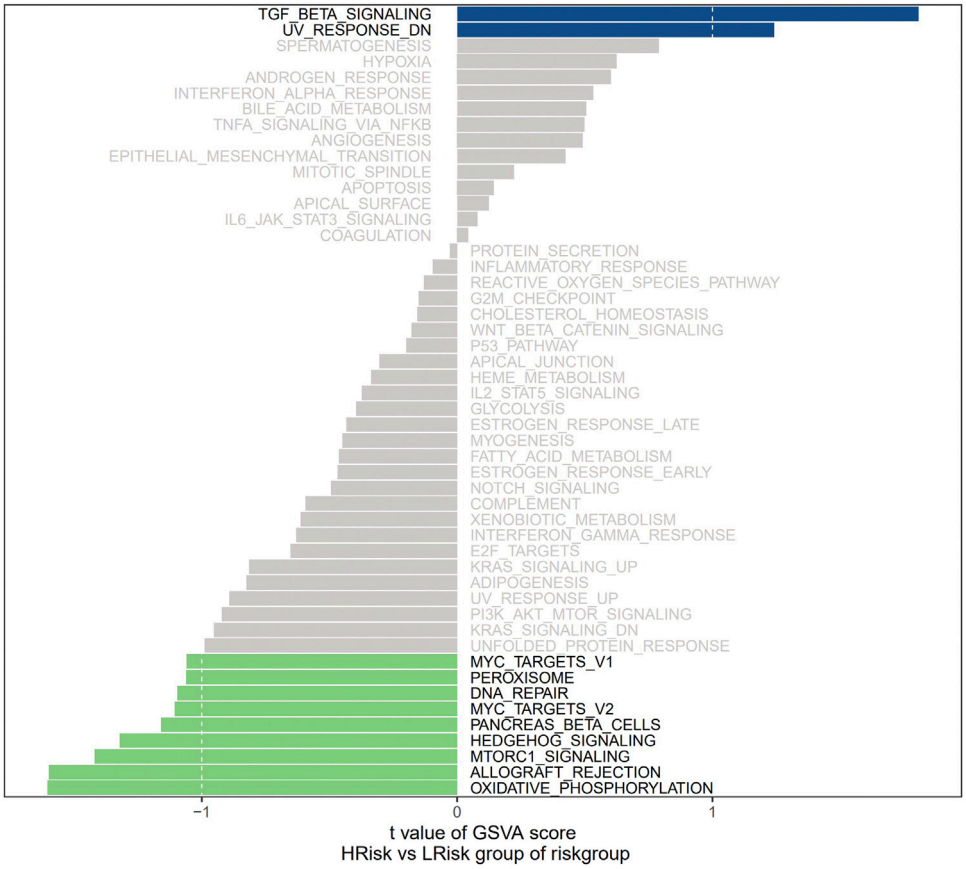


FIGURE 4
GSVA plot of riskgroup.

2015). In prostate cancer cell lines, the knockout of *CYLD* increased the proliferation, migration, colony formation, and invasion of cancer cells *in vitro* (Haq et al., 2022). JAK1 (Janus kinase 1) is a member of a class of protein-tyrosine kinases involved in autoimmune diseases and malignancies. The targeted inhibition of JAK1 expression by miR-20a-5p can decrease proliferation and invasion and improve the adhesion ability of endometrial cancer cells (He et al., 2021). Wen et al. (Wen et al., 2014) have also reported that JAK1/STAT3 plays a crucial role in ovarian cancer as a pro-oncogenic signaling pathway. The targeted inhibition of the JAK1/STAT3 pathway can effectively prevent the progression and metastasis of ovarian cancer. *APC* is a tumor suppressor gene. The frequency of APC germline mutations in patients with familial adenomatous polyposis (FAP)-related diseases, such as gastric fundus adenomatous polyposis, duodenal adenoma, desmoid tumors, and thyroid cancer is greater than 60% (Takao et al., 2021). Enhancer of rudimentary homolog (ERH) is a small, highly conserved protein. It binds to various factors involved in many cellular processes, such as pyrimidine metabolism, mitosis, and transcriptional regulation (Fujimura et al., 2012). The overexpression of ERH weakens the invasion and migration ability of gastric cancer cells, suggesting that it is a prognostic marker (Park et al., 2020). However, a study of ovarian cancer suggested that ERH may be a associated with a poor prognosis, and inhibiting ERH expression can promote cancer cell apoptosis and inhibit the metastasis and invasion of ovarian cancer cells by regulating the epithelial-mesenchymal transition (EMT) (Zhang et al., 2020). Cellular nucleic acid-binding protein (CNBP) is associated with cell proliferation and is highly expressed in various human tumors. The lncRNA SUMO1P3 enhances proliferation, invasiveness, and drug resistance in gastric cancer cell lines by directly binding to CNBP, resulting in high levels of c-myc and cyclinD1 (CCND1) (Guo et al., 2020). The Bcl-2 family is an important family of apoptosis regulatory proteins, with key roles in the apoptosis signal transduction pathway. Bax is a pro-apoptotic factor in the Bcl-2 family of proteins (Meng et al., 2019). In human retinoblastoma, the expression of anti-apoptotic Bcl-2 is significantly related to poor differentiation and strong invasiveness, and the lack of Bax expression is related to choroidal infiltration and lymph node metastasis (Singh et al.,

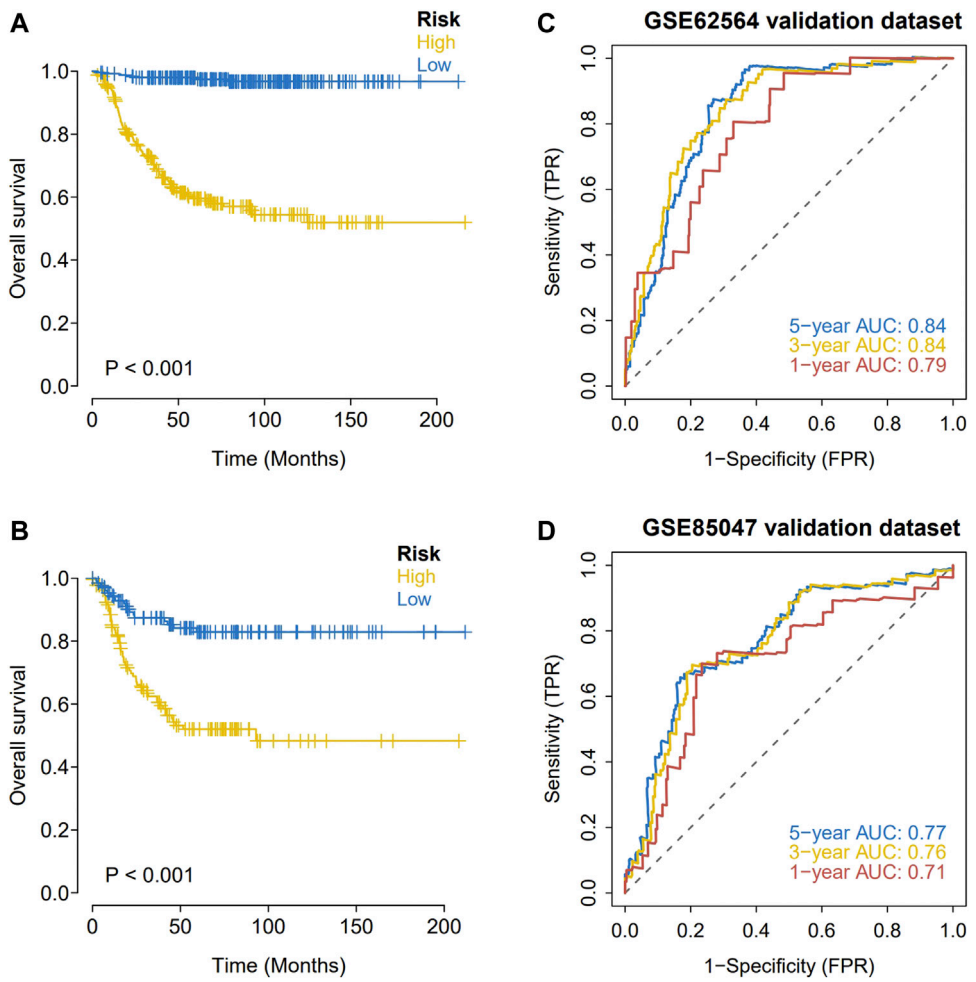


FIGURE 5
External validation of the hypoxia risk score. **(A,B)** Validation of the hypoxia risk score in GSE62564; **(C,D)** Validation of the hypoxia risk score in GSE85047.

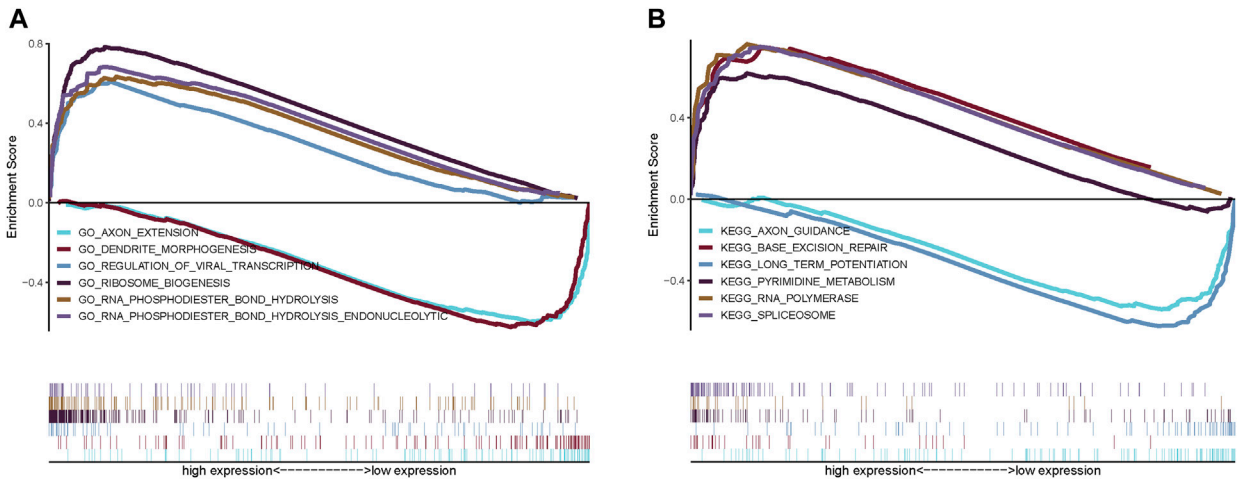


FIGURE 6
GSEA analysis of risk scores. **(A)** GO pathways; **(B)** KEGG pathways.

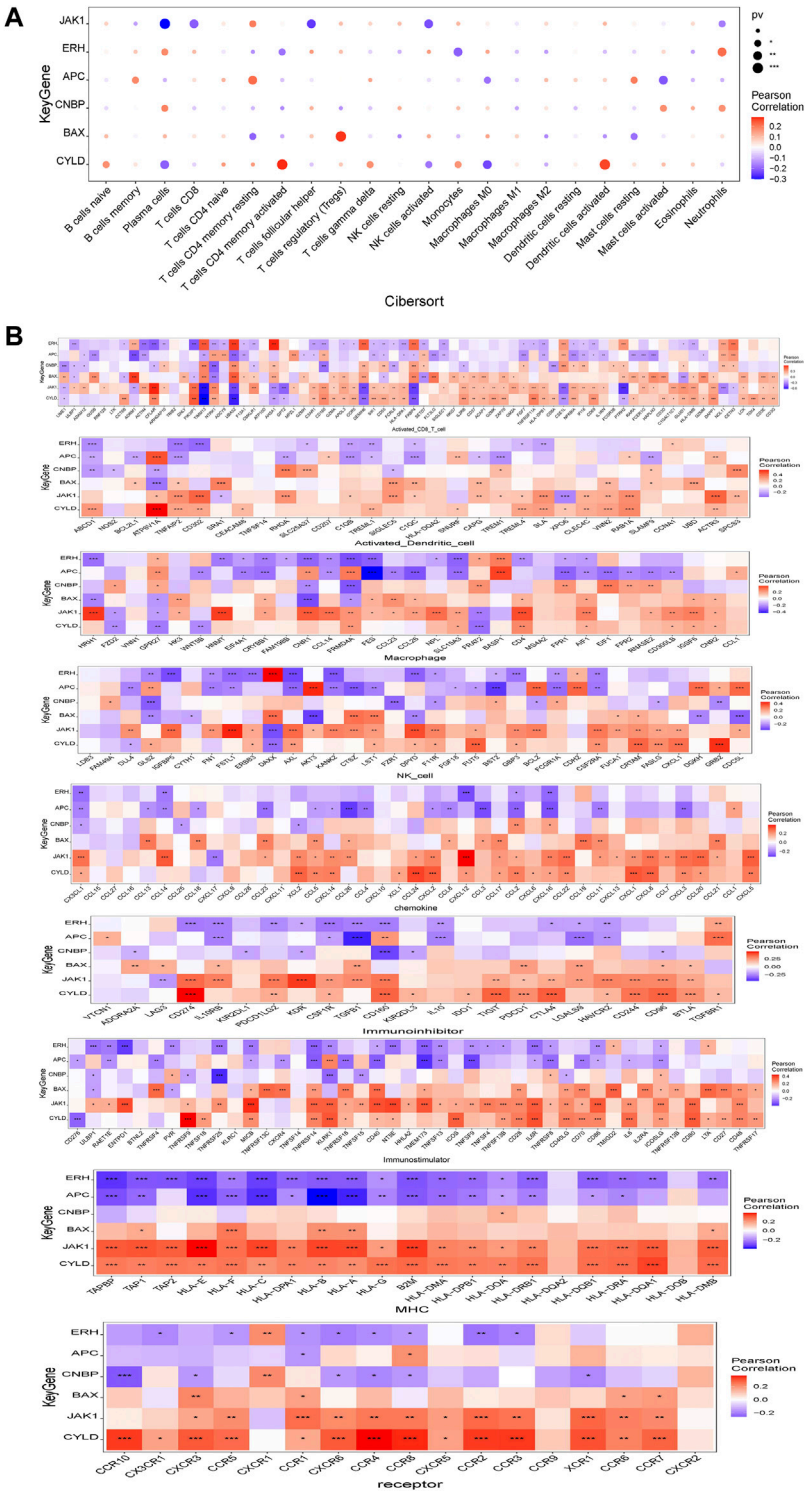


FIGURE 7
Relationship between model gene and immune infiltration. **(A)** Pearson correlation between model genes and 22 kinds of immune cells; **(B)** Pearson correlation analysis between model genes and various immune-related genes. *means $p < 0.05$; ** means $p < 0.01$; *** means $p < 0.001$.

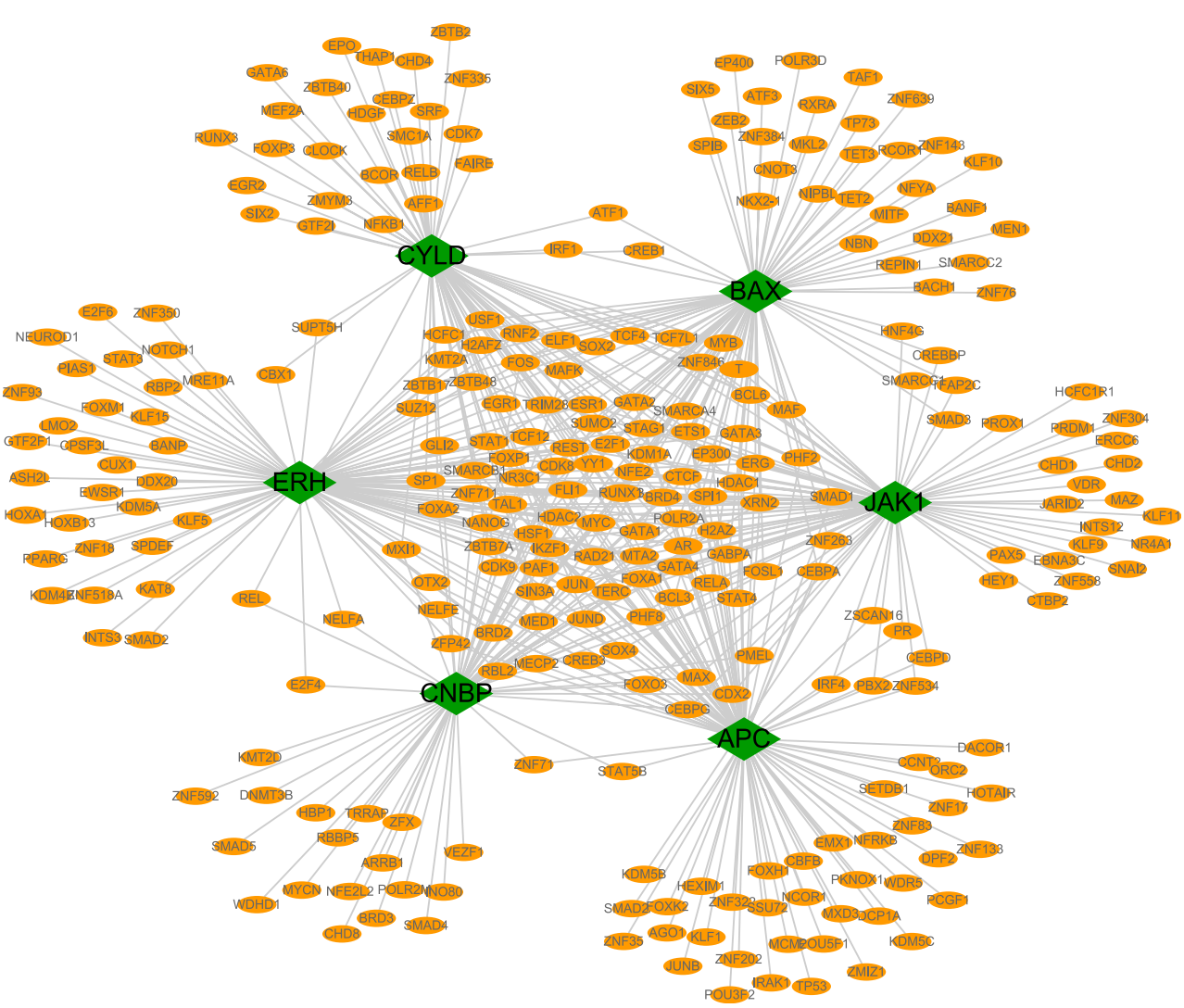


FIGURE 8
Transcriptional regulatory networks of model genes. The transcriptional regulatory network of core genes was based on cistrome database, and the species was set as HG38, and the transcription start site was set as 10 KB. Green represents model genes and orange represents associated transcription factors.

2015). These previous results indicate that the necroptosis-related genes identified in this study play important roles in various human tumors, further supporting their potential roles in the occurrence and development of NBL. However, further research is needed to explore the molecular mechanisms of action of these genes.

We analyzed the predictive value of the six gene-based model for OS. A Kaplan-Meier analysis showed that OS was significantly lower in the middle-and high-risk groups than in the low-risk group. An ROC curve analysis showed that the prognostic model has good stability and can effectively screen for patients with NBL with poor prognosis. The risk score was identified as an independent prognostic factor for NBL by univariate and multivariate analyses after stratification according to clinical parameters.

Clinical data have shown that immunotherapy with the disialoganglioside GD2 combined with granulocyte-macrophage colony-stimulating factor (GM-CSF) or interleukin-2 significantly improves prognosis in high-risk patients with NBL (Yu et al., 2010). Subsequently, the molecular events associated with NBL-related immune cell infiltration and immune responses in NBL have been a focus of research. In a study of the relationship between immune cell infiltration and prognosis in NBL, Schaafsma et al. (2021) discovered that a high abundance of naïve B cells, memory B cells, CD8^{+} T cells, and NK cells was significantly associated

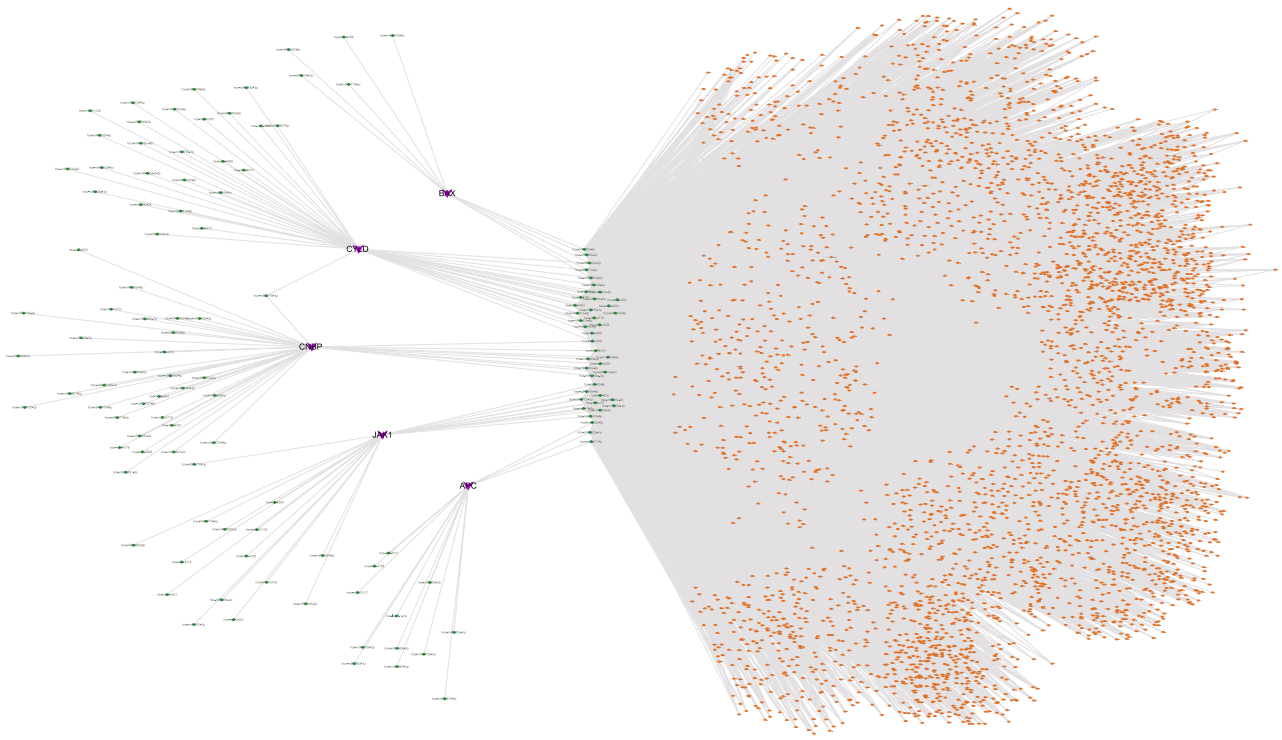


FIGURE 9
Model gene-related ceRNA network, purple represents mRNA, green represents miRNA, yellow represents lncRNA. Based on miRWalk, TargetScan and miRDB databases, using Cytoscape software visualization.

with a longer OS; conversely, high levels of $CD4^+$ T cell infiltration were negatively associated with OS. Our results revealed that B cells play an important role in the NBL tumor microenvironment, suggesting that B cells can be used as an independent variable to predict recurrence-free and overall survival. Batchu (Sai, 2020) have shown that low levels of $CD4^+$ naïve T cells and monocytes are associated with a reduced event-free survival. Tumor-associated macrophages (TAMs) closely resemble M2-polarized macrophages and are critical modulators of the tumor microenvironment. TAM aggregation in a variety of human tumors is associated with poor clinical outcomes, and TAMs can provide a favorable microenvironment for tumor progression (Liu and Joshi, 2020). TAMs can upregulate the expression of MYC *via* the signal transducer and activator of transcription 3 (STAT3) pathway, and this may explain the association between TAMs and a poor prognosis in patients with non-MYCN-amplified NBL (Liu and Joshi, 2020). In our prognostic model, the relative abundance of plasma cells in the low-risk group was significantly lower than that in the high-risk group, and the risk score was significantly correlated with plasma cells and CD4 memory resting T cells. The six genes in the prognostic model were closely associated with levels of immune cell infiltration and played an important role in the immune microenvironment. These results suggest that necroptosis influences tumorigenesis and tumor development by regulating immune cell infiltration.

We analyzed the specific signaling pathways differentiating the high- and low-risk groups based on the prognostic model. GSVA results showed that the differential pathways between the two groups were mainly signaling pathways such as TGF-beta signaling, mTORC1 signaling, and oxidative phosphorylation. By a GSEA, enriched GO terms were axon extension and dendrite morphogenesis; enriched KEGG pathways were base excision repair, long-term Potentiation, and axon guidance. Transforming growth factor-β (TGF-β) pathway plays an important role in cellular homeostasis by regulating cell growth inhibition, cellular senescence, differentiation, and apoptosis (Lin et al., 2005). The EMT is a transdifferentiation process in which epithelial cells lose polarity and contact inhibition and obtain mesenchymal characteristics, such as the fibroblast migration phenotype. The EMT plays an important role in human embryonic development and is also considered a pathological mechanism. Cancer cells can acquire migration and invasion abilities through the EMT, leading to tumor metastasis. In human NB cell lines, EMT is significantly up-

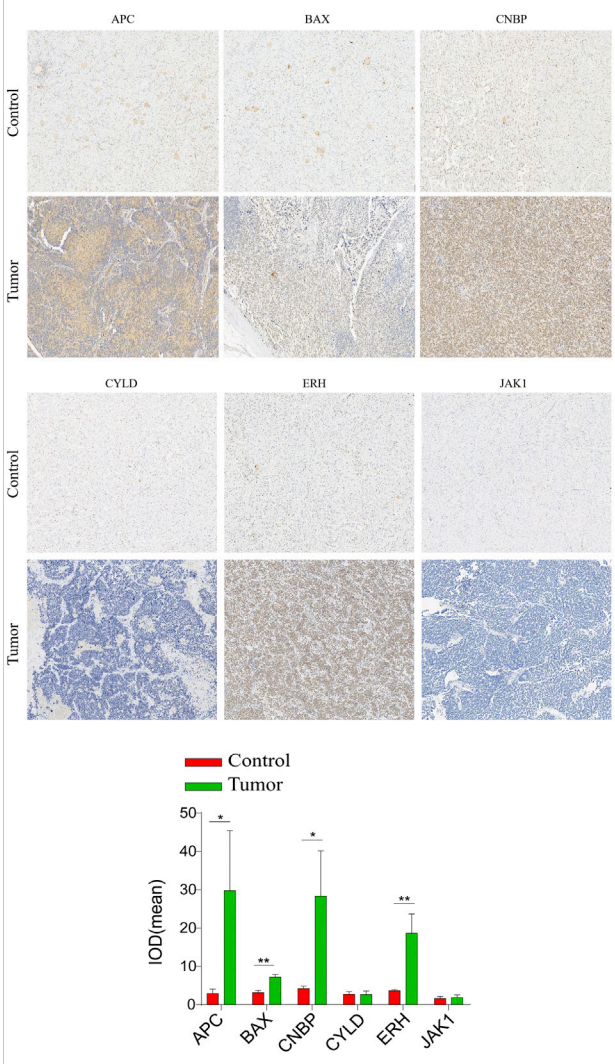


FIGURE 10
Validation of the expression of the prognostic Necroptosis in NBL and Ganglioneuroma (GN) by immunohistochemical staining analysis. The expressions of CYLD、JAK1、APC、ERH、CNBP, and BAX in NBL tissues and GN were detected by immunohistochemical staining (magnification ×100). Quantification of immunohistochemical staining for CYLD、JAK1、APC、ERH、CNBP, and BAX by Image Pro Plus software. *$p < 0.05$ vs. control group, **$p < 0.01$ vs. control group.

regulated via the TGF-β pathway, resulting in a more aggressive phenotype (Naiditch et al., 2015). MYCN plays an important role in NB. The MYC genes (*MYCN, c-myc,* and *L-myc*) drive tumorigenesis in part by the activation of the mammalian target of rapamycin (mTOR) pathway, a master regulator of translation and protein synthesis. Therefore, the effective inhibition of mTOR function represents a potential therapeutic strategy targeting MYCN in NB (Moreno-Smith et al., 2017). Oxidative phosphorylation (OXPHOS) is an important pathway for the survival and proliferation of tumor cells. Some compounds inhibit the growth of NB *in vivo* by inhibiting the activities of OXPHOS and mitochondrial respiratory complex I (Nagasaki-Maeoka et al., 2020). Base excision repair (BER) fixes the majority of endogenous DNA damage, including deamination, depurination, alkylation, and oxidative damage. Abnormalities in this pathway are strongly associated with tumorigenesis (Wallace et al., 2012). These findings show that the enriched pathways associated with the genes in the prognostic model are closely related to the development of NB.

We also studied the relationship between the risk score obtained by the newly developed model and clinical drug sensitivity. The risk score was significantly correlated with the sensitivity to AS601245, AZD0530, AZD6244, AZD6482, CHIR99021, CCT007093, and other drugs. These chemotherapeutic drugs have been studied extensively in human glioblastoma (GBM). AZD0530, a potent small-molecule inhibitor of Src family kinases, can enhance the radiosensitivity of GBM tumor cells (Yun et al., 2021). AZD6244, an inhibitor of MEK in the RAF/MEK/ERK pathway, inhibits proliferation in the GBM cell line (See et al., 2012). As an inhibitor of PI3Kβ in the PI3K/Akt pathway, AZD6482 can exert an anti-tumor effect by inhibiting proliferation and inducing apoptosis in human GBN tumor cells (Xu et al., 2019). CCT007093 can attenuate cell proliferation, migration, and invasion induced by UVC radiation in human GBM (Yang et al., 2014). The role of these drugs in children with NBL needs to be confirmed by further studies.

The results of this study provide new insights into the occurrence and development of NBL from the perspective of necroptosis. The prognostic model based on six necroptosis-related genes can effectively predict the prognosis of patients with NBL. In addition, the risk score obtained from the necroptosis model is associated with essential biological functions and has clinical value.

Conclusion

By a variety of bioinformatics analyses of high-throughput sequencing datasets, we systematically evaluated the molecular characteristics and prognostic value of necroptosis in NBL. Our results provide preliminary evidence for the complex biological functions and immunoregulatory effects of necroptosis-related genes. These necroptosis-related genes may be involved in the occurrence, development, invasion, and metastasis of NBL. We constructed a risk score model that can independently predict prognosis in NBL. Our results will aid in revealing the pathogenesis of NBL and in the identification of new biomarkers and provide a

basis for the development of therapeutic strategies targeting necroptosis.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

JC: wrote and edited the manuscript.

Acknowledgments

We thank all individuals participated in this study.

Conflict of interest

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

Publisher's note

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

References

Dai, X., Hakizimana, O., Zhang, X., Kaushik, A. C., and Zhang, J. (2020). Orchestrated efforts on host network hijacking: Processes governing virus replication. *Virulence* 11 (1), 183–198. doi:10.1080/21505594.2020.1726594

Fujimura, A., Kishimoto, H., Yanagisawa, J., and Kimura, K. (2012). Enhancer of rudimentary homolog (ERH) plays an essential role in the progression of mitosis by promoting mitotic chromosome alignment. *Biochem. Biophys. Res. Commun.* 423 (3), 588–592. doi:10.1016/j.bbrc.2012.06.018

Fusco, P., Esposito, M. R., and Tonini, G. P. (2018). Chromosome instability in neuroblastoma. *Oncol. Lett.* 16 (6), 6887–6894. doi:10.3892/ol.2018.9545

Gong, Y., Fan, Z., Luo, G., Yang, C., Huang, Q., Fan, K., et al. (2019). The role of necroptosis in cancer biology and therapy. *Mol. Cancer* 18 (1), 100. doi:10.1186/s12943-019-1029-8

Guo, Y., Wang, Y., Ma, Y., Chen, G., Yue, P., and Li, Y. (2020). Upregulation of lncRNA SUMO1P3 promotes proliferation, invasion and drug resistance in gastric cancer through interacting with the CNBP protein. *RSC Adv.* 10 (10), 6006–6016. doi:10.1039/c9ra09497k

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell.* 144 (5), 646–674. doi:10.1016/j.cell.2011.02.013

Haq, S., Sarodaya, N., Karapurkar, J. K., Suresh, B., Jo, J. K., Singh, V., et al. (2022). CYLD destabilizes NoxO1 protein by promoting ubiquitination and regulates prostate cancer progression. *Cancer Lett.* 525, 146–157. doi:10.1016/j.canlet.2021.10.032

He, Y., Ma, H., Wang, J., Kang, Y., and Xue, Q. (2021). miR-20a-5p inhibits endometrial cancer progression by targeting janus kinase 1. *Oncol. Lett.* 21 (5), 427. doi:10.3892/ol.2021.12688

Johnstone, R. W., Ruefli, A. A., and Lowe, S. W. (2002). Apoptosis: A link between cancer genetics and chemotherapy. *Cell.* 108 (2), 153–164. doi:10.1016/s0092-8674(02)00625-6

Kobayashi, T., Masoumi, K. C., and Massoumi, R. (2015). Deubiquitinating activity of CYLD is impaired by SUMOylation in neuroblastoma cells. *Oncogene* 34 (17), 2251–2260. doi:10.1038/onc.2014.159

Lin, H. K., Bergmann, S., and Pandolfi, P. P. (2005). Deregulated TGF-beta signaling in leukemogenesis. *Oncogene* 24 (37), 5693–5700. doi:10.1038/sj.onc.1208923

Liu, K. X., and Joshi, S. (2020). "Re-educating" tumor associated macrophages as a novel immunotherapy strategy for neuroblastoma. *Front. Immunol.* 11, 1947. doi:10.3389/fimmu.2020.01947

Meng, K., Yuan, G., Bao, H., Wang, L., Ma, R., Yu, B., et al. (2019). Interaction of HCCR-1 and Bax in breast cancer. *J. BUON* 24 (3), 1027–1037.

Moreno-Smith, M., Lakoma, A., Chen, Z., Tao, L., Scorsone, K. A., Schild, L., et al. (2017). p53 nongenotoxic activation and mTORC1 inhibition lead to effective combination for neuroblastoma therapy. *Clin. Cancer Res.* 23 (21), 6629–6639. doi:10.1158/1078-0432.CCR-17-0668

Nagasaki-Maeoka, E., Ikeda, K., Takayama, K. I., Hirano, T., Ishizuka, Y., Koshinaga, T., et al. (2020). Polyethylene glycol derivative 9bw suppresses growth of neuroblastoma cells by inhibiting oxidative phosphorylation. *Cancer Sci.* 111 (8), 2943–2953. doi:10.1111/cas.14512

Naiditch, J. A., Jie, C., Lautz, T. B., Yu, S., Clark, S., Voronov, D., et al. (2015). Mesenchymal change and drug resistance in neuroblastoma. *J. Surg. Res.* 193 (1), 279–288. doi:10.1016/j.jss.2014.07.018

Najafov, A., Chen, H., and Yuan, J. (2017). Necroptosis and cancer. *Trends Cancer* 3 (4), 294–301. doi:10.1016/j.trecan.2017.03.002

Negroni, A., Colantoni, E., Cucchiara, S., and Stronati, L. (2020). Necroptosis in intestinal inflammation and cancer: New concepts and therapeutic perspectives. *Biomolecules* 10 (10), 1431. doi:10.3390/biom10101431

Park, J. H., Park, M., Park, S. Y., Lee, Y. J., Hong, S. C., Jung, E. J., et al. (2020). ERH overexpression is associated with decreased cell migration and invasion and a good prognosis in gastric cancer. *Transl. Cancer Res.* 9 (9), 5281–5291. doi:10.21037/tcr-20-1498

Park, J. R., Eggert, A., Caron, H., et al. (2010). Neuroblastoma: Biology, prognosis, and treatment. *Hematol. Oncol. Clin. North Am.* 24 (1), 65–86. doi:10.1016/j.hoc.2009.11.011

Sai, B (2020). Immunological landscape of Neuroblastoma and its clinical significance[J]. *Cancer Treat. Res. Commun.* 26, 100274. doi:10.1016/j.ctarc.2020.100274

Schaafsma, E., Jiang, C., and Cheng, C. (2021). B cell infiltration is highly associated with prognosis and an immune-infiltrated tumor microenvironment in neuroblastoma. *J. Cancer Metastasis Treat.* 7 (34). doi:10.20517/2394-4722.2021.72

See, W. L., Tan, I. L., Mukherjee, J., Nicolaides, T., and Pieper, R. O. (2012). Sensitivity of glioblastomas to clinically available MEK inhibitors is defined by neurofibromin 1 deficiency. *Cancer Res.* 72 (13), 3350–3359. doi:10.1158/0008-5472.CAN-12-0334

Singh, L., Pushker, N., Saini, N., Sen, S., Sharma, A., Bakhshi, S., et al. (2015). Expression of pro-apoptotic Bax and anti-apoptotic Bcl-2 proteins in human retinoblastoma. *Clin. Exp. Ophthalmol.* 43 (3), 259–267. doi:10.1111/ceo.12397

Takao, M., Yamaguchi, T., Eguchi, H., Yamada, T., Okazaki, Y., Tomita, N., et al. (2021). APC germline variant analysis in the adenomatous polyposis phenotype in Japanese patients. *Int. J. Clin. Oncol.* 26 (9), 1661–1670. doi:10.1007/s10147-021-01946-4

Wallace, S. S., Murphy, D. L., and Sweasy, J. B. (2012). Base excision repair and cancer. *Cancer Lett.* 327 (1-2), 73–89. doi:10.1016/j.canlet.2011.12.038

Wen, W., Liang, W., Wu, J., Kowolik, C. M., Buettner, R., Scuto, A., et al. (2014). Targeting JAK1/STAT3 signaling suppresses tumor progression and metastasis in a peritoneal model of human ovarian cancer. *Mol. Cancer Ther.* 13 (12), 3037–3048. doi:10.1158/1535-7163.MCT-14-0077

Xu, P. F., Yang, J. A., Liu, J. H., Yang, X., Liao, J. M., Yuan, F. E., et al. (2019). PI3Kβ inhibitor AZD6482 exerts antiproliferative activity and induces apoptosis in human glioblastoma cells. *Oncol. Rep.* 41 (1), 125–132. doi:10.3892/or.2018.6845

Yang, L., Zhou, Z., Yao, D., Xu, W., and Zhao, H. (2014). ET-69 * specific wip1 inhibitor, cct007093 abrogate cell proliferation, migration and invasion induced by the uvc radiation in human glioblastoma cells. *Neuro. Oncol.* 16 (5), v94. doi:10.1093/neuonc/nou255.66

Yu, A. L., Gilman, A. L., Ozkaynak, M. F., London, W. B., Kreissman, S. G., Chen, H. X., et al. (2010). Anti-GD2 antibody with GM-CSF, interleukin-2, and isotretinoin for neuroblastoma. *N. Engl. J. Med.* 363 (14), 1324–1334. doi:10.1056/NEJMoa0911123

Yun, H. S., Lee, J., Kil, W. J., Kramp, T. R., Tofilon, P. J., and Camphausen, K. (2021). The radiosensitizing effect of AZD0530 in glioblastoma and glioblastoma stem-like cells. *Mol. Cancer Ther.* 20 (9), 1672–1679. doi:10.1158/1535-7163.MCT-20-0883

Zhang, D., Chu, Y. J., Song, K. J., Chen, Y. L., Liu, W., Lv, T., et al. (2020). Knockdown of enhancer of rudimentary homolog inhibits proliferation and metastasis in ovarian cancer by regulating epithelial-mesenchymal transition. *Biomed. Pharmacother.* 125, 109974. doi:10.1016/j.biopha.2020.109974

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 09 August 2022
DOI 10.3389/fgene.2022.953739

OPEN ACCESS

EDITED BY
Yilin Zhang,
The University of Chicago, United States

REVIEWED BY
Edmund Ui-Hang Sim,
Universiti Malaysia Sarawak, Malaysia
Hongru Li,
Xiangya Hospital, Central South University, China

*CORRESPONDENCE
Chun-Li Li,
lichunli2000@qq.com
Yao Hu,
huyao125@sina.com
Hong-Jun Xiao,
xhjent_whxh@hust.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 26 May 2022
ACCEPTED 30 June 2022
PUBLISHED 09 August 2022

CITATION
Zhou L-Q, Shen J-X, Zhou T, Li C-L, Hu Y and Xiao H-J (2022), The prognostic significance of β-Catenin expression in patients with nasopharyngeal carcinoma: A systematic review and meta-analysis. *Front. Genet.* 13:953739. doi: 10.3389/fgene.2022.953739

The prognostic significance of β-Catenin expression in patients with nasopharyngeal carcinoma: A systematic review and meta-analysis

Liu-Qing Zhou[1†], Jin-Xiong Shen[1†], Tao Zhou[1†], Chun-Li Li[2*], Yao Hu[3*] and Hong-Jun Xiao[1*]

[1]Department of Otorhinolaryngology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2]Department of Otorhinolaryngology, Wuhan First Hospital/Wuhan Hospital of Traditional Chinese and Western Medicine, Wuhan, China, [3]Department of Otorhinolaryngology, The Central Hospital of Wuhan, Wuhan, China

Background: β-Catenin has been recently identified as a promising novel therapeutic target and prognostic marker in different types of cancer. Here, we conduct a meta-analysis to better clarify the correlation between β-Catenin expression and survival outcomes in nasopharyngeal carcinoma (NPC) patients.

Patients/methods: Following the Preferred Reporting Items or Systematic Reviews Meta Analyses (PRISMA) 2020 guidelines, the PubMed, Embase, Web of Science, Cochrane Library, Chinese National Knowledge Infrastructure (CNKI) and Wanfang databases were systematically searched for relevant studies to explore the prognostic significance of β-Catenin in NPC. Pooled hazards ratios (HRs) and 95% confidence intervals (CIs) were used to estimate the association of β-Catenin expression with survival outcomes in NPC patients. Odd ratios (ORs) and 95% CIs for clinicopathological characteristics were also statistically analyzed.

Results: Eight studies involving 1,179 patients with NPC were ultimately included in the meta-analysis. Pooled analysis indicated that elevated β-Catenin expression was significantly associated with poor OS (HR = 2.45, 95% CIs: 1.45–4.16, p = 0.001) and poor DFS/PFS (HR 1.79, 95% CIs: 1.29–2.49, p = 0.000). Furthermore, β-cadherin was signifcantly associated with higher TMN stages (OR = 5.10, 95% CIs 2.93–8.86, p = 0.000), clinical stages (OR = 5.10, 95% CIs 2.93–8.86, p = 0.000) and lymph node metastasis (LNM) (OR = 5.01, 95% CIs 2.40–10.44, p = 0.000).

Conclusions: This study demonstrated that for NPC, patients with elevated β-Catenin expression are more likely to have poor survival.

KEYWORDS

β-catenin, nasopharyngeal carcinoma, prognosis, meta-analysis, os

Introduction

Nasopharyngeal carcinoma (NPC) is one of the most common types of head and neck tumors and shows remarkable differences in geographic and racial distribution (Stransky et al., 2011). NPC is prevalent in Southeast Asia, especially in Southern China, the Arctic region and North Africa (Chang and Adami, 2006). Risk factors for NPC include male sex, EBV infection, Cantonese ethnicity, salt-preserved fish consumption, low fresh fruit and vegetable intake, and smoking, among others. Irrespective of the progress in radiation therapy and potent chemotherapy, approximately 5%–15% local recurrence and 15%–30% distant metastasis rates remain the main causes of failure after NPC treatment (Lee et al., 2015). Clinical staging is essential for the prognosis of NPC; however, patients at the same clinical stage may have different prognoses. In general, the current staging system is inadequate to predict survival due to variations in treatment outcomes. Hence, it is necessary to identify more reliable prognostic factors to improve the prognosis of NPC.

β-Catenin was first characterized as a family of cell-cell adhesion molecules dependent on Ca^{2+} that are present in most cell types, and it was also shown to have more detailed specificity with regard to cell-cell aggregation patterns and segregation during development (Takeichi, 1990). *β*-Catenin is one of the hallmarks of the epithelial-mesenchymal transition, which is important for early tumor metastasis and invasion (Thiery and Sleeman, 2006). It also plays a crucial role in the Wnt/β-Catenin signaling pathway, which is one of the most important signaling pathways involved in many human malignancies, and might participate in the development of various cancers and tumors (Anastas and Moon, 2013). Indeed, aberrant activation of Wnt/β-Catenin signaling is found in various types of human cancer, including osteosarcoma, lung cancer, colorectal cancer, renal cell carcinoma, breast cancer, and hepatocellular cancer, among others (Kim et al., 2002; Hoang et al., 2004; Arai et al., 2014; Jang et al., 2015; Fu et al., 2016).

Numerous studies have focused on the identification of new prognostic markers that can be used for cancer monitoring and detection. An association between *β*-Catenin expression and survival has been shown in NPC (Wang et al., 2009; Luo et al., 2012; Xu et al., 2013; Pang et al., 2019). Although many studies have reported an association between *β*-Catenin expression and NPC patient survival, the results are still controversial and ambiguous. For example, Jin et al. (2019), Sun et al. (2017), Wang et al. (2017) found that *β*-Catenin is highly expressed in NPC and is a potential risk factor that leads to an unfavorable survival prognosis in these patients. However, contradictory results were reported by Hao et al. (2014), Galera-Ruiz et al. (2011), who found no association between *β*-Catenin and survival in NPC patients compared with normal controls. In this study, we conducted a meta-analysis based on PubMed, Embase, Web of Science, Cochrane Library, Chinese National Knowledge Infrastructure (CNKI) and Wanfang databases to statistically assess the association between *β*-Catenin and the prognosis of NPC patients.

Methods

Search strategy

Following the Preferred Reporting Items or Systematic Reviews Meta Analyses (PRISMA) 2020 guidelines, electronic searches for relevant studies were performed in the PubMed, Web of Science, EMBASE, Cochrane Library, Chinese National Knowledge Infrastructure (CNKI) and Wanfang database until 1 March 2022 (Page et al., 2021). The search terms of PubMed were "((((((((((((((((((((Nasopharyngeal Neoplasm) OR (Neoplasm, Nasopharyngeal)) OR (Neoplasms, Nasopharyngeal)) OR (Nasopharynx Neoplasms)) OR (Nasopharynx Neoplasm)) OR (Neoplasm, Nasopharynx)) OR (Neoplasms, Nasopharynx)) OR (Cancer of Nasopharynx)) OR (Nasopharynx Cancers)) OR (Nasopharyngeal Cancer)) OR (Cancer, Nasopharyngeal)) OR (Cancers, Nasopharyngeal)) OR (Nasopharyngeal Cancers)) OR (Nasopharynx Cancer)) OR (Cancer, Nasopharynx)) OR (Cancers, Nasopharynx)))) OR (Cancer of the Nasopharynx)) AND (((((prognosis) OR (outcome)) OR (recurrence)) OR (survival)) OR (mortality)) OR (progression))) AND ((Catenin, beta) OR (beta-Catenin)) " The EMTREE terms were as follows "('nasopharynx cancer'/exp OR rhinopharyngioma OR 'cancer, nasopharynx' OR 'epipharynx cancer' OR 'nasopharyngeal cancer' OR 'rhinopharynx cancer') AND (prognosis OR outcome OR recurrence OR survival OR mortality OR progression) AND ('beta catenin'/exp OR "catenin beta")." Furthermore, the reference lists of retrieved articles were manually searched for additional articles. If several publications reported the same patient populations, the most complete study was enrolled to avoid duplication.

Selection criteria

This meta-analysis was limited to publications about the association between NPC and *β*-Catenin. The inclusion criteria of the meta-analysis were as follows: 1) all patients diagnosed with NPC; 2) *β*-Catenin was evaluated in both samples of NPC and normal controls; 3) the study revealed the association between *β*-Catenin and survival of NPC; 4) sufficient statistical analysis, including hazard ratios (HRs), odds ratios (ORs) and their 95% confidence intervals (95% CIs) were reported. The exclusion criteria were as follows: 1) studies without sufficient data for meta-analysis; 2) abstracts, reviews, letters, expert opinions; 3) studies about cell lines, *in vivo*/vitro studies, and human xenografts. If several studies reported the same cohort, we used the most recent one in our meta-analysis.

Data extraction

First, we inspected duplicates and removed repeated papers. Then, we carefully perused the titles and abstracts of the papers. Finally, full articles were selected to include appropriate studies. Two researchers independently evaluated the literature using the inclusion and exclusion criteria (LQ Zhou and Y Hu). Any discrepancy in assessment was resolved by consulting with a third researcher (HJ Xiao). The authors of the studies were contacted by e-mail to request data or additional information for meta-analysis calculations. Eligible studies were reviewed by two reviewers (LQ Zhou and Y Hu) independently. The Newcastle–Ottawa Scale (NOS) (Peterson et al., 2011) was included to assess the quality of the included publications, and a star system (maximum is nine stars) was adopted to evaluate a study in three domains: comparability of study groups, selection of participants and ascertainment of outcomes of interest. Scores of NOS ≥6 indicated high-quality studies. Reporting recommendations for tumor marker prognostic studies (REMARK) were also applied to evaluate study quality in cancer-related meta-analyses (Sauerbrei et al., 2018).

The following information was extracted from each publication: 1) first author's name, year, cancer type, country of the population, patient age, sample size, publication journal; 2) survival data including overall survival (OS), disease-free survival (DFS) and progression-free survival (PFS) (OS was detected from the date of medical treatment to the date of the last follow-up or death of patient; PFS was detected from the date of treatment to the date of death or recurrence tumor from any cause; DFS was detected from the date of diagnosis to the date of relapse, progression, death, or last follow-up visit and similarly censored at last follow-up visit); 3) The number of patients in each group was divided according to the TMN stages, clinical stages, the presence or absence of lymph node metastasis (LNM), gender and the number of patients with high or low β-catenin expression in each group. 4) Methods and cut-off value (Table 1).

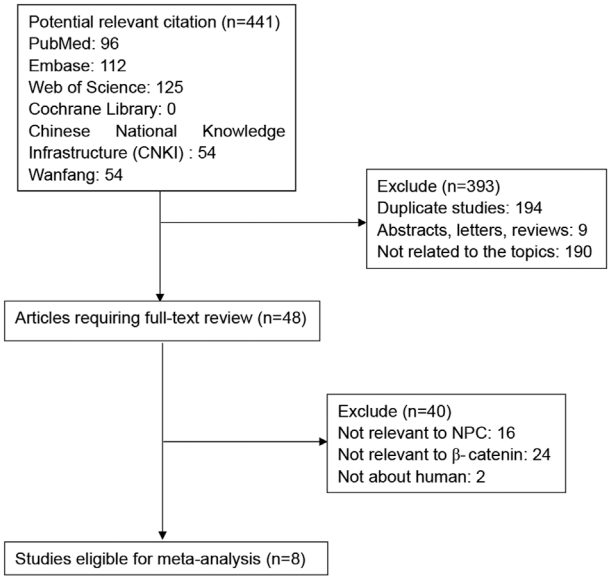


FIGURE 1
Flow diagram of the selection of relevant studies for the meta-analysis.

Statistical analysis

Pooled HRs, ORs and their 95% CIs were directly obtained or estimated by p values and other published data following Parmer's methods from the primary studies (Ambrosio et al., 2014). Statistical heterogeneity among the included studies was

TABLE 1 Characteristics of the studies examined in the meta-analysis. NR, not reported; IHC, Immunohistochemistry; RT-qPCR, Reverse transcription-quantitative polymerase chain reaction.

Author	Year	Country	Sample size	Age	Follow-up (month)	Survival analysis	Methods	Cut-off value	NOS/REMARK score
Hao	2014	Canada	279	51.7 (18-85)	48 (3-120)	OS, DFS	IHC	NR	6/15
Jin	2019	China	164	45.3 (24-70)	49.2 (9-60)	OS, DFS, DMFS, LRFS	IHC, RT-qPCR	50%	7/13
Pang	2019	China	175	NR (22-69)	NR (36-48)	OS	IHC	75%	7/14
Sun	2017	China	128	NR	NR	OS, PFS	IHC, RT-qPCR	50%	7/12
Wang	2009	China	111	47 (18-71)	65 (8-88)	PFS	IHC	50%	7/11
Wang	2017	China	163	NR	NR	OS	IHC	70%	8/11
Xu	2013	China	148	NR	78 (10-125)	OS	IHC	NR	7/12
Luo	2012	China	122	47.2 (15-73)	51.9 (8-92)	OS	IHC	50%	7/12

assessed by the χ^2-based Q test and I^2 test (Higgins et al., 2003). The fixed-effect model was used for analysis in the absence of significant heterogeneity between studies ($p > 0.10$, $I^2 = 0\%$); we adopted the random-effects model if significant heterogeneity was present. We also performed sensitivity analysis to investigate the influence of each individual study on the overall pooled results. Begg's test and Egger's test were applied to detect publication bias ($p > 0.05$ indicated no publication bias). All statistical analyses were performed using STATA statistical software version 12.0 (STATA, College Station, TX).

Results

Study selection and characteristics

As shown in Figure 1, a total of 312 potential publications were initially identified by searching the PubMed, Web of Science, EMBASE, Cochrane Library, Chinese National Knowledge Infrastructure (CNKI) and Wanfang databases. Following exclusion of duplicates (n = 194), abstracts, letters and reviews (n = 9), and studies not related to the topics (n =

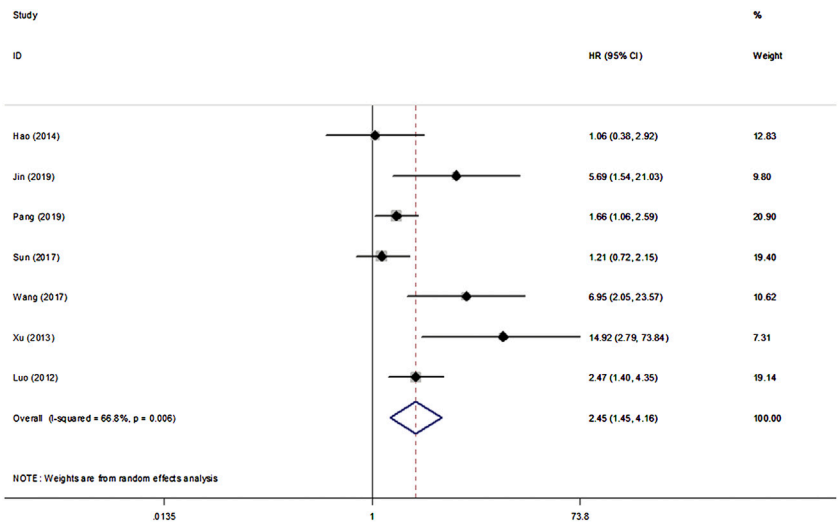


FIGURE 2
Forest plot indicating the association between β-Catenin expression and OS in NPC.

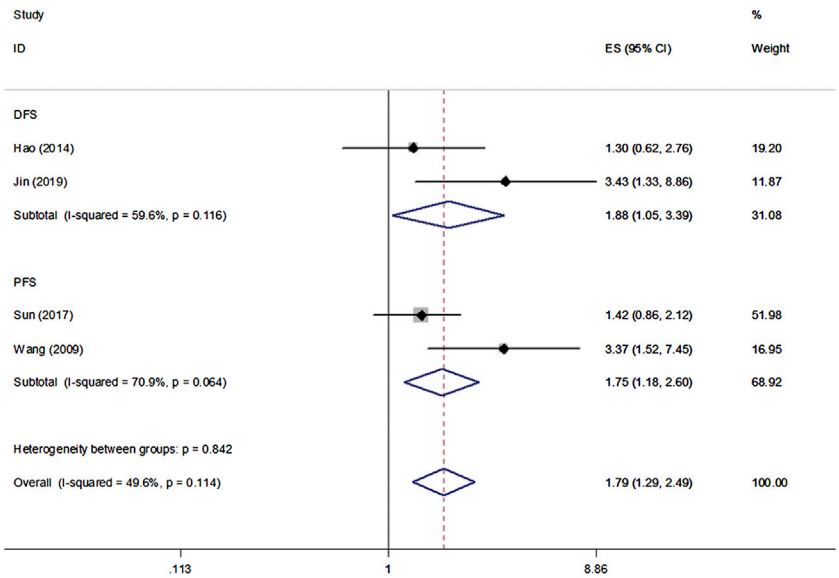


FIGURE 3
Forest plot examining the association between β-Catenin expression and DFS/PFS in NPC.

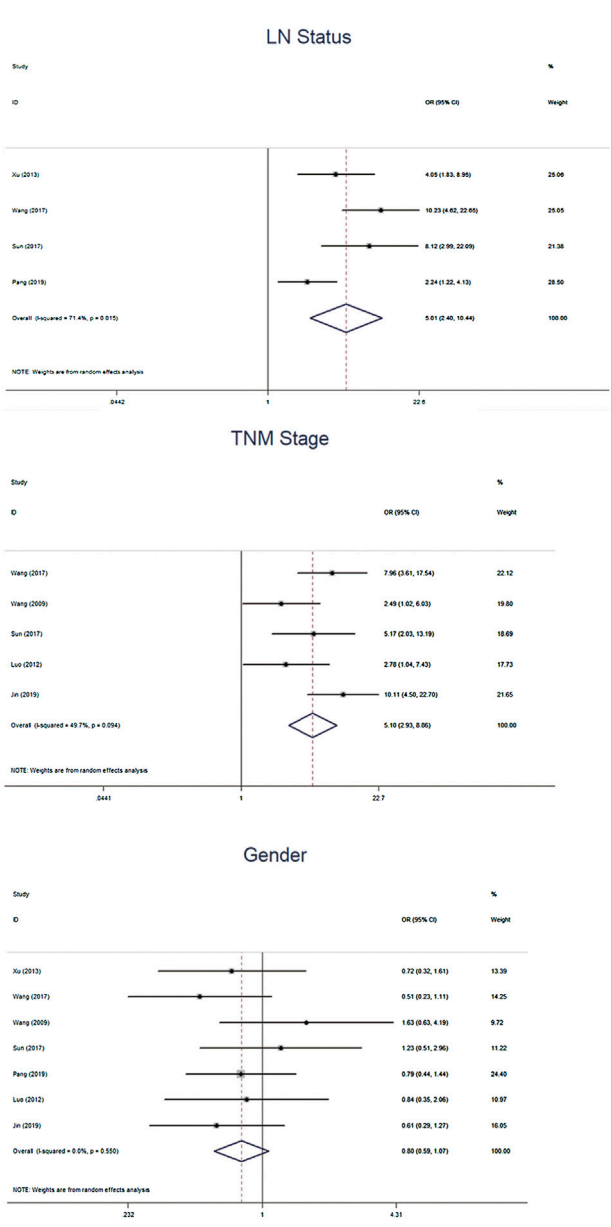


FIGURE 4
Forest plot examining the association between β-cadherin and NPC patients outcomes.

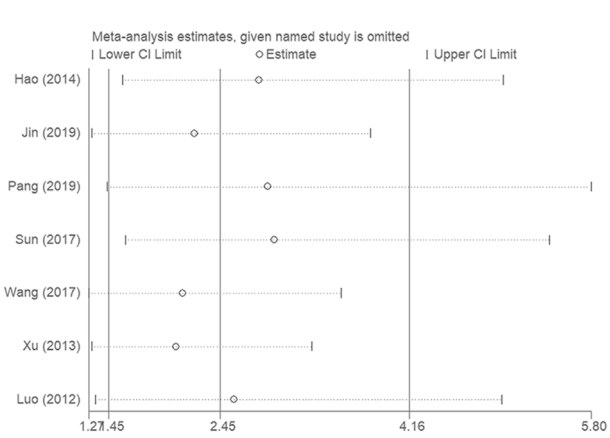


FIGURE 5
Sensitivity analyses were conducted to evaluate the impact of each single study on the overall effect.

190), the remaining potentially relevant studies (n = 48) were further identified by reading their full texts. 40 studies did not provide specific data regarding NPC or β-Catenin and therefore were excluded. Finally, eight studies between 2009 and 2020 with a total of 1179 NPC patients were included in our meta-analysis.

The study characteristics are summarized in Table 1. All of the eight publications involved >100 patients. Seven studies including 1,068 patients reported OS, 2 studies including 443 patients reported DFS, and 2 studies including 239 patients reported PFS. All HRs, ORs and 95%CIs values were directly reported in the original study. NOS scores for all publications were above 6, and REMARK scores were between 11-15.

Association between β-Catenin and survival in nasopharyngeal carcinoma patients

All eight studies in our analysis reported the association between β-Catenin and the OS, DFS and PFS of patients with NPC. Heterogeneity among the publications was significant based on the Q test (p < 0.1). Hence, the random-effect model was adopted and showed that β-Catenin was significantly associated with shorter OS in NPC (HR = 2.45, 95% CIs: 1.45–4.16, p = 0.001). Medium heterogeneity was noted between β-Catenin expression and OS (I^2 = 66.8%, $P_{heterogeneity}$ = 0.006) (Figure 2). Furthermore, two studies including 433 patients reported DFS, and two studies including 239 patients reported PFS. A significant correlation between β-Catenin and shorter DFS/PFS (HR = 1.79, 95% CIs: 1.29–2.49, p = 0.000) was observed, with low heterogeneity (I^2 = 49.6%, $P_{heterogeneity}$ = 0.114) (Figure 3).

Association between β-cadherin and nasopharyngeal carcinoma patients outcomes

We further calculated the pooled ORs and the 95% CIs to evaluate the association between β-catenin and NPC outcomes: gender (female vs male), TMN stage (T3–4 vs T1–2), clinical stage (T3–4 vs T1–2) and lymph node (LN) status (LNM vs No LNM). The pooled analysis showed that β-cadherin was signifcantly associated with higher TMN stages (OR = 5.10, 95% CIs 2.93–8.86, p = 0.000) and LNM (OR = 5.01, 95% CIs 2.40–10.44,

$p = 0.000$). However, β-cadherin was not signifcantly correlated with gender (OR = 0.80, 95% CIs 0.59–1.07, $p = 0.135$) (Figure 4).

Sensitivity analysis

Sensitivity analysis was conducted to evaluate the impact of each single study on the overall effect. As depicted in Figure 5, the analysis did not detect a single study that significantly altered the combined results. Overall, the pooled effect size of our meta-analytic results was stable and reliable.

Publication bias

Publication bias was assessed by using Begg's funnel plots and Egger's test. The results were quite symmetric, with those based on Begg's funnel plot ($p = 0.077$) and Egger's test ($p = 0.077$) revealing no publication bias among the studies (Figure 6).

Meta-regression analysis

Medium heterogeneity was noted between β-Catenin expression and OS (I2 = 66.8%, Pheterogeneity = 0.006). Hence, the meta regression analyses were used to explain statistical heterogeneity. The HR was not modified by the year of publication, female ratio (%), area, sample size, or quality score, this result does not fully explain the medium level of heterogeneity observed.

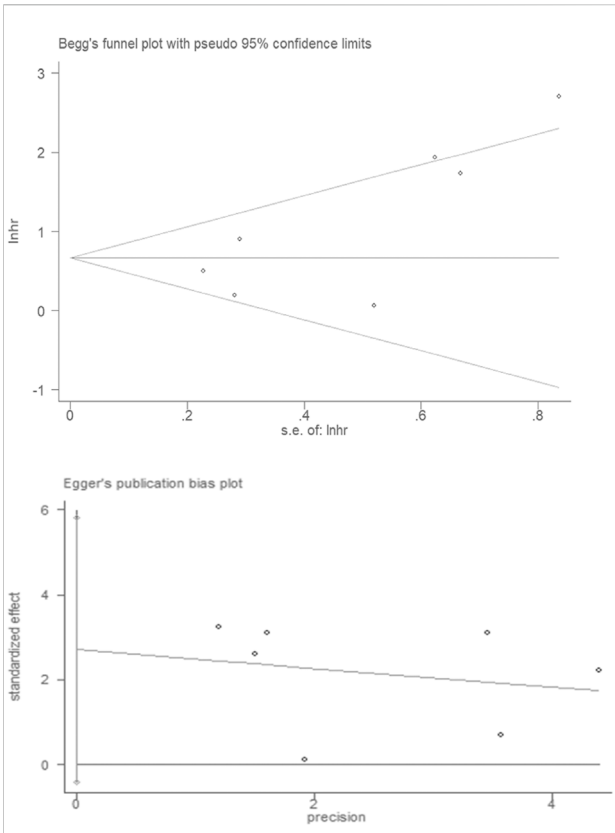


FIGURE 6
Publication bias in the enrolled studies. Publication bias was assessed using Begg's funnel plots and Egger's test.

Discussion

The present study is the first meta-analysis including eight published studies with 1,179 patients to provide useful information for clinical decision-making in NPC. β-Catenin was significantly associated with shorter OS in NPC patients, with HR values of 2.45. Significant correlation between β-Catenin and shorter DFS/PFS (HR 1.79) was also observed. Furthermore, our results also demostrated that β-cadherin was signifcantly associated with higher TMN stages, clinical stages and LNM. These results confirm the clinical value of β-Catenin in NPC. NPC tumor cells invade adjacent tissues or metastasize to regional lymph nodes at an early stage of NPC development (Wei and Mok, 2007), though the exact mechanism underlying the process remains unknown. It has been reported that cell–cell adhesion molecules, cytokines and the matrix metalloproteinase family may be involved in adjacent invasion and distant metastasis. β-Catenin is a key mediator in the cadherin-Catenin complex, which is essential for connecting the actin filaments of cells to the cell-cell interface at adherent junctions (Anastas and Moon, 2013); it is also a key mediator of canonical signaling in the Wnt/β-Catenin pathway. β-Catenin can accumulate in both the cytoplasm and nucleus (Khramtsov et al., 2010), and it helps to promote the progression of tumors by suppressing T-cell responses (Hong et al., 2015). Gene mutations or aberrant activation of Wnt receptors activate Wnt/β-Catenin signaling and trigger tumorigenesis in the skin, colon, liver, bone marrow, breast, and possibly other tissues (Fodde and Brabletz, 2007; Monga, 2015). In addition, β-Catenin plays roles in maintaining the stemness of normal intestinal cells, and high-level nuclear localization and cytoplasmic expression promote cancer cell proliferation and survival (Valkenburg et al., 2011). β-Catenin is high expressed when Wnt/β-catenin signal is aberrantly activated, it activates numerous Wnt pathway downstream proliferation signals, including c-Myc and cyclin D1 and finally accelerates cell cycle, facilitates cell proliferation and migration, which induced to poor diagnosis of NPC (Alamoud and Kukuruzinska, 2018).

Targeted therapies have produced striking benefits for cancer patients. The Wnt/β-Catenin pathway has been proven to play a key role in various kinds of carcinomas (Sanchez-Vega et al., 2018). Therefore, this signaling pathway is a preferable target for

fighting cancer. Although there are no drugs specifically inhibiting this signaling pathway approved for the clinic, intensive efforts have been made in signaling pathway development. Wnt/β-Catenin pathway inhibitors can be classified into five categories according to their properties: peptides, small molecules, antibodies, natural compounds and RNA interference (Cui et al., 2018). There are already some ongoing phase 1/2 trials with Wnt/β-Catenin pathway inhibitors in metastatic colorecta, head and neck cancers, breast cancers and some other cancers (Krishnamurthy and Kurzrock, 2018). These trials provide proof that in certain patients, cancer can be treated with Wnt/β-Catenin inhibitors. According to our results, Wnt/β-Catenin inhibitors may constitute therapeutics against NPC.

Nevertheless, the present meta-analysis contains several limitations. First, significant heterogeneity was noted in the association between β-Catenin and the OS of patients with NPC. The heterogeneity of the population was most likely due to differences in the baseline characteristics of the included patients (age, race, and tumor stage), the duration of follow-up, the method of mutation detection, and other parameters. A random-effects model was employed to minimize the effects of these differences. Second, the number of articles used for assessing the association between β-Catenin and the prognosis of NPC patients was limited in the present meta-analysis. Therefore, additional studies are required to produce accurate conclusions. Finally, our results may overestimate the prognostic significance of β-Catenin to some extent because the majority of the included studies reported positive results.

In summary, we searched electronic databases, and a total of 1,179 patients in eight studies were enrolled for meta-analysis, demonstrating that patients with elevated β-Catenin expression are more likely to have poorer prognosis. Taken together, our meta-analysis results suggest that β-Catenin has prognostic value for NPC. However, studies with larger sample sizes are needed to obtain more representative and precise results.

Author Contributions

YH, CL, and L-QZ collected and analyzed the data, wrote the paper. L-QZ, YH, JS, and H-JX analyzed the data and wrote the paper. LZ and H-JX conceived and designed this study, analyzed the data, wrote the paper. All authors read and approved the final manuscript.

Funding

This work was financially supported by the National Natural Science Foundation of China (81500792 and 81771005).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.953739/full#supplementary-material

SUPPLEMENTARY TABLE S1
Quality assessment based on the Recommendations for Tumor Marker Prognostic Studies (REMARK).

SUPPLEMENTARY TABLE S2
PRISMA_2020_checklist.

SUPPLEMENTARY TABLE S3
PRISMA_2015_checklist.

SUPPLEMENTARY TABLE S4
Quality assessment based on the Newcastle-Ottawa Scale (NOS).

References

Alamoud, K. A., and Kukuruzinska, M. A. (2018). Emerging insights into wnt/β-catenin signaling in head and neck cancer. *J. Dent. Res.* 97 (6), 665–673. doi:10.1177/0022034518771923

Ambrosio, M. R., De Falco, G., Gozzetti, A., Rocca, B. J., Amato, T., Mourmouras, V., et al. (2014). Plasmablastic transformation of a pre-existing plasmacytoma: A possible role for reactivation of epstein barr virus infection. *Haematologica* 99 (11), e235–7. doi:10.3324/haematol.2014.111872

Anastas, J. N., and Moon, R. T. (2013). WNT signalling pathways as therapeutic targets in cancer. *Nat. Rev. Cancer* 13 (1), 11–26. doi:10.1038/nrc3419

Arai, E., Sakamoto, H., Ichikawa, H., Totsuka, H., Chiku, S., Gotoh, M., et al. (2014). Multilayer-omics analysis of renal cell carcinoma, including the whole exome, methylome and transcriptome. *Int. J. Cancer* 135 (6), 1330–1342. doi:10.1002/ijc.28768

Chang, E. T., and Adami, H.-O. (2006). The enigmatic epidemiology of nasopharyngeal carcinoma. *Cancer Epidemiol. Biomarkers Prev.* 15 (10), 1765–1777. doi:10.1158/1055-9965.EPI-06-0353

Cui, C., Zhou, X., Zhang, W., Qu, Y., and Ke, X. (2018). Is β-catenin a druggable target for cancer therapy? *Trends biochem. Sci.* 43 (8), 623–634. doi:10.1016/j.tibs.2018.06.003

Fodde, R., and Brabletz, T. (2007). Wnt/beta-catenin signaling in cancer stemness and malignant behavior. *Curr. Opin. Cell Biol.* 19 (2), 150–158. doi:10.1016/j.ceb.2007.02.007

Fu, X., Li, H., Liu, C., Hu, B., Li, T., and Wang, Y. (2016). Long noncoding RNA AK126698 inhibits proliferation and migration of non-small cell lung cancer cells by targeting Frizzled-8 and suppressing Wnt/β-catenin signaling pathway. *Onco. Targets. Ther.* 9, 3815–3827. doi:10.2147/OTT.S100633

Galera-Ruiz, H., Rios, M. J., Gonzalez-Campora, R., de Miguel, M., Carmona, M. I., Moreno, A. M., et al. (2011). The cadherin-catenin complex in nasopharyngeal carcinoma. *Eur. Arch. Otorhinolaryngol.* 268 (9), 1335–1341. doi:10.1007/s00405-010-1464-z

Hao, D., Phan, T., Jagdis, A., Siever, J. E., Klimowicz, A. C., Laskin, J. J., et al. (2014). Evaluation of E-cadherin, β-catenin and vimentin protein expression using quantitative immunohistochemistry in nasopharyngeal carcinoma patients. *Clin. Invest. Med.* 37 (5), E320–E330. doi:10.25011/cim.v37i5.22012

Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj* 327 (7414), 557–560. doi:10.1136/bmj.327.7414.557

Hoang, B. H., Kubo, T., Healey, J. H., Sowers, R., Mazza, B., Yang, R., et al. (2004). Expression of LDL receptor related protein 5 (LRP5) as a novel marker for disease progression in high-grade osteosarcoma. *Int. J. Cancer* 109 (1), 106–111. doi:10.1002/ijc.11677

Hong, Y., Manoharan, I., Suryawanshi, A., Majumdar, T., Angus-Hill, M. L., Koni, P. A., et al. (2015). β-catenin promotes regulatory T-cell responses in tumors by inducing vitamin A metabolism in dendritic cells. *Cancer Res.* 75 (4), 656–665. doi:10.1158/0008-5472.CAN-14-2377

Jang, G.-B., Kim, J. Y., Cho, S. D., Park, K. S., Jung, J. Y., Lee, H. Y., et al. (2015). Blockade of Wnt/β-catenin signaling suppresses breast cancer metastasis by inhibiting CSC-like phenotype. *Sci. Rep.* 5, 12465. doi:10.1038/srep12465

Jin, P.-Y., Zheng, Z. H., Lu, H. J., Yan, J., Zheng, G. H., Zheng, Y. L., et al. (2019). Roles of β-catenin, TCF-4, and survivin in nasopharyngeal carcinoma: Correlation with clinicopathological features and prognostic significance. *Cancer Cell Int.* 19 (1), 48. doi:10.1186/s12935-019-0764-7

Khramtsov, A. I., Khramtsova, G. F., Tretiakova, M., Huo, D., Olopade, O. I., and Goss, K. H. (2010). Wnt/beta-catenin pathway activation is enriched in basal-like breast cancers and predicts poor outcome. *Am. J. Pathol.* 176 (6), 2911–2920. doi:10.2353/ajpath.2010.091125

Kim, H. C., Kim, H. J., and Kim, J. C. (2002). Reduced E-cadherin expression as a cause of distinctive signet-ring cell variant in colorectal carcinoma. *J. Korean Med. Sci.* 17 (1), 23–28. doi:10.3346/jkms.2002.17.1.23

Krishnamurthy, N., and Kurzrock, R. (2018). Targeting the Wnt/beta-catenin pathway in cancer: Update on effectors and inhibitors. *Cancer Treat. Rev.* 62, 50–60. doi:10.1016/j.ctrv.2017.11.002

Lee, A., Ma, B. B. Y., Ng, W. T., and Chan, A. T. C. (2015). Management of nasopharyngeal carcinoma: Current practice and future perspective. *J. Clin. Oncol.* 33 (29), 3356–3364. doi:10.1200/JCO.2015.60.9347

Luo, W., Fang, W., Li, S., and Yao, K. (2012). Aberrant expression of nuclear vimentin and related epithelial-mesenchymal transition markers in nasopharyngeal carcinoma. *Int. J. Cancer* 131 (8), 1863–1873. doi:10.1002/ijc.27467

Monga, S. P. (2015). β-Catenin signaling and roles in liver homeostasis, injury, and tumorigenesis. *Gastroenterology* 148 (7), 1294–1310. doi:10.1053/j.gastro.2015.02.056

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Int. J. Surg.* 88, 105906. doi:10.1016/j.ijsu.2021.105906

Pang, Q., Hu, W., Zhang, X., and Pang, M. (2019). Wnt/β-Catenin signaling pathway-related proteins (DKK-3, β-catenin, and c-MYC) are involved in prognosis of nasopharyngeal carcinoma. *Cancer biother. Radiopharm.* 34 (7), 436–443. doi:10.1089/cbr.2019.2771

Peterson, J., Shea, B. J., O'Connell, A. M., Peterson, J., Welch, V., Losos, M., et al. (2011). *The newcastle-ottawa scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses*. Ottawa: Ottawa Hospital Research Institute 2.1, 1–12.

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W. K., Luna, A., La, K. C., et al. (2018). Oncogenic signaling pathways in the cancer genome atlas. *Cell* 173 (2), 321–337. doi:10.1016/j.cell.2018.03.035e10

Sauerbrei, W., Taube, S. E., McShane, L. M., Cavenagh, M. M., and Altman, D. G. (2018). Reporting recommendations for tumor marker prognostic studies (REMARK): An abridged explanation and elaboration. *J. Natl. Cancer Inst.* 110 (8), 803–811. doi:10.1093/jnci/djy088

Stransky, N., Egloff, A. M., Tward, A. D., Kostic, A. D., Cibulskis, K., Sivachenko, A., et al. (2011). The mutational landscape of head and neck squamous cell carcinoma. *Science* 333 (6046), 1157–1160. doi:10.1126/science.1208130

Sun, H., Liu, M., Wu, X., Yang, C., Zhang, Y., Xu, Z., et al. (2017). Overexpression of N-cadherin and β-catenin correlates with poor prognosis in patients with nasopharyngeal carcinoma. *Oncol. Lett.* 13 (3), 1725–1730. doi:10.3892/ol.2017.5645

Takeichi, M. (1990). Cadherins: A molecular family important in selective cell-cell adhesion. *Annu. Rev. Biochem.* 59 (1), 237–252. doi:10.1146/annurev.bi.59.070190.001321

Thiery, J. P., and Sleeman, J. P. (2006). Complex networks orchestrate epithelial–mesenchymal transitions. *Nat. Rev. Mol. Cell Biol.* 7 (2), 131–142. doi:10.1038/nrm1835

Valkenburg, K. C., Graveel, C. R., Zylstra-Diegel, C. R., Zhong, Z., and Williams, B. O. (2011). Wnt/β-catenin signaling in normal and cancer stem cells. *Cancers (Basel)* 3 (2), 2050–2079. doi:10.3390/cancers3022050

Wang, F. L., Guo, X., Yuan, T. Z., Cao, S. M., Rao, H. L., Hou, J. H., et al. (2009). Expression and clinical significance of Wnt-1 and beta-catenin in nasopharyngeal carcinoma. *Ai Zheng* 28 (1), 72–75.

Wang, W., Wen, Q., Luo, J., Chu, S., Chen, L., Xu, L., et al. (2017). Suppression of β-catenin nuclear translocation by CGP57380 decelerates poor progression and potentiates radiation-induced apoptosis in nasopharyngeal carcinoma. *Theranostics* 7 (7), 2134–2149. doi:10.7150/thno.17665

Wei, W. I., and Mok, V. W. (2007). The management of neck metastases in nasopharyngeal cancer. *Curr. Opin. Otolaryngol. Head. Neck Surg.* 15 (2), 99–102. doi:10.1097/MOO.0b013e3280148a06

Xu, L., Jiang, Y., Zheng, J., Xie, G., Li, J., Shi, L., et al. (2013). Aberrant expression of β-catenin and E-cadherin is correlated with poor prognosis of nasopharyngeal cancer. *Hum. Pathol.* 44 (7), 1357–1364. doi:10.1016/j.humpath.2012.10.025

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 26 August 2022
DOI 10.3389/fgene.2022.952083

OPEN ACCESS

EDITED BY
Yilin Zhang,
The University of Chicago, United States

REVIEWED BY
Antonio Neme,
National Autonomous University of Mexico, Mexico
Hee Sung Kim,
Chung-Ang University, South Korea

*CORRESPONDENCE
Qiao-Ying Lv,
lvqiaoying@126.com
Xiao-Jun Chen,
xiaojunchen2013@sina.com

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

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 24 May 2022
ACCEPTED 26 July 2022
PUBLISHED 26 August 2022

CITATION
Hu J-L, Yierfulati G, Wang L-L, Yang B-Y, Lv Q-Y and Chen X-J (2022), Identification of potential models for predicting progestin insensitivity in patients with endometrial atypical hyperplasia and endometrioid endometrial cancer based on ATAC-Seq and RNA-Seq integrated analysis. *Front. Genet.* 13:952083. doi: 10.3389/fgene.2022.952083

Identification of potential models for predicting progestin insensitivity in patients with endometrial atypical hyperplasia and endometrioid endometrial cancer based on ATAC-Seq and RNA-Seq integrated analysis

Jia-Li Hu[1,2†], Gulinazi Yierfulati[1,2†], Lu-Lu Wang[1,2], Bing-Yi Yang[1,2], Qiao-Ying Lv[1,2]* and Xiao-Jun Chen[1,2]*

[1]Department of Gynecology, Obstetrics and Gynecology Hospital of Fudan University, Shanghai, China, [2]Shanghai Key Laboratory of Female Reproductive Endocrine Related Diseases, Shanghai, China

Objective: The aim of this study was to establish predictive models based on the molecular profiles of endometrial lesions, which might help identify progestin-insensitive endometrial atypical hyperplasia (EAH) or endometrioid endometrial cancer (EEC) patients before progestin-based fertility-preserving treatment initiation.

Methods: Endometrial lesions from progestin-sensitive (PS, n = 7) and progestin-insensitive (PIS, n = 7) patients were prospectively collected before progestin treatment and then analyzed by ATAC-Seq and RNA-Seq. Potential chromatin accessibility and expression profiles were compared between the PS and PIS groups. Candidate genes were identified by bioinformatics analyses and literature review. Then expanded samples (n = 35) were used for validating bioinformatics data and conducting model establishment.

Results: ATAC-Seq and RNA-Seq data were separately analyzed and then integrated for the subsequent research. A total of 230 overlapping differentially expressed genes were acquired from ATAC-Seq and RNA-Seq integrated analysis. Further, based on GO analysis, REACTOME pathways, transcription factor prediction, motif enrichment, Cytoscape analysis and literature review, 25 candidate genes potentially associated with progestin insensitivity were identified. Finally, expanded samples were used for data verification, and based on these data, three predictive models comprising 9 genes (*FOXO1*, *IRS2*, *PDGFC*, *DIO2*, *SOX9*, *BCL11A*, *APOE*, *FYN*, and *KLF4*) were established with an overall predictive accuracy above 90%.

Conclusion: This study provided potential predictive models that might help identify progestin-insensitive EAH and EEC patients before fertility-preserving treatment.

KEYWORDS
predictive models, progestin insensitivity, endometrial atypical hyperplasia, endometrioid endometrial cancer, ATAC-seq, RNA-seq, fertility-preserving treatment

Introduction

Endometrioid endometrial cancer (EEC) is one of the most common gynecological malignancies, with an increasing trend in new cancer cases and deaths each year (Siegel et al., 2022). Notably, EEC and its precancerous lesions, endometrial atypical hyperplasia (EAH), present a younger trend, and approximately half of young EEC and EAH patients are nulliparous when diagnosed (Trojano et al., 2019). Therefore, fertility-sparing treatment for these patients has attracted increasing attention in clinical research. Currently, high-dose progestin therapy is the main conservative strategy and achieves an approximately 70–80% complete response (CR) rate, and the median duration from treatment to CR is as long as six to 7 months (Gallos et al., 2012; Gunderson et al., 2012; Yang et al., 2019; Westin et al., 2021). However, there are still approximately 20–30% of cases are not sensitive to progestin and having to switch to second-line treatment or even receive definitive surgery. Identifying progestin-insensitive (PIS) cases accurately before progestin treatment initiation might aid clinicians in providing more efficient treatment for these patients and thus improve the overall outcome of fertility-preserving treatment.

There is still a lack of objective indicators predicting progestin sensitivity in EAH or EEC patients. Studies have shown that positive progesterone receptor (PR) expression in EAH and EEC tissues was associated with shorter CR time of fertility-sparing therapy (Yamazawa et al., 2007; Raffone et al., 2019; Wang et al., 2021). While the abnormal expression of other molecular markers, such as elevated dual-specificity phosphatase 6 or downregulated nuclear factor NF-E2-related factor or survivin, might be associated with progestin insensitivity (Zhang et al., 2015; Fan et al., 2017). However, there is less high-quality evidence of molecular markers that can be used to predict progestin response in EAH and EEC cases. Therefore, further studies are still needed to explore promising models for predicting progestin response in EAH and EEC cases.

To explore potential predictive models for predicting progestin insensitivity in EAH or EEC patients before receiving progestin-based fertility-preserving treatment, this study was designed based on assay for transposase-accessible chromatin sequencing (ATAC-Seq) and RNA sequencing (RNA-Seq) of EAH and EEC tissues. Based on ATAC-Seq and RNA-Seq integrated bioinformatics analyses and literature review, candidate genes were identified and further verified in another 35 cases for predictive model construction. Our study provided potential models for predicting progestin insensitivity in patients with EAH and EEC.

Materials and methods

Ethics statement

This is a retrospective study using samples prospectively collected from December 2017 to November 2020, in the Obstetrics and Gynecology Hospital of Fudan University, Shanghai, China (hereafter referred to as 'Ob&Gyn Hospital'). This study was approved by the Ethics Committees of Ob&Gyn Hospital (Approval NO. 2021-130). Patients were fully informed of the use of their medical data and pathological samples for scientific research, and signed informed consent forms.

Patient selection and tissue collection

Young patients diagnosed with EAH or well-differentiated EEC receiving progestin-based fertility-sparing treatment were prospectively registered. All patients were pathologically diagnosed with EAH or EEC for the first time by endometrial biopsy with or without hysteroscopy. Inclusion and exclusion criteria as well as treatment regimen and evaluation procedure were as previously reported (Yang et al., 2020). Briefly, patients received progestin-based treatment, hysteroscopic evaluation and endometrial biopsy every 3 months on average. Pathological diagnosis was confirmed by at least two experienced gynecological pathologists independently according to the World Health Organization (WHO) pathological classification (2020). If their opinions differed, a seminar was held in the pathological department for the final diagnosis.

'PIS' was defined as disease progression at any time during treatment, stable disease after 7 months of treatment, or did not achieve CR after 10 months of treatment (Zhou and Xu, 2021). Other patients who achieved CR within 10 months of treatment were regarded as 'PS'.

Endometrial lesions before progestin treatment initiation were prospectively collected through biopsy under hysteroscopy and stored at -80°C equipped with or without RNA preservation solution. Samples from 7 PIS patients and 7 PS patients were firstly collected for ATAC-Seq and RNA-Seq analyses from December 2017 to November 2019 (regarded as the 'Analysis Group'). Because the number of EAH or EEC patients receiving fertility preserving treatment is relatively low, we tried to collect as many patients as possible for validation to minimize possible bias caused by low case number. As a result, a total of 35 cases met the inclusion and exclusion criteria of this study were recruited from November 2019 to November 2020. These patients were regarded as

TABLE 1 General characteristics of the study population.

Variables	Analysis group				Construction group				
	Total	PS	PIS	*p value	Total	PS-C	sub-PS-C	PIS-C	+p value
Patients (n)	14	7	7	—	35	13	15	7	—
Diagnosis									1.000
EAH	7 (50)	4 (57.1)	4 (57.1)	1.000	25 (69.44)	10 (76.92)	10 (66.67)	5 (62.5)	—
EEC	7 (50)	3 (42.9)	3 (42.9)		11 (30.56)	3 (23.08)	5 (33.33)	3 (37.5)	—
Age at diagnosis (year)	31 (26–36)	34 (28–36)	30 (26–34)	0.097	32.5 (21–42)	34 (21–39)	30 (23–36)	34 (24–42)	0.2895
BMI (kg/m2)	28.26 (20.70–37.65)	28.13 (23.44–36.13)	28.40 (20.70–37.65)	0.710	28.09 (18.87–45.17)	26.15 (18.87–37.74)	28.04 (19.57–45.17)	29.94 (20.28–35.26)	0.880
HOMA-IR	4.15 (1.40–6.37)	4.41 (1.47–6.37)	3.53 (1.40–5.58)	0.535	3.16 (0.84–22.80)	4.12 (1.18–10.13)	3.23 (0.84.22.80)	2.35 (1.56–7.64)	0.647
MS§	8 (57.1)	4 (57.1)	4 (57.1)	1.000	15 (41.7)	5 (38.5)	6 (40.0)	4 (50.0)	0.830
Hypertension	3 (21.4)	2 (28.6)	1 (14.3)	1.000	3 (8.3)	1 (7.7)	2 (13.3)	0 (0.0)	0.782
Diabetes mellitus	0 (0.0)	0 (0.0)	0 (0.0)	—	4 (11.1)	1 (7.7)	2 (13.3)	1 (12.5)	1.000
Nulliparous	11 (78.6)	5 (71.4)	6 (85.7)	1.000	29 (80.6)	9 (69.2)	13 (86.7)	7 (87.5)	0.553
Progestin therapy									
MA	6	2 (28.6)	4 (57.1)		12 (33.3)	2 (15.4)	8 (53.3)	2 (25.0)	
MA + Metformin	4	2 (28.6)	2 (28.6)		12 (33.3)	4 (30.8)	6 (40.0)	2 (25.0)	
LNG-IUD	1	1 (14.3)	0 (0)		4 (11.1)	3 (23.1)	0 (0.0)	1 (12.5)	
MA + LNG-IUD	3	2 (28.6)	1 (14.3)		4 (11.11)	2 (15.4)	0 (0.0)	2 (25.0)	
MA + Rosuvastatin	-	-	-		4 (11.1)	2 (15.4)	1 (6.7)	1 (12.5)	
CR time (months)††	7.8 (3.7–29.5)	7.0 (3.7–7.9)	12.0 (6.0–29.5)	0.011	6.33 (3.07–13.23)	3.9 (3.07–4.90)	6.87 (5.87–8.1)	11.17 (10.53–13.23)‡‡	<0.0001

††Total treatment duration from initiation of conservative treatment to CR.

‡‡Note: CR time of one patient in PIS-C group was not included, because this patient did not achieve CR and underwent hysterectomy eventually.

§Diagnosis of MS meets at least three of the following criteria: 1) BP ≥ 130/85 mmHg or hypertension; 2) Waist circumference ≥80 cm; 3) Total cholesterol ≥1.7 mmol/L; 4) High density lipoprotein <1.04 mmol/L; 5) Fasting plasma glucose ≥5.6 mmol/L or type II diabetes mellitus.

*p value: comparison between PS group and PIS group in Analysis Group.

+p value: comparison between PS-C group, sub-PS-C group and PIS-C group in Construction Group.

Values are presented as median (range) or number (%).

PS, progestin-sensitive; PIS, progestin-insensitive; PS-C, progestin-sensitive in Construction Group; sub-PS-C, progestin-sub-sensitive in Construction Group; PIS-C, progestin-insensitive in Construction Group; EAH, endometrial atypical hyperplasia; EEC, endometrioid endometrial cancer; BMI, body mass index; HOMA-IR, homeostasis model assessment-insulin resistance; MS, metabolic syndrome; MA, megestrol acetate; LNG-IUD, levonorgestrel intrauterine device; CR, complete response.

'Construction Group' for validation and model construction. They were further classified as PS-C (achieved CR within 5 months of treatment, n = 13), sub-PS-C (achieved CR within 5–9 months of treatment, n = 15) and PIS-C (n = 7). The basic characteristics of the enrolled patients were shown in Table 1.

Library construction and ATAC-Seq analysis

ATAC-Seq was performed to analyze transposase accessible chromatin as previously described (Buenrostro et al., 2015). An improved ATAC-Seq protocol that reduces background and enables interrogation of frozen tissues was used for nuclei collection (Corces et al., 2017). Libraries were pooled at equimolar ratios with barcodes and sequenced on the BGISEQ-500 platform (BGI, Shenzhen, China).

Raw sequence reads were initially processed for quality control by FastQC. Before statistical analysis, ATAC-Seq read counts of different samples were normalized according to the methods described previously (Zhang and Parmigiani, 2020). In ATAC-Seq analysis, opening or closing peaks were chosen with $|\log_2$ fold change$|>0.5849$ and non-adjusted $p < 0.05$ (PIS vs. PS). The proportion of all reads in each sample was matched to the elements in the human genome according to functional and

positional information, including 3′ UTR, 5′UTR, distal intergenic, downstream, exon, intron, and promoter. Scatter plot showed the accessibility at each peak. Hierarchical cluster analysis was performed to assess chromatin accessibility with differential gene peaks.

Library construction and RNA-Seq analysis

RNA-Seq was performed to assess the expression of genes in tissue samples as described previously (Wang et al., 2018a). Libraries were generated on the BGIseq500 platform (BGI-Shenzhen, China). Fragments per kilobase per million reads (FPKM) was used to quantitatively estimate gene expression values (Trapnell et al., 2010). DESeq2 was used to analyze the raw count (Wang et al., 2010). Before statistical analysis, RNA-Seq read counts of different samples were normalized according to a previously reported method (Zhang and Parmigiani, 2020). Differential expression analysis was performed using the R DESeq2 package (v1.4.5) (Love et al., 2014). Genes with $|\log_2$ fold change$|>0.5849$ and non-adjusted $p < 0.05$ (PIS vs. PS) were defined as differentially expressed genes (DEGs) between the PIS and PS patients. A heatmap was drawn to cluster the DEGs. The DEGs were further analyzed by Gene Ontology (GO) and REACTOME pathways to determine the potential functions and pathways enriched by these DEGs using R packages. GO analysis included biological process (BP), molecular function (MF), and cellular components (CC).

Integration analysis of ATAC-Seq and RNA-Seq

ATAC-Seq and RNA-Seq profiles were analyzed after integration to accurately determine the potential center genes that can distinguish PIS from PS patients. The overlapping DEGs were defined as 1) the upregulated DEGs in RNA-Seq with an enhanced chromatin open region signal in ATAC-Seq and 2) the downregulated DEGs in RNA-Seq with an attenuated chromatin open region signal in ATAC-Seq (PIS vs. PS). A Venn diagram was generated to present the overlapping upregulated and downregulated DEGs. Scatter plots were used to evaluate the relationship between the transposase accessible chromatin and gene expression derived from ATAC-Seq and RNA-Seq data, respectively.

The candidate genes for predictive model construction were screened out based on ATAC-Seq and RNA-Seq integrated bioinformatics analyses and literature review, but not only based on the level of change between the two conditions. The bioinformatics analyses in this part included REACTOME pathways, Transcription factor (TF) prediction, Motif enrichment, and Functional protein-associated networks. 1) Based on overlapping DEGs by ATAC-Seq and RNA-Seq integrated analysis, top ten REACTOME pathways were enriched, and DEGs in the pathways potentially regulating progestin insensitivity were first screened out. 2) Potential TFs that regulate the expression of the overlapping DEGs were enriched by HOMER Software, and DEGs-encoding TFs with p value less than 0.05 were screened out. 3) Motif enrichment was performed to identify important TFs by using homer peak analysis software. The generated homer known TFs with p value less than 0.05 and more than 20% of target sequences with motifs enriched in chromatin regions were listed in Table 2, and their encoding genes among the overlapping DEGs were identified. 4) The interactions between proteins encoded by overlapping DEGs were analyzed using STRING (https://string-db.org/) and Cytoscape software (version 3.6.1). Central proteins were determined with both >4 connected lines and >0.4 combined score, and their encoding DEGs were identified. Furthermore, all the candidate genes screened out based on aforementioned bioinformatics analyses above, were comprehensively evaluated by literature review according to whether these candidate genes were involved in tumor initiation, progression and treatment resistance.

Validation of candidate genes in the expanded samples

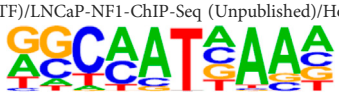
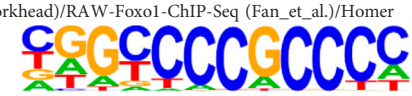
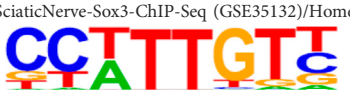
Endometrial samples from the Construction Group were analyzed by real-time quantitative PCR (RT-qPCR) for the expression of the twenty-five candidate genes. Each gene was analyzed in triplicate and normalized to the housekeeping gene *GAPDH*. Detailed primer sequences were listed in Supplementary Table S1. The value of the Δ cycle threshold (ΔCT) was used as the relative expression level of mRNA of the candidate genes compared to *GAPDH*. Then ΔCT values were normalized by SPSS Version 22.0 for subsequent analysis.

Statistics

Statistical analysis was calculated using GraphPad Prism Version 8.0 and SPSS Version 22.0. RT-qPCR data were presented as the mean ± standard error of the mean (SEM) and were calculated by unpaired t test, unless otherwise noted. A two-tailed p value less than 0.05 was considered statistically significant.

To determine which candidate genes could be used for predicting progestin insensitivity, predictive models were established using multinomial logistic regression (SPSS Version 22.0). The PS-C, sub-PS-C, and PIS-C groups were identified as the dependent variables. Normalized ΔCT values of candidate genes were stratified into low, medium, and high expression stratifications according to cutoff values (X-tile Version 3.6). Then, the expression stratification of candidate

TABLE 2 TFs binding homer known motifs enriched in chromatin region in response to progestin in PIS group compared to PS group from Analysis Group.

TFs	Binding motif	% Of target sequences with motif	*p* Value
NANOG			
	Nanog (Homeobox)/mES-Nanog-ChIP-Seq (GSE11724)/Homer	44.87%	1.00E−02
TGIF2			
	Tgif2 (Homeobox)/mES-Tgif2-ChIP-Seq (GSE55404)/Homer	39.74%	1.00E−02
NF1			
	NF1-halfsite (CTF)/LNCaP-NF1-ChIP-Seq (Unpublished)/Homer	29.49%	1.00E−05
HOXA9			
	Hoxa9 (Homeobox)/ChickenMSG-Hoxa9.Flag-ChIP-Seq (GSE86088)/Homer	29.17%	1.00E−02
FOXO1			
	Foxo1 (Forkhead)/RAW-Foxo1-ChIP-Seq (Fan_et_al.)/Homer	28.53%	1.00E−06
SP2			
	Sp2 (Zf)/HEK293-Sp2.eGFP-ChIP-Seq (Encode)/Homer	28.53%	1.00E−03
SOX10			
	Sox10 (HMG)/SciaticNerve-Sox3-ChIP-Seq (GSE35132)/Homer	26.60%	1.00E−08
SOX3			
	Sox3 (HMG)/NPC-Sox3-ChIP-Seq (GSE33059)/Homer	25.64%	1.00E−06
TWIST2			
	Twist2 (bHLH)/Myoblast-Twist2.Ty1-ChIP-Seq (GSE127998)/Homer	25.64%	1.00E−03
SOX6			
	Sox6 (HMG)/Myotubes-Sox6-ChIP-Seq (GSE32627)/Homer	25.00%	1.00E−07
SOX21			
	Sox21 (HMG)/ESC-SOX21-ChIP-Seq (GSE110505)/Homer	24.68%	1.00E−04
KLF5			

(Continued on following page)

TABLE 2 (*Continued*) TFs binding homer known motifs enriched in chromatin region in response to progestin in PIS group compared to PS group from Analysis Group.

TFs	Binding motif	% Of target sequences with motif	*p* Value
MAZ	KLF5 (Zf)/LoVo-KLF5-ChIP-Seq (GSE49402)/Homer 	23.72%	1.00E−02
TCF4	Maz (Zf)/HepG2-Maz-ChIP-Seq (GSE31477)/Homer 	23.08%	1.00E−02
AP-1	TCF4 (bHLH)/SHSY5Y-TCF4-ChIP-Seq (GSE96915)/Homer 	22.76%	1.00E−03
BHLHA15R	AP-1 (bZIP)/ThioMac-PU.1-ChIP-Seq (GSE21512)/Homer 	22.44%	1.00E−21
NEUROG2	BHLHA15 (bHLH)/NIH3T3-BHLHB8.HA-ChIP-Seq (GSE119782)/Homer 	22.44%	1.00E−04
ATF3	NeuroG2 (bHLH)/Fibroblast-NeuroG2-ChIP-Seq (GSE75910)/Homer 	22.12%	1.00E−02
SOX15	Atf3 (bZIP)/GBM-ATF3-ChIP-Seq (GSE33912)/Homer 	21.15%	1.00E−22
BATF	Sox15 (HMG)/CPA-Sox15-ChIP-Seq (GSE62909)/Homer 	20.83%	1.00E−09
	BATF (bZIP)/Th17-BATF-ChIP-Seq (GSE39756)/Homer	20.51%	1.00E−21

TFs, transcription factors; PIS, progestin-insensitive; PS, progestin-sensitive.

genes was identified as an independent variable. The PS-C group was regarded as the control group in the multinomial logistic regression method. The predictive accuracy of the established models to predict PS, sub-PS and PIS was analyzed. Model fitting was used to illustrate the reliability of the models.

Availability of supporting data

The raw data and processed data used in this study have been uploaded to the Gene Expression Omnibus repository under GEO accession number GSE201928 at https://www.ncbi.nlm.nih.gov/geo/.

Results

Comparison of chromatin accessibility between PIS and PS cases by ATAC-Seq

Flowchart of study design was shown in Figure 1A. Firstly, genomic chromatin accessibility was analyzed by ATAC-Seq

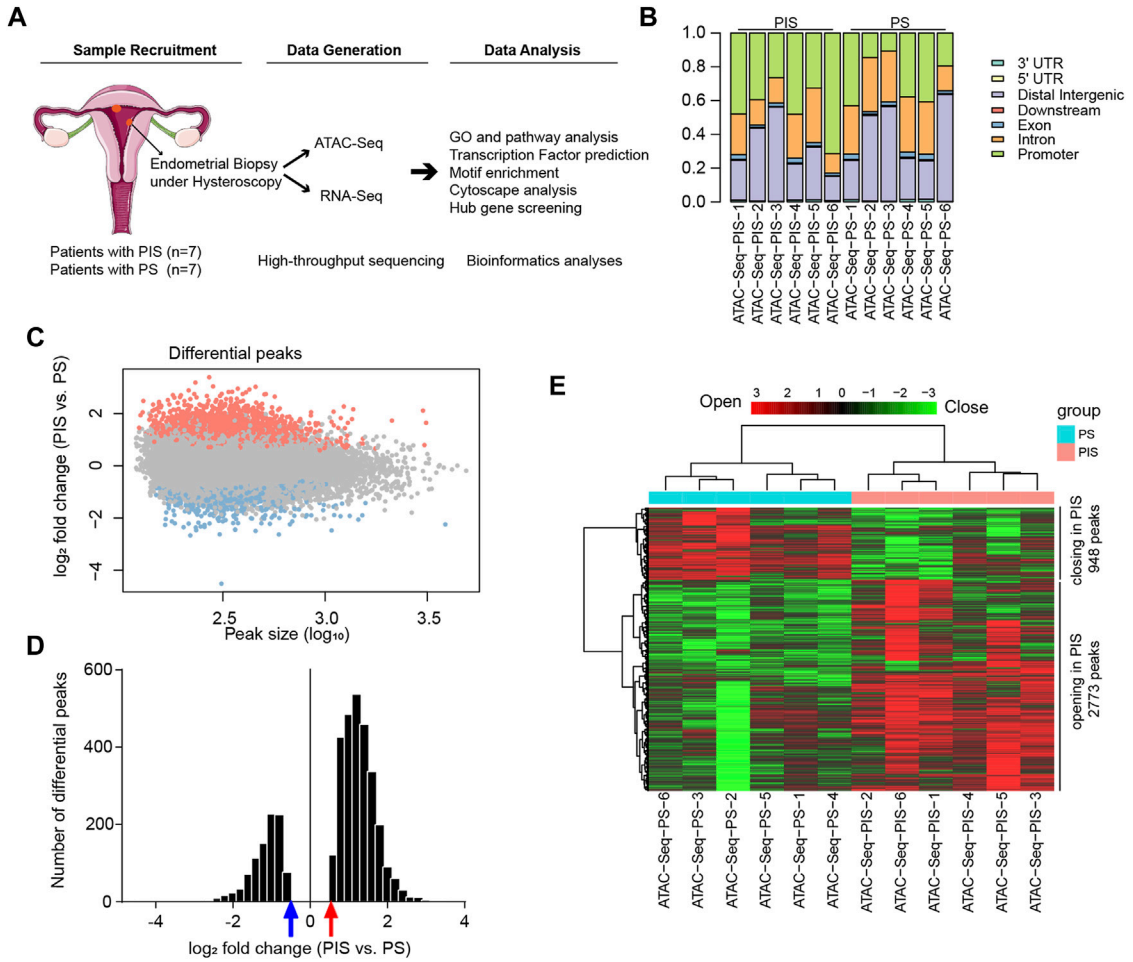


FIGURE 1
Landscape of genomic chromatin accessibility by ATAC-Seq. **(A)** Flowchart of study design. Endometrial lesions in the Analysis Group were collected for ATAC-Seq and RNA-Seq and further data analysis. **(B)** Genomic distribution of differential peaks. Bars with different colors and lengths represent different elements in the human genome and proportions, respectively. **(C)** Scatter plot of the chromatin accessibility at each peak in the PIS group compared to the PS group. The X-axis represents the peak size (log_{10}), and the Y-axis represents the log_2 fold change (PIS vs. PS) in ATAC-Seq analysis. The orange-red dots represent the opening peaks and the light blue dots represent the closing peaks in the PIS group compared to the PS group. **(D)** The histogram presents the distribution of log_2 fold change of the differential peaks (PIS vs. PS). The abscissa represents log_2 fold change of the differential peaks (PIS vs. PS) and the vertical axis represents the number of the differential peaks. Red arrow indicates log_2 fold change = 0.5849 while blue arrow indicates log_2 fold change = −0.5849. **(E)** Hierarchical cluster analysis of all the regulated opening and closing peaks in genes. Red plates represent opening peaks, while green plates indicate closing peaks in the PIS and PS groups. Abbreviations: PIS, progestin insensitive; PS, progestin sensitive; ATAC-Seq, assay for transposase-accessible chromatin sequencing; RNA-Seq, RNA sequencing; UTR, untranslated region.

using samples from the Analysis Group (PIS, $n = 7$ and PS, $n = 7$). Five patients from each group had both ATAC-Seq and RNA-Seq data. The remaining two patients in each group had only ATAC-Seq data or RNA-Seq data, respectively. In the ATAC-Seq results, the proportion of all reads in each sample was matched to the elements in the human genome according to functional and positional information. The accessibility of transcriptional sites was more abundant in the promoter region in the PIS group but more abundant in intron and distal intergenic sites in the PS group (Figure 1B). The accessibility of the other four sites, including the 3′ UTR, 5' UTR, downstream and exon, constituted a very small percentage of accessible transcriptional sites. After ATAC-Seq analysis, approximately 3721 differential opening or closing peaks were enriched, and most peaks were between 10^2 and 10^3 in size (Figure 1C). Additionally, distribution of 3721 differential peaks [log_2 fold change (PIS vs. PS)] were provided, and the results showed that PIS group had more opening differential peaks than the PS group (Figure 1D). In detail, 2773 opening peaks and 948 closing peaks were shown

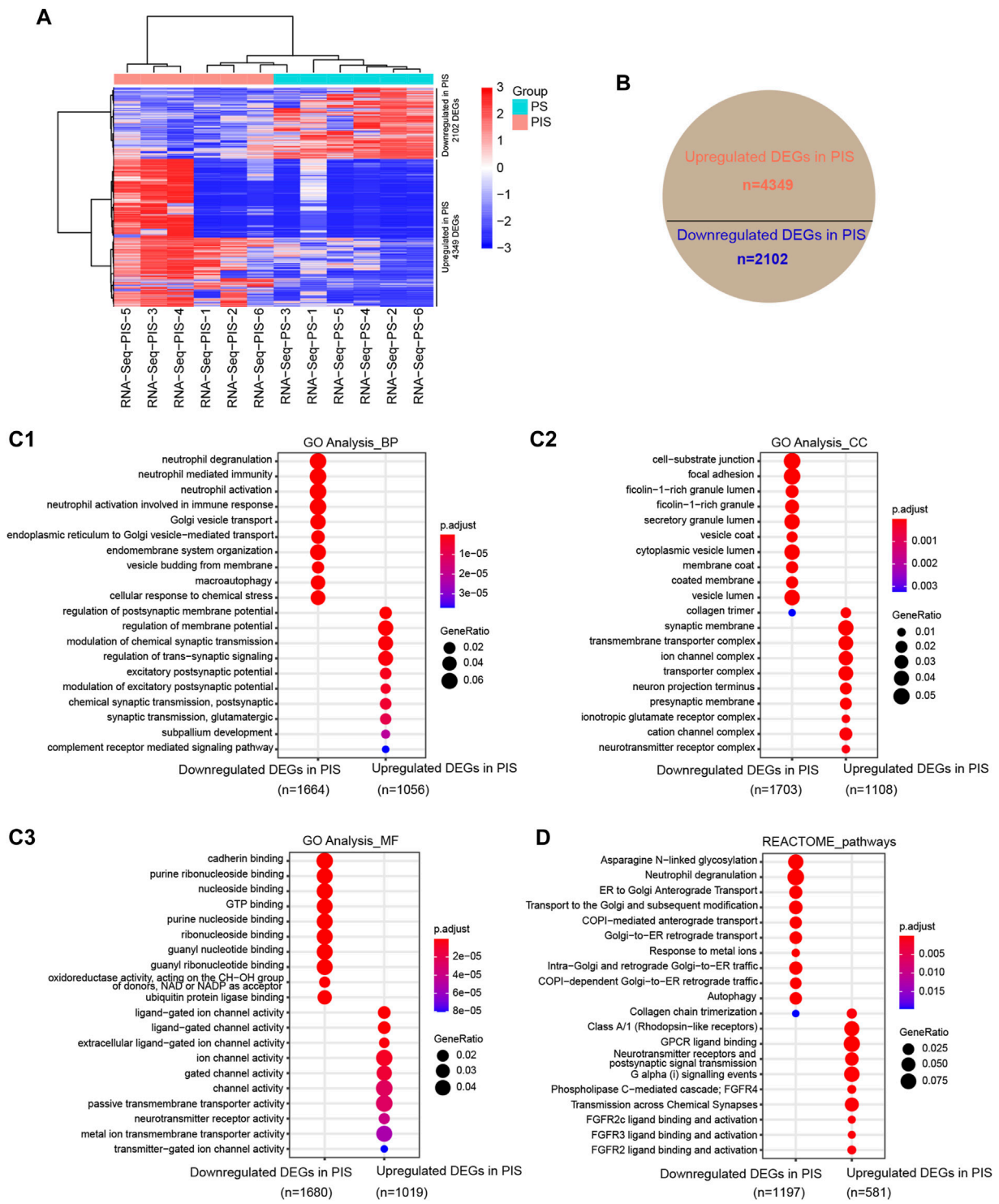


FIGURE 2
Expression profiles by RNA-Seq in PIS and PS patients with EAH and EEC. **(A)** Hierarchical cluster analysis of all DEGs annotated by FPKM by using DESeq2 normalization. The rows represent the 4349 upregulated and 2102 downregulated genes. Red grids represent upregulated genes while blue grids represent downregulated genes. **(B)** Statistical pie chart of upregulated and downregulated DEGs in the PIS group compared to the PS group. **(C)** Bubble diagram of the GO analysis of the upregulated and downregulated DEGs in the PIS group compared to the PS group, including BP (**C1**), CC (**C2**), and MF (**C3**). The top ten clusters with adjusted $p < 0.05$ were shown. **(D)** REACTOME pathway annotation of upregulated and downregulated DEGs in the PIS group compared to the PS group. The top ten enriched pathways with adjusted $p < 0.05$ were shown. Abbreviations: PIS, progestin insensitive; PS, progestin sensitive; DEGs, differentially expressed genes; GO, Gene Ontology; BP, biological process; CC, cellular components; MF, molecular function.

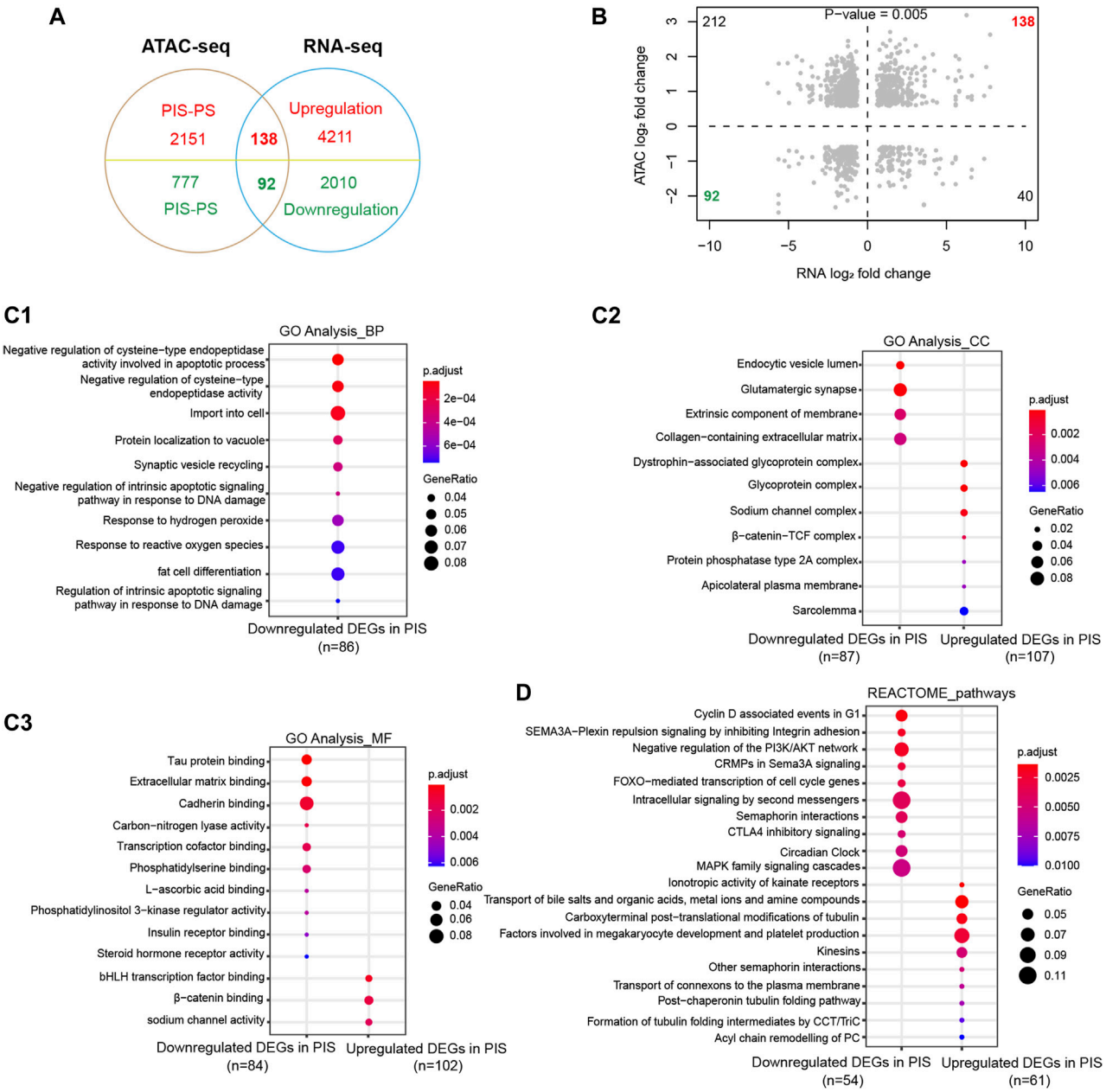


FIGURE 3
Enrichment analysis of DEGs integrated by ATAC-Seq and RNA-Seq. **(A)** Venn diagram of DEGs in RNA-Seq with differential opening and closing peaks in ATAC-Seq. **(B)** Chromatin accessibility correlates significantly with the 230 overlapping DEGs (Pearson analysis, $p = 0.005$). Dashed lines delineate the set of DEGs in RNA-Seq (X-axis) and differential opening or closing peaks in ATAC-Seq (Y-axis) between the PIS and PS groups. Shaded points in the upper right quadrant and lower left quadrant define the genes showing congruent chromatin accessibility and gene expression. **(C)** GO annotation of the upregulated and downregulated DEGs in the PIS group compared to the PS group was performed based on ATAC-Seq and RNA-Seq integration, including BP (**C1**), CC (**C2**), and MF (**C3**). The top ten clusters with adjusted $p < 0.05$ were shown. **(D)** REACTOME pathway annotation of the overlapping upregulated and downregulated DEGs in the PIS group by ATAC-Seq and RNA-Seq integration. The top ten enriched pathways with adjusted $p < 0.05$ were shown. Abbreviations: ATAC-Seq, assay for transposase-accessible chromatin sequencing; RNA-Seq, RNA sequencing; PIS, progestin insensitive; PS, progestin sensitive; DEGs, differentially expressed genes; GO, Gene Ontology; BP, biological process; CC, cellular components; MF, molecular function.

in the PIS group compared to the PS group by hierarchical cluster analysis (Figure 1E). The heatmap showed that a higher proportion of genes were transcriptionally active in PIS cases than in PS cases. Gene peaks in six samples of each group were hierarchically clustered into one group, illustrating the reliability and accuracy of ATAC-Seq data.

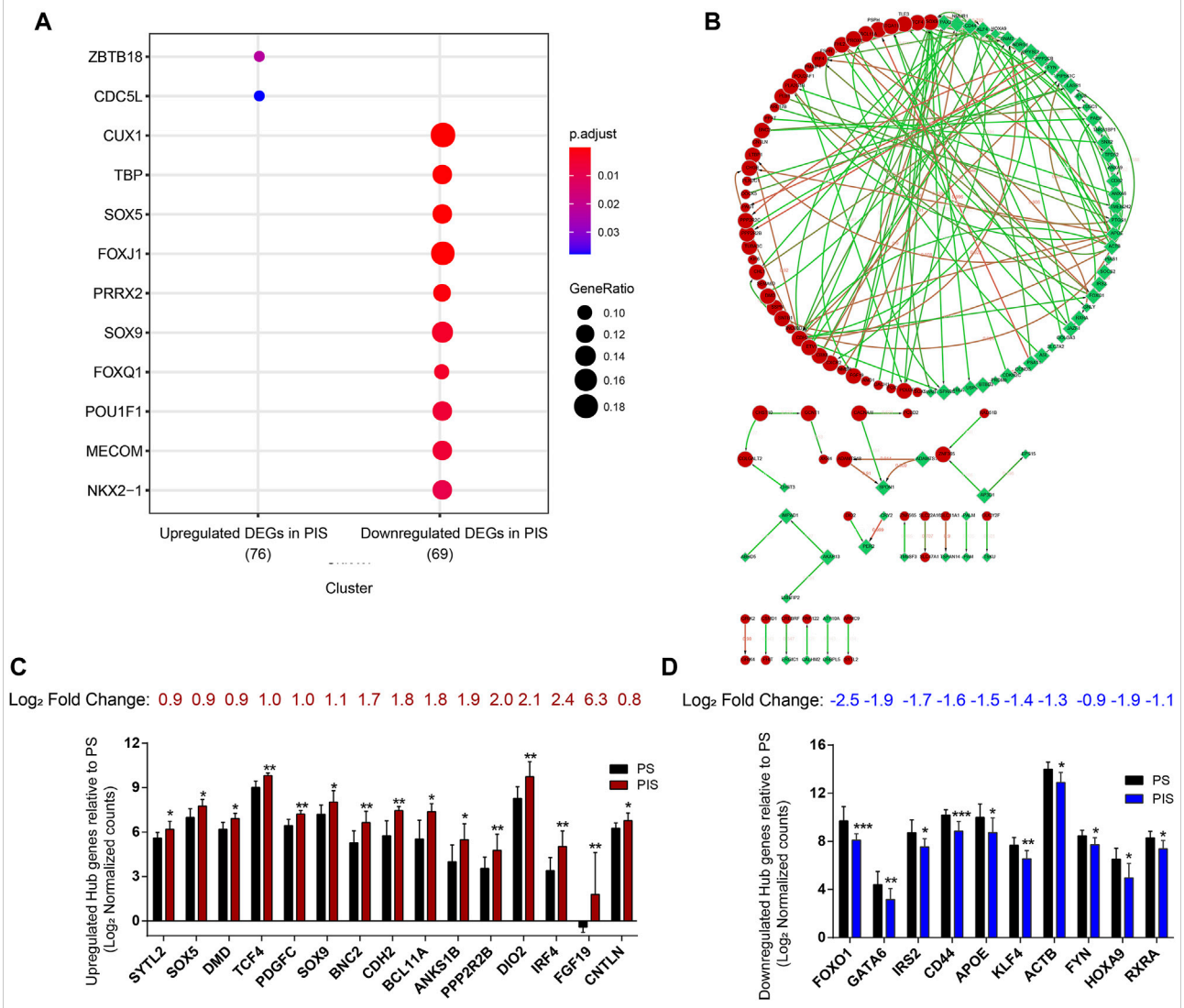


FIGURE 4
Bioinformatics analyses for screening the potential candidate genes. **(A)** Bubble diagram showing the transcription factor-binding sites clustered based on 230 overlapping DEGs. **(B)** Protein interaction networks were analyzed through STRING and Cytoscape analysis. Red nodes represent the upregulated overlapping DEGs, and green nodes represent the downregulated overlapping DEGs in the PIS group compared to the PS group. The larger the node is, the higher the connectivity is. Edges between two nodes indicate potential interactions between two proteins encoded by corresponding DEGs. The higher the value around the edges between connected nodes is, the higher the credibility is. **(C)** Log_2 normalized counts of 15 upregulated candidate genes in the PIS group relative to the PS group. **(D)** Log_2 normalized counts of 10 downregulated candidate genes in the PIS group relative to the PS group. Abbreviations: DEGs, differentially expressed genes; PIS, progestin insensitive; PS, progestin sensitive.

Comparison of expression profiles between PIS and PS cases by RNA-Seq

To compare the expression profiles between PIS and PS lesions in the Analysis Group, RNA-Seq was conducted and analyzed. DEGs were shown by hierarchical cluster analysis (Figure 2A). There were 4349 upregulated and 2102 downregulated DEGs in the PIS group compared to the PS group (Figure 2B). To identify whether these upregulated and downregulated DEGs in the PIS group were enriched in particular functions, GO annotation, including BP, CC, and MF categories, was performed (Figures 2C1,C2,C3). In the BP categories, downregulated DEGs in the PIS group were mainly enriched in neutrophil-associated activity, Golgi vesicle transport, endomembrane system organization, macroautophagy, and cellular response to chemical stress, while upregulated DEGs in the PIS group were mainly enriched in membrane potential and synaptic signaling-related

functions (Figure 2C1). In accordance with the results for the BP categories, the CC categories showed that downregulated DEGs in the PIS group were mainly enriched in granule lumen, vesicle lumen, cell-substrate junction, and focal adhesion, while upregulated DEGs in the PIS group were enriched in synaptic membrane and related transporter complex (Figure 2C2). In the MF categories, downregulated DEGs in the PIS group were enriched in various kinds of binding, including cadherin, nucleoside, GTP, and ubiquitin protein ligase binding, etc., while upregulated DEGs in the PIS group were enriched in channel, transmembrane transporter, and neurotransmitter receptor activity (Figure 2C3). Furthermore, REACTOME pathway annotation of the DEGs showed that downregulated DEGs in the PIS group were significantly enriched in pathways including asparagine N-linked glycosylation, neutrophil degranulation, autophagy, and transport between Golgi and endoplasmic reticulum (ER), while the upregulated DEGs in the PIS group were enriched in chemical and synaptic signal transmission, fibroblast growth factor receptor (FGFR), and G protein-coupled receptor (GPCR) (Figure 2D). RNA-Seq data demonstrated that expression profiles varied widely between PIS and PS cases.

Gene ontology and REACTOME analysis by ATAC-Seq and RNA-Seq integration

To further determine the specific functions and pathways related to progestin insensitivity, ATAC-Seq and RNA-Seq results were integrated for further analysis. By overlapping the results of ATAC-Seq and RNA-Seq, the PIS group had 138 upregulated DEGs with opening peaks and 92 downregulated DEGs with closing peaks in chromatin accessibility compared to the PS group (Figure 3A). Correlation analysis showed a significant positive correlation between expression profiles and chromatin accessibility of the above mentioned 230 overlapping DEGs (Figure 3B). To gain further insight into whether these 230 overlapping DEGs were engaged in specific functions and pathways, GO annotation and REACTOME pathways were performed. In BP categories, the overlapping downregulated DEGs in the PIS group mainly influenced cell import-transportation, negative regulation of cysteine-type endopeptidase activity, response to reactive oxygen species and fat cell differentiation (Figure 3C1). In CC categories, overlapping downregulated DEGs in the PIS group were enriched in glutamatergic synapse, extrinsic component of membrane, collagen-containing extracellular matrix, and endocytic vesicle lumen, while those overlapping upregulated DEGs in the PIS group were located in glycoprotein complex, sodium channel complex, β-catenin-TCF complex, and sarcolemma (Figure 3C2). Similarly, in MF categories, overlapping downregulated DEGs in the PIS group were enriched in extracellular matrix binding, cadherin binding, transcriptional cofactor binding and phosphatidylserine binding, while upregulated DEGs in the PIS group were associated with bHLH transcription factor binding, β-catenin binding, and sodium channel activity (Figure 3C3). Furthermore, REACTOME pathway analysis showed that these overlapping downregulated DEGs in the PIS group mainly influenced pathways including MAPK family signaling cascades, intracellular signaling by second messengers, negative regulation of the PI3K/AKT network, cyclin D-associated events in G1, and FOXO−mediated transcription of cell cycle genes, while upregulated DEGs in the PIS group were enriched in pathways including transport of bile salts and organic acids, metal ions and amine compounds, carboxyterminal post-translational modifications of tubulin, factors involved in megakaryocyte development and platelet production, and kinesins (Figure 3D). Taken together, these data suggested that the overlapping downregulated DEGs in the PIS group are responsible for signal transfer, the activity of transcription cofactors, DNA damage, cell apoptosis and cell cycle, while the overlapping upregulated DEGs in the PIS group are mainly responsible for substance transport and the regulation of cytoskeletal proteins.

Screening of candidate genes for predicting progestin insensitivity

To further screen candidate genes predicting progestin insensitivity, potential TFs that regulate the expression of the 230 overlapping DEGs were enriched by HOMER Software. The TFs identified included CUX1, TBP, SOX5, FOXJ1, PRRX2, SOX9, FOXQ1, POU1F1, MECOM, and NKX2-1 based on the 69 downregulated DEGs in the PIS group, while only ZBTB18 and CDC5Lwereidentified based on 76 upregulated DEGs in the PIS group (Figure 4A). Additionally, motif enrichment was performed by homer peak analysis based on the results of ATAC-Seq and RNA-Seq integration. The generated homer known TFs with more than 20% of target sequences with motifs enriched in chromatin regions (PIS vs. PS) included NANOG, TGIF2, NF1, HOXA9, FOXO1, SP2, SOX10, SOX3, TWIST2, SOX6, SOX21, KLF5, MAZ, TCF4, AP-1, BHLHA15R, NEUROG2, ATF3, SOX15, and BATF (Table 2). Additionally, the interactions between proteins encoded by DEGs were analyzed using STRING and Cytoscape software (Figure 4B). Potential candidate genes or central genes were screened out based on the principle that more connected lines had higher combined scores. The left part showed the proteins encoded by the upregulated DEGs in the PIS group, and the top four social proteins with more than 4 connected lines were encoded by *SOX9*, *CDH2*, *IRF4*, and *TCF4*, respectively. There were eight proteins in the right part that had more than

TABLE 3 Characteristics of different predictive models based on candidate genes.

No.	Candidate genes included in models	Model fitting (*p* value)	Pseudo R square	Predictive accuracy of PIS-C (%)	Predictive accuracy of sub-PS-C (%)	Predictive accuracy of sub-PS-C (%)	Overall accuracy of prediction (%)
1	BCL11A + SOX9+ApoE + FOXO1+FYN + KLF4+DIO2	<0.001	≥0.846	100	93.3	84.6	91.4
2	BCL11A + SOX9+ApoE + FOXO1+FYN + KLF4+IRS2+DIO2	<0.001	≥0.857	100	86.7	100.0	94.3
3	BCL11A + PDGFC + SOX9+ApoE + FYN + KLF4+IRS2+DIO2	<0.001	≥0.846	100	86.7	92.3	91.4
4	BCL11A + PDGFC + ApoE + FOXO1+FYN + KLF4+DIO2	<0.01	≥0.793	100	80.0	92.3	88.6
5	BCL11A + PDGFC + ApoE + FOXO1+FYN + KLF4+IRS2+DIO2	<0.01	≥0.828	100	80.0	92.3	88.6
6	BCL11A + ApoE + FOXO1+FYN + KLF4+DIO2	<0.001	≥0.743	100	80.0	84.6	85.7
7	BCL11A + SOX9+ApoE + FYN + KLF4+DIO2	<0.001	≥0.798	100	80.0	84.6	85.7
8	BCL11A + PDGFC + ApoE + FOXO1+FYN + IRS2+DIO2	<0.001	≥0.811	100	80.0	84.6	85.7
9	BCL11A + ApoE + FOXO1+FYN + KLF4+IRS2+DIO2	<0.01	≥0.748	100	80.0	76.9	82.9
10	BCL11A + PDGFC + SOX9+ApoE + FOXO1+KLF4+DIO2	<0.01	≥0.812	100	80.0	76.9	82.9
11	BCL11A + ApoE + FYN + KLF4+IRS2+DIO2	<0.01	≥0.698	100	80.0	69.2	80.0
12	BCL11A + PDGFC + ApoE + FYN + KLF4+IRS2+DIO2	<0.01	≥0.787	100	73.3	92.3	85.7
13	BCL11A + PDGFC + ApoE + FOXO1+FYN + DIO2	<0.01	≥0.761	100	73.3	92.3	85.7
14	BCL11A + PDGFC + ApoE + FOXO1+KLF4+DIO2	<0.01	≥0.602	100	73.3	84.6	82.9
15	BCL11A + ApoE + FYN + KLF4+DIO2	<0.01	≥0.682	100	73.3	76.9	80.0
16	BCL11A + SOX9+ApoE + FYN + KLF4+IRS2+DIO2	<0.01	≥0.812	100	73.3	100	88.6
17	BCL11A + PDGFC + FOXO1+KLF4+IRS2+DIO2	<0.01	≥0.61	100	66.7	84.6	80.0
18	BCL11A + ApoE + FOXO1+FYN + IRS2+DIO2	<0.01	≥0.675	100	66.7	76.9	77.1
19	BCL11A + PDGFC + FYN + KLF4+IRS2+DIO2	<0.001	≥0.726	100	60.0	92.3	80.0
20	BCL11A + ApoE + FYN + IRS2+DIO2	<0.01	≥0.634	100	60.0	86.4	77.1
21	PDGFC + ApoE + FOXO1+FYN + KLF4+DIO2	<0.05	≥0.528	100	53.3	69.2	68.6
22	ApoE + FOXO1+FYN + KLF4+IRS2+DIO2	<0.05	≥0.526	100	53.3	69.2	68.6
23	ApoE + FYN + KLF4+IRS2+DIO2	<0.05	≥0.479	100	46.7	61.5	62.9
24	PDGFC + FYN + KLF4+IRS2+DIO2	<0.05	≥0.457	100	40.0	76.9	65.7
25	PDGFC + ApoE + FYN + KLF4+DIO2	<0.05	≥0.480	100	40.0	69.2	62.9

No., number; PIS-C, progestin-insensitive in Construction Group; sub-PS-C, progestin-sub-sensitive in Construction Group; PS-C, progestin-sensitive in Construction Group.

4 connected lines in the downregulated DEGs in the PIS group, which were encoded by *CD44*, *ACTB*, *KLF4*, *APOE*, *SNAI2*, *FYN*, *PAX2*, and *FOXO1*, respectively. Finally, twenty-five candidate genes (*SYTL2*, *SOX5*, *DMD*, *TCF4*, *PDGFC*, *SOX9*, *BNC2*, *CDH2*, *BCL11A*, *ANKS1B*, *PPP2R2B*, *DIO2*, *IRF4*, *FGF19*, *FOXO1*, *GATA6*, *IRS2*, *CD44*, *APOE*, *KLF4*,

ACTB, *FYN*, *CNTLN*, *HOXA9*, and *RXRA*.) were screened out based on bioinformatics analyses and literature review. The expression levels of these genes were presented according to the RNA-Seq results (Figures 4C,D).

Establishment of potential models for predicting progestin insensitivity

Samples from the Construction Group (n = 35) were used for model construction. To construct models that can precisely predict the status of progestin sensitivity, 35 cases were further classified as PS-C (n = 13), sub-PS-C (n = 15) and PIS-C (n = 7), as shown in Table 1. Firstly, RT-qPCR was used to determine the expression of the 25 candidate genes in these 35 cases. As the CT values of *SYTL2*, *ANKS1B*, *PPP2R2B*, and *FGF19* exceeded 35, which suggested low gene expression and inaccurate analyses, these four genes were not included in the following analyses.

Predictive models were established by using multinomial logistic regression based on normalized ΔCT values of the remaining 21 genes according to different progestin sensitive conditions. The results in Table 3 showed that a total of 25 predictive models were generated with predictive accuracy of 100% for PIS-C patients, among which 11 models had predictive accuracy of more than 80% for sub-PS-C prediction ($p < 0.01$). Three models' overall predictive accuracy were higher than 90%, involving 9 candidate genes (*FOXO1*, *IRS2*, *PDGFC*, *DIO2*, *SOX9*, *BCL11A*, *APOE*, *FYN*, and *KLF4*) (Supplementary Figure S1).

Discussion

It is necessary to establish highly accurate predictive models for identifying PIS patients and helping provide individualized fertility-preserving treatment for EAH and EEC patients. In this study, through ATAC-Seq and RNA-Seq analyses of 14 cases and verification of candidate genes in 35 expanded samples, predictive models comprising nine genes (*FOXO1*, *IRS2*, *PDGFC*, *DIO2*, *SOX9*, *BCL11A*, *APOE*, *FYN*, and *KLF4*) were established. Our models provided new molecular markers that could be used in combination with the well-known PR status to help identify PIS patients prior to treatment initiation.

In this study, we found that the expression of *PDGFC*, *DIO2*, *SOX9*, and *BCL11A*was upregulated and *FOXO1*, *IRS2*, *APOE*, *FYN* and *KLF4* was downregulated in PIS endometrial lesions compared with PS endometrial lesions. These nine genes were all reported to play important roles in tumor progression or drug response. *PDGFC*-encoded platelet-derived growth factor C was reported to promote angiogenesis, cancer cell proliferation, invasion, and metastasis (Kim et al., 2021). *SOX9*- and *BCL11A*-encoded proteins were both involved in inducing tumor initiation, proliferation, migration, and chemoresistance (Yin et al., 2019; Jana et al., 2020). *DIO2*-encoded protein can catalyze the conversion of tetraiodothyronine to bioactive triiodothyronine. Triiodothyronine was reported to be associated with lipid accumulation and metabolism in adipose tissue, which contributes to obesity-related insulin resistance (Bradley et al., 2018). Previous studies showed that high expression levels of PDGFC, SOX9, BCL11A, and DIO2 were associated with poor response to chemotherapy in cancer cells and short survival time of various patients, which could be regarded as negative prognostic factors (Bradley et al., 2018; Yin et al., 2019; Jana et al., 2020; Kim et al., 2021). Carriers of the *DIO2* polymorphism were also reported to be predisposed to the development of endometrial cancer (Janowska et al., 2022). Furthermore, the inhibition of *SOX9* or *DIO2* has been reported to be a potential therapeutic strategy for cancer (Carrasco-Garcia et al., 2019; Kojima et al., 2019).

FOXO1, an important member of the *FOXO* subfamily in the *FOX* family, encodes a transcription factor and has been reported to be involved in various physiological processes, including inducing cancer cell cycle arrest and suppressing the migration and invasion of cancer cells (Xing et al., 2018). *FOXO1* was also identified as a progesterone target gene containing PR elements within the promoter regions (Yang et al., 2011). Downregulated FOXO1expression was found in progestin-resistant EC cells and was associated with progestin insensitivity in EC patients (Yang et al., 2011; Reyes et al., 2016; Wang et al., 2018b). *IRS2*, encoding a kind of insulin receptor substrate that is commonly phosphorylated by the receptor tyrosine kinase, was reported to promote cell proliferation, invasion and sphere formation of cancer cells (Shaw, 2011). However, *IRS2* amplification and high expression of IRS2 were potentially related to good response to chemotherapy (Lee et al., 2020). *APOE*, one of apolipoproteins, plays anti-immunosuppressive and anti-metastatic roles in tumorigenesis (Tavazoie et al., 2018). High expression of APOE was reported to be associated with good prognosis of thyroid cancer patients (Nan et al., 2021). *KLF4* encodes a transcription factor that acted as a tumor suppressor which inhibited cell cycle, promoted apoptosis and differentiation, and suppressed metastasis (Yan et al., 2016). Downregulated expression of KLF4 by promoter methylation modification was reported in EC tissues, which was associated with accelerated tumorigenesis, drug resistance and poor prognosis (Jia et al., 2012; Danková et al., 2018). *FYN* encodes a membrane-associated tyrosine kinase that promoted cell proliferation, migration and invasion and inhibited apoptosis of cancer cells (Saito et al., 2010). Overexpression of FYN was reported to be correlated with chemotherapy resistance and poor survival (Elias et al., 2014). However, the roles of 9 candidate genes in regulating progestin response needs further investigation.

The strength of our study is the use of ATAC-Seq together with RNA-Seq technology to help identify the upregulated or downregulated genes with simultaneous opening or closing chromatin accessibility which effectively improves the

accuracy of candidate gene screening. The improved ATAC-Seq protocol used in this work could further reduce background disturbances from different individuals to improve the accuracy of the analysis (Corces et al., 2017). Thirty-five patients with various progestin sensitive conditions were used for further data verification and construction of potential predictive models with an overall predictive accuracy above 90%. There are some limitations in the study. First, the sample size was not large enough to address tissue heterogeneity. Second, integration of ATAC-Seq and RNA-Seq can be used to analyze the epigenetic and transcriptional changes in genes, but post-transcriptional and post-translational regulatory levels cannot be analyzed.

In conclusion, the predictive models we provided may be useful in identifying progestin insensitive EAH and EEC patients before initiating fertility-sparing therapy. The accuracy of our predictive models requires more samples validation and molecular mechanism exploration.

Data availability statement

The data presented in the study are deposited in the Gene Expression Omnibus repository, accession number GSE201928 at https://www.ncbi.nlm.nih and we have released the accession.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committees of Obstetrics and Gynecology Hospital of Fudan University (Approval NO. 2021-130). The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

QL and XC contributed to the study design and manuscript revision. Experiments were performed by JH and GY. GY and QL contributed to patient selection and sample collection. JH and QL contributed to the literature search, figures drawing, and tables construction. This manuscript was written by JH. LW contributed to language editing. BY contributed to statistical analysis. All co-authors have critically reviewed the manuscript and approved the final version for submission.

Funding

This work was supported by the Science and Technology Commission of Shanghai Municipality (Grant NO. 20Z11900700), National Natural Science Foundation of China (Grant No. 82071611), and Shanghai "Rising Stars of Medical Talents" Youth Development Program, Youth Medical Talents-Specialist Program (Grant NO. [2022]65).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.952083/full#supplementary-material

References

Bradley, D., Liu, J., Blaszczak, A., Wright, V., Jalilvand, A., Needleman, B., et al. (2018). Adipocyte DIO2 expression increases in human obesity but is not related to systemic insulin sensitivity. *J. Diabetes Res.* 2018, 2464652. doi:10.1155/2018/2464652

Buenrostro, J. D., Wu, B., Chang, H. Y., and Greenleaf, W. J. (2015). ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* 109 (21), 21–29. doi:10.1002/0471142727.mb2129s109

Carrasco-Garcia, E., Álvarez-Satta, M., García-Puga, M., Ribeiro, M. L., Arevalo, S., Arauzo-Bravo, M., et al. (2019). Therapeutic relevance of SOX9 stem cell factor in gastric cancer. *Expert Opin. Ther. Targets* 23, 143–152. doi:10.1080/14728222.2019.1559826

Corces, M. R., Trevino, A. E., Hamilton, E. G., Greenside, P. G., Sinnott-Armstrong, N. A., Vesuna, S., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* 14, 959–962. doi:10.1038/nmeth.4396

Danková, Z., BRANý, D., DVORSKá, D., ŇACHAJOVá, M., Fiolka, R., GRENDáR, M., et al. (2018). Methylation status of KLF4 and HS3ST2 genes as predictors of endometrial cancer and hyperplastic endometrial lesions. *Int. J. Mol. Med.* 42, 3318–3328. doi:10.3892/ijmm.2018.3872

Elias, D., Vever, H., Lænkholm, A. V., Gjerstorff, M. F., Yde, C. W., Lykkesfeldt, A. E., et al. (2014). Gene expression profiling identifies FYN as an important

molecule in tamoxifen resistance and a predictor of early recurrence in patients treated with endocrine therapy. *Oncogene* 34, 1919–1927. doi:10.1038/onc.2014.138

Fan, R., Wang, W., Wei, W., and Zheng, W. (2017). Mechanism of progestin resistance in endometrial precancer/cancer through Nrf2-survivin pathway. *Am. J. Transl. Res.* 9, 1483–1491.

Gallos, I. D., Yap, J., Rajkhowa, M., Luesley, D. M., Coomarasamy, A., and Gupta, J. K. (2012). Regression, relapse, and live birth rates with fertility-sparing therapy for endometrial cancer and atypical complex endometrial hyperplasia: A systematic review and metaanalysis. *Am. J. Obstet. Gynecol.* 207, 266–312. doi:10.1016/j.ajog.2012.08.011

Gunderson, C. C., Fader, A. N., Carson, K. A., and Bristow, R. E. (2012). Oncologic and reproductive outcomes with progestin therapy in women with endometrial hyperplasia and grade 1 adenocarcinoma: A systematic review. *Gynecol. Oncol.* 125, 477–482. doi:10.1016/j.ygyno.2012.01.003

Jana, S., Madhu Krishna, B., Singhal, J., Horne, D., Awasthi, S., Salgia, R., et al. (2020). SOX9: The master regulator of cell fate in breast cancer. *Biochem. Pharmacol.* 174, 113789. doi:10.1016/j.bcp.2019.113789

Janowska, M., Potocka, N., Paszek, S., Skrzypa, M., Zulewicz, K., Kluz, M., et al. (2022). An assessment of GPX1 (rs1050450), DIO2 (rs225014) and SEPP1 (rs7579) gene polymorphisms in women with endometrial cancer. *Genes (Basel)* 13, 188. doi:10.3390/genes13020188

Jia, Y., Zhang, W., Liu, H., Peng, L., Yang, Z., and Lou, J. (2012). Inhibition of glutathione synthesis reverses Krüppel-like factor 4-mediated cisplatin resistance. *Cancer Chemother. Pharmacol.* 69, 377–385. doi:10.1007/s00280-011-1708-7

Kim, S., You, D., Jeong, Y., Yoon, S. Y., Kim, S. A., and Lee, J. E. (2021). Inhibition of platelet-derived growth factor C and their receptors additionally increases doxorubicin effects in triple-negative breast cancer cells. *Eur. J. Pharmacol.* 895, 173868. doi:10.1016/j.ejphar.2021.173868

Kojima, Y., Kondo, Y., Fujishita, T., Mishiro-Sato, E., Kajino-Sakamoto, R., Taketo, M. M., et al. (2019). Stromal iodothyronine deiodinase 2 (DIO 2) promotes the growth of intestinal tumors in Apc Δ716 mutant mice. *Cancer Sci.* 110, 2520–2528. doi:10.1111/cas.14100

Lee, M. S., Jung, K., Song, J. Y., Sung, M. J., Ahn, S. B., Lee, B., et al. (2020). IRS2 amplification as a predictive biomarker in response to ceritinib in small cell lung cancer. *Mol. Ther. - Oncolytics* 16, 188–196. doi:10.1016/j.omto.2019.12.009

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Nan, B.-Y., Xiong, G.-F., Zhao, Z.-R., Gu, X., Huang, X.-S., and Fedele, M. (2021). Comprehensive identification of potential crucial genes and miRNA-mRNA regulatory networks in papillary thyroid cancer. *BioMed Res. Int.* 2021, 1–25. doi:10.1155/2021/6752141

Raffone, A., Travaglino, A., Saccone, G., Mollo, A., De Placido, G., Insabato, L., et al. (2019). Should progesterone and estrogen receptors be assessed for predicting the response to conservative treatment of endometrial hyperplasia and cancer? A systematic review and meta-analysis. *Acta Obstet. Gynecol. Scand.* 98, 976–987. doi:10.1111/aogs.13586

Reyes, H. D., Carlson, M. J., Devor, E. J., Zhang, Y., Thiel, K. W., Samuelson, M. I., et al. (2016). Downregulation of FOXO1 mRNA levels predicts treatment failure in patients with endometrial pathology conservatively managed with progestin-containing intrauterine devices. *Gynecol. Oncol.* 140, 152–160. doi:10.1016/j.ygyno.2015.10.023

Saito, Y. D., Jensen, A. R., Salgia, R., and Posadas, E. M. (2010). Fyn. *Cancer* 116, 1629–1637. doi:10.1002/cncr.24879

Shaw, L. M. (2011). The insulin receptor substrate (IRS) proteins. *Cell Cycle* 10, 1750–1756. doi:10.4161/cc.10.11.15824

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2022). Cancer statistics, 2022. *CA A Cancer J. Clin.* 72, 7–33. doi:10.3322/caac.21708

Tavazoie, M. F., Pollack, I., Tanqueco, R., Ostendorf, B. N., Reis, B. S., Gonsalves, F. C., et al. (2018). LXR/ApoE activation restricts innate immune suppression in cancer. *Cell* 172, 825–840. e18. doi:10.1016/j.cell.2017.12.026

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi:10.1038/nbt.1621

Trojano, G., Olivieri, C., Tinelli, R., Damiani, G. R., Pellegrino, A., and Cicinelli, E. (2019). Conservative treatment in early stage endometrial cancer: A review. *Acta Biomed.* 90, 405–410. doi:10.23750/abm.v90i4.7800

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26, 136–138. doi:10.1093/bioinformatics/btp612

Wang, Y., Zeng, X., and Liu, W. (2018a). De novo transcriptomic analysis during Lentinula edodes fruiting body growth. *Gene* 641, 326–334. doi:10.1016/j.gene.2017.10.061

Wang, Y., Zhang, L., Che, X., Li, W., Liu, Z., and Jiang, J. (2018b). Roles of SIRT1/FoxO1/SREBP-1 in the development of progestin resistance in endometrial cancer. *Arch. Gynecol. Obstet.* 298, 961–969. doi:10.1007/s00404-018-4893-3

Wang, Y., Zhou, R., Zhang, X., Liu, H., Shen, D., and Wang, J. (2021). Significance of serum and pathological biomarkers in fertility-sparing treatment for endometrial cancer or atypical hyperplasia: A retrospective cohort study. *BMC Women's Health* 21, 252. doi:10.1186/s12905-021-01383-5

Westin, S. N., Fellman, B., Sun, C. C., Broaddus, R. R., Woodall, M. L., Pal, N., et al. (2021). Prospective phase II trial of levonorgestrel intrauterine device: Nonsurgical approach for complex atypical hyperplasia and early-stage endometrial cancer. *Am. J. Obstet. Gynecol.* 224, 191–e15. doi:10.1016/j.ajog.2020.08.032

Xing, Y. Q., Li, A., Yang, Y., Li, X. X., Zhang, L. N., and Guo, H. C. (2018). The regulation of FOXO1 and its role in disease progression. *Life Sci.* 193, 124–131. doi:10.1016/j.lfs.2017.11.030

Yamazawa, K., Hirai, M., Fujito, A., Nishi, H., Terauchi, F., Ishikura, H., et al. (2007). Fertility-preserving treatment with progestin, and pathological criteria to predict responses, in young women with endometrial cancer. *Hum. Reprod.* 22, 1953–1958. doi:10.1093/humrep/dem088

Yan, Y., Li, Z., Kong, X., Jia, Z., Zuo, X., Gagea, M., et al. (2016). KLF4-Mediated suppression of CD44 signaling negatively impacts pancreatic cancer stemness and metastasis. *Cancer Res.* 76, 2419–2431. doi:10.1158/0008-5472.can-15-1691

Yang, B., Xu, Y., Zhu, Q., Xie, L., Shan, W., Ning, C., et al. (2019). Treatment efficiency of comprehensive hysteroscopic evaluation and lesion resection combined with progestin therapy in young women with endometrial atypical hyperplasia and endometrial cancer. *Gynecol. Oncol.* 153, 55–62. doi:10.1016/j.ygyno.2019.01.014

Yang, B. Y., Gulinazi, Y., Du, Y., Ning, C. C., Cheng, Y. L., Shan, W. W., et al. (2020). Metformin plus megestrol acetate compared with megestrol acetate alone as fertility-sparing treatment in patients with atypical endometrial hyperplasia and well-differentiated endometrial cancer: A randomised controlled trial. *BJOG Int. J. Obstet. Gy* 127, 848–857. doi:10.1111/1471-0528.16108

Yang, S., Thiel, K. W., and Leslie, K. K. (2011). Progesterone: The ultimate endometrial tumor suppressor. *Trends Endocrinol. Metabolism* 22, 145–152. doi:10.1016/j.tem.2011.01.005

Yin, J., Xie, X., Ye, Y., Wang, L., and Che, F. (2019). BCL11A: A potential diagnostic biomarker and therapeutic target in human diseases. *Biosci. Rep.* 39, BSR20190604. doi:10.1042/BSR20190604

Zhang, H., Yan, L., Bai, Y., Li, C., Guo, Q., Wang, C., et al. (2015). Dual-specificity phosphatase 6 predicts the sensitivity of progestin therapy for atypical endometrial hyperplasia. *Gynecol. Oncol.* 136, 549–553. doi:10.1016/j.ygyno.2014.11.008

Zhang, Y., Parmigiani, G., and Johnson, W. E. (2020). ComBat-seq: Batch effect adjustment for RNA-seq count data. *Nar. Genom Bioinform* 2, lqaa078. doi:10.1093/nargab/lqaa078

Zhou, S., Xu, Z., Yang, B., Guan, J., Shan, W., Shi, Y., et al. (2021). Characteristics of progestin-insensitive early stage endometrial cancer and atypical hyperplasia patients receiving second-line fertility-sparing treatment. *J. Gynecol. Oncol.* 32, e57. doi:10.3802/jgo.2021.32.e57

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 26 August 2022
DOI 10.3389/fgene.2022.948353

OPEN ACCESS

EDITED BY
Peter Hart,
Roosevelt University College of Pharmacy, United States

REVIEWED BY
Hanyu Shen,
Wuxi Huishan District People's Hospital, China
Fuxun Zhang,
Sichuan University, China

*CORRESPONDENCE
Yuanfeng Zhang,
304100@cqmu.edu.cn

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 19 May 2022
ACCEPTED 19 July 2022
PUBLISHED 26 August 2022

CITATION
Li J, Jia Y, Tang L, Zhang R and Zhang Y (2022), Identification of a chromatin regulator signature and potential prognostic ability for adrenocortical carcinoma. *Front. Genet.* 13:948353. doi: 10.3389/fgene.2022.948353

Identification of a chromatin regulator signature and potential prognostic ability for adrenocortical carcinoma

Junwu Li[1†], Yuanzhen Jia[†,2], Lin Tang[2], Ronggui Zhang[3] and Yuanfeng Zhang[1*]

[1]Department of Urology, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China, [2]Department of Rheumatology and Immunology, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China, [3]Department of Urology, Chongqing Emergency Medical Center, Chongqing, China

Objective: Adrenocortical carcinoma (ACC) is a rare malignant tumor. Chromatin regulators (CRs) can drive epigenetic changes, which have been considered as one of the most vital hallmarks of tumors. This study aimed to explore the CR signature for ACC in order to clarify the molecular basis of ACC's pathogenic mechanism and provide novel methods to diagnose and treat ACC clinically.

Methods: This study obtained transcriptome sequencing datasets of ACC patients and sequencing data on normal adrenal tissues in TCGA and GTEx databases, respectively. Meanwhile, prognostic genes were selected through Lasso and Cox regression analyses. Using the transcriptome sequencing datasets of ACC patients downloaded from the GEO database to finish validation, we performed Kaplan–Meier (KM) analysis for evaluating the differential survival between low- and high-risk groups. Then, this work constructed the risk model for predicting ACC prognosis. TIMER 2.0 was employed to assess the differences in immune infiltration between the two groups. Furthermore, this work adopted the R package "pRRophetic" for exploring and estimating the sensitivity of patients to different chemotherapeutic agents.

Results: A 5-CR model was established to predict ACC survival, and the CR signature was confirmed as a factor in order to independently predict ACC patient prognosis. In addition, a nomogram composed of the risk score and clinical T stage performed well in the prediction of patients' prognosis. Differentially expressed CRs (DECRs) were mostly associated with the cell cycle, base excision repair, colon cancer, gene duplication, homologous recombination, and other signaling pathways for the high-risk group. As for the low-risk group, DECRs were mainly enriched in allograft rejection, drug metabolism of cytochrome P450, metabolism of xenogeneic organisms by cytochrome P450, retinol metabolism, and other signaling pathways. According to TIMER analysis, the immune infiltration degrees of endothelial cells, M2 macrophages, myeloid dendritic cells, $CD4^+$ Th1 cells, NKT cells, and M0 macrophages showed significant statistical differences between the high- and low-risk groups, and high infiltration levels of M0 and

M2 macrophages were more pronounced in higher T stage (T3 and T4), N stage (N1), and clinical stages (III and IV). In addition, high-risk cases exhibited higher sensitivity to etoposide and doxorubicin. Additionally, low-risk patients had significantly decreased expression of RRM1 compared with high-risk cases, suggesting the better effect of mitotane treatment.

Conclusion: This study identified the DECRs, which might be related to ACC genesis and progression. The pathways enriched by these DECRs were screened, and these DECRs were verified with excellent significance for estimating ACC survival. Drug sensitivity analysis also supported the current clinical treatment plan. Moreover, this study will provide reliable ideas and evidence for diagnosing and treating ACC in the clinic.

KEYWORDS

chromatin regulator, adrenocortical carcinoma, prognosis, diagnosis, treatment

Introduction

ACC represents an uncommon malignant cancer, which has an annual morbidity of around 1-2/1,000,000 people (Else et al., 2014). It is also a frequently seen primary adrenal gland cancer (Chandrasekar et al., 2019), accounting for 6.8% of primary adrenal tumors (Lam, 1992), and it ranks second place among endocrine organ cancers, only second to thyroid cancer (TC) (Abe and Lam, 2021). ACC displays a high malignancy grade, and the 5-year survival rate is only 10%–20% in accordance with the statistics (Libé, 2015). ACC can occur at any age, with two peaks in childhood and the age of 50–70 years, and is more common in women (Fassnacht et al., 2009). ACC has rapid development, strong invasiveness, and dismal survival. Many patients have developed local invasion or distant metastasis (DM) when they are diagnosed. Based on the reports, the 5-year survival rates of stage I-IV ACC are 82%, 58%, 55%, and 13%, respectively (Allolio and Fassnacht, 2006). Recent epidemiological studies have indicated that the incidence of ACC increases year by year over the past 40 years, without any improvement in patient survival (Aufforth and Nilubol, 2014).

Epigenetic alterations are considered a vital hallmark of cancer. They are driven *via* CRs, the integral regulatory elements in epigenetics (Lu et al., 2018). According to their roles in epigenetics, CRs are mainly divided into three categories, namely, DNA methylating agents, histone modifiers, and chromatin remodeling agents (Plass et al., 2013). CRs are closely associated with each other. Further research shows that abnormal CR levels are related to various biological processes, such as inflammation (Marazzi et al., 2018), apoptosis (Li et al., 2020a), autophagy (Chu et al., 2020), and proliferation (Chen et al., 2020). This indicates that CR dysregulation may possibly generate disease occurrence, such as cancer. In recent years, an increasing number of studies have been conducted to screen key prognostic genes for ACC by bioinformatics analysis. However, CRs, as a key point of epigenetics, have not received corresponding attention. Therefore, this study aimed to explore the CR signature in ACC and further examine their functions in ACC prognosis with the purpose of clarifying ACC molecular basis and offering novel methods to diagnose and treat ACC in the clinic.

Methods and materials

Data acquisition

The transcriptome sequencing dataset for 79 ACC cases was downloaded from TCGA database (https://portal.gdc.cancer.gov). As normal samples were not included in TCGA-ACC, UCSC Xena was applied to obtain sequencing data on 128 normal adrenal tissue samples from the GTEx database. Thereafter, the top 100 CR-encoding genes with the greatest impact on ACC patients were obtained from the Facer database (http://bio-bigdata.hrbmu.edu.cn/FACER/). As a validation cohort, we downloaded the GSE10927 dataset with transcriptome sequencing data on 33 ACC cases, 22 adrenocortical adenoma (ACA) cases, and 10 normal adrenal tissue samples from the GEO database (https://ncbi.nlm.nih.gov) in order to confirm the differential expression of CR-encoding genes. In addition, we also downloaded the transcriptome sequencing and prognosis data on 23 ACC patients from the GSE33371 dataset to verify the reliability of the prognosis prediction model.

Differential analysis

All data were corrected to the log2 (FPKM+0.001) format for further comparison. Meanwhile, "Limma" in the R package was adopted for correcting the offset of datasets and performing differential analysis. The absolute value of logFC greater than 1 and $p < 0.05$ were applied as the thresholds to select differentially expressed genes (DEGs). Afterward, up-and downregulated genes were, respectively, explored, and the DECRs in ACC were obtained after intersecting with CR-encoding genes.

Construction of the prognosis prediction model

Univariate Cox regression was conducted to analyze DECRs' effect on prognosis, and the significant prognostic genes ($p < 0.05$) screened were later incorporated into Lasso regression analysis, followed by the construction of the prognosis prediction model. Thereafter, based on the median risk score, patients were classified into a low- or high-risk group. Subsequently, receiver operating characteristic (ROC) curves were plotted to assess whether the prognostic model was of high prognostic power. Afterward, univariate and multivariate COX regression analyses were conducted for assessing the effect of risk scores on ACC survival. In addition, we also utilized the R package "rms" for drawing the nomogram of risk scores for ACC patients and the 1-, 3-, and 5-year calibration curves. The model C-index was also calculated, and the effect of DECRs on overall survival (OS) was assessed by adopting Kaplan–Meier (KM) survival analysis.

Functional enrichment analysis

The enrichment of DECRs in Gene Ontology_biological process (GO_BP), cellular component (GO_CC), and molecular function (GO_MF) pathways was assessed using the R package "enrichplot," respectively. Furthermore, GSEA software was employed to explore the significantly different GO_BP, GO_CC, GO_MF, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (FDR < 0.25) between low- and high-risk patients.

Immune functional analysis

The infiltration levels of immune cells within TCGA-ACC cancer tissues under seven algorithms were obtained from TIMER 2.0. The differences between low- and high-risk patients were evaluated.

Drug sensitivity analysis

Mitotane is currently the most common and effective agent used for adjuvant therapy after ACC surgery and metastatic ACC. The expression of RRM1 in the tumor is a good predictor of the efficacy of mitotane therapy, and its low expression usually indicates the response to mitotane therapy. Therefore, the expression levels of RRM1 in ACC patients were extracted in order to compare the mitotane response in high- and low-risk patients. In addition, etoposide, doxorubicin, and cisplatin are also the commonly used chemotherapeutic agents for metastatic ACC. As a result, "pRRophetic" of the R package was utilized for predicting chemotherapeutic sensitivity based on the whole-transcriptome information of patients.

Results

Establishment of a chromatin regulator signature

After intersecting TCGA-ACC dataset with the ACC CR dataset, a total of 20 DECRs were screened, among which 12 showed downregulation whereas 8 exhibited upregulation (Figure 1A). According to the abovementioned dysregulated CRs, univariate Cox regression was adopted for exploring their prognostic significance. As a result, only 8 out of these 20 DECRs showed a prognostic value (Figures 1B–J). Later, this work utilized Lasso Cox regression for constructing the prognosis prediction characteristics for ACC patients. The risk model based on five genes (*TAF5*, *EMHT1*, *AURKB*, *SETD5*, and *HDAC2*) was successfully constructed (Figure 2A). For verification, we employed the expression data extracted from the GSE10927 dataset of the GEO database to intersect with the ACC CR dataset with the aim of performing expression differences. The result proved that these CRs have significant expression differences between ACC and other tissues including ACA and normal adrenal tissues (Supplementary Figure S1). Then, we determined the risk score by correlation coefficients of the 5 DECRs: Risk score = (0.0637 × TAF5 level) + (0.2699 × EMHT1 level) + (0.2068 × AURKB level) + (0.0418 × SETD5 level) + (0.0482 × HDAC2 level) (Table 1). Finally, ACC cases were classified into two (low- or high-risk) groups, in accordance with the median risk score. As a result, high-risk patients showed an obviously increased death proportion compared with low-risk counterparts ($p < 0.001$), suggesting the negative correlation of the risk score with patient survival (Figures 2B,C). Based on ROC analysis, the CR signature achieved a 0.889 prognostic accuracy in TCGA dataset (Figure 2D). The results of the validation cohort also proved the significant difference between high- and low-risk groups ($p < 0.05$), and the ROC analysis indicated a 0.857 prognostic accuracy of the CR signature (Figures 2E,F).

Independent prognostic indicators of chromatin regulator signature

Univariate and multivariate Cox regression analyses were conducted to demonstrate the feasibility of the CR signature in order to independently predict prognosis. According to the results of univariate regression, the clinical stage, clinical T stage, and risk score showed significant relation to ACC survival ($p < 0.001$). Upon multivariate regression, the clinical T stage and risk score remained significantly associated with ACC survival ($p < 0.05$) (Table 2). In the validation cohort, univariate Cox regression also showed that the risk score was notably associated with ACC survival ($p < 0.05$) (Table 3). All the

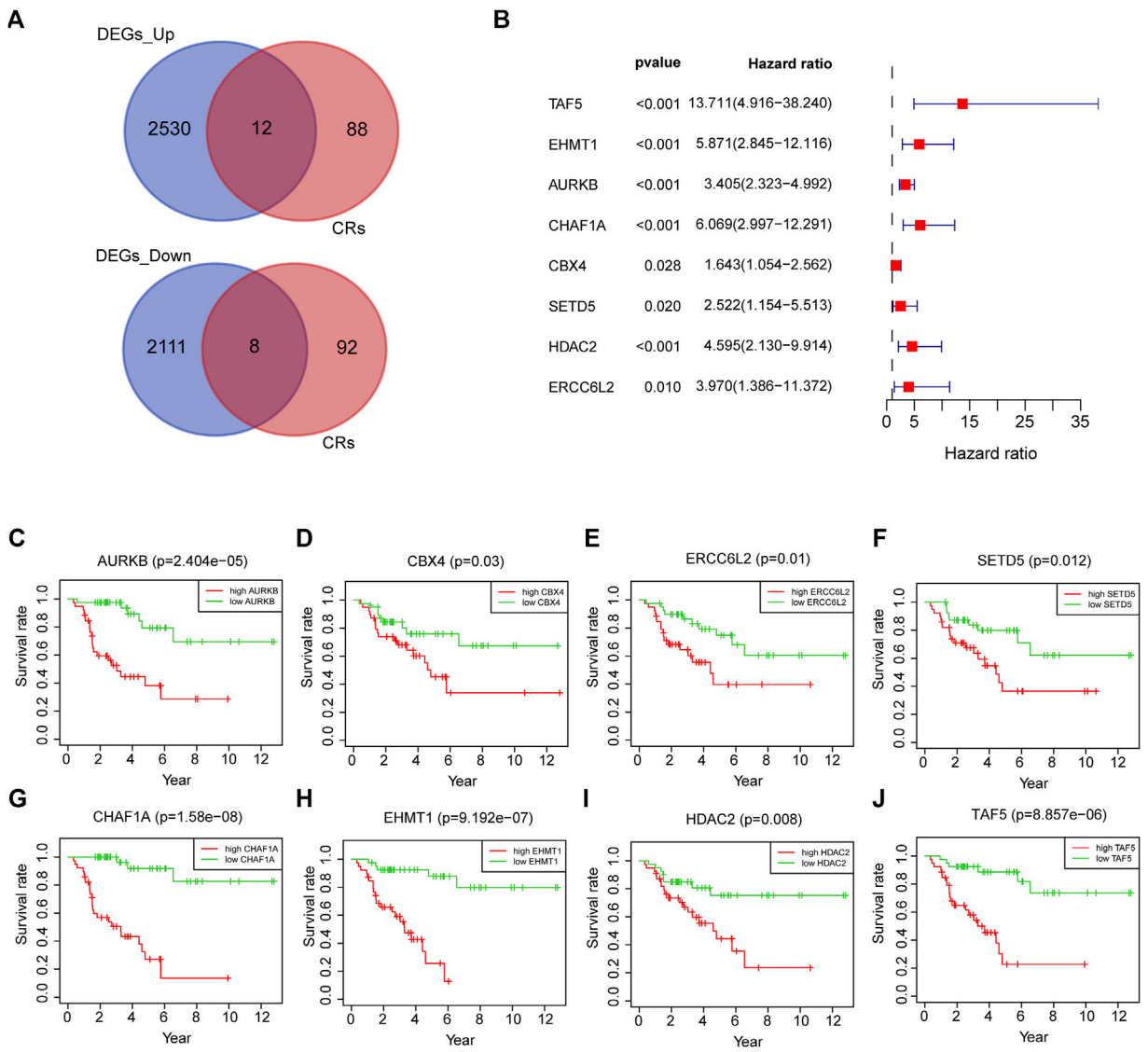


FIGURE 1
20 DECRs selected from the intersection of DEGs and CRs **(A)**. Univariate Cox regression of eight DECRs that showed a prognostic value **(B)**. KM survival analysis of eight DECRs that showed a prognostic value **(C–J)**.

aforementioned results indicated that the CR signature was the independent prognostic indicator for ACC patients.

Relationship of chromatin regulator signature with clinical features

This study utilized a chi-squared test to explore the involvement of CR prognostic features in ACC occurrence and progression. As a result, the clinical T stage ($p < 0.001$) and clinical stage ($p < 0.001$) were significantly different between high- and low-risk groups, while no difference was detected in gender or clinical N stage ($p > 0.05$) (Figures 3A,B). In addition, further subgroup analyses were performed to investigate whether the CR signature was significant for prognosis prediction. According to the obtained results, the CR signature exhibited excellent performance in predicting I–III ($p = 0.02$), I–IV ($p < 0.001$), II–III ($p = 0.017$), II–IV ($p < 0.001$), T1–T3 ($p = 0.046$), T1–T4 ($p < 0.001$), and T2–T4 ($p < 0.001$) stages, while the CR signature performed poorly concerning its prognosis prediction performance at T1–T2 and I–II stages ($p > 0.05$) (Figures 3C,D).

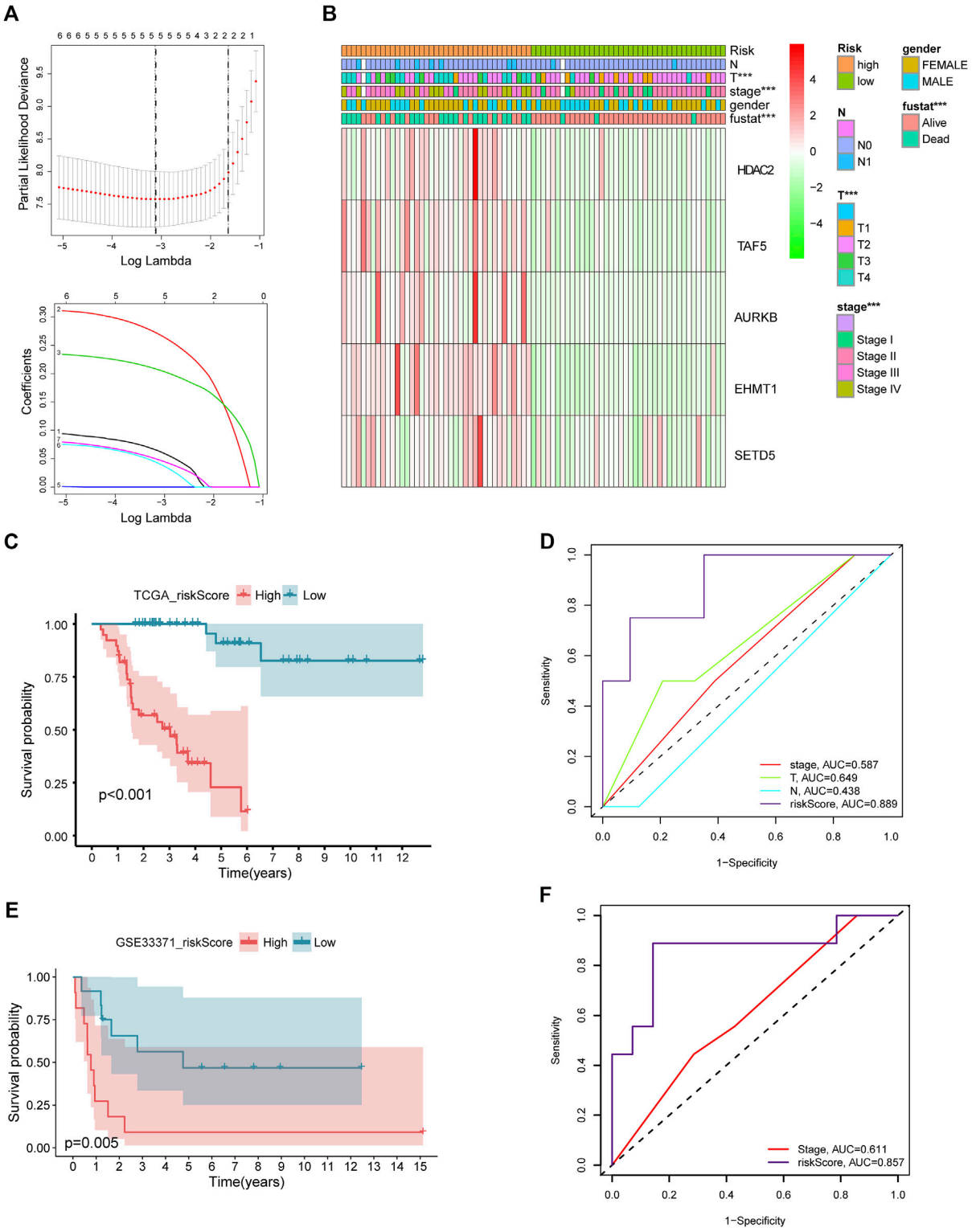


FIGURE 2
Lasso Cox regression of DECRs **(A)**. Expression of five selected DECRs in different clinical features groups **(B)**. KM survival analysis of low- and high-risk patients in test cohort **(C)**. ROC analysis of clinical stage, clinical T stage, clinical N stage and riskscore in test cohort **(D)**. KM survival analysis of low- and high-risk patients in validation cohort **(E)**. ROC analysis of stage and riskscore in validation cohort **(F)**.

TABLE 1 Correlation coeffcients of five selected DECRs.

Gene	Coef
TAF5	0.063716491
EHMT1	0.269857731
AURKB	0.206839715
SETD5	0.041749529
HDAC2	0.048168878

Construction and verification of the nomogram

Different prognostic indicators were incorporated into the nomogram to graphically assess survival probabilities for different patients in the preoperative stage. The nomogram incorporating the clinical T stage and risk score was constructed to better predict 1-, 3-, and 5-year patient prognosis (Figure 4A). Based on the calibration curve, there was good consistency between the measured patient survival and the estimated survival (Figures 4B–D). In addition, the nomogram achieved a C-index of 0.929, proving its good predictive power. During validation, due to missing the clinical T stage date of patients, we constructed a nomogram based only on the risk score and used it to predict 1-, 3-, and 5-year patient prognosis in the validation cohort. The calibration curve also showed good consistency between the measured patient survival and the estimated survival. At the same time, the C-index of the nomogram in the validation cohort was 0.726 (Figures 5A–D).

Functional annotation and gene set enrichment analyses

This study performed GO and KEGG analyses to explore possible functions of DECRs. According to BP analysis results, these 20 DECRs were significantly related to histone modification, peptidyl-lysine modification, and covalent chromatin modification. Based on the analysis of CC, these 20 DECRs were mainly associated with PcG protein complexes, nuclear chromatin, and RNA polymerase II. MF analysis demonstrated that the 20 DECRs were mainly enriched in the histone methyltransferase activity, histone–lysine N-methyltransferase activity, and protein–lysine N-methyltransferase activity (Figure 6). Furthermore, KEGG analysis showed that the signaling pathways including base excision repair, cell cycle, colon cancer, gene duplication, and homologous recombination were enriched in the high-risk group. Meanwhile, the low-risk group was associated with signaling pathways such as allograft rejection, drug metabolism of cytochrome P450, metabolism of xenogeneic organisms by cytochrome P450, and retinol metabolism (Figures 7A,B). At the same time, to better clarify the molecular basis of CR signature, gene set enrichment analysis

TABLE 2 Univariate and multivariate regression of CR signature and other clinical features in test cohort.

Characteristic	Univariate analysis		Multivariate analysis	
	Hazard ratio (95% CI)	*p*-value	Hazard ratio (95% CI)	*p*-value
Gender	1.056 (0.490–2.276)	0.890	—	—
Stage	2.903 (1.844–4.569)	<0.001	1.198 (0.481–2.983)	0.699
T	3.364 (2.098–5.393)	<0.001	3.222 (1.302–7.971)	0.011
N	2.058 (0.774–5.472)	0.148	—	—
Risk score	1.006 (1.004–1.008)	<0.001	1.005 (1.003–1.008)	<0.001

TABLE 3 Univariate and multivariate regression of CR signature and other clinical features in validation cohort.

Characteristic	Univariate analysis		Multivariate analysis	
	Hazard ratio (95% CI)	*p*-value	Hazard ratio (95% CI)	*p*-value
Gender	1.358 (0.467–3.954)	0.574	—	—
Stage	1.700 (1.036–2.790)	0.036	2.589 (1.353–4.954)	0.004
Risk score	12,191.428 (18.539–8,017,342.094)	0.004	1676261.622 (370.856–7576636516.957)	<0.001

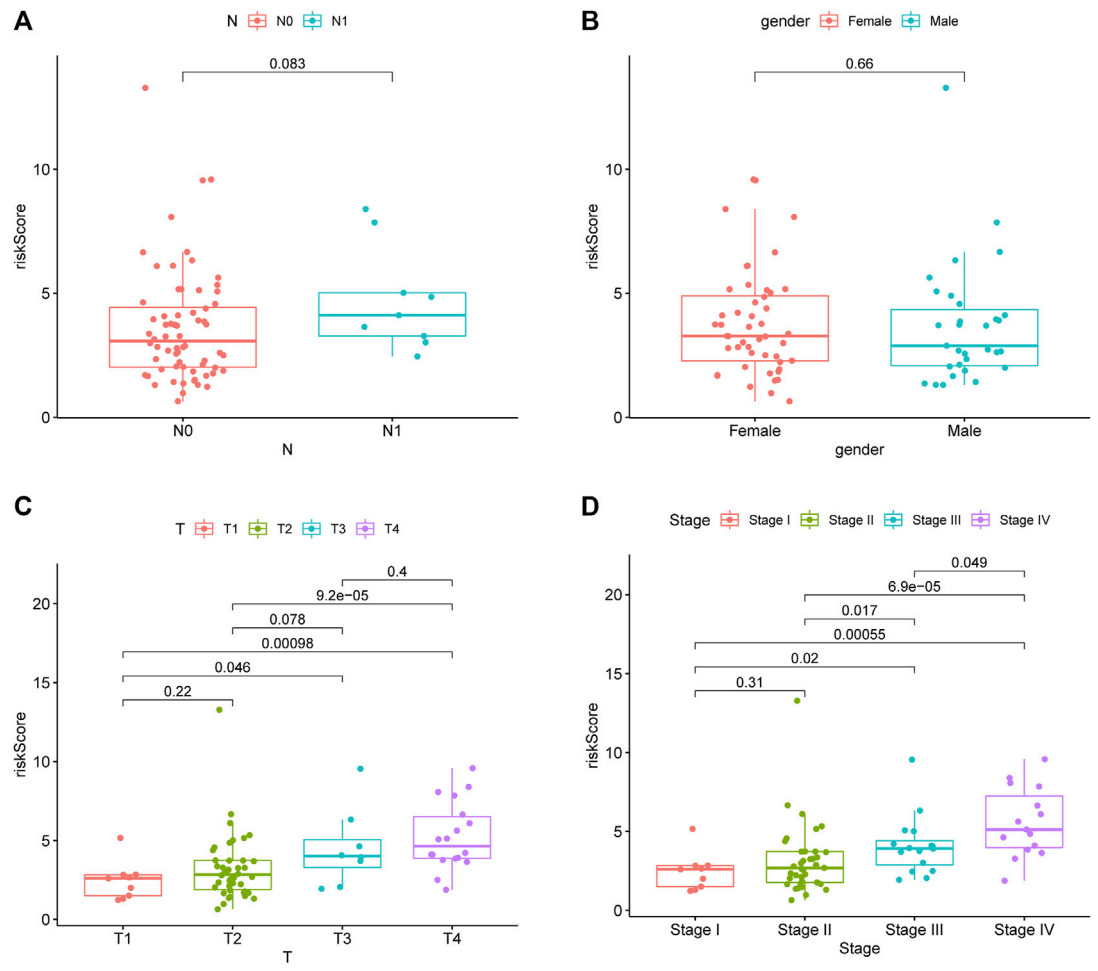


FIGURE 3
Riskscore in different gender, clinical stage, clinical T stage and clinical N stage.

(GSEA) was conducted. As a result, high-risk patients were associated with chromosome segregation, chromosome region, transcriptional binding, and other pathways, whereas low-risk patients were mostly associated with antigen processing and internalization, originated antigen presentation, luminal side of the endoplasmic reticulum, fatty acid binding, and other pathways (Figures 7C–H).

Immune infiltration analysis of chromatin regulator signature

According to the TIMER analysis, relations of the CR signature with immune infiltration were shown by a heat map (Figure 8). As a result, the immune infiltration degrees of endothelial cells, M2 macrophages, myeloid dendritic cells, $CD4^+$ Th1 cells, NKT cells, and M0 macrophages exhibited significant statistical differences between the high- and low-risk groups, and high infiltration levels of M0 and M2 macrophages were more pronounced in a higher T stage (T3 and T4), N stage (N1), and clinical stages (III and IV).

Drug sensitivity test

To improve the therapeutic efficacy in ACC cases, this study explored the difference in common chemotherapeutic agent sensitivity in ACC. Based on the results of the GDSC database analysis, for high-risk patients, their IC50 values of etoposide and doxorubicin increased compared with low-risk patients, suggesting the higher sensitivity of high-risk patients to these drugs (Figure 9). Meanwhile, RRM1 levels were significantly elevated among high-risk patients compared with low-risk patients ($p < 0.001$), which indicated the better curative effect of mitotane on low-risk cases.

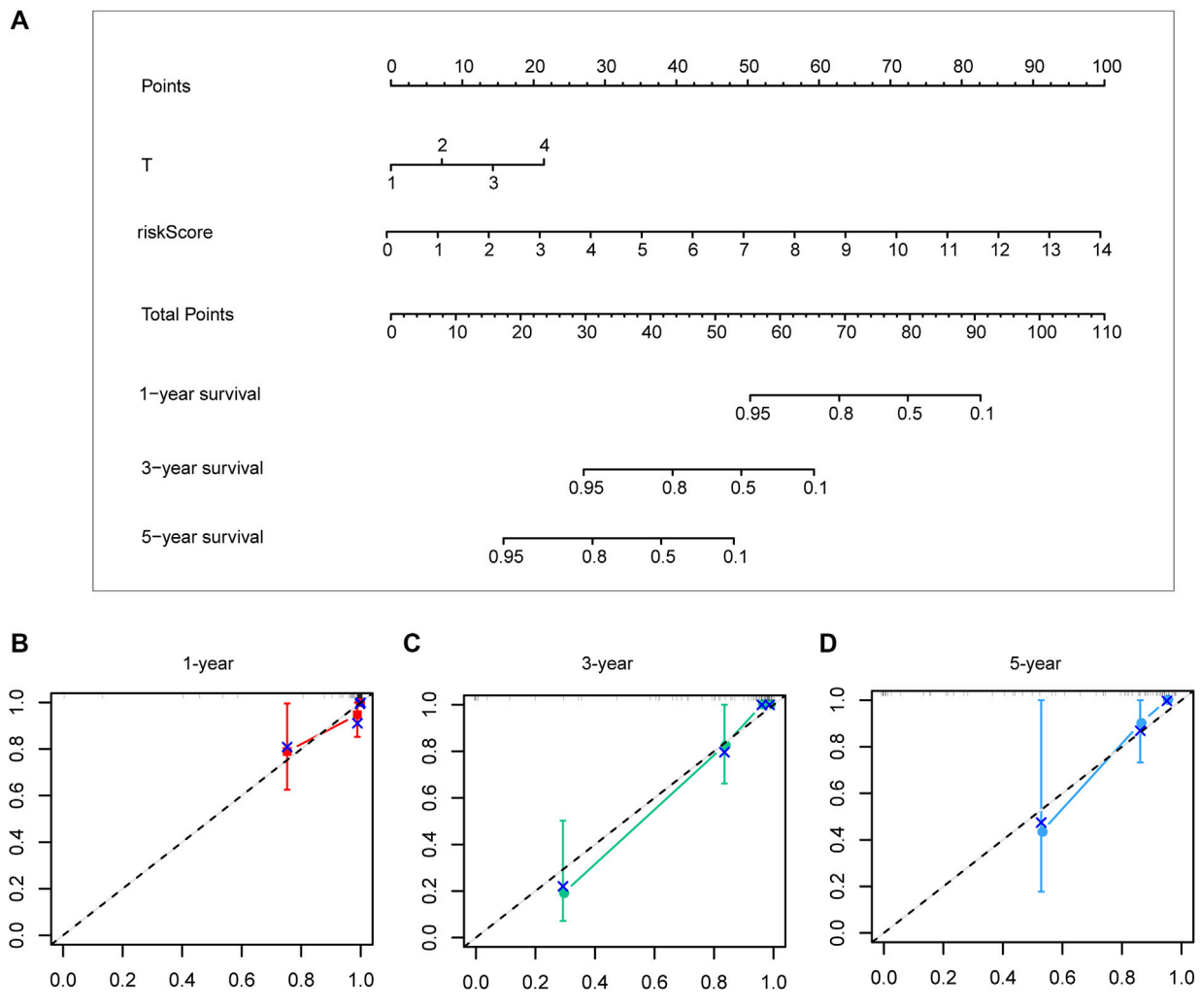


FIGURE 4
A nomogram constructed by clinical T stage and riskscore in test cohort **(A)**. Calibration curve predicting 1-, 3-, and 5-year patient prognosis **(B–D)**.

Discussion

Previous studies have shown that most ACC is sporadic of unknown origin, while a minority can be attributed to some hereditary neoplastic syndromes, including Li-Fraumeni syndrome, Lynch syndrome, MEN-1, and familial adenomatous polyposis (Vaidia et al., 2019). According to whether the tumor has an endocrine function, ACC is categorized into functional and non-functional types. Functional ACC can be diagnosed easier, and it clinically manifests as hypercortisolism, Cushing's syndrome, and primary hyperaldosteronism, while non-functional ACC often manifests as nonspecific tumor-induced symptoms due to its insidious onset. Because ACC does not show any obvious early onset characteristics, 70% of such cases are already in stages III-IV when they are diagnosed (Bharwani et al., 2011; Fay et al., 2014). Currently, the treatment methods for ACC include surgery, chemotherapy, and radiotherapy, but none of them can achieve ideal therapeutic effects (Allolio and Fassnacht, 2006). The discovery of novel predictive factors for the diagnosis and prognosis of ACC will help clinicians assess the risk for patients and formulate the targeted treatment strategies. With the development of information technology, study on the diagnostic and prognostic markers for ACC has gradually emerged in recent years. For instance, He ZJ screened 15 key genes (*CXCR6*, *SELL*, *P2RY13*, *GNG8*, *OMD*, *ABI3BP*, *OGN*, *FBLN1*, *LOXL1*, *ELN*, *CTSK*, *HGF*, *SH3GL3*, *F13A1*, and *GTPBP2*) based on the mRNA-seq sequencing data and the stem cell index established according to TCGA-ACC mRNA profiles. In addition, they

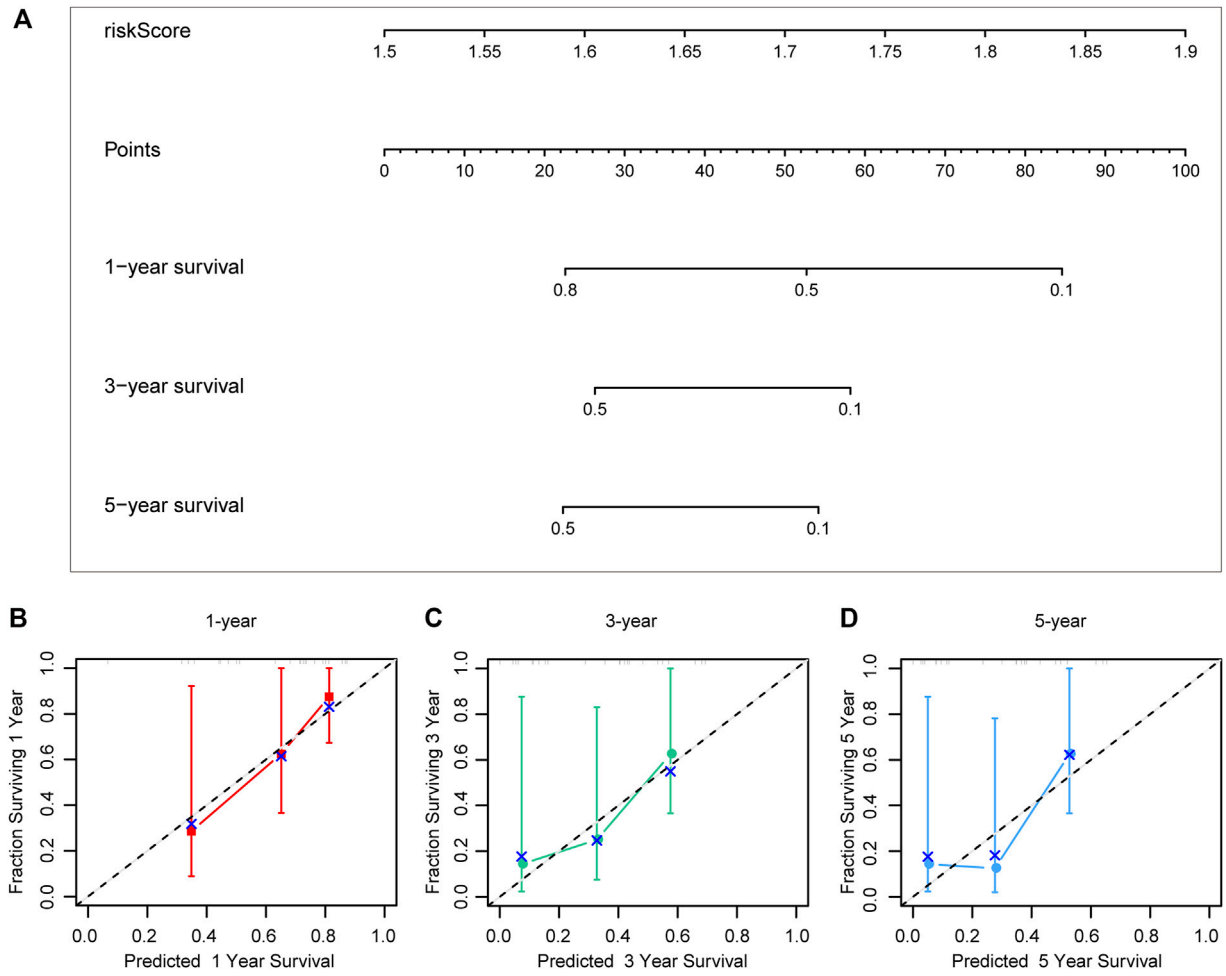


FIGURE 5
A nomogram constructed by riskscore in validation cohort **(A)**. Calibration curve predicting 1-, 3-, and 5-year patient prognosis **(B–D)**.

also pointed out that GTPBP2 was the only key gene with prognostic significance (He, 2021). After that, Hu DF et al. employed the GEO database to screen the differential core genes that were upregulated (*RACGAP1*, *CCNB1*, *TYMS*, *MAD2L1*, *NCAPG*, and *CDK1*) and downregulated (*IGF1*, *CXCL12*, *TLR4*, *TGFBR2*, and *HGF*). However, the authors did not investigate the value of these genes. Qi et al. (2021) and Zhou et al. (2022) analyzed more updated GEO database samples on the aforementioned basis. As a result, *CCNB1*, *CCNA2*, *CDK1*, *BUB1B*, *MAD2L1*, *RRM2*, *TPX2*, *AURKA*, *TOP2A*, *ZWINT*, and *NCAPG* were found to be closely related to prognosis. Nevertheless, further exploration of their value was lacking. In other words, the data-supported reliable clinical diagnostic and prognostic indicators are still needed for ACC. Recently, CRs have been increasingly suggested to make different effects on carcinogenesis, while little existing research has systemically examined CRs and explored their clinical value for ACC.

Based on the aforementioned starting point, this study selected a total of eight DECRs with upregulation and 12 DECRs with downregulation. According to univariate Cox regression, eight of them showed a prognostic value, and five genes were successfully constructed by adopting the Lasso Cox model. Subsequently, we verified the expression differences of the aforementioned genes by using the GEO database. The result demonstrated that *TAF5*, *EMHT1*, *AURKB*, and *SETD5* all showed significant expression differences in line with the results except *HDAC2*. After discussion, we believed that the difference in *HDAC2* may be caused by the small sample size. Then, correlation coefficients of five DECRs were determined to calculate the risk score as follows (0.0637 × TAF5 expression) + (0.2699 × EMHT1 expression) + (0.2068 × AURKB expression) + (0.0418 × SETD5 expression) + (0.0482 × HDAC2 expression). According to the median risk score, ACC cases were classified into a low- or high-risk group. As a result, high-risk patients showed significantly increased deaths compared with low-risk

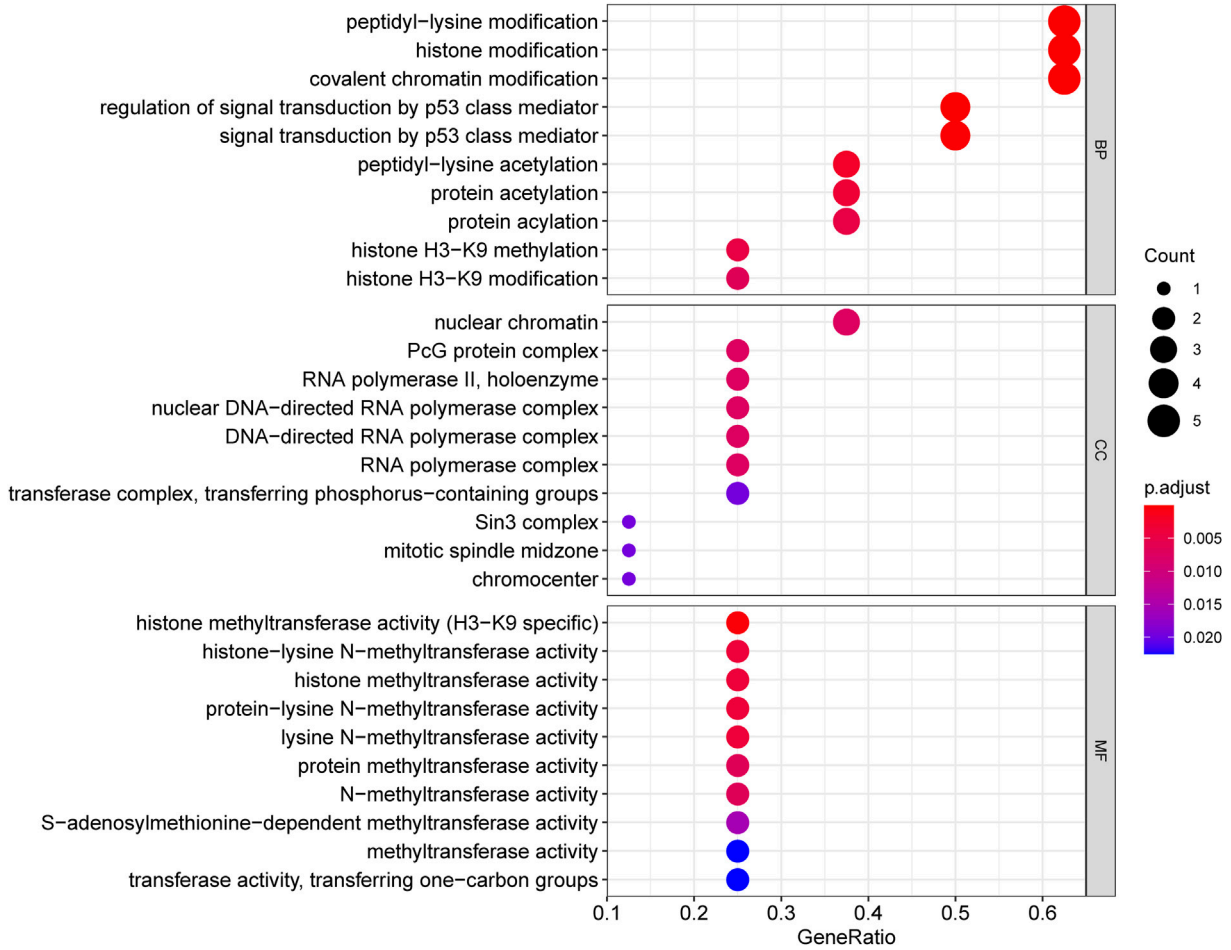


FIGURE 6
GO_BP, CC and MF analysis of 20 DECRs.

counterparts ($p < 0.001$), suggesting a negative correlation between the risk score and prognosis. Based on ROC analysis, the CR signature achieved a prognostic accuracy of 0.889. The results of the validation cohort were consistent with this finding, and the ROC analysis showed the prognostic accuracy of the CR signature was 0.857 in the validation cohort, which reflected the superior prognostic value of the CR signature. Afterward, univariate and multivariate Cox regression analyses were conducted. As a result, the CR signature independently predicted ACC prognosis. According to the chi-squared test, the clinical T stage ($p < 0.001$) and clinical stage ($p < 0.001$) were significantly different between the two groups, while age and clinical N stage did not exhibit any difference ($p > 0.05$). Furthermore, our study proved that the CR signature exhibited excellent performance in predicting I–III ($p = 0.02$), I–IV ($p < 0.001$), II–III ($p = 0.017$), II–IV ($p < 0.017$) T1–T3 ($p = 0.046$), T1–T4 ($p < 0.001$), and T2–T4 ($p < 0.001$) stages. Finally, the nomogram incorporating the clinical T stage and risk score was constructed. According to the calibration curve, the measured patient survival showed high consistency with the estimated one. Our nomogram achieved a C-index of 0.929, confirming its good prediction performance. Because one of the main purposes of our research was to explore the prognostic ability of the CR signature, although the clinical T stage data on patients in the validation cohort were missing, we still constructed a nomogram based on the risk score for validation. The results also showed good consistency between the measured patient survival and estimated survival. At the same time, the C-index of the nomogram in the validation cohort was 0.726. It is of note that previous studies have revealed that surgical methods, surgical margins, pathological features, and Ki-67 proliferation index are also associated with poor prognosis in ACC. Nevertheless, databases including TCGA and GEO cannot provide detailed data on the corresponding aspects of patients. This study concentrated on building a preoperative, less traumatic predictive risk model. Therefore, only the clinical T

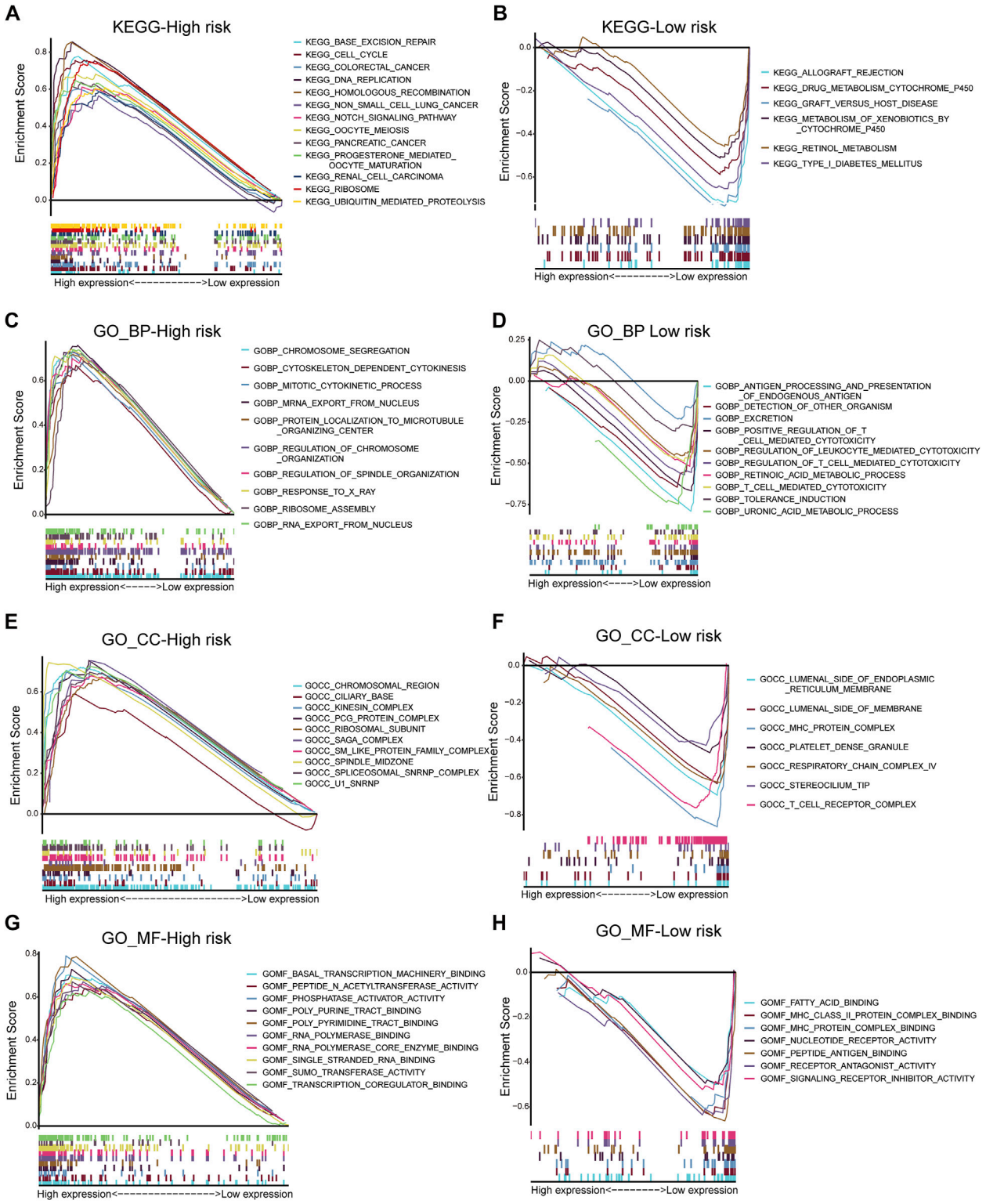


FIGURE 7
KEGG analysis of low- and high-risk groups **(A,B)**. GSEA analysis of BP, CC and MF in low- and high-risk groups **(C–H)**.

FIGURE 8
Heat map of CR signature with immune infiltration by TIMER analysis.

stage and risk score were included to build the nomogram. In addition, the results also confirmed the excellent predictive ability of the model.

As the 100-kDa subunit of the universal transcription factor TFIID, human TAF5 makes a vital effect on assembling the 1.2-MDa TFIID complex. In a study on human papillomavirus (HPV), in the context of oral squamous cell carcinoma (OSCC), *TAF5* and other genes showed high enrichment into HPV-positive somatic mutations, which mostly influence the HPV oncoprotein-targeted host pathways including pRB and p53 pathways. They also play important roles in disrupting the host's defense against viral infection and are potentially involved in nuclear factor-kappa B (NF-kB) and interferon (IFN) signaling (Gillison et al., 2019). Lee J and colleagues explored the effect of EHMT1 on lung cancer. According to the obtained results, EHMT1 was significantly related to apoptosis and the cell cycle process and had an important impact on regulating the apoptosis and cell cycle of tumor cells by regulating the expression of *CDKN1A* (Lee et al., 2021). Watson et al. (2019) also confirmed in their study on high-grade serous ovarian cancer (HGCOC) that disruption of EHMT1/2 sensitized HGSOC cells to PARP inhibitors (PARPi). In addition, the authors also proposed a potential mechanism through DNA damage and cell cycle dysregulation (Watson et al., 2019). As a pan-cancer marker, AURKB is related to different tumor occurrences and development, including hepatocellular carcinoma (HCC) (Yang et al., 2022), bladder cancer (BLCA) (Tang and Wang, 2019), breast cancer (BRCA) (Zhang et al., 2021), lung adenocarcinoma (LUAD) (Ding et al., 2019), and osteosarcoma (Shan, 2021), exhibiting certain prognostic significance. Wang et al. (2020) identified SETD5 as a major driver of resistance to MEK1/2 (MEKi) in pancreatic ductal adenocarcinoma (PDAC), revealing that SETD5 was a key mediator of acquired resistance to MEKi therapy in PDAC. In addition, Chen et al. (2021) also confirmed that SETD5 promoted the cancer stem cell properties of non-small cell lung cancer (NSCLC) by attenuating the PI3K/Akt/mTOR pathway activation. Currently, the abnormal expression of *HDAC2* in

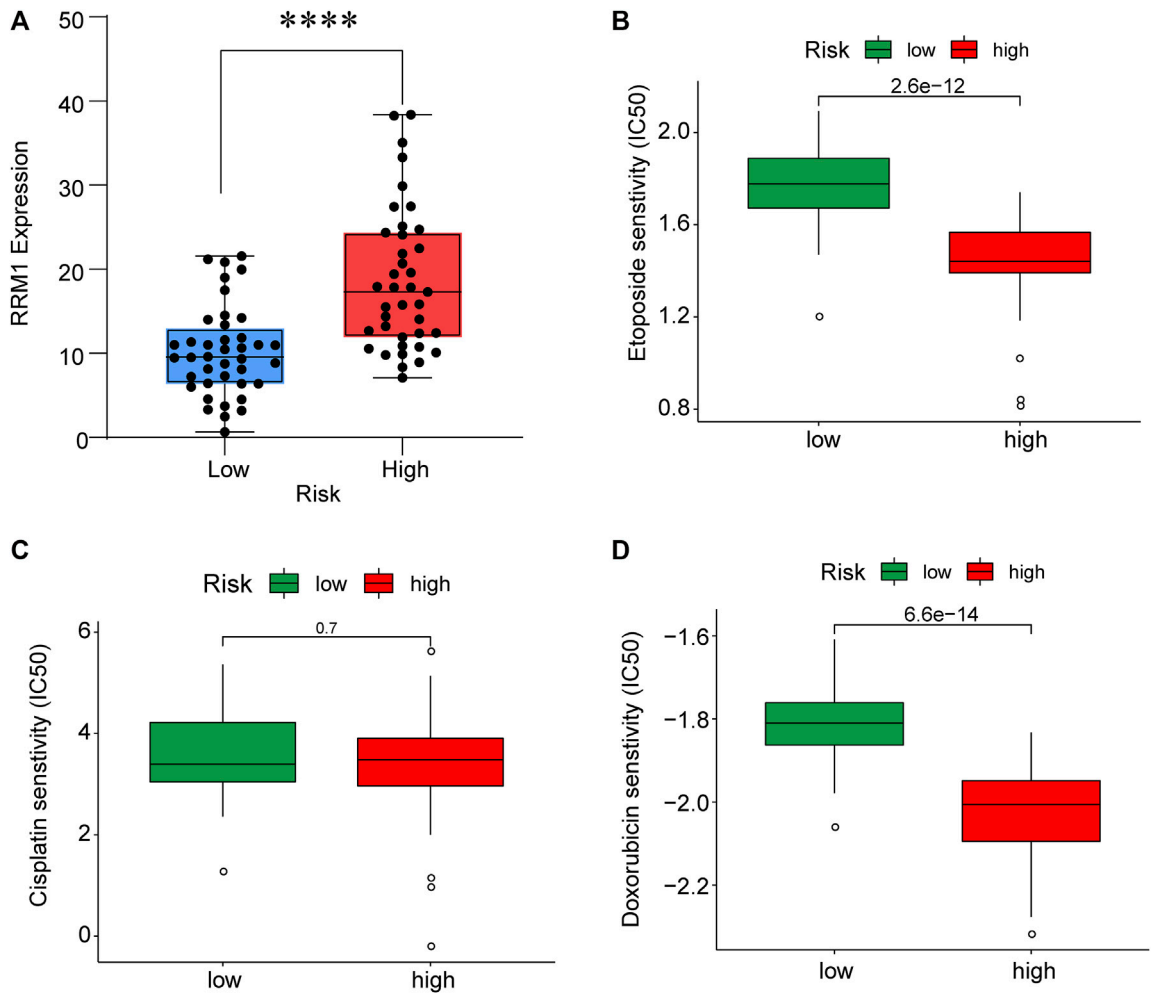


FIGURE 9
RRM1 expression in low- and high-risk groups **(A)**. Etoposide, cisplatin and doxorubicin senstivity in low- and high-risk groups **(B–D)**.

different cancers has been widely confirmed, which is associated with cancer proliferation, invasion, migration, and drug resistance. HDAC2 also participates in tumor metabolism and influences the clinical diagnosis, treatment, and prognosis of cancers. In tumor cells, *HDAC2* acts as both a tumor-promoting gene and a tumor suppressor gene. In addition, its specific role is related to its target genes and pathways involved in various malignant tumors.

This study combined these five key CR genes at the ACC level for the first time and verified their unique prognostic and diagnostic significance. Based on GO and KEGG analyses, the BP analysis revealed the significant involvement of 20 DECRs in covalent chromatin modification, peptidyl-lysine modification, and histone modification. Analysis of CC revealed the significant enrichment of 20 DECRs in nuclear chromatin, PcG protein complexes, and RNA polymerase II. MF analysis showed that the 20 DECRs were mainly located in the histone methyltransferase activity, histone-lysine N-methyltransferase activity, and protein–lysine N-methyltransferase activity. Moreover, KEGG analysis showed that the signaling pathways including the base excision repair, cell cycle, colon cancer, gene duplication, and homologous recombination were enriched in the high-risk group. Apart from that, signaling pathways such as allograft rejection, drug metabolism of cytochrome P450, metabolism of xenogeneic organisms by cytochrome P450, retinol metabolism, and other signaling pathways were mostly related to low-risk patients. Based on GSEA results, the high-risk group was mainly associated with chromosome segregation, chromosome region, transcriptional binding, and other pathways, whereas low-risk groups were mostly related to antigen processing and presentation of endogenous antigens, luminal side of the endoplasmic reticulum, and fatty acid binding. Considering that ACC is a type of malignant endocrine tumor, this study also attempted to explore the relationship between these five

DECRs and endocrine function. It is interesting to find that these five CRs are rarely discovered to be involved in some key endocrine metabolic pathways in previous studies. Among the enriched pathways, only cytochrome P450-related pathways have been shown to regulate aldosterone biosynthesis and participate in the pathogenesis of primary hyperaldosteronism (Bassett et al., 2004; Zennaro et al., 2012). According to other pathway enrichment results, it could be speculated that CRs more probably promote the occurrence and development of ACC by influencing cell division, cell cycle, and other links, rather than changing the level of hormone metabolism.

Through TIMER analysis, this study proved that the immune infiltration degrees of endothelial cells, M2 macrophages, myeloid dendritic cells, $CD4^+$ Th1 cells, NKT cells, and M0 macrophages exhibited significant statistical differences between the high- and low-risk groups, and high infiltration levels of M0 and M2 macrophages were more pronounced in the higher T stage (T3, T4), N stage (N1), and clinical stages (III, IV). Macrophages, also known as tumor-associated macrophages (TAMs), block tumor immunity by producing immunosuppressive molecules and inducing immune tolerance, thereby generating a tumor microenvironment (TME) favorable for immune heterogeneity. Studies have proved that TAMs are involved in various biological events, including epithelial-mesenchymal transition, immune escape, tumor angiogenesis, and cancer metastasis (Li et al., 2020b), which are also likely to be the main mechanisms that these CRs affect the poor prognosis of ACC in terms of immune infiltration. According to our sensitivity difference analysis of common chemotherapeutic agents in ACC, for high-risk patients, their IC50 values of etoposide and doxorubicin increased compared with those of low-risk patients, suggesting the higher drug sensitivity of high-risk patients. Mitotane is currently the most commonly used and effective oral drug for the treatment of ACC (Author Anonymous, 2021). The low expression of the *RRM1* gene has been confirmed to be related to mitotane efficacy (Tang et al., 2020). Therefore, this study investigated *RRM1* expression based on the aforementioned analyses. As a result, high-risk patients had markedly increased RRM1 expression relative to low-risk counterparts ($p < 0.001$), indicating that mitotane had a better therapeutic efficacy in low-risk cases. Certainly, certain limitations should be noted in this work. For example, the mechanism by which these CRs regulate ACC cell biology should be further verified through further experiments. In addition, more multi-center clinical trials are also needed for verifying that our prognosis prediction model is practicable. There are still some challenges to be encountered, given the clinical rarity of ACC.

Conclusion

To conclude, this study identified the DECRs that were possibly related to ACC genesis and progression, screened pathways enriched by these DECRs, and verified the excellent value of these DECRs in prognosis prediction for ACC cases. Moreover, the drug sensitivity of DECRs was also analyzed. Although more investigations are warranted for verifying our conclusions, this study provides reliable ideas and evidence for the clinical diagnosis and treatment of ACC.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials; further inquiries can be directed to the corresponding author.

Author contributions

JL, YJ, and YZ designed the study. JL, YJ, LT, and RZ drafted the manuscript. JL, YJ, and YZ prepared the table and figure. All authors participated in the revision of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Research Program of Natural Science Foundation in Chongqing (cstc2021jcyj-msxmX0484), the National Natural Science Foundation of China (No.81801507), and the Kuanren Talent Program of Second Affiliated Hospital of Chongqing Medical University (KY2019Y004).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.948353/full#supplementary-material

References

Abe, I., and Lam, A. K. (2021). Anaplastic thyroid carcinoma: Updates on WHO classification, clinicopathological features and staging. *Histol. Histopathol.* 36 (3), 239–248. doi:10.14670/HH-18-277

Allolio, B., and Fassnacht, M. (2006). Clinical review: Adrenocortical carcinoma: clinical update. *J. Clin. Endocrinol. Metab.* 91 (6), 2027–2037. doi:10.1210/jc.2005-2639

Aufforth, R. D., and Nilubol, N. (2014). Emerging therapy for adrenocortical carcinoma. *Int. J. Endocr. Oncol.* 1 (2), 173–182. doi:10.2217/ije.14.13

Bassett, M. H., White, P. C., and Rainey, W. E. (2004). The regulation of aldosterone synthase expression. *Mol. Cell. Endocrinol.* 217, 67–74. doi:10.1016/j.mce.2003.10.011

Bharwani, N., Rockall, A. G., Sahdev, A., Gueorguiev, M., Drake, W., Grossman, A. B., et al. (2011). Adrenocortical carcinoma: the range of appearances on CT and MRI. *AJR. Am. J. Roentgenol.* 196 (6), W706–W714. doi:10.2214/AJR.10.5540

Chandrasekar, T., Goldberg, H., Klaassen, Z., Wallis, C. J. D., Woon, D. T. S., Herrera-Caceres, J. O., et al. (2019). The who, when, and why of primary adrenal malignancies: Insights into the epidemiology of a rare clinical entity. *Cancer* 125 (7), 1050–1059. doi:10.1002/cncr.31916

Chen, J., Wang, F., Xu, H., Xu, L., Chen, D., Wang, J., et al. (2020). Long non-coding RNA SNHG1 regulates the wnt/β-catenin and PI3K/AKT/mTOR signaling pathways via EZH2 to affect the proliferation, apoptosis, and autophagy of prostate cancer cell. *Front. Oncol.* 10, 552907. doi:10.3389/fonc.2020.552907

Chen, Q., Sun, Z., Li, J., Zhang, D., Guo, B., and Zhang, T. (2021). SET domain-containing protein 5 enhances the cell stemness of non-small cell lung cancer via the PI3K/Akt/mTOR pathway. *J. Environ. Pathol. Toxicol. Oncol.* 40 (2), 55–63. doi:10.1615/JEnvironPatholToxicolOncol.2021036991

Chu, Y., Chen, W., Peng, W., Liu, Y., Xu, L., Zuo, J., et al. (2020). Amnion-derived mesenchymal stem cell exosomes-mediated autophagy promotes the survival of trophoblasts under hypoxia through mTOR pathway by the downregulation of EZH2. *Front. Cell Dev. Biol.* 8, 545852. doi:10.3389/fcell.2020.545852

Ding, H., Ye, Y., An, H., Gao, Q., and Zhong, Y. (2019). Expression of key genes in lung adenocarcinoma and its prognostic significance. *Chin. J. Comp. Med.* 29 (10), 54–60. doi:10.3969/j.issn.1671-7856.2019.10.010

Else, T., Kim, A. C., Sabolch, A., Raymond, V. M., Kandathil, A., Caoili, E. M., et al. (2014). Adrenocortical carcinoma. *Endocr. Rev.* 35 (2), 282–326. doi:10.1210/er.2013-1029

Fassnacht, M., Johanssen, S., Quinkler, M., Bucsky, P., Willenberg, H. S., Beuschlein, F., et al. (2009). Limited prognostic value of the 2004 international union against cancer staging classification for adrenocortical carcinoma: proposal for a revised TNM classification. *Cancer* 115 (2), 243–250. doi:10.1002/cncr.24030

Fay, A. P., Elfiky, A., Teló, G. H., McKay, R. R., Kaymakcalan, M., Nguyen, P. L., et al. (2014). Adrenocortical carcinoma: the management of metastatic disease. *Crit. Rev. Oncol. Hematol.* 92 (2), 123–132. doi:10.1016/j.critrevonc.2014.05.009

Gillison, M. L., Akagi, K., Xiao, W., Jiang, B., Pickard, R. K. L., Li, J., et al. (2019). Human papillomavirus and the landscape of secondary genetic alterations in oral cancers. *Genome Res.* 29 (1), 1–17. doi:10.1101/gr.241141.118

He, Z. (2021). Bioinformatic analysis identifies prognostic genes associated with cancer stem cell properties in adrenocortical carcinoma. *Zeng Guohua, editor-in-chief* 2022, 1–67. doi:10.27043/d.cnki.ggzyc.2021.000317

Lam, K. Y. (1992). Adrenal tumours in Chinese. *Virchows Arch. A Pathol. Anat. Histopathol.* 421 (1), 13–16. doi:10.1007/BF01607133

Lee, J., Kim, K., Ryu, T. Y., Jung, C. R., Lee, M. S., Lim, J. H., et al. (2021). EHMT1 knockdown induces apoptosis and cell cycle arrest in lung cancer cells by increasing CDKN1A expression. *Mol. Oncol.* 15 (11), 2989–3002. doi:10.1002/1878-0261.13050

Li, H., Ji, Z., Zhang, B., Zhang, Y., Lu, L., Deng, J., et al. (2021). Expert consensus on diagnosis and treatment of adrenocortical carcinoma. *Mod. J. urology* 26 (11), 902–908. doi:10.3969/j.issn.1009-8291.2021.11.002

Li, T., Yang, J., Yang, B., Zhao, G., Lin, H., Liu, Q., et al. (2020). Ketamine inhibits ovarian cancer cell growth by regulating the lncRNA-PVT1/EZH2/p57 Axis. *Front. Genet.* 11, 597467. doi:10.3389/fgene.2020.597467

Li, Z., Liu, F. Y., and Kirkwood, K. L. (2020). The p38/MKP-1 signaling axis in oral cancer: Impact of tumor-associated macrophages. *Oral Oncol.* 103, 104591. doi:10.1016/j.oraloncology.2020.104591

Libé, R. (2015). Adrenocortical carcinoma (ACC): diagnosis, prognosis, and treatment. *Front. Cell Dev. Biol.* 3, 45. doi:10.3389/fcell.2015.00045

Lu, J., Xu, J., Li, J., Pan, T., Bai, J., Wang, L., et al. (2018). FACER: comprehensive molecular and functional characterization of epigenetic chromatin regulators. *Nucleic Acids Res.* 46 (19), 10019–10033. doi:10.1093/nar/gky679

Marazzi, I., Greenbaum, B. D., Low, D., and Guccione, E. (2018). Chromatin dependencies in cancer and inflammation. *Nat. Rev. Mol. Cell Biol.* 19 (4), 245–261. doi:10.1038/nrm.2017.113

Plass, C., Pfister, S. M., Lindroth, A. M., Bogatyrova, O., Claus, R., and Lichter, P. (2013). Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nat. Rev. Genet.* 14 (11), 765–780. doi:10.1038/nrg3554

Qi, E., Zhou, Y., and Dai, Y. (2021). Using bioinformatics methods to explore key genes in adrenocortical carcinoma. *J. Chongqing Med. Univ.* 46 (04), 444–449. doi:10.13406/j.cnki.cyxb.002458

Shan, Z. (2021). Bioinformatics methods for analysis of key gene expression and identification of osteosarcoma. *Li Fuguang, Editor-in-chief.* 2022, 1–65. doi:10.27466/d.cnki.gzzdu.2021.003470

Tang, Y., and Wang, S. (2019). Research progress of AURKB in bladder cancer. *J. Qiqihar Med. Coll.* 40 (12), 1521–1524. doi:10.3969/j.issn.1002-1256.2019.12.031

Tang, Y., Zhang, X., Liu, X., Mei, D., Fu, Q., Li, J., et al. (2020). Pharmacological effects and safety of mitotane in the treatment of adrenocortical carcinoma. *J. Clin. Pharmacother.* 18 (10), 68–71. doi:10.3969/j.issn.1672-3384.2020.10.015

Vaidia, A., Nehs, M., and Kilbridge, K. (2019). Treatment of adrenocortical carcinoma. *Surg. Pathol. Clin.* 12 (4), 997–1006. doi:10.1016/j.path.2019.08.010

Wang, Z., Hausmann, S., Lyu, R., Li, T. M., Lofgren, S. M., Flores, N. M., et al. (2020). SETD5-Coordinated chromatin reprogramming regulates adaptive resistance to targeted pancreatic cancer therapy. *Cancer Cell* 37 (6), 834–849. e13. doi:10.1016/j.ccell.2020.04.014

Watson, Z. L., Yamamoto, T. M., McMellen, A., Kim, H., Hughes, C. J., Wheeler, L. J., et al. (2019). Histone methyltransferases EHMT1 and EHMT2 (GLP/G9A) maintain PARP inhibitor resistance in high-grade serous ovarian carcinoma. *Clin. Epigenetics* 11 (1), 165. doi:10.1186/s13148-019-0758-2

Yang, Y., Xi, Z., and Zhang, L. (2022). Expression of Aurora kinase B and tumor protein 53 in hepatocellular carcinoma and their relationship with clinicopathological features and prognosis. *Liver* 27 (02), 188–192. doi:10.14000/j.cnki.issn.1008-1704.2022.02.006

Zennaro, M. C., Jeunemaitre, X., and Boulkroun, S. (2012). Integrating genetics and genomics in primary aldosteronism. *Hypertension* 60, 580–588. doi:10.1161/HYPERTENSIONAHA.111.188250

Zhang, L., Lu, J., Song, W., Liu, X., Liu, F., Wang, X., et al. (2021). Screening of breast cancer-related prognostic genes based on TCGA database and its clinical significance. *J. Qiqihar Med. Coll.* 42 (21), 1841–1845. doi:10.3969/j.issn.1002-1256.2021.21.001

Zhou, N., Fan, R., Zhang, S., Gu, C., and Lu, W. (2022). Screening and prognostic analysis of key genes in adrenocortical adenocarcinoma based on GEO database. *Mod. J. urology* 27 (02), 161–165. doi:10.3969/j.issn.1009-8291.2022.02.016

Glossary

ACC adrenocortical carcinoma

ACA adrenocortical adenoma

CRs chromatin regulators

KM Kaplan–Meier analysis

MPCs myeloid progenitor cells

NK natural killer cells

TC thyroid cancer

DM distant metastasis

DEGs differentially expressed genes

DECRs differentially expressed CRs

ROC receiver operating characteristic curves

OS overall survival

GO_BP Gene Ontology_biological process

GO_CC Gene Ontology_cellular component

GO_MF Gene Ontology_molecular function

KEGG Kyoto Encyclopedia of Genes and Genomes pathways

HPV human papillomavirus

OSCC oral squamous cell carcinoma

NF-kB Nuclear factor-kappa B

IFN interferon signaling

HGCOC high-grade serous ovarian cancer

PARPi PARP inhibitors

HCC hepatocellular carcinoma

BLCA bladder cancer

BRCA breast cancer

LUAD lung adenocarcinoma

PDAC pancreatic ductal adenocarcinoma

NSCLC non-small cell lung cancer

TAM tumor-associated macrophage

TME tumor microenvironment

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 29 August 2022
DOI 10.3389/fgene.2022.972043

OPEN ACCESS

EDITED BY
Peter Hart,
Roosevelt University College of Pharmacy, United States

REVIEWED BY
Jiangbo Wei,
The University of Chicago, United States
Lei Gu,
Max Planck Institute for Heart and Lung Research, Germany

*CORRESPONDENCE
Xuyong Sun,
sxywnyzyx@163.com

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 17 June 2022
ACCEPTED 05 August 2022
PUBLISHED 29 August 2022

CITATION
Yang X, Yang F, Lan L, Wen N, Li H and Sun X (2022), Diagnostic and prognostic value of m5C regulatory genes in hepatocellular carcinoma.
Front. Genet. 13:972043.
doi: 10.3389/fgene.2022.972043

Diagnostic and prognostic value of m5C regulatory genes in hepatocellular carcinoma

Xiawei Yang[1], Feng Yang[2], Liugen Lan[3,4,5,6], Ning Wen[3,4,5,6], Haibin Li[3,4,5,6] and Xuyong Sun[1,3,4,5,6]*

[1]Graduate School, Guangxi Medical University, Nanning, China, [2]Department of Gynocology, The Second Affiliated Hospital of Guangxi Medical University, Nanning, China, [3]Transplant Medical Center, The Second Affiliated Hospital of Guangxi Medical University, Nanning, China, [4]Guangxi Key Laboratory of Organ Donation and Transplantation, Nanning, China, [5]Guangxi Key Laboratory for Transplantation Medicine, Nanning, China, [6]Guangxi Transplantation Medicine Research Center of Engineering Technology, Nanning, China

Background: A high mortality rate makes hepatocellular carcinoma (HCC) one of the most common types of cancer globally. 5-methylcytosine (m5C) is an epigenetic modification that contributes to the prognosis of several cancers, but its relevance to HCC remains unknown. We sought to determine if the m5C-related regulators had any diagnostic or prognostic value in HCC.

Methods: M5C regulatory genes were screened and compared between HCC and normal tissue from The Cancer Genome Atlas (TCGA)and Gene Expression Omnibus (GEO) databases. Least absolute shrinkage and selection operator method (LASSO) and univariate Cox regression analysis of differentially expressed genes were then performed to identify diagnostic markers. A LASSO prognostic model was constructed using M5C regulatory genes with prognostic values screened by TCGA expression data. HCC patients were stratified based on risk score, then clinical characteristics analysis and immune correlation analysis were performed for each subgroup, and the molecular functions of different subgroups were analyzed using both Gene Set Enrichment Analysis (GSEA) and Gene Set Variation Analysis (GSVA). The prognostic model was evaluated using univariate and multivariate Cox analyses as well as a nomogram. Molecular typing was performed according to m5C regulatory genes and immune checkpoint genes expression respectively, and clinical characterization and immune correlation analysis were performed for each subgroup.

Results: M5C regulatory genes are expressed differently in HCC patients with different clinical and pathological characteristics, and mutations in these genes are frequent. Based on five m5C regulators (NOP2, NSUN2, TET1, YBX1, and DNMT3B), we constructed a prognostic model with high predictive ability. The risk score was found to be an independent prognostic indicator. Additionally, risk scores can also be applied in subgroups with different clinical characteristics as prognostic indicators.

Conclusion: The study combined data from TCGA and GEO for the first time to reveal the genetic and prognostic significance of m5C-related regulators in

HCC, which provides new directions for identifying predictive biomarkers and developing molecularly targeted therapies for HCC.

KEYWORDS

hepatocellular carcinoma, HCC, 5-methylcytosine, m5C, biomarkers, prognosis

Introduction

Hepatocellular carcinoma (HCC) ranks sixth in the cancer incidence worldwide and ranks third in cancer-related deaths (de Martel et al., 2020), and it is a major public health issue. Despite significant advancements in therapy, the 5-year survival rate for advanced HCC is still dismal due to the cancer's late detection, susceptibility to metastasis, and high recurrence rate. Although some biomarkers, including alpha-fetoprotein (AFP) and heat shock protein 90 (Hsp90), have proven to be useful, the search for early diagnosis biomarkers and effective therapies for HCC patients is urgent.

There is growing evidence that post-transcriptional modifications of RNA are important in different cancers (Cheng et al., 2018; Barbieri and Kouzarides, 2020; Begik et al., 2020; Chu et al., 2022), which provides ideas for developing new treatment modalities. There have been 170 types of modifications identified thus far (Boccaletto et al., 2018), such as N6-methyladenosine (m6A), 5-methylcytosine (m5C) (Wang et al., 2013), 7-methylguanosine, and pseudouridylation (Roundtree et al., 2017; Shi et al., 2020). However, their functions remain widely unknown due to technical limitations in accurate localization throughout the genome (Cohn, 1960; Bauer et al., 2016). There are many post-transcriptional modifications, but the most common is a reversible modification called m5C, which serves different functions in different RNA types (Chow et al., 2007; Squires et al., 2012; Huang et al., 2019; Trixl and Lusser, 2019; He et al., 2020a; Cui et al., 2020). M5C modification involves adenosine methyltransferases ("writers"), demethylases ("erasers"), and "readers" for protein recognition and binding. The "writers" include NSUN1-NSUN7, DNMT1, DNMT2, DNMT3a, and DNMT3b, "erasers" include TET1, TET2, TET3, and ALKBH1, and among the "readers" are ALYREF and YBX1. Abnormal modification of m5C has been connected to many abnormal states, for example mitochondrial dysfunction, abnormal embryogenesis and neurodevelopment, tumorigenesis, and tumor cell proliferation and migration (Navarro et al., 2021; Walworth et al., 2021). It has also been suggested that m5C modification can even alter the fate of cancer cells (Yang et al., 2020), and can be utilized as a biomarker for the prognosis of many kinds of cancers (Gama-Sosa et al., 1983; Chellamuthu and Gray, 2020). One study comprehensively explored and systematically profiled the expression features of m5C-related regulators in HCC and proved the m5C modification patterns play a crucial role in the tumor immune microenvironment and prognosis of HCC (Liu et al., 2022b). In spite of the fact that anomalous RNA m5C modification has been detailed to play numerous capacities in HCC (He et al., 2020b; Sun et al., 2020), the relationship between m5C regulatory genes and HCC is still poorly understood, and the diagnostic and prognostic value of m5C regulatory genes for HCC is unknown.

This study screened and compared the expression characteristics of the m5C regulators in HCC samples with those in normal samples using the expression matrix from TCGA and GEO databases. Univariate Cox as well as LASSO regression analyses were employed to discover diagnostic markers. Then five m5C regulatory genes with prognostic value were screened by using the data from TCGA to construct a prognostic model. To find out if m5C regulatory genes are valuable for diagnosis and prognosis in HCC, researchers performed molecular typing based on m5C regulatory gene and immune checkpoint gene expression, and immune correlate analyses and clinical characteristic analyses were also performed for each subgroup.

Materials and methods

Acquired data and identified differentially expressed genes

We obtained Gene expression data from TCGA database (Hutter and Zenklusen, 2018) (https://portal.gdc.cancer.gov/) and the GSE76427 dataset (Grinchuk et al., 2018) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76427) in the GEO database (Barrett et al., 2007) (https://www.ncbi.nlm.nih.gov/geo/). The TCGA database contains expression data (Table 1), copy number variants (CNVs), single nucleotide polymorphisms (SNPs), and relevant clinicopathological features for 374 HCC samples and 50 paraneoplastic samples.

TABLE 1 Baseline data.

Data	Normal	Tumor
TCGA	50 (11.8%)	374 (88.2%)
GSE76427	52 (31.1%)	115 (68.9)

The microarray platform for GSE76427 (sample size: disease group 115/control group 52) (Table 1) dataset is Illumina HumanHT-12 V4.0 expression beadchip, and gene set related to m5C regulators was obtained by Cui et al.'s study (Cui et al., 2021; Wang et al., 2021; Liu et al., 2022a). We first used "sva" package (Leek et al., 2012) to preprocess the downloaded TCGA and GEO dataset expression matrices, including: data background adjustment and normalization, and output the expression of intersecting genes in the two datasets separately. The Perl language was then applied to extract the expression of m5C regulator genes in both datasets. To determine the validity of the grouping, we did a principal component analysis (PCA) and visualized with the help of "ggplot2" package. Subsequently, by using "limma" package, we determined DEGs between HCC and normal liver tissue at $p < 0.05$.

Copy number variant and single nucleotide polymorphism analyses

GISTIC 2.0 was used to find genes with significant amplifications or deletions (Mermel et al., 2011) with thresholds of $p > 0.1$ and $p < 0.05$. Mutsig2 was used to search genes with significant mutations using a threshold of $p < 0.05$.

Predictive model construction and validation

We used m5C regulator genes to construct a prediction model. The "survival" R package helped us separate HCC patients into high- and low-risk groups, then we identified significant RNA regulator genes through univariate Cox analysis, and visualized through R package "forestplot." The R package "glmnet" was used to perform the LASSO regression analysis (Friedman et al., 2010) on the training cohort, and overfitting was prevented by tenfold cross-validation. Lastly, according to the LASSO regression coefficients, the scoring system was constructed, which prognostic grouping was performed accordingly. With the help of the "survival" package in R, we compared the overall survival of both groups. To evaluate the stability of the model, we performed ROC curves and calculated AUC for different survival times and different clinical traits using the "survival" package. Key genes were obtained by intersecting differentially expressed m5C-related regulators from the TCGA and GEO data set, and prognosis-related genes from our prognostic model. Afterwards, we validated the expression of key genes in different subgroups. Supplementary Figure S1 shows the technology roadmap of the study.

On the basis of risk scores and clinical characteristics, we constructed a nomogram for predicting HCC patients' survival probabilities. Afterwards, the discriminative power of the nomogram was measured by calibration curve and C-index value obtained from bootstrap analysis (1,000 replicates). The interactive nomogram was drawn using the R package "regplot".

GenSet enrichment analysis and gene set variation analysis enrichment analysis

Gene Set Enrichment Analysis (GSEA) allows us to examine the distribution of genes within predefined gene sets in a gene list which arranged according to their phenotype correlation, and thus determine how they contribute to the phenotype (Subramanian et al., 2005). The MSigDB database (http://www.gsea-msigdb.org/gsea/index.jsp) provided "c2.kegg.v7.4.symbols" and "c5.go.v7.4.symbols" gene sets (Liberzon et al., 2015). The R package "clusterprofiler" (Yu et al., 2012) can be used to perform GSEA analysis for those two gene sets in high and low-risk groups, where a p value less than 0.05 qualifies as statistically significant.

Gene Set Variation Analysis (GSVA) is a non-parametric, unsupervised method for evaluating gene set enrichment in transcriptomes. Through the conversion expression matrices of genes into expression matrices of gene sets, it is possible to assess the enriched metabolic pathways in different samples. GSVA analyses on the two gene sets mentioned above in different groups was conduct with "GSVA" package (Hänzelmann et al., 2013) and visualized using the "pheatmap" package.

Immune infiltration in hepatocellular carcinoma

By using gene expression profiles, ESTIMATE R package predicted stromal and immune cell scores, and calculated their numbers for the analysis of HCC tumor purity in this study. We further compared the ESTIMATE scores among cancer and para-cancer groups, and among high and low-risk groups.

Molecular isoform construction

Based on "ConsensusClusterPlus" package (Wilkerson and Hayes, 2010) we clustered cancer and para-cancer samples from TCGA and GEO databases into different groups by m5C regulator genes expression in each sample. The parameters were set to 50 replicates and a resampling rate of 80% (pItem = 0.8). To determine the validity of the groupings, a PCA was carried out, and the results were plotted using the "ggplot2" package. We also analyzed the correlation between prognostic models, molecular subtypes, and clinicopathological features based on TCGA data. Additionally, we examined the correlation between different subgroups and risk scores, and the expression of key genes in different subgroups.

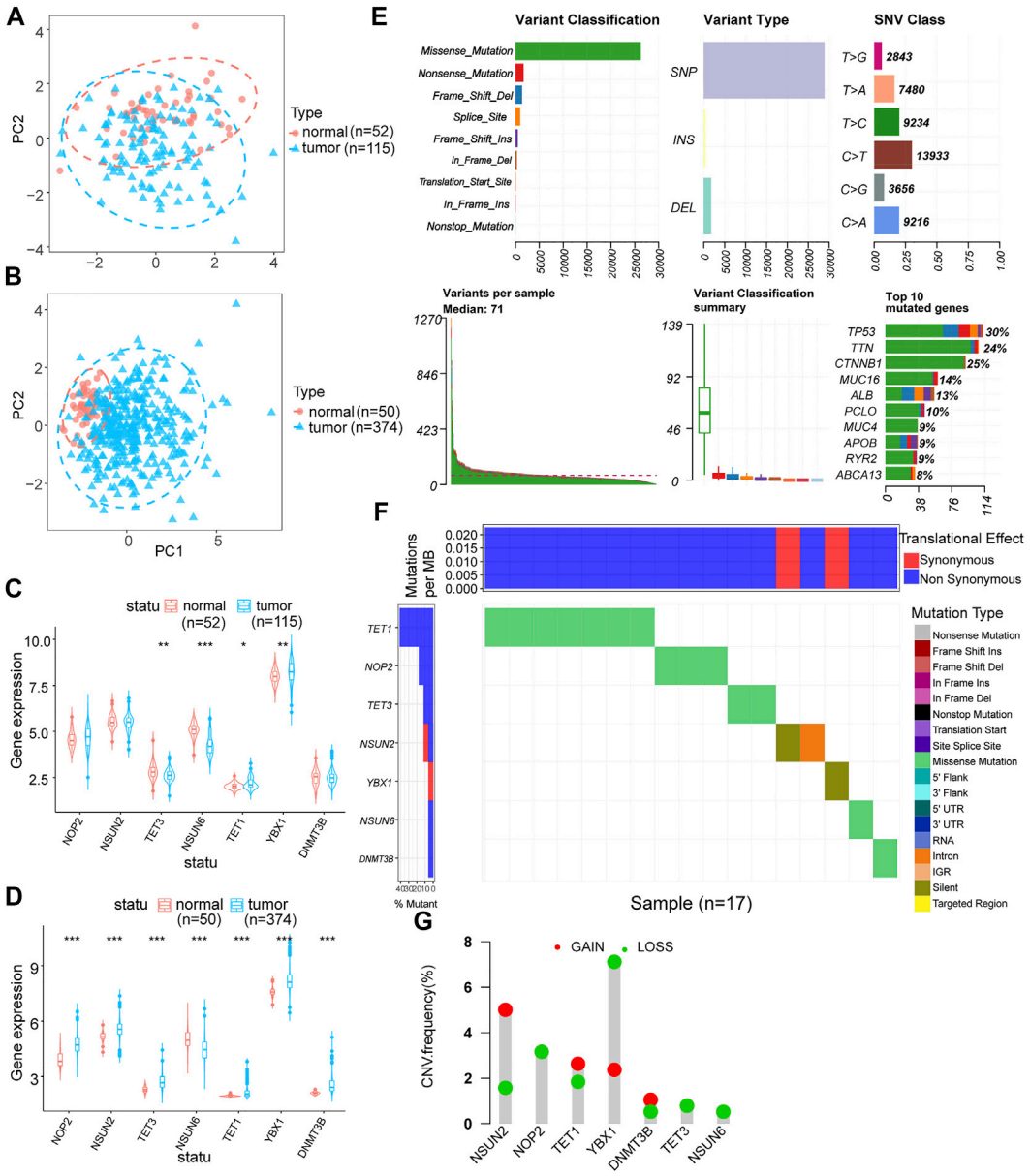


FIGURE 1
m5C regulator genes analysis. **(A,B)**: PCA analysis of GEO and TCGA expression matrices after data correction, blue represents tumor samples (GEO: n = 115, TCGA: n = 374) and red represents control samples (GEO: n = 52, TCGA: n = 50); **(C,D)**: differential expression analysis of m5C regulator genes in GEO and TCGA expression matrices after data correction, blue represents tumor samples (GEO: n = 115, TCGA: n = 374) and red represents control samples (GEO: n = 52, TCGA: n = 50); **(E)**: mutation profile of m5C regulator genes in hepatocellular carcinoma; **(F)**: m5C-related regulators SNV mutation category and frequency; **(G)**: m5C regulator genes CNV amplification and deletion.

Immune infiltration analysis

CIBERSORT is a deconvolution algorithm that utilizes linear support vector regression to evaluate the expression matrices of immunocellular subtypes, and now is increasingly being used for immune infiltration characterization analysis in non-tumor tissues (Ge et al., 2021). Infiltration analysis of immune cells in HCC patients using RNA-Seq data can be an important guide in disease research and treatment prognosis prediction, etc. (Newman et al., 2019). With the CIBERSORT algorithm, this study compared immune cell infiltration levels between different prognostic model subgroups and different molecular subtype groupings, to examine how immune cells infiltration relates to different models.

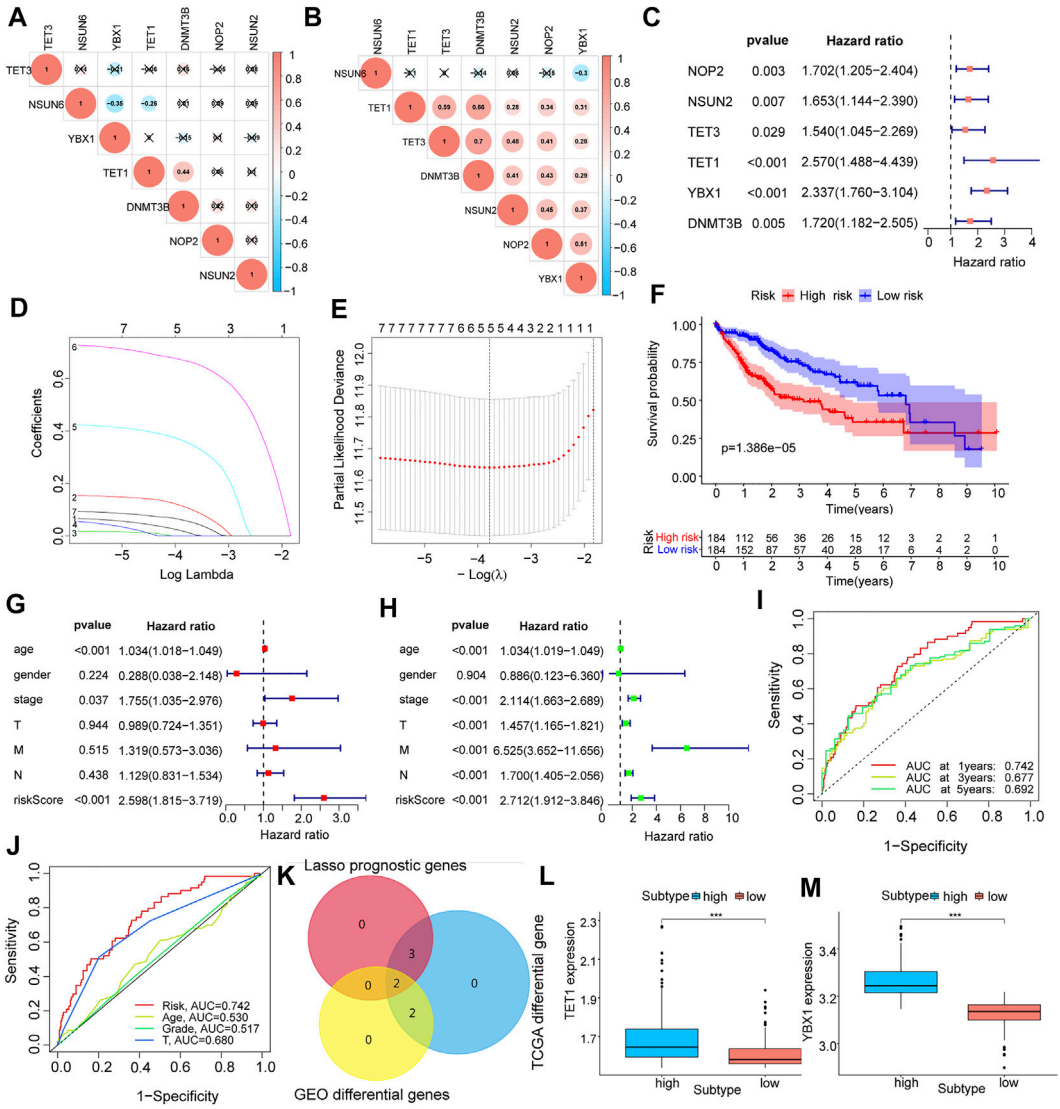


FIGURE 2
Expression characteristics and prognostic model construction of m5C regulator genes in hepatocellular carcinoma. **(A,B)**: m5C regulator genes co-expression analysis in the corrected GEO **(A)** and TCGA **(B)** expression matrices; **(C)**: identify m5C regulator genes associated with prognosis by univariate COX regression analysis, forest plots show the screened genes; **(D,E)**: show the regression coefficients in the LASSO regression algorithm and the cross-validation in the proportional risk model to adjust the parameter, finalize the best parameter(λ) to screen the most relevant genes for hepatocellular carcinoma; **(F)**: survival analysis of different LASSO subgroups; **(G,H)**: multivariate and univariate analysis of risk scores combined with clinical factors such as patient age, gender, and TNM stage; **(I,J)**: AUC analysis of prognostic model and clinical characteristics; **(K)**: Venn diagram mapping of differential genes in GEO and TCGA liver cancer samples and the intersection of genes screened out by LASSO; **(L,M)**: expression of key genes TET1 and YBX1 in each LASSO subgroups.

Molecular isotype construction of immune checkpoint genes

Immune checkpoint genes were obtained from a review [34]. We clustered cancer and para-cancer samples of the TCGA by the expression level of immune checkpoint genes using the R package "ConsensusClusterPlus" with 50 repetitions and a pItem = 0.8. To determine the validity of the grouping. PCA was used to analyze the genes expression levels, and "ggplot2" package visualize the results. The expression of key m5C regulator genes was also assessed in different subgroups. Using correlation analysis, we examined whether key m5C regulator genes play a role in HCC through immune cell infiltration.

TABLE 2 Univariate Cox regression analysis.

Id	HR	HR.95L	HR.95H	*p*-value
NOP2	1.70	1.21	2.40	2.54E-03
NSUN2	1.65	1.14	2.39	7.47E-03
TET3	1.54	1.05	2.27	2.90E-02
NSUN6	0.92	0.71	1.20	5.60E-01
TET1	2.57	1.49	4.44	7.12E-04
YBX1	2.34	1.76	3.10	4.52E-09
DNMT3B	1.72	1.18	2.50	4.63E-03

Statistical analysis

R version 4.0.2 was used for calculations and statistical analysis (https://www.r-project.org). Student's t-tests (normally distributed variables) and Mann-Whitney U-tests (nonnormally distributed variables) were used for the comparison of continuous variables between two groups. All statistical p values all had a two-sided significance with $p < 0.05$.

Results

Expression characteristics of m5C regulator genes in hepatocellular carcinoma

We performed PCA analysis on the corrected datasets from GEO and TCGA, the results suggested a good correction effect (Figures 1A,B). Referring to Cui et al.'s study (Cui et al., 2021; Wang et al., 2021; Liu et al., 2022a), we selected the seven most common m5C regulator genes (NOP2, NSUN2, TET3, NSUN6, TET1, YBX1, and DNMT3B) as the subjects. In the GEO dataset, four of the seven m5C regulator genes (TET3, NSUN6, TET1, and YBX1) were differentially expressed (Figure 1C), while all seven m5C regulator genes had significant differential expression in the TCGA dataset (Figure 1D). Figure 1E lists the overall m5C regulator genes SNP

mutations in HCC samples situation, and Figure 1F shows the mutation types of different m5C regulator genes most closely associated with the development of HCC. We used CNV data from TCGA to identify significantly missing or amplified m5C regulator genes. Among the m5C regulator genes, YBX1 had the highest deletion frequency and the lowest amplification frequency (Figure 1G).

Construction of prognostic model of m5C regulator genes and screening of key m5C regulator genes

Using co-expression analysis (Figures 2A,B) and univariate COX regression analysis (Figure 2C; Table 2), we assessed the effects of m5C regulator genes on HCC tissues. In co-expression analysis, TET1 and DNMT3B showed a significant positive correlation, and regression analysis screened six genes, including NOP2, NSUN2, TET3, TET1, YBX1, and DNMT3B, were associated with HCC. We constructed a LASSO prognostic model containing five genes, including NOP2, NSUN2, TET1, YBX1, and DNMT3B (Figures 2D,E), and a median risk score was used to separate HCC patients into two groups. It was demonstrated that low-risk patients lived significantly longer (Figure 2F). We evaluated COX regressions based on risk scores and clinical traits (age, gender, and TNM stage) using univariate and multivariate models (Figures 2G,H). Using AUC, we validated the LASSO prognostic model, and demonstrated that risk scores were highly predictive for 1-year, 3-years, and 5-years survival (Figures 2I,J). To further screen the key m5C regulator genes, we performed an intersection between DEGs from GEO and TCGA dataset and the key genes identified by LASSO modal, and finally obtained two of them, TET1, and YBX1 (Figure 2K), and it suggested that both two genes were higher expressed in high-risk group (Figures 2L–M). In combination with risk scores and clinical information, a nomogram (Figure 3A) and its calibration curve were constructed (Figure 3B), and we observed that sample's risk scores tended to increase with the progression of T-stage and grade (Figures 3C,D), which is consistent with our previous predictions.

Evaluation of prognostic model for m5C regulator genes

We performed a GSVA analysis of the molecular functions for the different groups classified by the LASSO model. Low-risk group focused on functions relating to platelet dense granule lumen, regulation of fibrinolysis, blood coagulation intrinsic pathway, and protein activation cascade according to GO analysis (Figure 3E; Supplementary Table S1). KEGG analysis revealed it focused on olfactory transduction, nitrogen metabolism, histidine metabolism, serine and threonine metabolism (Figure 3E; Supplementary Table S1). We also performed GSEA analysis (Supplementary Figures S1A–D, Supplementary Table S2). As shown by GO analysis, the high-risk group was related to functions such as actin filament organization, actin polymerization or depolymerization, adaptive immune response, αβT cell activation, and anatomical structure homeostasis (Supplementary Figure S1A), while the low-risk group was linked to functions such as bile acid secretion, drug transmembrane transport, fatty acid β oxidation using acyl-CoA dehydrogenase, negative regulation of triglyceride metabolic process, and neurotransmitter catabolic process (Supplementary Figure S1B). According to KEGG analysis,

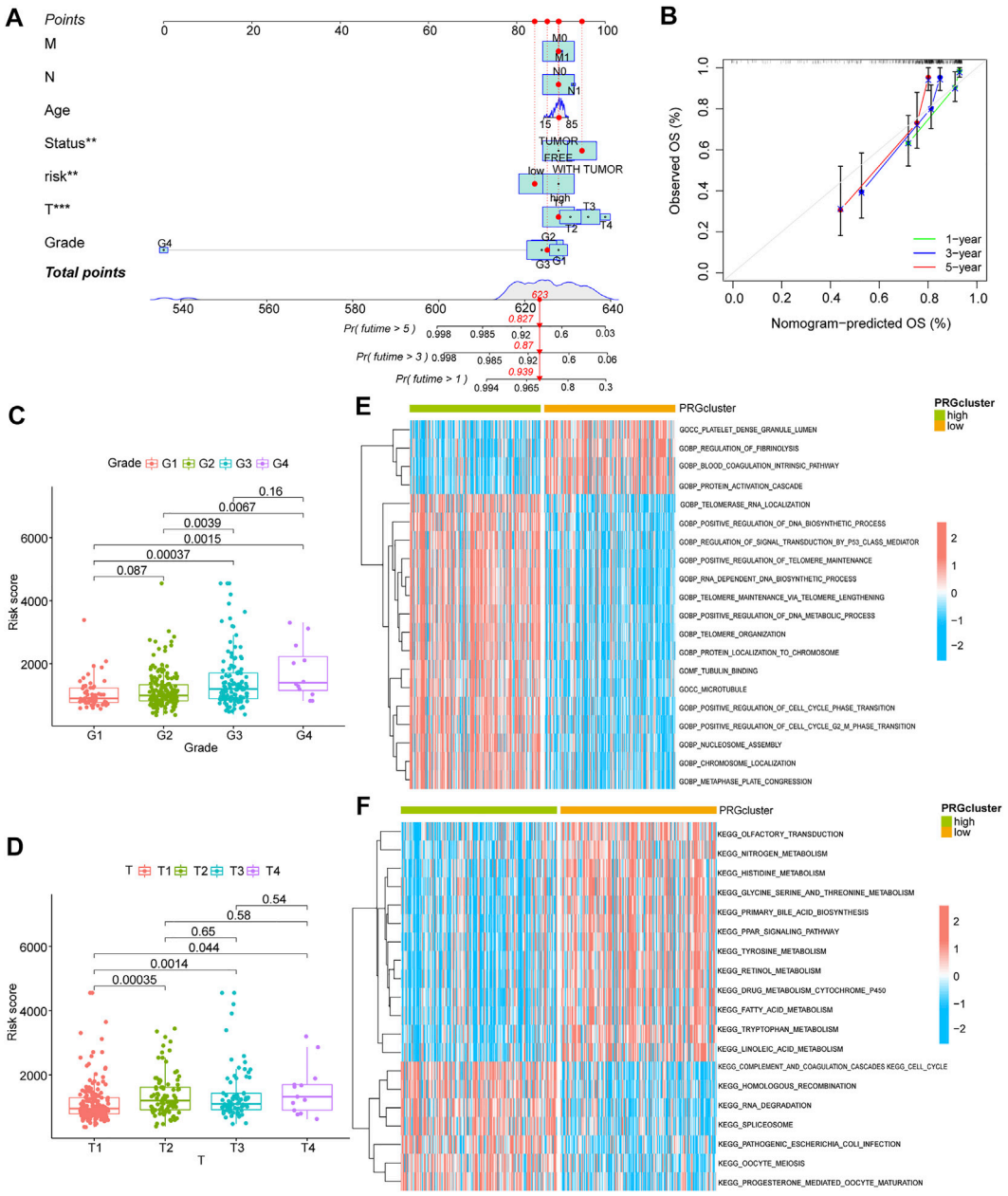


FIGURE 3
Clinical analysis and GSVA analysis of the prognostic model. **(A)** risk score combined with clinical indicators to draw nomogram; **(B)** comparison of predicted survival time and actual survival time using nomogram; **(C,D)**: correlation analysis for G-stage and T-stage, respectively; **(E,F)**: GSVA-GO analysis **(E)** and GSVA-KEGG analysis **(F)** for high and low-risk group.

pathways of high-risk group appeared to be enriched in Chemokine signaling pathway, cell adhesion molecules cams, Cell cycle, spliceosome, and Fc gamma r mediated phagocytosis (Supplementary Figure S1C). For low-risk group, pathways were enriched in beta alanine metabolism, histidine metabolism, linoleic acid metabolism, primary bile acid biosynthesis, and renin angiotensin system (Supplementary Figure S1D). We scored each subgroup using the ESTIMATE algorithm, and found a higher immune score in the high-risk group (Supplementary Figure S1E), but a lower stromal score, immune score, and ESTIMATE total score in the tumor group (Supplementary Figure S1F).

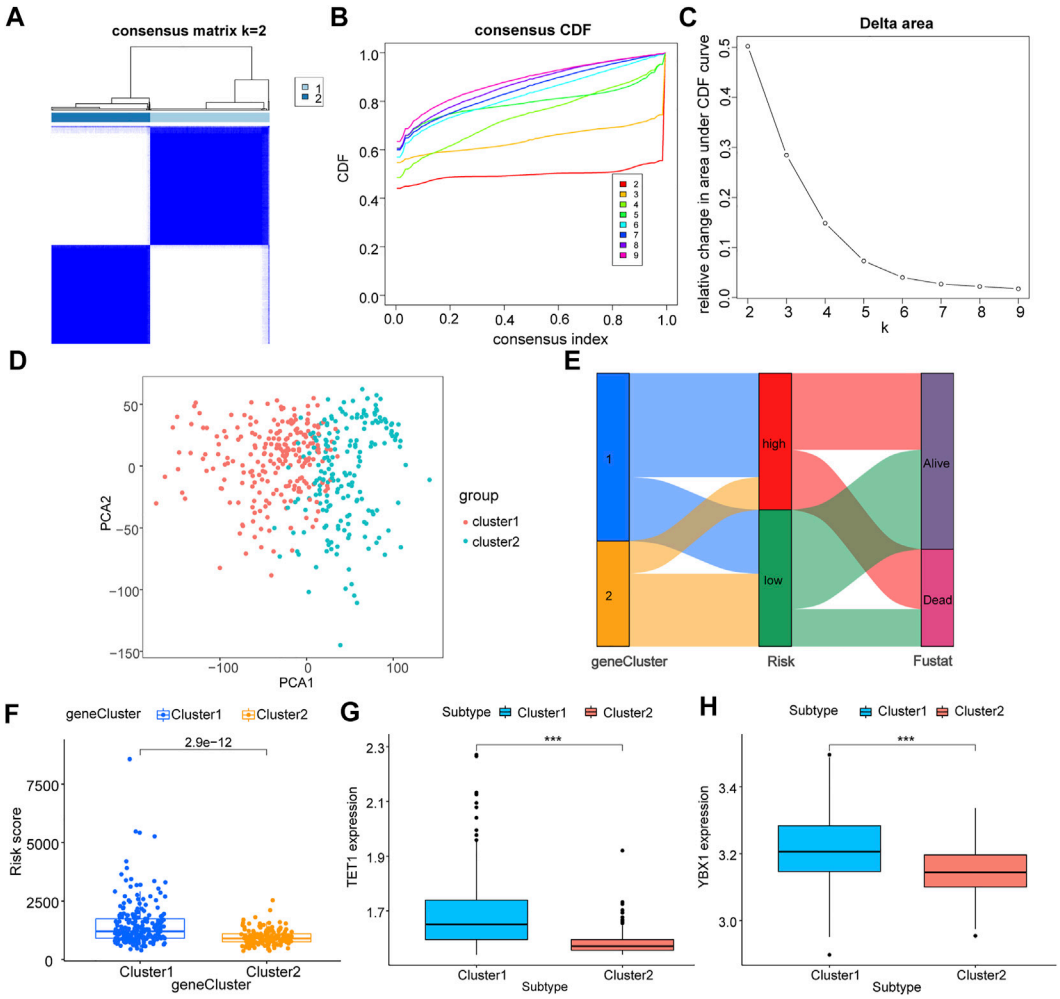


FIGURE 4
Correlation analysis of m5C regulator genes with molecular subtypes of TCGA liver cancer. All samples of TCGA were clustered according to their expression level of m5C regulator genes; **(A)**: sample size after grouping; **(B)**: change in area under the CDF curve (k = 2–9); **(C)**: change of delta area plot when k = 2 to k = 9; D: PCA analysis of cluster1 and cluster2, where cluster 1 is in red and cluster 2 is in blue; **(E)**: Sankey diagram combining survival status and LASSO model grouping; **(F)**: difference in risk scores of different groupings, cluster 1 in blue and cluster 2 in orange; **(G,H)**: differential expression of key m5C regulator genes TET1 **(G)** and YBX1 **(H)** in different groupings, cluster 1 in blue and cluster 2 in red.

Molecular typing of m5C regulator genes and correlation analysis

In an effort to a better understood for the biological characteristics of m5C regulator genes in HCC patients, TCGA samples were clustered according to their expression level. Two subtypes of samples were identified (1: n = 232; 2: n = 192, Figures 4A–C), which PCA result showed high separation quality (Figure 4D), and in combination with the survival information of HCC patients and the grouping information of LASSO model, we constructed a Sankey diagram (Figure 4E). Cluster1 shows a significantly higher risk score compared to cluster2 (Figure 4F), confirming again the previous results. The differential analysis indicated the two key genes, TET1 and YBX1 were significantly higher expressed in cluster1 ($p < 0.05$, Figures 4G,H).

We validated the previous results using the GEO expression matrix and samples were also classified into two isoforms (I: n = 95; 2: II = 72, Figures 5A–C). PCA result showed a higher quality of isolation (Figure 5D), and above two key genes were also present in cluster I with significantly higher expression ($p < 0.05$, Figures 5E,F).

Correlation analysis between m5C regulator genes and immune infiltration

Through CIBERSORT, we calculated the infiltration degree of 22 immune cell types in two groups classified by the LASSO model

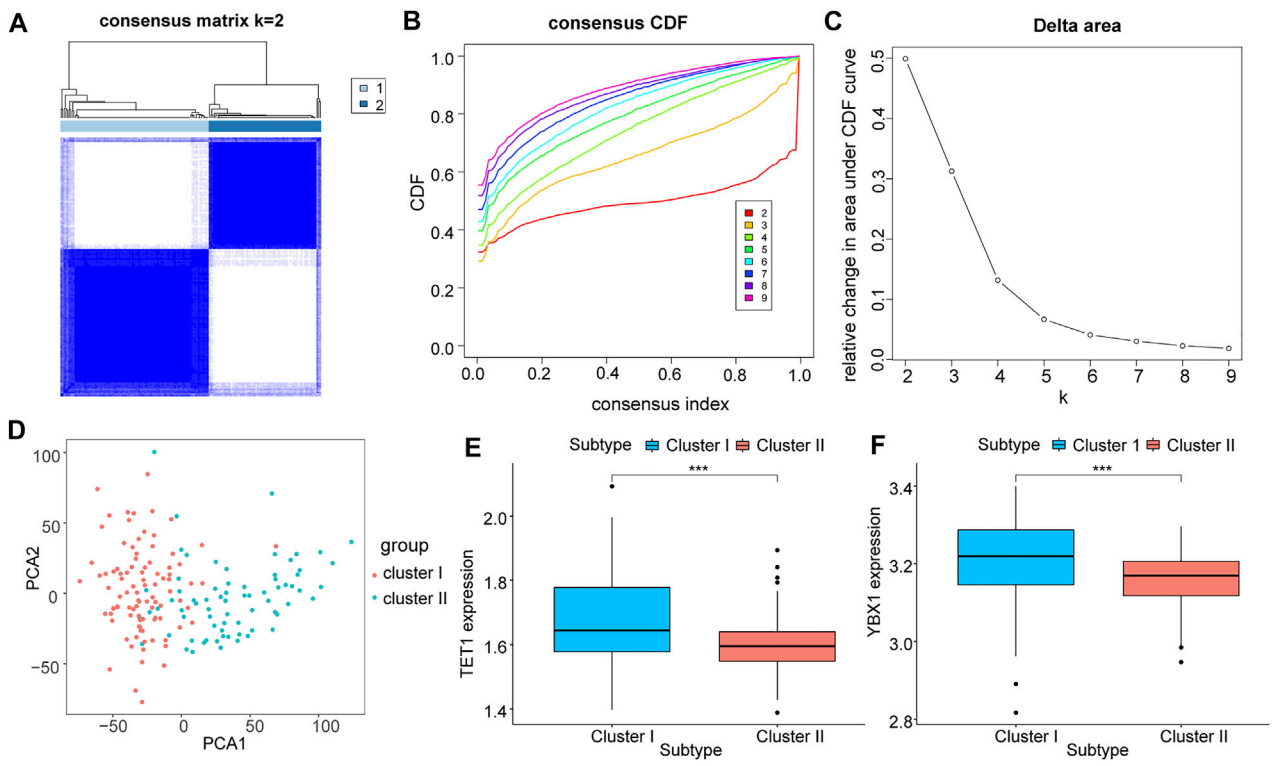


FIGURE 5
Correlation analysis of m5C regulatory genes with molecular subtypes of GEO liver cancer. **(A-C)**: All samples of GEO were classified by the expression level of m5C regulator genes; **(D)**: PCA analysis under different groupings; **(E,F)**: differential expression of key m5C regulator genes TET1 **(E)** and YBX1 **(F)** in different groupings, where cluster I is in blue and cluster II is in red.

to compare their variability of immune infiltration. A significant difference was observed in the infiltration degree of six kinds of immune cells when using the wilcox.test algorithm (Figure 6), namely activated CD4 T cells、resting CD4 T cells, resting NK cells, M0 Macrophages, resting dendritic cells, and resting mast cells. Among them, four immune cells types were $p < 0.001$, one kind was $p < 0.01$, and another kind was $p < 0.05$. Additionally, nine kinds of immune cells showed a difference in their infiltration degree between two subtypes of molecular typing (Supplementary Figure S2), namely activated CD4 T cells, T gamma delta cells, naive B cells, M0 Macrophages, resting CD4 T cells, Monocytes, M2 Macrophages, T follicular helper cells, and Tregs cells, and six of them were $p < 0.001$, one was $p < 0.01$ and two were $p < 0.05$.

Molecular isotype construction of immune checkpoint genes

We used significantly differentially express immune checkpoint genes to conduct hierarchical clustering of all HCC samples again to find out the correlation between these genes and m5C. Among all samples, two subtypes were identified (A: $n = 332$; B: $n = 92$, Figures 7A–C). The PCA result showed a high quality of separation (Figure 7D), and differential analysis showed that TET1 and YBX1 were significantly differentially expressed in different subgroups ($p < 0.01$, Figures 7E,F.

Correlation analysis between key m5C regulator genes and immune cells

Using correlation analysis between key m5C regulator genes and the immune microenvironment, we examined the potential correlation between m5C regulators and immunotherapy efficacy. Combining CIBERSORT results with key m5C regulator genes, we found a positive correlation between TET1 and the infiltration level of various kinds of innate or acquired immune cells, such as M0 macrophages, resting dendritic cells, and T follicular helper cells, while a significant negative correlation with M1 macrophages, M2 macrophages, and resting mast. There was a positive correlation between YBX1 and resting dendritic cells and M0 macrophages, but a negative correlation with Tregs and CD4 T cells (Supplementary Figure S3).

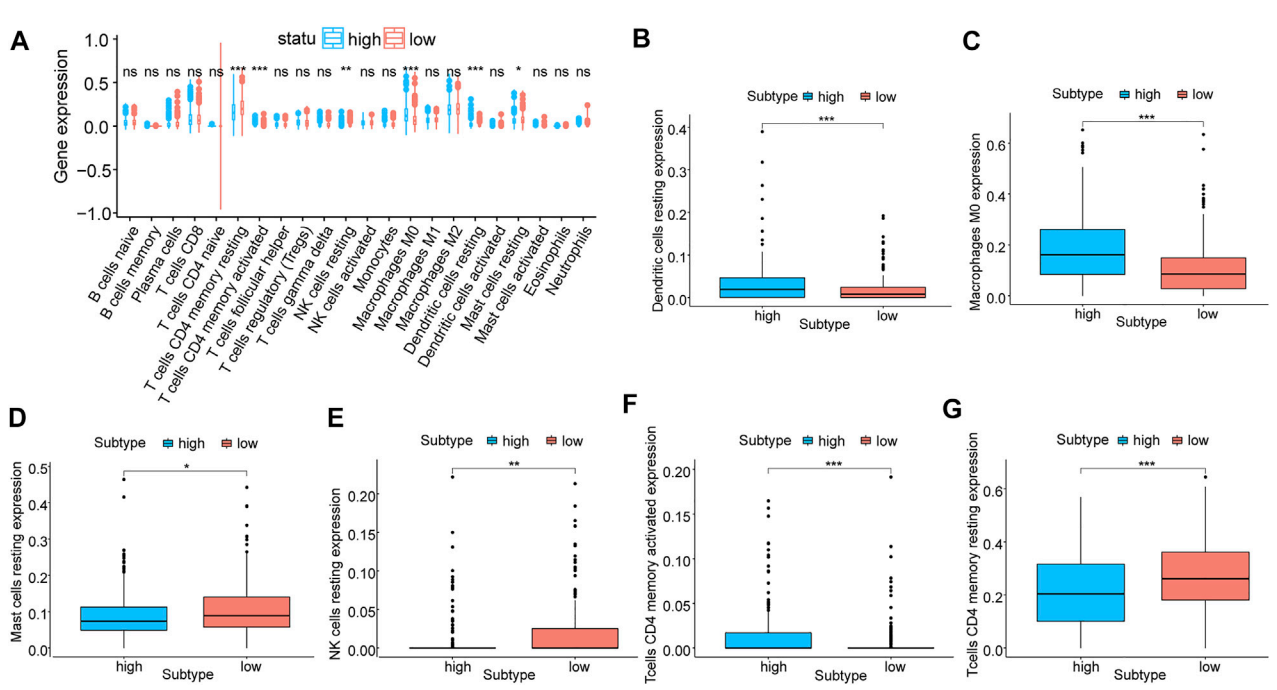


FIGURE 6
Correlation analysis of prognostic models and immune cells. **(A)** Differential analysis of the degree of immune cell infiltration in prognostic model subgroups, blue for high-risk and red for low-risk; **(B–G)**: Differential analysis of the degree of infiltration of six types of immune cells, including resting dendritic cells, M0 Macrophages, resting mast cells, resting NK cells, activated CD4 T cells, and resting CD4 T cells.

Discussion

Ongoing studies have showed that RNA modification contributes to tumorigenesis and tumor progression, and there is growing evidence that m5C regulator genes may serve as potential biomarkers for cancer prediction (Huang et al., 2021a; Huang et al., 2021b; Cui et al., 2021; Xue et al., 2021). It has been suggested that 5 mC methylation influences the development of HCC including clinical stage, progression, and prognosis (Villanueva et al., 2015; Hlady et al., 2019), but the relationship between m5C-related RNA modification and HCC is still poorly understood. In order to test whether these genes can provide prognostic clues for HCC and assist in its initiation and progression, we need to focus on their aberrant expression in HCC. This study confirmed that m5C regulator genes was differentially expressed between HCC and normal samples.

The difference of m5C regulator genes expression levels between tumor and paraneoplastic tissues suggested that these genes may be associated with the carcinogenesis and progression of HCC. MeRIP-seq was used in one study to analyze the m5C modification in tumor and paraneoplastic tissues, and it was found that m5C modification peaks were more abundant and higher in mRNA of HCC tissues, which reconfirmed the relevance of m5C in this disease (Zhang et al., 2020). Aberrant gene methylation is strongly associated with HCC, both in frequency and amount (Nishida et al., 2008).

We constructed a LASSO regression model, which showed satisfactory predictive performance. Similarly, He et al. (2020b) utilized TCGA data developed a two-gene signature of m5C regulators (NSUN4 and ALYREF) with HCC prognostic value based on the LASSO and multivariate Cox regression models. Also demonstrate that the role of m5C related regulators in HCC are dysregulated and associated with patient survival. The methodology we used is largely similar, the major difference being is that we analyzed GEO data combined with the TCGA analysis. In fact, our study proves that utilizing multiple datasets and analytic approaches may identify important gene signatures that would otherwise not be identified using a single dataset/approach. Ultimately this may improve the validity of the findings and be a stronger indication to evaluate these genes in experimental and clinical settings.

For a comprehensive analysis, we performed GSVA and GSEA analyses. “Adaptive immune response” and “cell cycle” et al. are found related to hepatocarcinogenesis and progression. M5C-related RNA modifications impact mRNA translation, transport, and stability, and m5C regulator genes appeared to be associated with “spliceosomes” in this study, suggesting their importance in RNA processing.

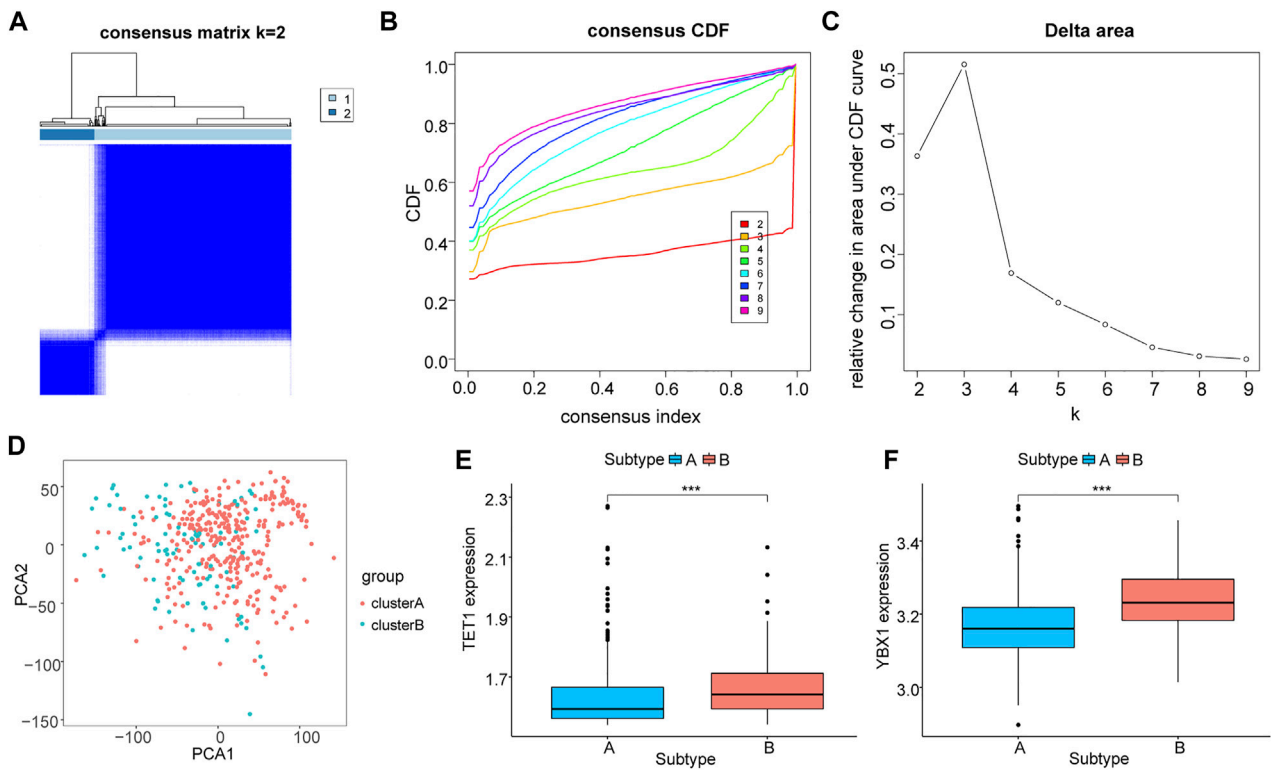


FIGURE 7
Molecular typing based on immune checkpoint genes. **(A–C)**: Cluster grouping based on immune checkpoint genes, **(A)**: sample size after grouping; **(B)**: change in area under the CDF curve (k = 2–9); **(C)**: change of delta area plot when k = 2 to k = 9; **(D)**: PCA analysis of cluster A and B; E-F: differential expression level of TET1 **(E)** and YBX1 **(F)** in different groupings, cluster **(A)** in blue and cluster **(B)** in red.

Tumor cells are the drivers of tumor development, but they can't function alone during tumor progression without the tumor microenvironment (TME). Blood vessels, fibroblasts, immune cells, extracellular matrix, and signaling molecules are all components of the TME which contribute to tumorigenesis and tumor progression. Evidence suggests m5C-related regulators are associated with the tumor immune microenvironment (Geng et al., 2021). Numerous tumors have been studied to correlate tumor immune cell infiltration with clinical outcome (Ishigami et al., 2000; Villegas et al., 2002; Hamanishi et al., 2007; Sharma et al., 2007; Zhu et al., 2009; Mahmoud et al., 2011), however, we do not yet know how m5C modification affects the immune system in HCC. Here, we describe the infiltration characteristics of TME cells in different model groupings and perform immune scoring, which shed light on the molecular mechanism of HCC and new clues for prognosis prediction.

As a result of its aggressiveness, metastasis, and refractoriness, HCC has a high mortality rate and poor prognosis (Ioannou, 2021). While medical technology continues to advance and therapeutic approaches vary, there are still no ideal therapeutic targets or targeted interventions for HCC because its molecular mechanisms of carcinogenesis and development are still unclear (Jiří et al., 2020). It has been shown that azacytidine can reduce cancer cells proliferation by inhibiting m5C modification (Esteller and Pandolfi, 2017), suggesting that reducing m5C modification may contribute to cancer treatment. Ultimately, different RNA epigenetic modifications mediated by regulatory factors provide new idea for finding potential therapeutic targets.

From the perspective of combined multi-omics analysis, we explored the expression profiling of m5C-related genes in HCC, correlation prognostic model construction and evaluation, molecular typing and correlation analysis, immune cell infiltration correlation analysis, immune checkpoint gene molecular subtype construction, and immune cell correlation analysis. Other functions, limited by the length of this study, we really did not study, but we intend to verify other biological functions of m5C through the experimental perspective by doing experiments such as WB, PCR and IHC.

Conclusion

The study combined data from TCGA and GEO for the first time to reveal the genetic and prognostic significance of m5C-related regulators in HCC, which provides new directions for identifying predictive biomarkers and developing molecularly targeted therapies for HCC.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

The work was designed and carried out by XS, and XY. FY conducted the bioinformatics analyses. The manuscript was written and approved final version by all authors.

Funding

National Natural Science Foundation of China (81670596).

Acknowledgments

Publicly available databases including GEO and TCGA are gratefully acknowledged by the authors.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.972043/full#supplementary-material

References

Barbieri, I., and Kouzarides, T. (2020). Role of RNA modifications in cancer. *Nat. Rev. Cancer* 20 (6), 303–322. doi:10.1038/s41568-020-0253-2

Barrett, T., Troup, D., Wilhite, S., Ledoux, P., Rudnev, D., Evangelista, C., et al. (2007). NCBI GEO: Mining tens of millions of expression profiles--database and tools update. *Nucleic Acids Res.* 35, D760–D765. doi:10.1093/nar/gkl887

Bauer, T., Trump, S., Ishaque, N., Thürmann, L., Gu, L., Bauer, M., et al. (2016). Environment-induced epigenetic reprogramming in genomic regulatory elements in smoking mothers and their children. *Mol. Syst. Biol.* 12 (3), 861. doi:10.15252/msb.20156520

Begik, O., Lucas, M. C., Liu, H., Ramirez, J. M., Mattick, J. S., and Novoa, E. M. (2020). Integrative analyses of the RNA modification machinery reveal tissue- and cancer-specific signatures. *Genome Biol.* 21 (1), 97. doi:10.1186/s13059-020-02009-z

Boccaletto, P., Machnicka, M. A., Purta, E., Piatkowski, P., Baginski, B., Wirecki, T. K., et al. (2018). Modomics: A database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* 46 (D1), D303–d307. doi:10.1093/nar/gkx1030

Chellamuthu, A., and Gray, S. G. (2020). The RNA methyltransferase NSUN2 and its potential roles in cancer. *Cells* 9 (8), E1758. doi:10.3390/cells9081758

Cheng, J. X., Chen, L., Li, Y., Cloe, A., Yue, M., Wei, J., et al. (2018). RNA cytosine methylation and methyltransferases mediate chromatin organization and 5-azacytidine response and resistance in leukaemia. *Nat. Commun.* 9 (1), 1163. doi:10.1038/s41467-018-03513-4

Chow, C. S., Lamichhane, T. N., and Mahto, S. K. (2007). Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications. *ACS Chem. Biol.* 2 (9), 610–619. doi:10.1021/cb7001494

Chu, Z., Gu, L., Hu, Y., Zhang, X., Li, M., Chen, J., et al. (2022). STAG2 regulates interferon signaling in melanoma via enhancer loop reprogramming. *Nat. Commun.* 13 (1), 1859. doi:10.1038/s41467-022-29541-9

Cohn, W. E. (1960). Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: Isolation, structure, and chemical characteristics. *J. Biol. Chem.* 235, 1488–1498. doi:10.1016/s0021-9258(18)69432-3

Cui, M., Qu, F., Wang, L., Liu, X., Yu, J., Tang, Z., et al. (2021). m5C RNA methyltransferase-related gene NSUN4 stimulates malignant progression of hepatocellular carcinoma and can be a prognostic marker. *Cancer Biomark.* 33, 389–400. doi:10.3233/cbm-210154

Cui, X. L., Nie, J., Ku, J., Dougherty, U., West-Szymanski, D. C., Collin, F., et al. (2020). A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. *Nat. Commun.* 11 (1), 6161. doi:10.1038/s41467-020-20001-w

de Martel, C., Georges, D., Bray, F., Ferlay, J., and Clifford, G. M. (2020). Global burden of cancer attributable to infections in 2018: A worldwide incidence analysis. *Lancet. Glob. Health* 8 (2), e180–e190. doi:10.1016/s2214-109x(19)30488-7

Esteller, M., and Pandolfi, P. P. (2017). The epitranscriptome of noncoding RNAs in cancer. *Cancer Discov.* 7 (4), 359–368. doi:10.1158/2159-8290.Cd-16-1292

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33 (1), 1–22. doi:10.18637/jss.v033.i01

Gama-Sosa, M. A., Slagel, V. A., Trewyn, R. W., Oxenhandler, R., Kuo, K. C., Gehrke, C. W., et al. (1983). The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res.* 11 (19), 6883–6894. doi:10.1093/nar/11.19.6883

Ge, Y., Chen, Z., Fu, Y., Xiao, X., Xu, H., Shan, L., et al. (2021). Identification and validation of hub genes of synovial tissue for patients with osteoarthritis and rheumatoid arthritis. *Hereditas* 158 (1), 37. doi:10.1186/s41065-021-00201-0

Geng, Q., Wei, Q., Shen, Z., Zheng, Y., Wang, L., Xue, W., et al. (2021). Comprehensive analysis of the prognostic value and immune infiltrates of the three-m5C signature in colon carcinoma. *Cancer Manag. Res.* 13, 7989–8002. doi:10.2147/cmar.S331549

Grinchuk, O., Yenamandra, S., Iyer, R., Singh, M., Lee, H., Lim, K., et al. (2018). Tumor-adjacent tissue co-expression profile analysis reveals pro-oncogenic ribosomal gene signature for prognosis of resectable hepatocellular carcinoma. *Mol. Oncol.* 12 (1), 89–113. doi:10.1002/1878-0261.12153

Hamanishi, J., Mandai, M., Iwasaki, M., Okazaki, T., Tanaka, Y., Yamaguchi, K., et al. (2007). Programmed cell death 1 ligand 1 and tumor-infiltrating CD8+ T lymphocytes are prognostic factors of human ovarian cancer. *Proc. Natl. Acad. Sci. U. S. A.* 104 (9), 3360–3365. doi:10.1073/pnas.0611533104

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). Gsva: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

He, Y., Shi, Q., Zhang, Y., Yuan, X., and Yu, Z. (2020a). Transcriptome-Wide 5-methylcytosine functional profiling of long non-coding RNA in hepatocellular carcinoma. *Cancer Manag. Res.* 12, 6877–6885. doi:10.2147/cmar.S262450

He, Y., Yu, X., Li, J., Zhang, Q., Zheng, Q., and Guo, W. (2020b). Role of m(5)C-related regulatory genes in the diagnosis and prognosis of hepatocellular carcinoma. *Am. J. Transl. Res.* 12 (3), 912–922.

Hlady, R. A., Zhao, X., Pan, X., Yang, J. D., Ahmed, F., Antwi, S. O., et al. (2019). Genome-wide discovery and validation of diagnostic DNA methylation-based biomarkers for hepatocellular cancer detection in circulating cell free DNA. *Theranostics* 9 (24), 7239–7250. doi:10.7150/thno.35573

Huang, T., Chen, W., Liu, J., Gu, N., and Zhang, R. (2019). Genome-wide identification of mRNA 5-methylcytosine in mammals. *Nat. Struct. Mol. Biol.* 26 (5), 380–388. doi:10.1038/s41594-019-0218-x

Huang, Z., Li, J., Chen, J., and Chen, D. (2021a). Construction of prognostic risk model of 5-methylcytosine-related long non-coding RNAs and evaluation of the characteristics of tumor-infiltrating immune cells in breast cancer. *Front. Genet.* 12, 748279. doi:10.3389/fgene.2021.748279

Huang, Z., Pan, J., Wang, H., Du, X., Xu, Y., Wang, Z., et al. (2021b). Prognostic significance and tumor immune microenvironment heterogenicity of m5C RNA methylation regulators in triple-negative breast cancer. *Front. Cell. Dev. Biol.* 9, 657547. doi:10.3389/fcell.2021.657547

Hutter, C., and Zenklusen, J. (2018). The cancer genome Atlas: Creating lasting value beyond its data. *Cell.* 173 (2), 283–285. doi:10.1016/j.cell.2018.03.042

Ioannou, G. N. (2021). Epidemiology and risk-stratification of NAFLD-associated HCC. *J. Hepatol.* 75 (6), 1476–1484. doi:10.1016/j.jhep.2021.08.012

Ishigami, S., Natsugoe, S., Tokuda, K., Nakajo, A., Che, X., Iwashige, H., et al. (2000). Prognostic value of intratumoral natural killer cells in gastric carcinoma. *Cancer* 88 (3), 577–583. doi:10.1002/(sici)1097-0142(20000201)88:3<577:aid-cncr13>3.0.co;2-v

Jiří, T., and Igor, K. (2020). Hepatocellular carcinoma future treatment options. *Klin. Onkol.* 33, 26–29. doi:10.14735/amko20203S26

Leek, J., Johnson, W., Parker, H., Jaffe, A., and Storey, J. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinforma. Oxf. Engl.* 28 (6), 882–883. doi:10.1093/bioinformatics/bts034

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell. Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Liu, T., Zhang, J., Lin, C., Liu, G., Xie, G., Dai, Z., et al. (2022). Molecular characterization clinical and immunotherapeutic characteristics of m5C regulator NOP2 across 33 cancer types. *Front. Cell. Dev. Biol.* 10, 839136. doi:10.3389/fcell.2022.839136

Liu, Y., Zheng, S., Wang, T., Fang, Z., Kong, J., and Liu, J. (2022). Identification of the expression patterns and potential prognostic role of 5-methylcytosine regulators in hepatocellular carcinoma. *Front. Cell. Dev. Biol.* 10, 842220. doi:10.3389/fcell.2022.842220

Mahmoud, S. M., Paish, E. C., Powe, D. G., Macmillan, R. D., Grainge, M. J., Lee, A. H., et al. (2011). Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *J. Clin. Oncol.* 29 (15), 1949–1955. doi:10.1200/jco.2010.30.5037

Mermel, C., Schumacher, S., Hill, B., Meyerson, M., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 12 (4), R41. doi:10.1186/gb-2011-12-4-r41

Navarro, I. C., Tuorto, F., Jordan, D., Legrand, C., Price, J., Braukmann, F., et al. (2021). Translational adaptation to heat stress is mediated by RNA 5-methylcytosine in *Caenorhabditis elegans*. *Embo J.* 40 (6), e105496. doi:10.15252/embj.2020105496

Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* 37 (7), 773–782. doi:10.1038/s41587-019-0114-2

Nishida, N., Nagasaka, T., Nishimura, T., Ikai, I., Boland, C. R., and Goel, A. (2008). Aberrant methylation of multiple tumor suppressor genes in aging liver, chronic hepatitis, and hepatocellular carcinoma. *Hepatology* 47 (3), 908–918. doi:10.1002/hep.22110

Roundtree, I. A., Evans, M. E., Pan, T., and He, C. (2017). Dynamic RNA modifications in gene expression regulation. *Cell.* 169 (7), 1187–1200. doi:10.1016/j.cell.2017.05.045

Sharma, P., Shen, Y., Wen, S., Yamada, S., Jungbluth, A. A., Gnjatic, S., et al. (2007). CD8 tumor-infiltrating lymphocytes are predictive of survival in muscle-invasive urothelial carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* 104 (10), 3967–3972. doi:10.1073/pnas.0611618104

Shi, H., Chai, P., Jia, R., and Fan, X. (2020). Novel insight into the regulatory roles of diverse RNA modifications: Re-Defining the bridge between transcription and translation. *Mol. Cancer* 19 (1), 78. doi:10.1186/s12943-020-01194-6

Squires, J. E., Patel, H. R., Nousch, M., Sibbritt, T., Humphreys, D. T., Parker, B. J., et al. (2012). Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.* 40 (11), 5023–5033. doi:10.1093/nar/gks144

Subramanian, A., Tamayo, P., Mootha, V., Mukherjee, S., Ebert, B., Gillette, M., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Sun, Z., Xue, S., Zhang, M., Xu, H., Hu, X., Chen, S., et al. (2020). Aberrant NSUN2-mediated m(5)C modification of H19 lncRNA is associated with poor differentiation of hepatocellular carcinoma. *Oncogene* 39 (45), 6906–6919. doi:10.1038/s41388-020-01475-w

Trixl, L., and Lusser, A. (2019). The dynamic RNA modification 5-methylcytosine and its emerging role as an epitranscriptomic mark. *Wiley Interdiscip. Rev. RNA* 10 (1), e1510. doi:10.1002/wrna.1510

Villanueva, A., Portela, A., Sayols, S., Battiston, C., Hoshida, Y., Méndez-González, J., et al. (2015). DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. *Hepatology* 61 (6), 1945–1956. doi:10.1002/hep.27732

Villegas, F. R., Coca, S., Villarrubia, V. G., Jiménez, R., Chillón, M. J., Jareño, J., et al. (2002). Prognostic significance of tumor infiltrating natural killer cells subset CD57 in patients with squamous cell lung cancer. *Lung Cancer* 35 (1), 23–28. doi:10.1016/s0169-5002(01)00292-6

Walworth, N. G., Lee, M. D., Dolzhenko, E., Fu, F. X., Smith, A. D., Webb, E. A., et al. (2021). Long-term m5C methylome dynamics parallel phenotypic adaptation in the cyanobacterium trichodesmium. *Mol. Biol. Evol.* 38 (3), 927–939. doi:10.1093/molbev/msaa256

Wang, Q., Gu, L., Adey, A., Radlwimmer, B., Wang, W., Hovestadt, V., et al. (2013). Tagmentation-based whole-genome bisulfite sequencing. *Nat. Protoc.* 8 (10), 2022–2032. doi:10.1038/nprot.2013.118

Wang, X., Han, Y., Li, J., Hong, D., Xue, Z., Huang, H., et al. (2021). Multi-omics analysis of copy number variations of RNA regulatory genes in soft tissue sarcoma. *Life Sci.* 265, 118734. doi:10.1016/j.lfs.2020.118734

Wilkerson, M., and Hayes, D. (2010). ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinforma. Oxf. Engl.* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Xue, C., Zhao, Y., Li, G., and Li, L. (2021). Multi-omic analyses of the m(5)C regulator ALYREF reveal its essential roles in hepatocellular carcinoma. *Front. Oncol.* 11, 633415. doi:10.3389/fonc.2021.633415

Yang, T., Low, J. J. A., and Woon, E. C. Y. (2020). A general strategy exploiting m5C duplex-remodelling effect for selective detection of RNA and DNA m5C methyltransferase activity in cells. *Nucleic Acids Res.* 48 (1), e5. doi:10.1093/nar/gkz1047

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhang, Q., Zheng, Q., Yu, X., He, Y., and Guo, W. (2020). Overview of distinct 5-methylcytosine profiles of messenger RNA in human hepatocellular carcinoma and paired adjacent non-tumor tissues. *J. Transl. Med.* 18 (1), 245. doi:10.1186/s12967-020-02417-6

Zhu, L. Y., Zhou, J., Liu, Y. Z., and Pan, W. D. (2009). Prognostic significance of natural killer cell infiltration in hepatocellular carcinoma. *Ai Zheng* 28 (11), 1198–1202. doi:10.5732/cjc.009.10284

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 21 October 2022
DOI 10.3389/fgene.2022.991706

OPEN ACCESS

EDITED BY
Yilin Zhang,
The University of Chicago, United States

REVIEWED BY
Taobo Hu,
Peking University People's Hospital, China
Shengchun Liu,
First Affiliated Hospital of Chongqing Medical University, China
Peter Hart,
Roosevelt University College of Pharmacy, United States

*CORRESPONDENCE
Adriana Aguilar-Lemarroy,
adry.aguilar.lemarroy@gmail.com
Luis F. Jave-Suárez,
lfjave@live.com.mx

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 11 July 2022
ACCEPTED 11 October 2022
PUBLISHED 21 October 2022

CITATION
Barrón-Gallardo CA, Garcia-Chagollán M, Morán-Mendoza AJ, Delgadillo-Cristerna R, Martínez-Silva MG, Villaseñor-García MM, Aguilar-Lemarroy A and Jave-Suárez LF (2022), A gene expression signature in HER2+ breast cancer patients related to neoadjuvant chemotherapy resistance, overall survival, and disease-free survival.
Front. Genet. 13:991706.
doi: 10.3389/fgene.2022.991706

A gene expression signature in HER2+ breast cancer patients related to neoadjuvant chemotherapy resistance, overall survival, and disease-free survival

Carlos A. Barrón-Gallardo[1†], Mariel Garcia-Chagollán[2†], Andres J. Morán-Mendoza[3], Raul Delgadillo-Cristerna[4], María G. Martínez-Silva[5], María M. Villaseñor-García[6], Adriana Aguilar-Lemarroy[6]* and Luis F. Jave-Suárez[6]*

[1]Programa de Doctorado en Ciencias Biomédicas, Centro Universitario de Ciencias de La Salud, Universidad de Guadalajara, Guadalajara, Mexico, [2]Instituto de Investigación en Ciencias Biomédicas (IICB), Centro Universitario de Ciencias de La Salud, Universidad de Guadalajara, Guadalajara, Mexico, [3]Hospital de Ginecología, Centro Médico Nacional de Occidente, Instituto Mexicano Del Seguro Social (IMSS), Guadalajara, Mexico, [4]Departamento de Radiología e Imagen, Centro Médico Nacional de Occidente, Instituto Mexicano Del Seguro Social (IMSS), Guadalajara, Mexico, [5]Departamento de Anatomía Patológica, Centro Médico Nacional de Occidente, Instituto Mexicano Del Seguro Social (IMSS), Guadalajara, Mexico, [6]División de Inmunología, Centro de Investigación Biomédica de Occidente, Instituto Mexicano Del Seguro Social (IMSS), Guadalajara, Mexico

Breast cancer ranks first in terms of mortality and incidence rates worldwide among women. The HER2+ molecular subtype is one of the most aggressive subtypes; its treatment includes neoadjuvant chemotherapy and the use of a HER2 antibody. Some patients develop resistance despite positive results obtained using this therapeutic strategy. Objective. To identify prognostic markers for treatment and survival in HER2+ patients. Methods. Patients treated with neoadjuvant chemotherapy were assigned to sensitive and resistant groups based on their treatment response. Differentially expressed genes (DEGs) were identified using RNA-seq analysis. KEGG pathway, gene ontology, and interactome analyses were performed for all DEGs. An enrichment analysis Gene set enrichment analysis was performed. All DEGs were analyzed for overall (OS) and disease-free survival (DFS). Results. A total of 94 DEGs were related to treatment resistance. Survival analysis showed that 12 genes (ATF6B, DHRS13, DIRAS1, ERAL1, GRIN2B, L1CAM, IRX3, PRTFDC1, PBX2, S100B, SLC9A3R2, and TNXB) were good predictors of disease-free survival, and eight genes (GNG4, IL22RA2, MICA, S100B, SERPINF2, HLA-A, DIRAS1, and TNXB) were good predictors of overall survival (OS). Conclusion: We highlighted a molecular expression signature that can differentiate the treatment response, overall survival, and DFS of patients with HER2+ breast cancer.

KEYWORDS

breast cancer, neoadjuvant chemotherapy, RNA-seq, biomarkers, bioinformatics, overall survival, disease free survival

Introduction

Breast cancer is a heterogeneous disease characterized by abnormal and uncontrolled growth of malignant breast cells. Among all types of cancer, this disease ranks first in mortality and incidence rates in women over 25 years of age worldwide (Sung et al., 2021). In 2000, Perou et al. reported different molecular expression patterns in patients with breast cancer, and these patterns were subsequently used to classify breast cancer into distinct molecular subtypes (Perou et al., 2000; Sorlie et al., 2003). According to this classification, cancer cells that express human epidermal growth factor 2 (ERBB, formerly HER2) and not estrogen receptors (ER) are identified as the HER2+ molecular subtype, which represents 15%–30% of breast cancer patients, is an aggressive phenotype, and a predictor of poor outcome (Ban et al., 2020).

The treatment of HER2+ breast cancer includes the administration of chemotherapy and trastuzumab, a monoclonal antibody against the HER2 receptor (Abal et al., 2003; Harbeck and Gnant, 2017; Waks and Winer, 2019). Conventional neoadjuvant chemotherapy involves anthracyclines followed by taxane application. Anthracyclines work by joining DNA and suppressing the binding of DNA polymerase, thereby preventing DNA replication (McGowan et al., 2017). Taxanes affect mitotic spindle formation by binding to tubulin dimers, thereby preventing the division of tumor cells (Yardley, 2013; Harbeck and Gnant, 2017). Furthermore, adding trastuzumab in conventional chemotherapy helps block HER2 receptor-induced cell growth signaling (Maximiano et al., 2016). Despite the positive results obtained with this therapeutic strategy, some patients develop resistance. The molecular mechanisms underlying resistance are not fully understood; therefore, there is a lack of predictive biomarkers that are helpful in the prognosis and prediction of chemotherapy response (Iwamoto et al., 2020).

This study aimed to evaluate the transcriptome of HER2+ breast cancer patients and, according to their response to chemotherapy (sensitivity or resistance), to identify differentially expressed genes (DEGs) that could be useful in predicting patient outcomes after neoadjuvant chemotherapy treatment.

Materials and methods

Sample selection, chemotherapy treatment, and study design

Patients aged 18 years and older with a diagnosis of breast cancer, HER2+/PR-/ER-, tumor size >2 cm, and positive nodes, candidates to receive neoadjuvant chemotherapy, and without previous therapy against cancer were recruited for this study. Patients with metastatic cancer, those with insufficient breast cancer biopsy tissue for pathological analysis, or those with RNA extraction were excluded. All participants provided written informed consent prior to enrolment. The data were deposited in the Gene Expression Omnibus (GEO) repository under the number GSE162187. Samples were separated by pathologic response into two groups: pathological complete response (pCR) was considered the sensitive group, and those in the non-pCR group were considered the resistant group.

Additionally, we used and analyzed data from the GSE163882 study, and HER2+/PR-/ER-samples were selected. The results obtained from both databases were compared.

Finally, from the TCGA breast ductal carcinoma database, HER+/PR-/ER-breast cancer samples were selected for analysis of overall survival.

Ethics and informed consent statements

The study was conducted in accordance with the guidelines of the Declaration of Helsinki and the ethical standards of the institutional and/or national research committee. This study was approved by the Ethical and Research Committee of the Instituto Mexicano del Seguro Social (IMSS) (number R-2013-785-061). Informed consent was obtained from all subjects involved in the study.

Quality control, alignment, and differential expression

The FASTQ files were analyzed with the Flexbar software tool version 3.5.0 (https://github.com/seqan/flexbar/releases/tag/v3.5.0) (Dodt et al., 2012; Roehr et al., 2017) to remove Illumina adapters and to filter reads by a Phred score >30. To quantify the RNA-seq data, a pseudo-alignment was performed using Kallisto software version 0.46.1 (https://pachterlab.github.io/kallisto/download.html) (Bray et al., 2016) with the default parameters and the GRCh38 human genome reference (GRCh38. p12). The DESeq2 package version 1.28.1 (https://bioconductor.riken.jp/packages/3.0/bioc/html/DESeq2.html) (Love et al., 2014) was used for the analysis of abundance tables and the identification of differentially expressed genes (DEGs) for comparing resistant and sensitive samples (set as the reference group). The Ensembl database was used for the annotation of genes. To decrease the false discovery rate, the Benjamini–Hochberg correction test was applied to obtain adjusted p-values.

Enrichment and interaction analysis

We selected all DEGs ($p < 0.05$) obtained from GSE162187 for analysis with the KEGG Mapper (https://www.

genome.jp/kegg/tool/map_pathway1.html) (Kanehisa and Sato, 2020) and the DAVID v6.8 web tools (https://david.ncifcrf.gov/home.jsp) to identify pathways implicated in treatment response.

The Panther database v.16.0 (Mi et al., 2021) web tool was used for Gene Ontology enrichment analysis using Fisher's exact test and false discovery rate (FDR), with a threshold of $p < 0.05$, which was considered to be significant for each of the three categories, that is, molecular function, cellular component, and biological process.

Gene set enrichment analysis (GSEA) was performed using the pre-ranked DEGs list. GSEA software v4.2.2 was used for analysis (Subramanian et al., 2005). A molecular signature database (MsigDB v7.4) was used, taking the nine collections (C1:C9, and H) for enrichment analysis (Subramanian et al., 2005; Liberzon et al., 2011; Liberzon et al., 2015).

To perform an interactome analysis, the DEGs were filtered by adjusted p-value <0.05 and analyzed using STRING-DB v11.0 software (https://string-db.org/) (Szklarczyk et al., 2019), the confidence score was set up at 0.7 to represent protein–protein associations.

Principal component analysis and heatmap representation

The geometric mean of the counts for each gene was used as a normalization factor. Once normalized, the principal component analysis (PCA) and heatmap representation were performed with the prcomp package and heatmap functions, respectively, with the default parameters in R v.4.0.2 ("Taking off Again") using as variables the normalized counts of the DEGs with an adjusted p-value of <0.05.

Survival analysis

Furthermore, a database of 109 patients obtained from the TCGA breast ductal carcinoma study with HER+/ER-/PR-was analyzed (TCGA Research Network: https://www.cancer.gov/tcga) at 60 and 120 months to analyze overall survival (OS) and disease-free survival (DFS). Gene expression levels were determined according to normalized Log2-read counts for each gene. Median and quartiles were used for determining high- and low-expression groups. DEGs with an adjusted p-value < 0.05 were analyzed. Statistical significance was set at $p < 0.05$.

Results

A previous study was conducted to determine biomarkers of response to neoadjuvant chemotherapy in patients with breast cancer (GSE162187) (Barron-Gallardo et al., 2022). In this study, HER2+ samples were taken and included in the RNA-seq analysis; five samples were from patients categorized as resistant to treatment, and three samples were from patients sensitive to treatment. This small subset was used as training data. The results were validated using the GSE163882 dataset, which included information from 222 patients with breast cancer. Patients were over 33 years old; the mean ages for the resistant and sensitive groups were 52.2 (±12.15) and 62 (±6.24) years, respectively, with no statistical differences. The diagnostic status of all the patients was invasive ductal carcinoma breast cancer. The histological grades for tumor biopsies based on SBR (Scarff–Bloom–Richardson) parameters were five SBRII, two SBRIII, and one with non-available information (Supplementary Table S1).

The transcriptomic pattern was studied to determine the variables that specifically discriminated HER2+ patients according to their neoadjuvant treatment response. Despite having two groups defined by their pathological response, principal component analysis (PCA) with all genes detected by RNA-seq showed that the samples did not form specific clusters. Moreover, the distribution of the samples followed a heterogeneous pattern, indicating that HER2+ breast cancer patients may have high variability in gene expression (Supplementary Figure S1A).

Determination of DEGs related to treatment resistance in the training data

HER2+ patients were grouped into sensitive and resistant to neoadjuvant chemotherapy groups to obtain DEGs related to treatment response. The transcriptomes of both groups were compared, and a total of 1383 genes were observed to be differentially expressed ($p < 0.05$), of which 719 were sub-expressed and 664 were overexpressed in the resistant group as compared with the sensitive group (Figure 1A). To diminish the inclusion of false-positive DEGs, the Benjamini–Hochberg post hoc test was applied; among the 1383 DEGs, only 94 maintained statistical significance with an adjusted p-value (p-adj) <0.05 (45 subexpressed and 49 overexpressed genes) (Figure 1B).

Thereafter, we investigated whether this set of 94 DEGs could adequately classify the samples as sensitive and resistant to treatment. Therefore, principal component analysis (PCA) with only 94 genes was performed again. The results showed two clusters defined by principal component 1 (PC1) with 55.35% and principal component 2 (PC2) with 15.79% of the data variance (Supplementary Figure S1B). Therefore, the selection of the 94 DEGs included genes capable of clustering HER2+ patients into resistant and sensitive groups. As shown in Supplementary Table S2, from the 94 DEGs, the top 10 overexpressed genes were HLA-DQA1, TRIM26, IGHJ6, AGPAT1, IGHV1-69-2, PBX2, HLA-DRB1, PRRC2A,

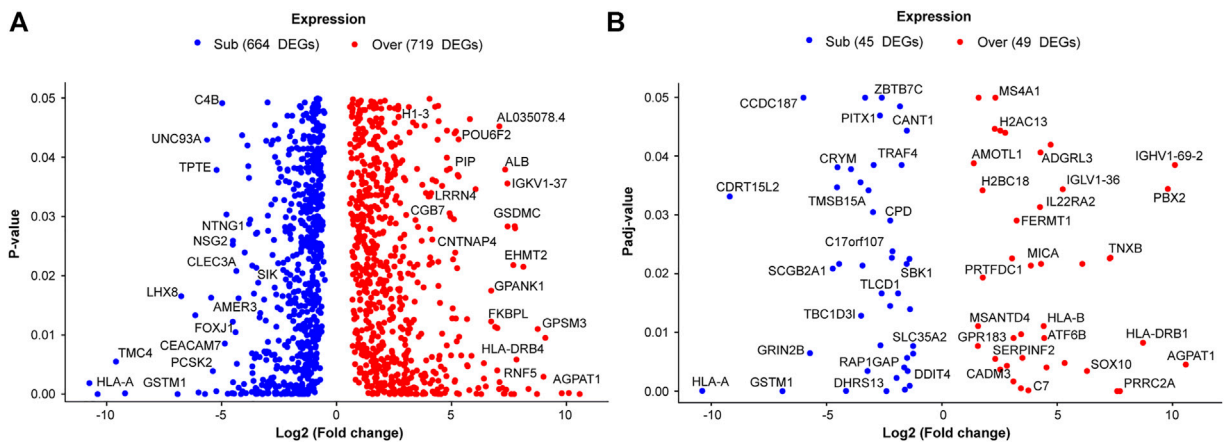


FIGURE 1
Differentially expressed genes (DEGs) related to neoadjuvant chemotherapy outcomes in HER2+ patients with breast cancer. Each point represents every DEG in the resistant group compared with the sensitive. Overexpressed genes are colored red, subexpressed genes are represented by blue points. The *x*-axis shows the value of the log2 (fold change), and the *y*-axis represents **(A)** the *p*-value and **(B)** the adjusted *p*-value.

LRRC37A3, and TNXB, and the top 10 underexpressed genes were HLA-A, CDRT15L2, GSTM1, CCDC187, GRIN2B, SCGB2A1, GNG4, SBSN, CRISP3, and ZG16B.

To visualize the expression pattern (color density) and distribution (clustering) of the 94 DEGs, heatmap analysis was performed (Figure 2). The column dendrogram results showed two clusters belonging to the sensitive and resistant groups. The row dendrogram shows four clusters of genes with similar expression patterns.

Pathways and enrichment analysis of DEGs related to chemotherapy resistance

The 94 DEGs were analyzed using the KEGG Mapper search pathway tool and DAVID v6.8. Among the 318 KEGG pathways, seven were statistically modulated (FDR <0.05) (Figure 3A), including graft-versus-host disease, allograft rejection, type I diabetes mellitus, autoimmune thyroid disease, viral myocarditis, antigen processing and presentation, and cell adhesion molecules. In addition, GO analysis results showed that the biological processes enriched by DEGs were related to the interferon-gamma-mediated signaling pathway. The cellular components in which the DEGs were included were associated with the MHC class II protein complex, luminal side of the endoplasmic reticulum membrane, endoplasmic reticulum (ER)-to-Golgi transport vesicle membrane, and extracellular space. Finally, the modulated molecular functions were MHC class II receptor activity and peptide-antigen binding (Figure 3B).

Enrichment analysis showed a total of 40 gene sets enriched with a p-value < 0.05 (35 positively and five negatively), which belongs to C1 (1 enriched set), C2 (6 enriched sets), C3 (2 enriched sets), C5 (19 enriched sets), C7 (7 enriched sets), and C8 collections (5 enriched sets). The C4, C6, and H collections did not contain enriched sets. From the enriched sets, we found two related to therapy resistance (Massarweh tamoxifen resistance and Creighton endocrine therapy resistance gene sets) and three related to the immune system, such as GOBP immune response, Goldrath antigen response, and CHR6P21, which is a location for genes related to the immune system (HLA-DQA1, HLA-DRB1, HLA-B, and MICA) (Supplementary Table S2).

Determination of interactions clusters between DEGs

An analysis of 94 DEGs was performed to determine the molecular interactions between them. The STRING-DB tool was used to set an interaction score with high confidence (0.7). The results showed 18 edges (genes) distributed among seven clusters: one with five genes, one with three genes, and five with two genes. Among the seven clusters, the cluster with five edges was related to the interferon-gamma-mediated signaling pathway, MHC class I/II-like antigen recognition protein, and cell adhesion molecules, including the HLA-A, HLA-DQA1, TRIM26, HLA-B, and HLA-DRB1 genes (Figure 3C).

Evaluation of DEGs for survival prediction

To evaluate whether the expression of the 94 DEGs was related to survival prediction, measured as DFS or OS, we analyzed the 94 DEGs individually using a database with

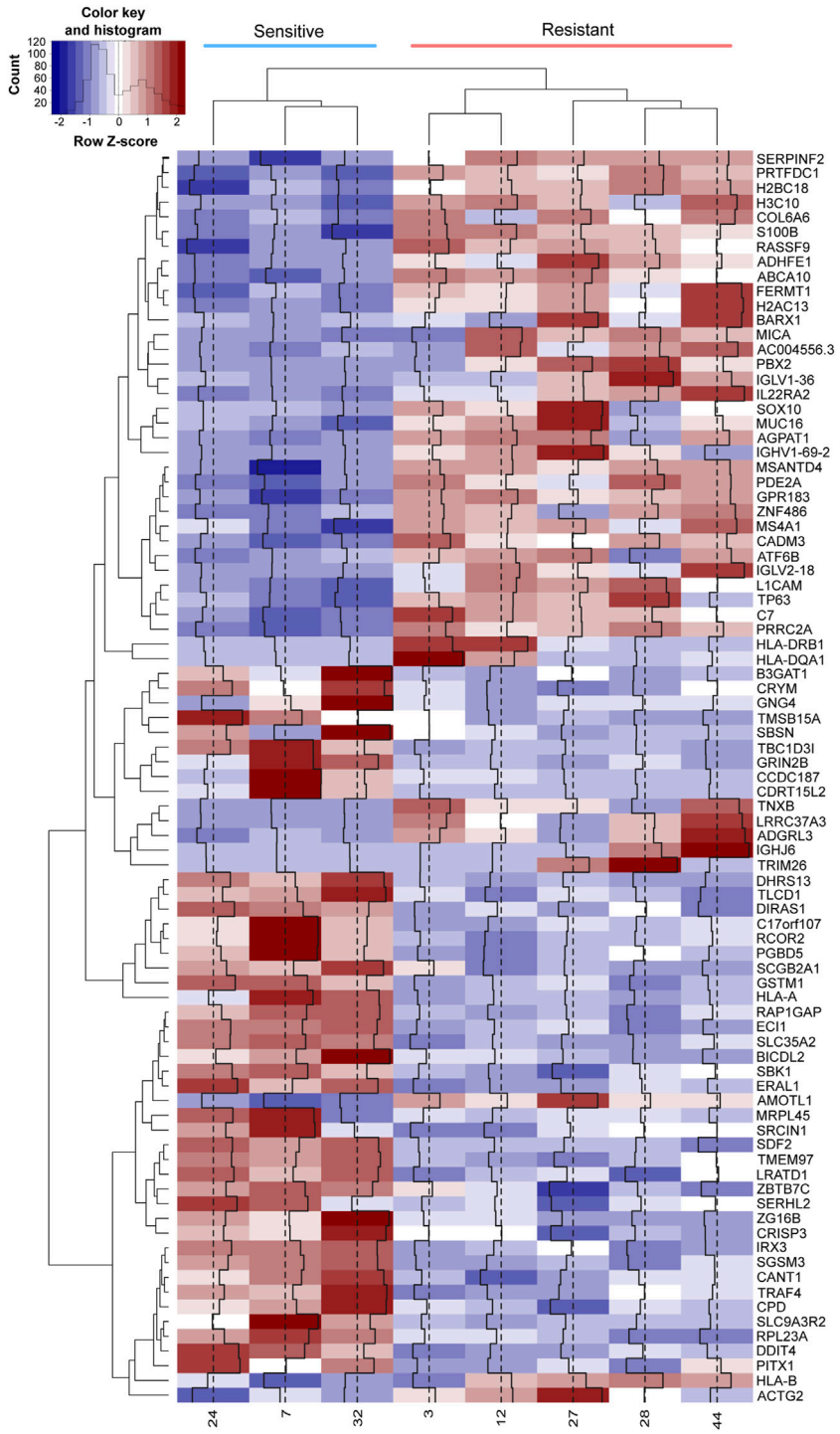


FIGURE 2
Expression patterns of DEGs in resistance and sensitivity. The row Z-score of the normalized read counts of DEGs with p-adj < 0.05 are plotted in the heatmap. The red color indicates a row Z-score >0, and the blue indicates a row Z-score <0. Columns represent each patient, and each row represents a gene. The dendrogram at the top of the heatmap clusters the patients according to their gene expression pattern, while the dendrogram at the left side of the heatmap groups the genes with similar expression patterns. Columns 1 to 3 represent sensitive patients, and columns 4 to 8 represent resistant patients.

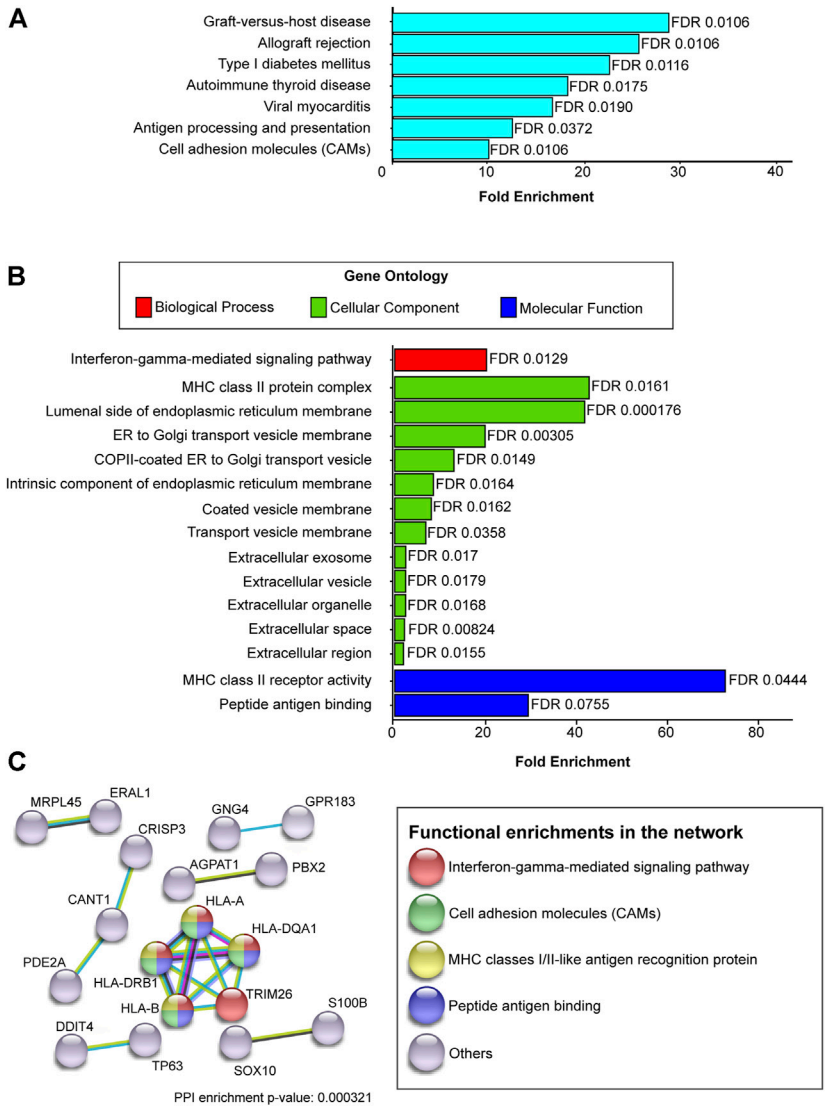


FIGURE 3
KEGG, GO, DEGs interaction related to chemotherapy resistance. The 94 DEGs with a p-adj < 0.05 were analyzed to know their contribution to KEGG pathways, gene ontology, and the interaction clusters **(A)** KEGG enrichment analysis **(B)** GO enrichment analysis. Each bar represents the fold enrichment value for KEGG and GO. The *x*-axis plots the fold enrichment values, and the *y*-axis shows the pathway's name or GO terms. In GO enrichment analysis, the plot is divided into three categories: biological process (red bars), cellular component (green bars), and molecular function (blue bars). **(C)** Interactome analysis. Only DEGs that interact with each other were plotted in the graph. Network nodes represent proteins encoded by DEGs; colors represent the category to which encoded proteins belong; edges represent protein-protein interactions. Line colors indicate the type of interaction reported.

expression information of 109 patients with HER+/ER-/PR-, and data were obtained from TCGA breast ductal carcinoma study. A total of 12 DEGs predicted the DFS. The high expression of ATF6B, DHRS13, DIRAS1, ERAL1, GRING2B, IRX3, PRTFDC1, and PBX2 was found to be an excellent prognostic of DFS at 5 years; on the other hand, a high expression of L1CAM was associated with lower DFS at 5 years (Figure 4). We found that low expression of TNXB and SLC9A3R2 and high expression of S100B were associated with better DFS in the long term (10 years) (Figure 5).

According to the OS analysis, groups with high expression of GNG4, IL22RA2, S100B, and SERPINF2 were associated with better OS at 5 years; the same was true for HLA-A and DIRAS1 at 10 years. In contrast, high expression of MICA and TNXB was related to lower OS times at 5 and 10 years, respectively (Figure 6).

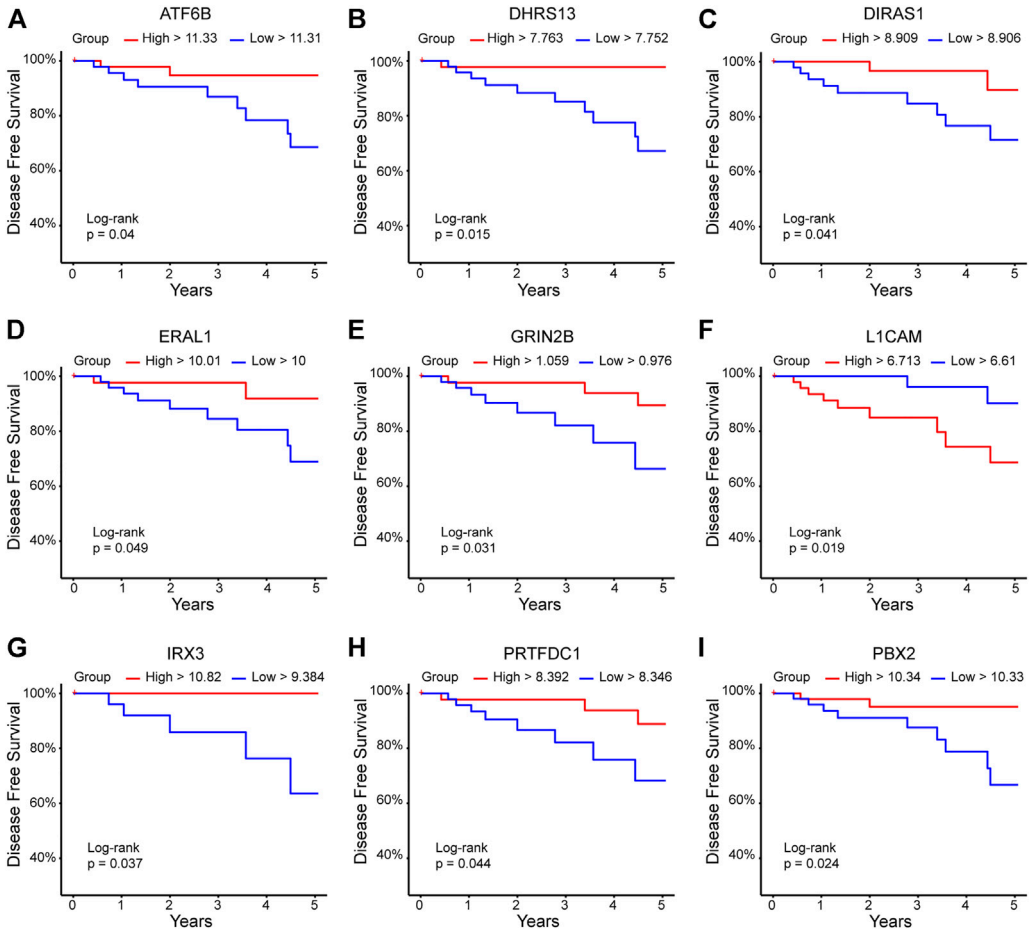


FIGURE 4
DEGs related to the prediction of DFS at 5 years. The 94 DEGs (p-adj<0.05) were analyzed in the TCGA ductal breast cancer database (https://www.cancer.gov/tcga). The red line indicates the high expression group, and the blue line represents the low expression group. *Y*-axis shows the DFS percentage; *X*-axis shows the time in years. **(A)** ATF6B, **(B)** DHRS13, **(C)** DIRAS1, **(D)** ERAL1, **(E)** GRIN2B, **(F)** L1CAM, **(G)** IRX3, **(H)** PRTFDC1, **(I)** PBX2.

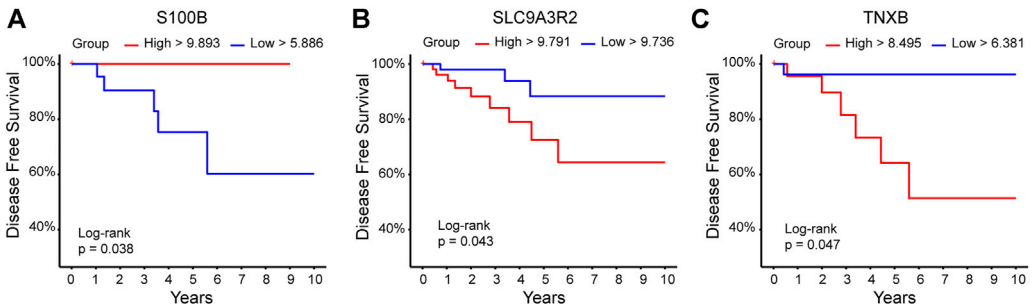


FIGURE 5
DEGs related to the prediction of DFS at 10 years. DEGs that meet the criteria of p-adj<0.05 were analyzed using the TCGA ductal breast cancer database (https://www.cancer.gov/tcga) to determine their association with DFS at 10 years. The red line indicates the high expression group, the blue line represents the low expression group. *Y*-axis shows the disease-free survival percentage; *X*-axis shows the time in years. **(A)** S100B, **(B)** SLC9A3R2, **(C)** TNXB.

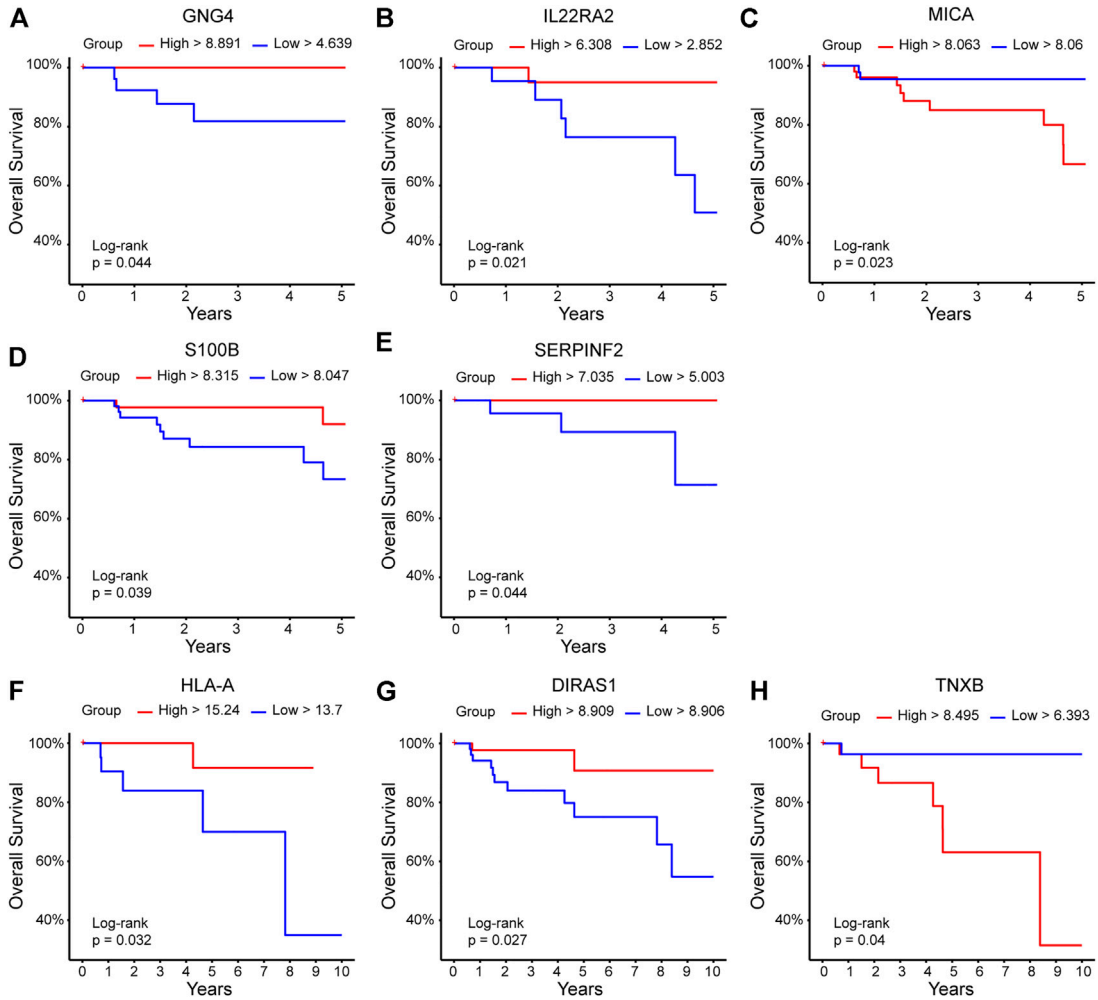


FIGURE 6
DEGs related to overall survival. DEGs with p-adj<0.05 were contrasted with overall survival data at 5 and 10 years in the TCGA ductal breast cancer database (https://www.cancer.gov/tcga). The red line indicates the high expression group, and the blue line represents the low expression group. *Y*-axis shows the overall survival percentage; *X*-axis shows the time in years. Panels for **(A)** GNG4, **(B)** IL22RA2, **(C)** MICA, **(D)** S100B, and **(E)** SERPINF2 represent OS at 5 years, while panels for **(F)** HLA-A, **(G)** DIRAS1, and **(H)** TNXB showed OS data for 10 years.

Odds ratio analysis was performed to determine if there is any difference at the end point of five of 10 years in OS or DFS. The results of the odds ratio analysis showed a similar prognostic pattern for each gene compared with the results obtained from the log-rank analysis, except for GRIN2B, PRTFDC1, SLC9A3R2 in DFS and HLA-A, IL22RA2 in OS whose p values were greater than 0.05 (Supplementary Table S3).

Furthermore, univariate and multivariate Cox analyses were performed. The results show some genes in which the expression can be considered a predictor variable associated with survival time. In univariate cox analysis for DFS, the coefficients were negative for DHRS13, GRIN2B and positive for L1CAM with $p < 0.05$. When applying the univariate cox analysis for OS, DIRAS had a negative coefficient, and MICA had a positive coefficient $p < 0.05$ (Table 1). We performed multivariate cox analysis using age, pathologic stage, radiation therapy, and the expression level as variables. The results show that as higher the pathologic stage, the hazard to disease recurrence increases for ATF6B, DHRS13, DIRAS1, ERAL1, GRIN2B, L1CAM, PBX2, PRTFDC1, SLC9A3R2, and TNXB, furthermore, increase the risk of death when analyzed DIRAS1, IL22RA, MICA, and S100B. Alongside, neoadjuvant radiation therapy was correlated with decrease recurrence risk in ATF6B, DHRS13, DIRAS1, LICAM, and SLC9A3R2, and decreases dead risk when analyzed S100B, MICA, and DIRAS1 (Table 2).

TABLE 1 Univariate cox regression analysis for expression variable.

Gene	Variable	Coef	exp (coef)	se (coef)	z	Pr (>|z|)	Survival
DHRS13	Expression	-2.1257	0.1193	1.0494	-2.026	0.0428	DFS
GRIN2B	Expression	-1.3827	0.2509	0.6886	-2.008	0.0446	DFS
L1CAM	Expression	1.6424	5.1677	0.7829	2.098	0.0359	DFS
DIRAS1	Expression	-1.5443	0.2135	0.7695	-2.007	0.0448	OS
MICA	Expression	1.6142	5.0238	0.7847	2.057	0.0397	OS

Coef = coefficient; exp = Exponential; se = standard error.

TABLE 2 Multivariate Cox Regression Analysis including clinical variables

Survival type	Gene	Variable	Coef	exp (coef)	se (coef)	p-value
DFS	ATF6B	Pathologic_stage	0.56021	1.75105	0.19235	0.0036
		Radiation_therapy	-1.52251	0.21816	0.76688	0.0471
	DHRS13	Pathologic_stage	0.76044	2.13923	0.22323	0.0007
		Radiation_therapy	-1.56215	0.20968	0.71944	0.0299
	DIRAS1	Pathologic_stage	0.73385	2.08308	0.21171	0.0005
		Radiation_therapy	-1.64476	0.19306	0.72105	0.0225
	ERAL1	Pathologic_stage	0.67782	1.96958	0.19343	0.0005
	GRIN2B	Pathologic_stage	0.71306	2.04023	0.21839	0.0011
	L1CAM	Pathologic_stage	0.56388	1.75748	0.17428	0.0012
		Radiation_therapy	-1.95488	0.14158	0.74922	0.0091
	PBX2	Pathologic_stage	0.57663	1.78003	0.19418	0.003
	PRTFDC1	Expression	-2.26336	0.104	1.10961	0.0414
		Pathologic_stage	0.64406	1.90419	0.1972	0.0011
	S100B	Expr_quant	3.647	38.37	1.088	0.0008
	SLC9A3R2	Expression	3.08808	21.9349	1.21697	0.0112
		Pathologic_stage	0.51109	1.66711	0.18757	0.0064
		Radiation_therapy	-2.09806	0.12269	0.80884	0.0095
		Expr_quant	-1.21336	0.2972	0.58311	0.0375
	TNXB	Pathologic_stage	0.56011	1.75086	0.2683	0.0368
OS	DIRAS1	Pathologic_stage	1.12604	3.08344	0.29716	0.0002
		Radiation_therapy	-3.9416	0.01942	1.08224	0.0003
	IL22RA2	Pathologic_stage	0.9143	2.49502	0.42703	0.0323
	MICA	Expression	4.936,011	139.214	1.840,017	0.0073
		Pathologic_stage	1.125,939	3.08311	0.275,097	4E-05
		Radiation_therapy	-3.643,406	0.02616	1.172,328	0.0019
		Expr_quant	-1.656,961	0.19072	0.803,941	0.0393
	S100B	Pathologic_stage	0.873,639	2.39561	0.23499	0.0002
		Radiation_therapy	-3.068291	0.0465	0.944,318	0.0012

Analysis of the 94 DEGs in other studies highlight similar DEGs as possible biomarkers

Finally, to evaluate whether the data from GSE162187 (94 DEGs between resistant and sensitive patients) has a consistent expression with other studies, we analyzed the data of the study GSE163882, which aimed to predict pCR to neoadjuvant therapy in breast cancer patients. Data from GSE162187 were used as training data, and data from the GSE163882 dataset were used as corroboration data. This analysis discard 84 DEGs and identified 10 DEGs in common (ATF6B, ERAL1, CRYM, MUC16, SOX10, MICA, PDE2A, TMEM97, SDF2, and BICDL2) that could discriminate patient

outcomes. Therefore, these 10 DEGs were considered possible biomarkers of pCR and neoadjuvant chemotherapy response. Moreover, three DEGs (ATF6B, ERAL1, and MICA) have a strong correlation with DFS and OS.

Discussion

One fundamental aspect of treating breast cancer patients is the knowledge of their molecular subtypes. This information has *per se* a prognostic value for predicting patient treatment response (von Minckwitz et al., 2012), which can be evaluated according to the criteria for the diagnosis of pCR. Achieving pCR has been associated with better overall survival (Broglio et al., 2016; Spring et al., 2020); however, the intrinsic factors involved in pCR have not been clarified. There is still controversy on whether standard adjuvant therapy increases pCR (Mauri et al., 2005). The percentage of patients who achieve pCR ranges between 27%–47% (Muller et al., 2021; Xin et al., 2021). In this study, 37.5% of patients achieved pCR. Therefore, HER2+ breast cancer patients were categorized as sensitive (pCR achieved) or resistant (pCR did not achieve) to neoadjuvant chemotherapy and were used as an RNA-seq strategy to identify predictors of pCR. It should be considered that the diagnosis of HER2+ breast cancer was because more than 10% of the tumor cells present detectable HER2 expression; therefore, there is a large percentage of cells that do not express HER2. This highlights the heterogeneity of this breast cancer subtype (Ng et al., 2015; Chen et al., 2020). When the transcriptional profiles of all patients were compared to determine clusters, this heterogeneity was emphasized (Supplementary Figure S1A). The implications of molecular differences in HER2+ breast cancer patients are not fully understood and may be relevant to prognosis and treatment response.

The DEGs found in HER2+ breast cancer patients sensitive and resistant to neoadjuvant chemotherapy were mainly related to plasma membranes, vesicles, and extracellular space and were involved in different biological processes, such as cellular response to chemical stimulus, cell adhesion, and signal regulation. Variations in the protein components of the extracellular matrix have been reported in breast tumors of different origins (Borghesi et al., 2021). In addition, extracellular components such as the extracellular matrix, vesicles, and plasma membranes can be modified by cancer-associated fibroblasts, leading to a tumor microenvironment involved in cancer development and drug resistance (Mashouri et al., 2019; Helal-Neto et al., 2020; Lugo-Cintron et al., 2020).

One of the most enriched pathways is related to cell adhesion molecules involved in tight junctions of epithelial and endothelial cells, such as claudins, which participate in epithelial-mesenchymal transition (EMT) and chemoresistance (Hewitt et al., 2006; Agarwal et al., 2009; Gowrikumar et al., 2019). According to KEGG enrichment analysis, the AMPK signaling pathway is involved in the resistance process; this pathway is considered a double-edged sword that protects and promotes cancer progression (Jeon and Hay, 2015). Sensitization of breast cancer cells to chemotherapy by activating AMPK signaling by CTAB has been observed (Pan et al., 2019). Similarly, histological evaluations have reported altered AMPK signaling in breast cancer samples (Hadad et al., 2009), and this pathway is considered a therapeutic target for breast cancer treatment (Hadad et al., 2008). However, it has been hypothesized that once cancer has developed, AMPK promotes the survival of cancer cells by protecting them against DNA damage, nutritional stress, and hypoxia (Russell and Hardie, 2020). Further studies are needed to delineate the role of the AMPK pathway in breast cancer and the development of chemotherapy resistance.

Another enriched pathway was cyclic guanosine 3,5-monophosphate (cGMP) and protein kinase G (PKG). The cGMP-PKG pathway has been associated with the modulation of apoptosis and growth inhibition in MCF-7 and MDA-MB-468 breast cancer cell lines (Fallahian et al., 2011). An essential component of this pathway is the protein kinase cGMP-dependent 2 (PRKG2), which was found to be downregulated in the resistant group in this study. Our results correlate with those of Karami-Tehrani et al. (Karami-Tehrani et al., 2012), who observed lower expression of PRKG2 protein in breast tumor samples. In addition, it has been reported that PRKG2 inhibits EGF-induced MAPK/c-Jun N-terminal kinase (JNK) signal transduction in human breast cancer cells (Lan et al., 2012) and also inhibits the activation of EGFR and HER2 in gastric cancer cells (Zhu et al., 2016; Lan et al., 2019). PRKG2 inhibits the migration, invasion, and proliferation of cancer cells and activates CREB, which modulates anti-apoptotic genes, such as BCL2 (Shankar et al., 2010), which are overexpressed in the resistant group, thereby contributing to the survival of cancer cells in the resistant group.

In this study, many DEGs related to resistance were identified. With the dimensional reduction, samples clustered better, highlighting the possibility of using these genes to predict the response to treatment. An interesting finding in our results was a group of DEGs that interacted with each other, including HLA-A, HLA-DQA1, HLA-DRB1, HLA-B, and TRIM26, which are components of the MHC protein complex, except TRIM26. These DEGs were found to be overexpressed in the resistant group. The upregulation of classical and non-classical HLA-I molecules has been reported to acquire a "protective" phenotype in melanoma cells (Balsamo et al., 2012). HLA molecules play a role in self-recognition by immune cells, which is essential for hematopoietic and healthy cells to avoid their destruction, and the loss, alteration, or absence of HLA molecules can cause susceptibility to NK cell attack (Ljunggren and Karre, 1990; Moretta et al., 2004). HLA molecules interact with inhibitory receptors such as killer cell immunoglobulin-like receptors (KIRs), leukocyte immunoglobulin-like receptors (LIRs), and

natural killer group 2A (NKG2A) on the NK surface, avoiding its activation (Khan et al., 2020). Overexpression of HLA by cancer cells has been reported as a mechanism for evading the immune response of NK cells and is termed immune checkpoint inhibition (Bi and Tian, 2019). From this group of genes, a variant of HLA-A (ENSEMBL ID ENSG00000235657) was observed subexpressed in the resistance group. In addition, low expression of this gene was associated with a worse prognosis for OS. This gene has already been reported to predict treatment response and OS (Sinn et al., 2021; Barron-Gallardo et al., 2022).

ATF6B has two ensemble IDs (ENSG00000228628 and ENSG00000213676). ENSG00000228628 ID was overexpressed. Nevertheless, ENSG00000213676, which corresponds to the primary assembly of this gene, was found to be sub-expressed, and low expression was related to lower OS and worse DFS. Variants of this gene have been associated with an increased risk of breast cancer development (Dierssen-Sotos et al., 2018).

DIRAS1 was found to be sub-expressed in the resistance group. Subexpression of this gene was correlated with lower OS. This gene has tumor-suppressive activity by binding to SmgGDS, which blocks the interactions of small GTPases, such as Rho and K-Ras4B. The expression of DIRAS1 is downregulated in most types of breast cancer (Bergom et al., 2016).

Other sub-expressed genes in resistant treatment and lower DFS were GRIN2B, GNG4, and IRX3. GRIN2B is involved in breast cancer progression and acts as a promoter of CpG islands (Park et al., 2011; Park et al., 2012). GNG4 is hypermethylated in breast cancer; however, when comparing all molecular subtypes, the HER2 subtype shows the highest expression levels for this gene (Fernandez-Nogueira et al., 2016; Mao et al., 2021). IRX3 plays an important role in obesity and type 2 diabetes; however, it plays an important role in the adaptability of tumor cells to metabolic challenges, a process that has a parallelism with the development of chemotherapeutic resistance (Singh et al., 2016).

A set of genes that showed high expression in the resistant group, which were related to lower OS and worse DFS, were L1CAM, MICA, and TNXB. The expression of L1CAM is increased in luminal B breast cancer, and its expression is related to disease recurrence and higher levels of Ki-67 expression (Moisini et al., 2021). A soluble form of L1CAM has been found in HER2-enriched primary breast cancer patients (Wu et al., 2018). There are reports that inhibition of L1CAM reverses cisplatin resistance in triple-negative breast cancer cells (Zhang et al., 2022). MICA is overexpressed in breast cancer when compared to normal tissue and is considered an indicator of poor prognosis (Madjd et al., 2007). It is an activation ligand of NK cells, which induces the lysis of cells that express it. However, there is a soluble form of MICA (sMICA) that decreases the expression and presentation of NKG2D, a natural cytotoxic receptor in natural killer cells, thus sMICA helps cancer cells to evade immune cell attack (Pan et al., 2017) and contributing to a worse prognosis in cancer (Roshani et al., 2016). In this study, high expression of MICA was observed in the resistant and lower OS groups; however, further studies are needed to determine the role of MICA or sMICA in chemotherapy resistance. In the case of TNXB, the expression of this gene has been analyzed in breast cancer, and a correlation between high TNXB expression and good survival prognosis has been found (Liot et al., 2020). Its expression decreases at late stages, major tumor grade, and node status of the disease (Liot et al., 2020), however, its expression in the HER2 molecular subtype and in relation to chemotherapy resistance has not been evaluated.

In contrast, genes with high expression but related to better OS and DFS were IL22RA2, PRTFDC1, PBX2, S100B, SERPINF2, DHRS13, ERAL1, and SLC9A3R2. IL22RA2 expression decreases in luminal A, B, and triple-negative breast cancers (Fu et al., 2015); however, but HER2+ breast cancer has not been reported. PRTFDC1 has been associated with the triple-negative basal-like immune-suppressed breast cancer subtype (TNBC-BLIS), which is considered one of the worst prognoses (Yin et al., 2020). The most highly expressed gene is PBX2. This gene was found to be upregulated in breast lesions and has been proposed along with other genes as a candidate biomarker for distinguishing breast cancer lesions (Hou et al., 2020). It has been showed that the overexpression of PBX2 increases the tumorigenic properties of SkBr3 breast cancer cell line when transfected with HoxB7 (Fernandez et al., 2008). S100B expression has been negatively correlated with lymph node metastasis (Wang et al., 2021), inhibition of cell migration, better overall survival in luminal B breast cancer patients, and being a good distant metastases-free survival biomarker (Yen et al., 2018). SERPINF2 is differentially expressed in breast cancer tissues compared with normal tissues (Malvia et al., 2019). The protein product of SERPINF2 has been found in the serum of breast cancer patients when evaluating treatment response; however, this protein appeared in both resistant and sensitive groups (Chantada-Vazquez et al., 2022).

Finally, DHRS13, ERAL1, and SLC9A3R2 could predict treatment response and survival; however, there are no reports related to breast cancer and its possible function in this disease.

Conclusion

This study underlines a molecular expression pattern related to the response of patients with HER2-positive breast cancer to neoadjuvant chemotherapy. Differentially expressed genes highlight the involvement of pathways, such as extracellular components, adhesion molecules, and immune responses, in the process of resistance to chemotherapy. Some differentially expressed genes can be

used as biomarkers of overall survival and disease-free survival in breast cancers.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ Supplementary Material.

Ethics statement

The studies involving human participants were reviewed and approved by Comite Nacional de Investigación Científica del Instituto Mexicano del Seguro Social. The patients/participants provided their written informed consent to participate in this study.

Author contributions

CB-G and MG-C. contributed to sample recruitment and processing. CB-G performed the bioinformatic analysis, interpretation of data, and drafting of the manuscript. MG-C. contributed to data analysis and funding acquisition. RD-C, AM-M, MM-S, and MV-G were involved in patient recruitment, obtaining clinical information, and data interpretation. AA-L. and LJ-S. conceived the study, advised, analyzed the results, contributed to funding acquisition, and wrote and revised the manuscript. All authors substantively revised the manuscript, suggested modifications, and approved the final version.

Funding

This work was funded by the Fondo de Investigación en Salud, Instituto Mexicano del Seguro Social with grant number FIS/IMSS/PROT/PRIO/14/030 to LJ-S.

Acknowledgments

CAB-G is grateful for a scholarship from Consejo Nacional de Ciencia y Tecnología (CONACyT)- Mexico.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.991706/full#supplementary-material.

References

Abal, M., Andreu, J. M., and Barasoain, I. (2003). Taxanes: Microtubule and centrosome targets, and cell cycle dependent mechanisms of action. *Curr. Cancer Drug Targets* 3 (3), 193–203. doi:10.2174/1568009033481967

Agarwal, R., Mori, Y., Cheng, Y., Jin, Z., Olaru, A. V., Hamilton, J. P., et al. (2009). Silencing of claudin-11 is associated with increased invasiveness of gastric cancer cells. *PLoS One* 4 (11), e8002. doi:10.1371/journal.pone.0008002

Balsamo, M., Vermi, W., Parodi, M., Pietra, G., Manzini, C., Queirolo, P., et al. (2012). Melanoma cells become resistant to NK-cell-mediated killing when exposed to NK-cell numbers compatible with NK-cell infiltration in the tumor. *Eur. J. Immunol.* 42 (7), 1833–1842. doi:10.1002/eji.201142179

Ban, M., Petric Mise, B., and Vrdoljak, E. (2020). Early HER2-positive breast cancer: Current treatment and novel approaches. *Breast Care (Basel)* 15 (6), 560–569. doi:10.1159/000511883

Barron-Gallardo, C. A., Garcia-Chagollan, M., Moran-Mendoza, A. J., Delgadillo-Cristerna, R., Martinez-Silva, M. G., Aguilar-Lemarroy, A., et al. (2022). Transcriptomic analysis of breast cancer patients sensitive and resistant to chemotherapy: Looking for overall survival and drug resistance biomarkers. *Technol. Cancer Res. Treat.* 21, 15330338211068965. doi:10.1177/15330338211068965

Bergom, C., Hauser, A. D., Rymaszewski, A., Gonyo, P., Prokop, J. W., Jennings, B. C., et al. (2016). The tumor-suppressive small GTPase DiRas1 binds the noncanonical guanine nucleotide exchange factor SmgGDS and antagonizes SmgGDS interactions with oncogenic small GTPases. *J. Biol. Chem.* 291 (12), 6534–6545. doi:10.1074/jbc.M115.696831

Bi, J., and Tian, Z. (2019). NK cell dysfunction and checkpoint immunotherapy. *Front. Immunol.* 10, 1999. doi:10.3389/fimmu.2019.01999

Borghesi, J., Giancoli Kato Cano da Silva, M., de Oliveira Pimenta Guimaraes, K., Mario, L. C., de Almeida da Anunciacao, A. R., Silveira Rabelo, A. C., et al. (2021). Evaluation of immunohistopathological profile of tubular and solid canine mammary carcinomas. *Res. Vet. Sci.* 136, 119–126. doi:10.1016/j.rvsc.2021.02.004

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34 (5), 525–527. doi:10.1038/nbt.3519

Broglio, K. R., Quintana, M., Foster, M., Olinger, M., McGlothlin, A., Berry, S. M., et al. (2016). Association of pathologic complete response to neoadjuvant therapy in HER2-positive breast cancer with long-term outcomes: A meta-analysis. *JAMA Oncol.* 2 (6), 751–760. doi:10.1001/jamaoncol.2015.6113

Chantada-Vazquez, M. D. P., Conde-Amboage, M., Grana-Lopez, L., Vazquez-Estevez, S., Bravo, S. B., and Nunez, C. (2022). Circulating proteins associated with response and resistance to neoadjuvant chemotherapy in HER2-positive breast cancer. *Cancers (Basel)* 14 (4), 1087. doi:10.3390/cancers14041087

Chen, B., Zhang, G., Wei, G., Wang, Y., Guo, L., Lin, J., et al. (2020). Heterogeneity of genomic profile in patients with HER2-positive breast cancer. *Endocr. Relat. Cancer* 27 (3), 153–162. doi:10.1530/ERC-19-0414

Dierssen-Sotos, T., Palazuelos-Calderon, C., Jimenez-Moleon, J. J., Aragones, N., Altzibar, J. M., Castano-Vinyals, G., et al. (2018). Reproductive risk factors in breast cancer and genetic hormonal pathways: A gene-environment interaction in the MCC-Spain project. *BMC Cancer* 18 (1), 280. doi:10.1186/s12885-018-4182-3

Dodt, M., Roehr, J. T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR-flexible barcode and adapter processing for next-generation sequencing platforms. *Biol. (Basel)* 1 (3), 895–905. doi:10.3390/biology1030895

Fallahian, F., Karami-Tehrani, F., Salami, S., and Aghaei, M. (2011). Cyclic GMP induced apoptosis via protein kinase G in oestrogen receptor-positive and -negative breast cancer cell lines. *FEBS J.* 278 (18), 3360–3369. doi:10.1111/j.1742-4658.2011.08260.x

Fernandez, L. C., Errico, M. C., Bottero, L., Penkov, D., Resnati, M., Blasi, F., et al. (2008). Oncogenic HoxB7 requires TALE cofactors and is inactivated by a dominant-negative Pbx1 mutant in a cell-specific manner. *Cancer Lett.* 266 (2), 144–155. doi:10.1016/j.canlet.2008.02.042

Fernandez-Nogueira, P., Bragado, P., Almendro, V., Ametller, E., Rios, J., Choudhury, S., et al. (2016). Differential expression of neurogenes among breast cancer subtypes identifies high risk patients. *Oncotarget* 7 (5), 5313–5326. doi:10.18632/oncotarget.6543

Fu, J., Khaybullin, R., Zhang, Y., Xia, A., and Qi, X. (2015). Gene expression profiling leads to discovery of correlation of matrix metalloproteinase 11 and heparanase 2 in breast cancer progression. *BMC Cancer* 15, 473. doi:10.1186/s12885-015-1410-y

Gowrikumar, S., Singh, A. B., and Dhawan, P. (2019). Role of claudin proteins in regulating cancer stem cells and chemoresistance-potential implication in disease prognosis and therapy. *Int. J. Mol. Sci.* 21 (1), E53. doi:10.3390/ijms21010053

Hadad, S. M., Baker, L., Quinlan, P. R., Robertson, K. E., Bray, S. E., Thomson, G., et al. (2009). Histological evaluation of AMPK signalling in primary breast cancer. *BMC Cancer* 9, 307. doi:10.1186/1471-2407-9-307

Hadad, S. M., Fleming, S., and Thompson, A. M. (2008). Targeting AMPK: A new therapeutic opportunity in breast cancer. *Crit. Rev. Oncol. Hematol.* 67 (1), 1–7. doi:10.1016/j.critrevonc.2008.01.007

Harbeck, N., and Gnant, M. (2017). Breast cancer. *Lancet* 389 (10074), 1134–1150. doi:10.1016/S0140-6736(16)31891-8

Helal-Neto, E., Barcellos-de-Souza, P., Morgado-Diaz, J., Barja-Fidalgo, C., and Barja-Fidalgo, C. (2020). Extracellular matrix derived from high metastatic human breast cancer triggers epithelial-mesenchymal transition in epithelial breast cancer cells through αvβ3 integrin. *Int. J. Mol. Sci.* 21 (8), E2995. doi:10.3390/ijms21082995

Hewitt, K. J., Agarwal, R., and Morin, P. J. (2006). The claudin gene family: Expression in normal and neoplastic tissues. *BMC Cancer* 6, 186. doi:10.1186/1471-2407-6-186

Hou, H., Lyu, Y., Jiang, J., Wang, M., Zhang, R., Liew, C. C., et al. (2020). Peripheral blood transcriptome identifies high-risk benign and malignant breast lesions. *PLoS One* 15 (6), e0233713. doi:10.1371/journal.pone.0233713

Iwamoto, T., Kajiwara, Y., Zhu, Y., and Iha, S. (2020). Biomarkers of neoadjuvant/adjuvant chemotherapy for breast cancer. *Chin. Clin. Oncol.* 9 (3), 27. doi:10.21037/cco.2020.01.06

Jeon, S. M., and Hay, N. (2015). The double-edged sword of AMPK signaling in cancer and its therapeutic implications. *Arch. Pharm. Res.* 38 (3), 346–357. doi:10.1007/s12272-015-0549-z

Kanehisa, M., and Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.* 29 (1), 28–35. doi:10.1002/pro.3711

Karami-Tehrani, F., Fallahian, F., and Atri, M. (2012). Expression of cGMP-dependent protein kinase, PKGIα, PKGIβ, and PKGII in malignant and benign breast tumors. *Tumour Biol.* 33 (6), 1927–1932. doi:10.1007/s13277-012-0453-9

Khan, M., Arooj, S., and Wang, H. (2020). NK cell-based immune checkpoint inhibition. *Front. Immunol.* 11, 167. doi:10.3389/fimmu.2020.00167

Lan, T., Chen, Y., Sang, J., Wu, Y., Wang, Y., Jiang, L., et al. (2012). Type II cGMP-dependent protein kinase inhibits EGF-induced MAPK/JNK signal transduction in breast cancer cells. *Oncol. Rep.* 27 (6), 2039–2044. doi:10.3892/or.2012.1726

Lan, T., Pang, J., Wang, Z., Wang, Y., Qian, H., Chen, Y., et al. (2019). Type II cGMP-dependent protein kinase phosphorylates EGFR at threonine 669 and thereby inhibits its activation. *Biochem. Biophys. Res. Commun.* 518 (1), 14–18. doi:10.1016/j.bbrc.2019.07.126

Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27 (12), 1739–1740. doi:10.1093/bioinformatics/btr260

Liot, S., Aubert, A., Hervieu, V., Kholti, N. E., Schalkwijk, J., Verrier, B., et al. (2020). Loss of tenascin-X expression during tumor progression: A new pan-cancer marker. *Matrix Biol. Plus* 6-7, 100021. doi:10.1016/j.mbplus.2020.100021

Ljunggren, H. G., and Karre, K. (1990). In search of the 'missing self': MHC molecules and NK cell recognition. *Immunol. Today* 11 (7), 237–244. doi:10.1016/0167-5699(90)90097-s

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8

Lugo-Cintron, K. M., Gong, M. M., Ayuso, J. M., Tomko, L. A., Beebe, D. J., Virumbrales-Munoz, M., et al. (2020). Breast fibroblasts and ECM components modulate breast cancer cell migration through the secretion of MMPs in a 3D microfluidic Co-culture model. *Cancers (Basel)* 12 (5), E1173. doi:10.3390/cancers12051173

Madjd, Z., Spendlove, I., Moss, R., Bevin, S., Pinder, S. E., Watson, N. F., et al. (2007). Upregulation of MICA on high-grade invasive operable breast carcinoma. *Cancer Immun.* 7, 17.

Malvia, S., Bagadi, S. A. R., Pradhan, D., Chintamani, C., Bhatnagar, A., Arora, D., et al. (2019). Study of gene expression profiles of breast cancers in Indian women. *Sci. Rep.* 9 (1), 10018. doi:10.1038/s41598-019-46261-1

Mao, X. H., Ye, Q., Zhang, G. B., Jiang, J. Y., Zhao, H. Y., Shao, Y. F., et al. (2021). Identification of differentially methylated genes as diagnostic and prognostic biomarkers of breast cancer. *World J. Surg. Oncol.* 19 (1), 29. doi:10.1186/s12957-021-02124-6

Mashouri, L., Yousefi, H., Aref, A. R., Ahadi, A. M., Molaei, F., and Alahari, S. K. (2019). Exosomes: Composition, biogenesis, and mechanisms in cancer metastasis and drug resistance. *Mol. Cancer* 18 (1), 75. doi:10.1186/s12943-019-0991-5

Mauri, D., Pavlidis, N., and Ioannidis, J. P. (2005). Neoadjuvant versus adjuvant systemic treatment in breast cancer: A meta-analysis. *J. Natl. Cancer Inst.* 97 (3), 188–194. doi:10.1093/jnci/dji021

Maximiano, S., Magalhaes, P., Guerreiro, M. P., and Morgado, M. (2016). Trastuzumab in the treatment of breast cancer. *BioDrugs* 30 (2), 75–86. doi:10.1007/s40259-016-0162-9

McGowan, J. V., Chung, R., Maulik, A., Piotrowska, I., Walker, J. M., and Yellon, D. M. (2017). Anthracycline chemotherapy and cardiotoxicity. *Cardiovasc. Drugs Ther.* 31 (1), 63–75. doi:10.1007/s10557-016-6711-0

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L. P., Mushayamaha, T., et al. (2021). PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* 49 (D1), D394–D403. doi:10.1093/nar/gkaa1106

Moisini, I., Zhang, H., D'Aguiar, M., Hicks, D. G., and Turner, B. M. (2021). L1CAM expression in recurrent estrogen positive/HER2 negative breast cancer: A novel biomarker worth considering. *Appl. Immunohistochem. Mol. Morphol.* 29 (4), 287–292. doi:10.1097/PAI.0000000000000909

Moretta, L., Bottino, C., Pende, D., Vitale, M., Mingari, M. C., and Moretta, A. (2004). Different checkpoints in human NK-cell activation. *Trends Immunol.* 25 (12), 670–676. doi:10.1016/j.it.2004.09.008

Muller, C., Schmidt, G., Juhasz-Boss, I., Jung, L., Huwer, S., Solomayer, E. F., et al. (2021). Influences on pathologic complete response in breast cancer patients after neoadjuvant chemotherapy. *Arch. Gynecol. Obstet.* 304 (4), 1065–1071. doi:10.1007/s00404-021-06018-6

Ng, C. K., Martelotto, L. G., Gauthier, A., Wen, H. C., Piscuoglio, S., Lim, R. S., et al. (2015). Intra-tumor genetic heterogeneity and alternative driver genetic alterations in breast cancers with heterogeneous HER2 gene amplification. *Genome Biol.* 16, 107. doi:10.1186/s13059-015-0657-6

Pan, J., Shen, J., Si, W., Du, C., Chen, D., Xu, L., et al. (2017). Resveratrol promotes MICA/B expression and natural killer cell lysis of breast cancer cells by suppressing c-Myc/miR-17 pathway. *Oncotarget* 8 (39), 65743–65758. doi:10.18632/oncotarget.19445

Pan, Y., Zhang, Y., Chen, Q., Tao, X., Liu, J., and Xiao, G. G. (2019). CTAB enhances chemo-sensitivity through activation of AMPK signaling cascades in breast cancer. *Front. Pharmacol.* 10, 843. doi:10.3389/fphar.2019.00843

Park, S. Y., Kwon, H. J., Choi, Y., Lee, H. E., Kim, S. W., Kim, J. H., et al. (2012). Distinct patterns of promoter CpG island methylation of breast cancer subtypes are associated with stem cell phenotypes. *Mod. Pathol.* 25 (2), 185–196. doi:10.1038/modpathol.2011.160

Park, S. Y., Kwon, H. J., Lee, H. E., Ryu, H. S., Kim, S. W., Kim, J. H., et al. (2011). Promoter CpG island hypermethylation during breast cancer progression. *Virchows Arch.* 458 (1), 73–84. doi:10.1007/s00428-010-1013-6

Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., et al. (2000). Molecular portraits of human breast tumours. *Nature* 406 (6797), 747–752. doi:10.1038/35021093

Roehr, J. T., Dieterich, C., and Reinert, K. (2017). Flexbar 3.0 - SIMD and multicore parallelization. *Bioinformatics* 33 (18), 2941–2942. doi:10.1093/bioinformatics/btx330

Roshani, R., Boroujerdnia, M. G., Talaiezadeh, A. H., and Khodadadi, A. (2016). Assessment of changes in expression and presentation of NKG2D under influence of MICA serum factor in different stages of breast cancer. *Tumour Biol.* 37 (5), 6953–6962. doi:10.1007/s13277-015-4584-7

Russell, F. M., and Hardie, D. G. (2020). AMP-activated protein kinase: Do we need activators or inhibitors to treat or prevent cancer? *Int. J. Mol. Sci.* 22 (1), E186. doi:10.3390/ijms22010186

Shankar, E., Krishnamurthy, S., Paranandi, R., and Basu, A. (2010). PKCepsilon induces Bcl-2 by activating CREB. *Int. J. Oncol.* 36 (4), 883–888. doi:10.3892/ijo_00000566

Singh, B., Kinne, H. E., Milligan, R. D., Washburn, L. J., Olsen, M., and Lucci, A. (2016). Important role of FTO in the survival of rare panresistant triple-negative inflammatory breast cancer cells facing a severe metabolic challenge. *PLoS One* 11 (7), e0159072. doi:10.1371/journal.pone.0159072

Sinn, B. V., Loibl, S., Hanusch, C. A., Zahm, D. M., Sinn, H. P., Untch, M., et al. (2021). Immune-related gene expression predicts response to neoadjuvant chemotherapy but not additional benefit from PD-L1 inhibition in women with early triple-negative breast cancer. *Clin. Cancer Res.* 27 (9), 2584–2591. doi:10.1158/1078-0432.CCR-20-3113

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., et al. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. U. S. A.* 100 (14), 8418–8423. doi:10.1073/pnas.0932692100

Spring, L. M., Fell, G., Arfe, A., Sharma, C., Greenup, R., Reynolds, K. L., et al. (2020). Pathologic complete response after neoadjuvant chemotherapy and impact on breast cancer recurrence and survival: A comprehensive meta-analysis. *Clin. Cancer Res.* 26 (12), 2838–2848. doi:10.1158/1078-0432.CCR-19-3492

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 1, 209–249. doi:10.3322/caac.21660

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47 (D1), D607–D613. doi:10.1093/nar/gky1131

von Minckwitz, G., Untch, M., Blohmer, J. U., Costa, S. D., Eidtmann, H., Fasching, P. A., et al. (2012). Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. *J. Clin. Oncol.* 30 (15), 1796–1804. doi:10.1200/JCO.2011.38.8595

Waks, A. G., and Winer, E. P. (2019). Breast cancer treatment: A review. *JAMA* 321 (3), 288–300. doi:10.1001/jama.2018.19323

Wang, C., Xu, K., Deng, F., Liu, Y., Huang, J., Wang, R., et al. (2021). A six-gene signature related with tumor mutation burden for predicting lymph node metastasis in breast cancer. *Transl. Cancer Res.* 10 (5), 2229–2246. doi:10.21037/tcr-20-3471

Wu, J. D., Hong, C. Q., Huang, W. H., Wei, X. L., Zhang, F., Zhuang, Y. X., et al. (2018). L1 cell adhesion molecule and its soluble form sL1 exhibit poor prognosis in primary breast cancer patients. *Clin. Breast Cancer* 18 (5), e851–e861. doi:10.1016/j.clbc.2017.12.011

Xin, L., Zhang, H., Zhang, S., Cheng, Y. J., Liu, Q., Xu, L., et al. (2021). [Docetaxel, carboplatin plus trastuzumab as neoadjuvant setting in patients with early-stage human epidermal growth factor receptor 2 positive breast cancer: A retrospective analysis]. *Zhonghua Wai Ke Za Zhi* 59 (3), 222–227. doi:10.3760/cma.j.cn112139-20201122-00811

Yardley, D. A. (2013). nab-Paclitaxel mechanisms of action and delivery. *J. Control. Release* 170 (3), 365–372. doi:10.1016/j.jconrel.2013.05.041

Yen, M. C., Huang, Y. C., Kan, J. Y., Kuo, P. L., Hou, M. F., and Hsu, Y. L. (2018). S100B expression in breast cancer as a predictive marker for cancer metastasis. *Int. J. Oncol.* 52 (2), 433–440. doi:10.3892/ijo.2017.4226

Yin, L., Duan, J. J., Bian, X. W., and Yu, S. C. (2020). Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Res.* 22 (1), 61. doi:10.1186/s13058-020-01296-5

Zhang, L. Y., Shen, Z. X., and Guo, L. (2022). Inhibiting L1CAM reverses cisplatin resistance of triple negative breast cancer cells by blocking AKT signaling pathway. *Cancer Invest.* 40 (4), 313–324. doi:10.1080/07357907.2021.2016801

Zhu, M., Yao, X., Wu, M., Qian, H., Wu, Y., and Chen, Y. (2016). Type II cGMP-dependent protein kinase directly inhibits HER2 activation of gastric cancer cells. *Mol. Med. Rep.* 13 (2), 1909–1913. doi:10.3892/mmr.2015.4688

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 28 October 2022
DOI 10.3389/fgene.2022.1002699

OPEN ACCESS

EDITED BY
Peter Hart,
Roosevelt University College of Pharmacy, United States

REVIEWED BY
Lin Wu,
Central South University, China
Nan Ma,
First Affiliated Hospital, Guangxi Medical University, China

*CORRESPONDENCE
Caicun Zhou,
drcaicunzhou@163.com
Chunyu Li,
lichunyu@tmu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 25 July 2022
ACCEPTED 18 October 2022
PUBLISHED 28 October 2022

CITATION
Tang Z, Wang Q, Chen P, Guo H, Shi J, Pan Y, Li C and Zhou C (2022), Computational recognition of LncRNA signatures in tumor-associated neutrophils could have implications for immunotherapy and prognostic outcome of non-small cell lung cancer. *Front. Genet.* 13:1002699. doi: 10.3389/fgene.2022.1002699

Computational recognition of LncRNA signatures in tumor-associated neutrophils could have implications for immunotherapy and prognostic outcome of non-small cell lung cancer

Zhuoran Tang[1†], Qi Wang[2†], Peixin Chen[1], Haoyue Guo[1], Jinpeng Shi[1], Yingying Pan[1], Chunyu Li[3]* and Caicun Zhou[2]*

[1]Tongji University Medical School Cancer Institute, Tongji University, Shanghai, China, [2]Department of Medical Oncology, Tongji University Affiliated Shanghai Pulmonary Hospital, Tongji University Medical School Cancer Institute, Tongji University, Shanghai, China, [3]Department of Integrated Traditional Chinese and Western Medicine, International Medical School, Tianjin Medical University, Tianjin, China

Cancer immune function and tumor microenvironment are governed by long noncoding RNAs (lncRNAs). Nevertheless, it has yet to be established whether lncRNAs play a role in tumor-associated neutrophils (TANs). Here, a computing framework based on machine learning was used to identify neutrophil-specific lncRNA with prognostic significance in squamous cell carcinoma and lung adenocarcinoma using univariate Cox regression to comprehensively analyze immune, lncRNA, and clinical characteristics. The risk score was determined using LASSO Cox regression analysis. Meanwhile, we named this risk score as "TANlncSig." TANlncSig was able to distinguish between better and worse survival outcomes in various patient datasets independently of other clinical variables. Functional assessment of TANlncSig showed it is a marker of myeloid cell infiltration into tumor infiltration and myeloid cells directly or indirectly inhibit the anti-tumor immune response by secreting cytokines, expressing immunosuppressive receptors, and altering metabolic processes. Our findings highlighted the value of TANlncSig in TME as a marker of immune cell infiltration and showed the values of lncRNAs as indicators of immunotherapy.

KEYWORDS

non-small cell lung cancer, tumor-associated neutrophils, long noncoding RNA, immunotherapy, computational recognition

Introduction

Lung cancer is related with high mortality rates in China with non-small cell lung cancer (NSCLC) accounting for >80% of lung cancers (Zhu et al., 2017). The administration of immune checkpoint inhibitors (ICIs) in cancer therapy has had remarkable results (Yue et al., 2018; Dolladille et al., 2020; Galluzzi et al., 2020). For advanced non-small cell lung cancer (NSCLC), several clinical trials have confirmed that as first- or second-line treatment, ICIs are superior to platinum-based chemotherapy (Ko et al., 2018; Vansteenkiste et al., 2019; Chen et al., 2020). However, only 20%–40% of advanced NSCLC patients achieve sustained clinical benefits from PD-(L)1 inhibitor therapy, with most patients having primary or acquired resistance to immunotherapy (Socinski, 2014). Moreover, those who do not respond to immunotherapy may suffer immune-related adverse events (IRAE) and the high costs of anti-PD-(L)1 monoclonal antibody therapy (Khoja et al., 2017; Das and Johnson, 2019; Schoenfeld et al., 2019). Thus, effective biomarkers that distinguish potential responders from non-responders, and indicate patient clinical response in real-time are urgently needed to improve treatment outcomes.

The TME is comprised of a complex cell population that includes tissue-resident lymphocytes, fibroblasts, endothelial cells, and neurons that are present before tumorigenesis, as well as blood-derived cells recruited to tumor sites (Butturini et al., 2019). Immune cells are the main cellular components in tumors. Tumor-infiltrating myeloid cells, including tumor-associated macrophages (TAM), regulatory dendritic cells, tumor-associated neutrophils (TAN), myeloid-derived suppressor cells (MDSC), as well as tolerogenic dendritic cells (TOL-DC), facilitate the formation of immunosuppressive microenvironments (Schupp et al., 2019). These cells directly or indirectly inhibit the antitumor immune response by secreting cytokines, expressing immunosuppressive receptors, and altering metabolic processes, leading to tumor immune escape. Tumor-associated neutrophils (TANs) are a key part of tumor-infiltrating myeloid cells and are regularly detected in the TME. Clinically, TANs can be used to predict treatment outcomes and immunotherapy response (Nielsen et al., 2021). Transcriptomic studies have identified gene expression biomarkers as well as signatures for quantitative assessment of TANs, as well as for stratification based on prognoses and immunotherapeutic response (Lecot et al., 2019; Wu and Zhang, 2020).

Long non-coding RNA (lncRNAs) influence almost all biological processes and pathways, and their dysregulation is associated with various diseases. Additionally, lncRNAs have wide functional diversity due to their influence on gene expression levels at transcriptional, post-transcriptional and epigenetic levels (Rinn and Chang, 2012; Fatica and Bozzoni, 2014; Marchese et al., 2017; Bao et al., 2020). The correlation between lncRNAs and immune function has been reported. Recent studies have shown that lncRNAs are abundant with cell type specificity in various immune cell subsets (Rinn and Chang, 2012; Atianand et al., 2017; Chen et al., 2017; Zhou et al., 2017; Zhou et al., 2018). LncRNAs expression pattern has been correlated with infiltrations of immune cells into the TME (Hu et al., 2013; Ranzani et al., 2015; Sage et al., 2018; Wang et al., 2018; Zhao et al., 2021). Nevertheless, neutrophil-specific lncRNAs as well as their significance in assessing TANs and prediction of clinical outcomes and immunotherapeutic responses require further study.

Here, a computational framework is proposed for determining neutrophil-specific lncRNA expression levels and lncRNA signatures for TANs (TANLncSig) *via* integrative immune, lncRNA, and clinical profiling analyses. The TANLncSig's ability to predict clinical outcome and response to immunotherapy by NSCLC patients was also investigated.

Materials and methods

Neutrophil-specific long noncoding RNAs screening

The data set can be obtained from the GEO database with series accession number GSE28490 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE28490), These included chip data on the expression of nine human immune cells (neutrophils, monocytes, B cells, eosinophils, CD4 T cells, NK cells, mDCs, CD8 T cells, and pDCs). The GEO2R tool from GEO was used for differential expression analysis. Using adjusted p = <0.05 and logFC >1 as cutoff thresholds identified 17 lncRNAs with high neutrophil-specific expression.

Construction of risk scoring model

Clinical data and TCGA RNA-seq datasets for LUSC and LUAD were downloaded by the UCSC Xena browser (https://xenabrowser.net/). Lusc-LINC01272-neutrophils malignant/Luad-LINC01272-neutrophils malignant results from single cell sequencing datasets. First, a monovariate Cox regression analysis was used to find neutrophil-specific lncRNAs with prognostic value in LUSC and LUAD, and LASSO Cox regression was used to determine their risk scores. The multivariate Cox regression analysis (age, risk score, tumor stage, gender), Kaplan-Meier (KM) survival analysis and 3, 5, and 10 years survival AUCs were used to evaluate risk score.

Correlation analysis between risk score and tumor clinical phenotype

Multivariate ANOVA was used to analyze differences between neutrophil-specific, highly expressed lncRNA and risk score in LUSC and LUAD samples at various TNM stages.

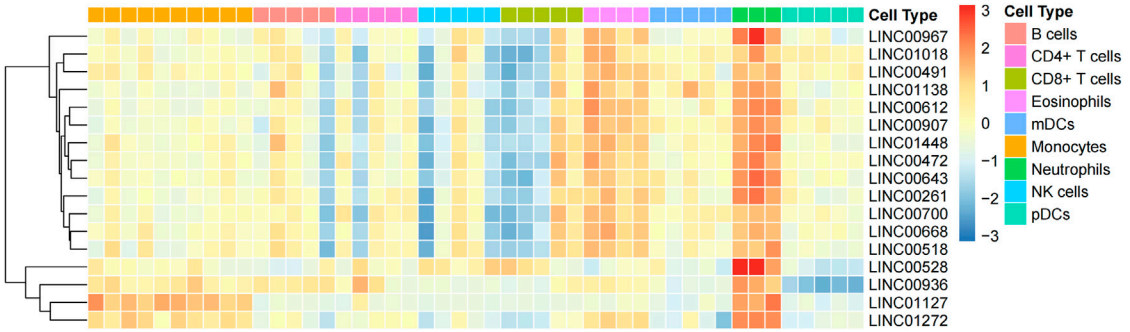


FIGURE 1
Using machine learning, 17 neutrophil-specific, highly expressed lncRNAs were identified. Differential expressions of lncRNAs between neutrophils and various immune cell types.

Analysis of risk score related pathways

In LUSC and LUAD samples, genes with mean expression levels >1 were identified and their correlation with risk score analyzed. 1,000 genes with the highest absolute correlation coefficient value were selected from those with positive correlation coefficients (>0, p = <0.05) and those with negative correlation coefficients (<0, p = <0.05). ClusterProfiler for R was used to analyze GO terms of biological process (BP), Molecular function (MF), cellular component (CC), and KEGG pathway enrichment analyses. After gene enrichment, the adjusted p-value < 0.05 and the smallest TOP10 was selected for mapping.

Development of tumor-associated neutrophils-derived long noncoding RNAs signature to judge the prognosis of immunotherapy for non-small cell lung cancer using machine learning

Pearson correlation analysis was used to determine correlations between risk score, neutropen-specific lncRNAs, and the expression of common immune checkpoint inhibitors and correlation heat maps drawn, with * denoting $p \leq 0.01$ while + denotes $p \leq 0.05$.

Results

Prognostic significance of neutrophil-specific long noncoding RNAs

To recognize neutrophil-specific lncRNAs, dataset GSE28490 was downloaded from GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE28490). This dataset includes chip data on expressions of nine human immune cells ($CD4^+$ T cells, neutrophils, monocytes, B cells, eosinophils, $CD8^+$ T cells, NK cells, mDCs, and pDCs). Using GEO2R, 17 lncRNA specifically highly-expressed in neutrophils (p =<0.05, log2>1) were identified. These neutrophil-specific lncRNAs are referred to as TAN-associated lncRNAs (TANlncRNA) (Figure 1).

Construction of a risk score based on neutrophil-specific long noncoding RNAs for prognosis prediction

To develop a neutrophil-specific lncRNA risk score for predicting prognosis, the TCGA NA-SEQ dataset, TCGA lung squamous cell carcinoma (LUSC) as well as adenocarcinoma (LUAD) gene expression data, clinical features, and prognosis data were downloaded from UCSC Xena. First, univariate Cox regression analyses were used to establish neutrophil-specific lncRNAs with prognostic value in LUSC and LUAD. The final signature named TANlncSig (Table 1). This analysis identified three lncRNAs with prognostic value in LUSC (LINC01272, LINC00261, LINC00668, p = <0.05). Using these three lncRNAs, the expression value of lncRNA was weighted using multivariate Cox regression coefficient to obtain risk scores *via* the formula: risk score = 0.09 * LNC00668 + 0.17 * LNC00261. Then, TANlncSig scores for every patient in the discovery dataset were determined, after which the 542 patients were grouped into the high (n = 271) or low (n = 271) risk groups. Low risk group patients were found to have longer overall survival (OS) relative to the high-risk group patients (p = 0.039, ≤0.05, Figure 2A). Multivariate Cox regression analyses revealed that risk score (p < 0.001), stage (p < 0.001), age (p = 0.037, ≤0.05), and gender (p = 0.007, ≤0.01) significantly affected the prognostic outcomes of LUSC patients. The p-value and hazard ratio of TANlncSig were better than those of stage and age (Figure 2B). That said, TANlncSig has the potential to be a good predictor of

TABLE 1 Detailed information of six lncRNAs in the TANlncSig.

	Ensembl ID	Gene symbol	Location (GRCh37/hg19)	HR	Lower 0.95	Upper 0.95	*p*-value
LUAD	ENSG00000259974	LINC00261	*chr20:22,541,191–22,559,280*	0.8726407	0.7771	0.98	0.021359
	ENSG00000269220	LINC00528	*chr22:18,260,056–18,262,247*	0.5049413	0.2662	0.9577	0.036422
	ENSG00000253138	LINC00967	*chr8:67,104,349–67,109,554*	0.0026941	9.72E-06	0.7471	0.039252
LUSC	ENSG00000259974	LINC00261	*chr20:22,541,191–22,559,280*	1.2676382	1.1	1.461	0.001087
	ENSG00000265933	LINC00668	*chr18:6,925,473–6,929,868*	0.8529687	0.7512	0.9685	0.01412
	ENSG00000224397	LINC01272	*chr20:48,884,015–48,896,333*	1.14005	1.022	1.272	0.018818

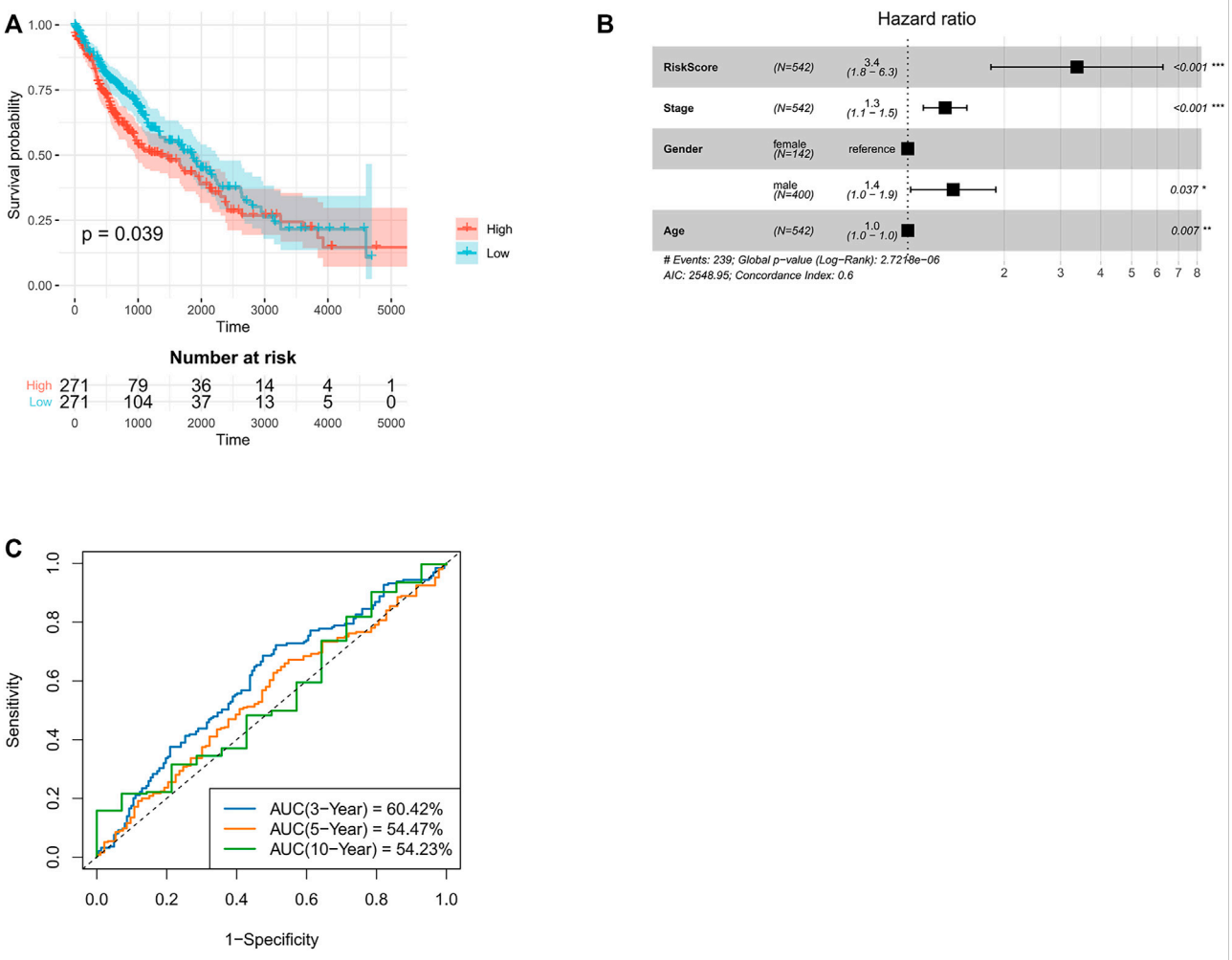


FIGURE 2
Validation of TANlncSig in TCGA lung squamous cell carcinoma discovery and testing datasets. **(A)** Kaplan-Meier survival analyses of lung squamous cell carcinoma patients. **(B)** Multivariate Cox regression analyses of patients with LUSC based on TCGA dataset. **(C)** ROC curve analyses of patients with LUSC using the TCGA dataset.

efficacy. The predictive capacity of TANlncSig was authenticated using the TCGA internal testing dataset and revealed the 3-, 5-, and 10-year OS rates for low-risk group patients to be 60.42, 54.47, and 54.23%, respectively (Figure 2C). Indicating that risk score significantly correlates with OS in LUSC.

Similar analysis was done for LUAD. First, three lncRNAs with prognostic values (LINC00528, LINC00967, and LINC00261) were identified using univariate Cox analysis. Using the above three lncRNAs, lncRNAs expression value was weighted by multivariate Cox

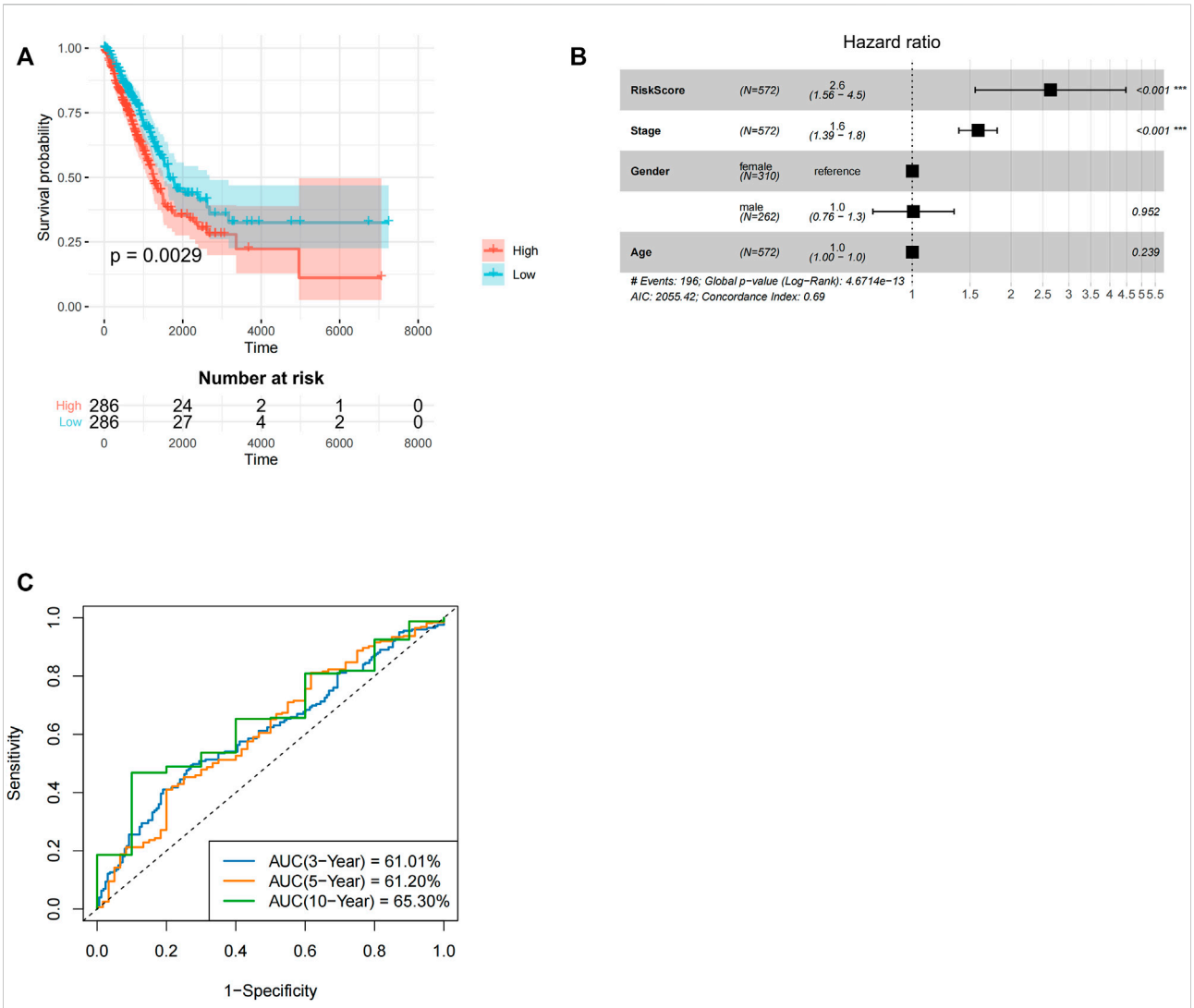


FIGURE 3
Development and subsequent validation of TANlncSig in the lung adenocarcinoma TCGA testing as well as discovery datasets. **(A)** Kaplan-Meier survival analyses for LUAD patients. **(B)** An analysis of the TCGA dataset of LUAD patients using multivariate Cox regression. **(C)** ROC curve analyses of patients with LUAD in the TCGA dataset.

regression coefficient to determine risk score using the formula: risk score = −5.32 * LINC00967-0.16 * LINC00261-0.74 * LINC00528. Patients with LUAD in the low-risk group had longer OS relative to high-risk group LUAD patients (p = 0.0029, ≤0.01, Figure 3A). Cox multivariate regression analyses revealed that risk score (p < 0.001) and stage (p < 0.001) significantly correlated with LUAD prognosis. In lung adenocarcinoma, the p-value and hazard ratio of TANlncSig were equally better than those of stage and age (Figure 3B). The 3-, 5-, and 10-year OS rates in low-risk group patients were 61.01, 61.20, and 65.30%, respectively (Figure 3C). These results indicate that risk scores in the LUAD dataset significantly correlate with patients' OS.

Correlation analysis between risk score and tumor clinical phenotype

Clinical phenotypic correlation analysis of single prognostic lncRNA and risk score (tumor stage, T, N, and M staging) was performed in lung adenocarcinoma as well as squamous cell carcinoma. According to statistical analysis, the risk score in different tumor stages of lung squamous cell carcinoma showed significant statistical differences, and the statistical results showed that p = 0.0013, <0.01 (Figure 4A). The risk score in different tumor stages of lung adenocarcinoma also showed significant statistical differences (p = 0.0081, <0.01) (Figure 4B).

The TNM staging system is the most widely used tumor staging system, worldwide. T denotes tumor sizes and local invasion range,

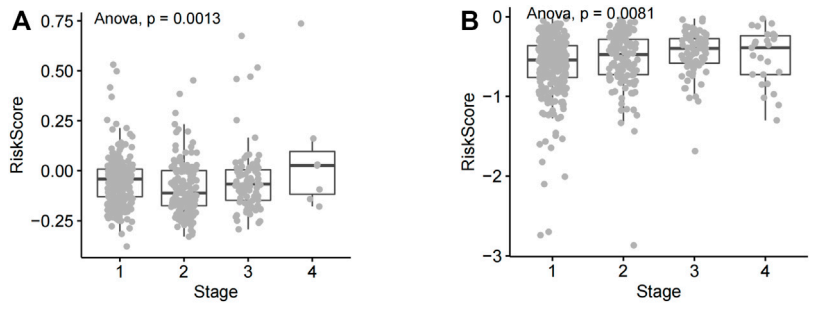


FIGURE 4
Analysis of risk score differences across NSCLC tumor stages. **(A)** The risk score of lung squamous cell carcinoma patients at various disease stages. **(B)** Risk scores of different lung adenocarcinoma stages.

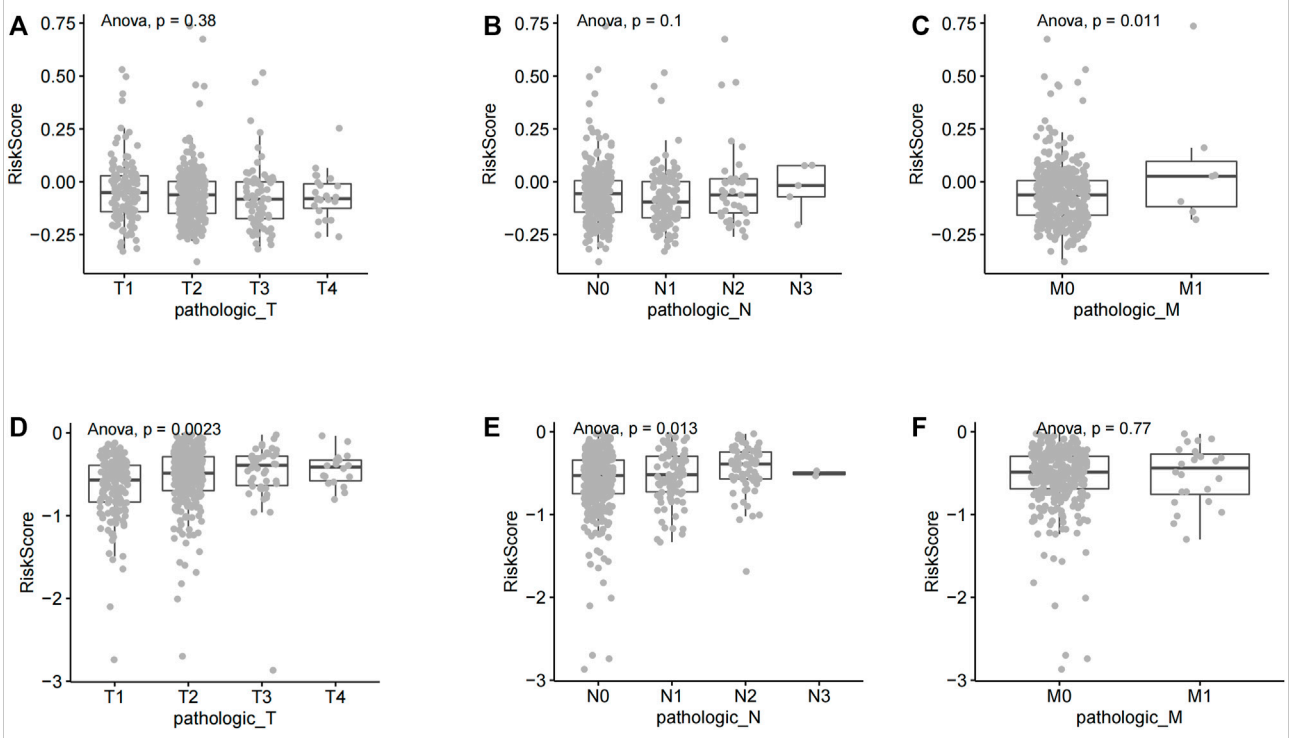


FIGURE 5
Correlation between riskscore and different TNM stages of non-small cell lung cancer. **(A)** A comparative analysis of risk scores in LUSC T staging. **(B)** A comparative analysis of risk scores in LUSC N staging. **(C)** A comparative analysis of risk scores in LUSC M staging. **(D)** A comparative analysis of risk scores in LUAD T staging **(E)** A comparative analysis of risk scores in LUAD N staging. **(F)** A comparative analysis of risk scores in LUAD M staging.

N denotes lymph node involvement, and M denotes distant metastasis. TNM staging has great clinical value in prognosis prediction (Ficarra et al., 2007; Moch et al., 2009). The risk score lack of significance in different T stages and N stages of lung squamous cell carcinoma (Figures 5A,B). The risk score has significant statistical difference in different M stages of lung squamous cell carcinoma (p = 0.011, <0.05) (Figure 5C). The risk score has significant statistical difference in different T stages (T1, T2, T3, and T4 stages) of lung adenocarcinoma (p = 0.0023, <0.01) (Figure 5D). Similarly, the risk score has significant statistical difference in different N stages (N0, N1, N2, and N3 stages) of lung adenocarcinoma (p = 0.013, <0.05) (Figure 5E). The risk score lack of significance in different M stages of lung adenocarcinoma (Figure 5F).

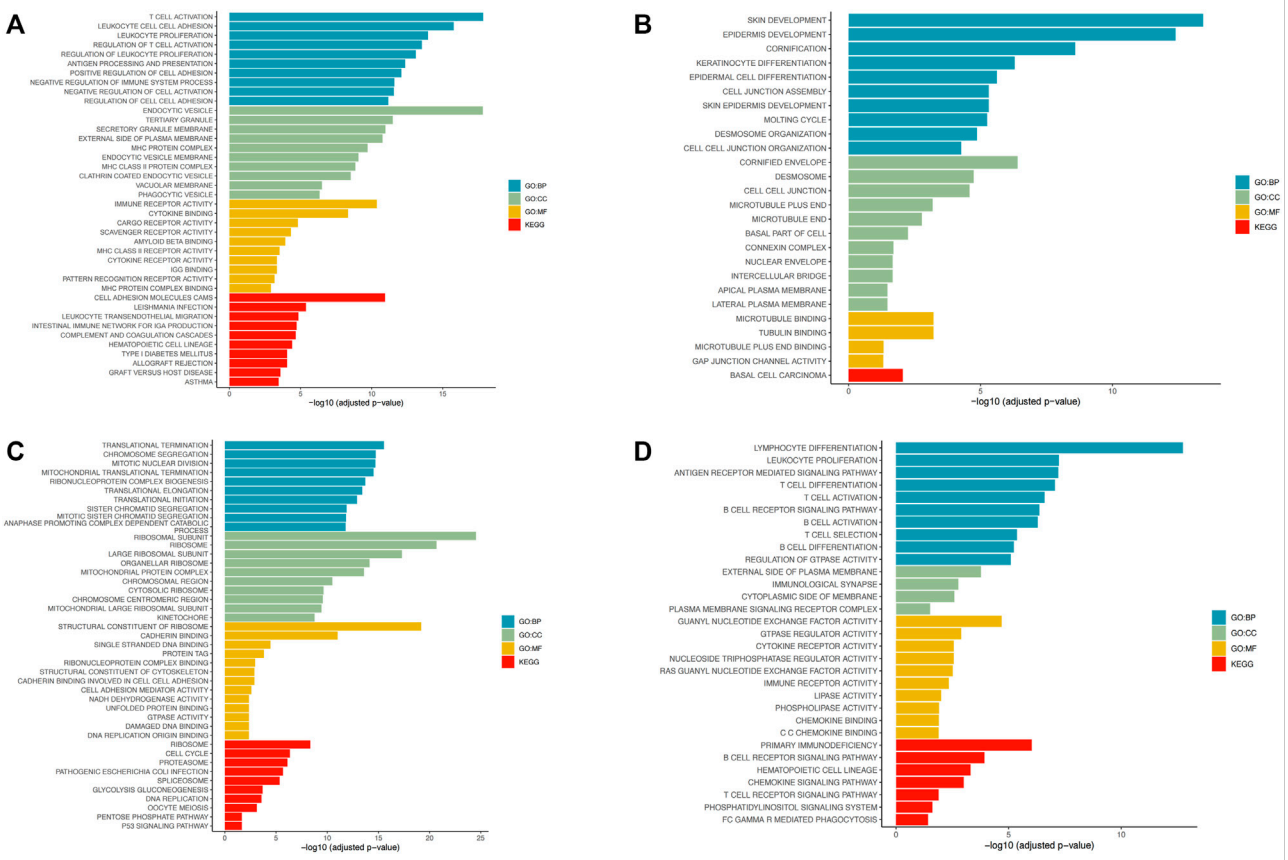


FIGURE 6
Pathway enrichment analysis using R's cluster profiler. **(A)** Enrichment of positive correlation gene pathways in lung squamous cell carcinoma. **(B)** Enrichment of negative correlation gene pathways in lung squamous cell carcinoma. **(C)** Enrichment of positively correlation gene pathways in lung adenocarcinoma. **(D)** Enrichment of negative correlation gene pathways in lung adenocarcinoma.

Riskscore correlation pathway analysis

In LUSC and LUAD samples, genes with average expression levels >1 were identified and their risk scores analyzed. 1,000 genes with the largest absolute correlation coefficient values were selected from positive (correlation coefficient >0, $p \leq 0.05$) and negative (correlation coefficient <0, $p \leq 0.05$) and correlation genes and pathway enrichment analysis done using cluster profiler on R. In LUSC, positive correlation genes are mainly associated with biological processes (BP) associated with T-cell activation, leukocyte proliferation, and leukocyte cell-cell adhesion. For cellular component (CC) they were enriched in endocytic vesicle, tertiary granule, and secretory granule membrane. For molecular function (MF), they were enriched in immune receptor activity and cytokine binding. KEGG pathway analysis revealed enrichment mainly for cell adhesion molecules cams (Figure 6A). Negative correlation genes in lung squamous cell carcinoma are mainly enriched for biological processes (BP) associated with skin development, epidermis development, and cornification. For cellular component (CC), they were enriched for cornified envelope, desmosome, and cell-cell junction. For molecular function (MF), they were enriched for microtubule binding and tubulin binding. For KEGG pathways, they were enriched for basal cell carcinoma (Figure 6B). Positive correlation genes in lung adenocarcinoma were mainly enriched in biological processes (BP) associated with translational termination and adenocarcinoma. For cellular component (CC), they were enriched for ribosomal subunits, ribosome and large ribosomal subunit. For molecular function (MF) they were enriched for structural constituent of ribosome and cadherin binding. For KEGG pathways, they were enriched for ribosome and cell cycle (Figure 6C). Genes associated with negative correlations in LUAD are involved in biological processes (BPs) associated with lymphocyte differentiation, leukocyte proliferation, and antigen receptor-mediated signaling. For cellular component (CC), they were enriched for external side of plasma membrane and immunological synapse. For molecular functions (MFs), they were enriched for guanyl-nucleotide exchange factor activity. For KEGG pathways, they were

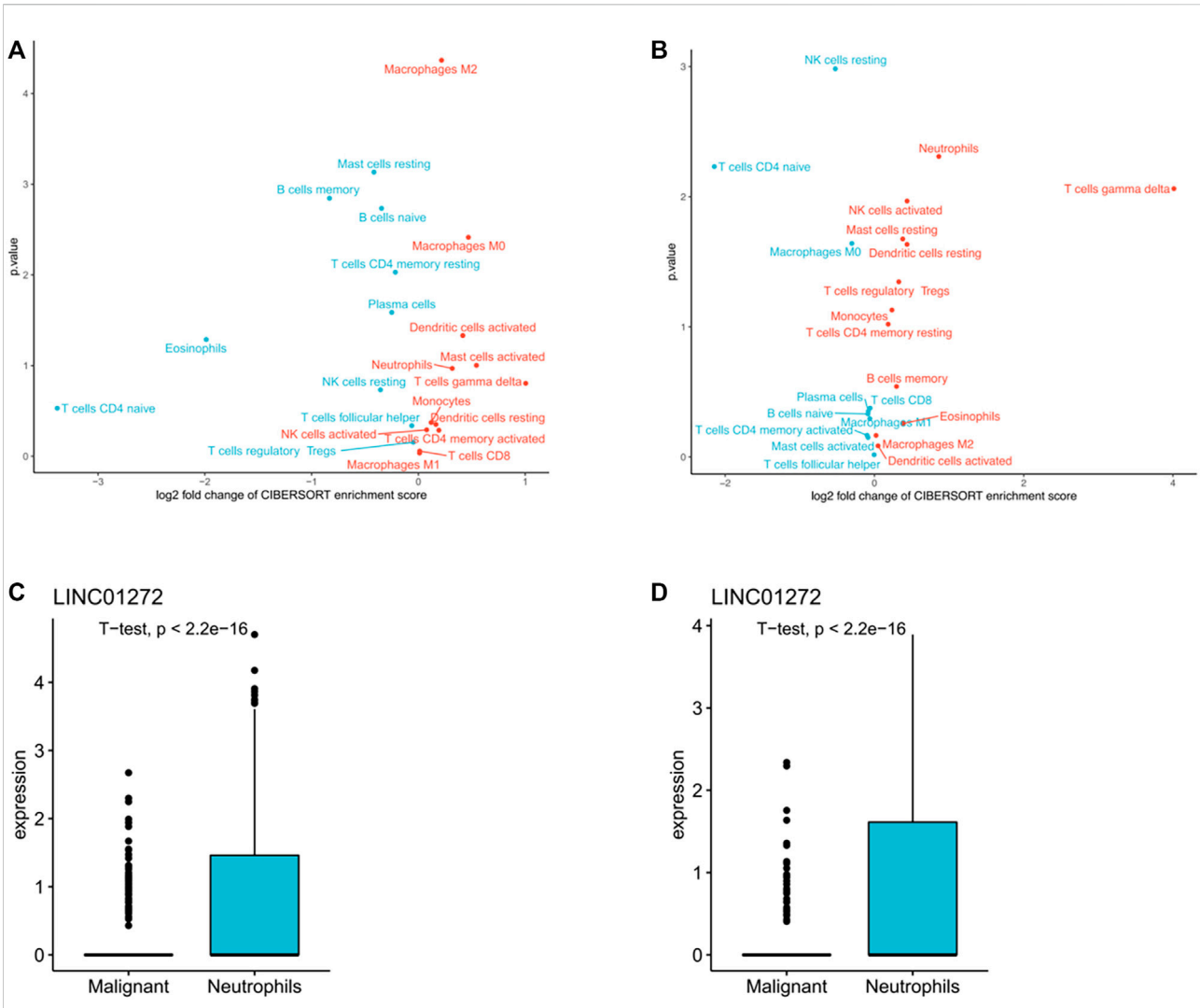


FIGURE 7
(A,B) An analysis of tumors with high and low TANlncSig based on the NES score from the GSEA in LUSC and LUAD provided the volcano plots for the enrichment of immune cell subpopulations. **(C,D)** The boxplots show lncRNA expression in both LUSC and LUAD neutrophil cell lines.

enriched for primary immunodeficiency and B-cell receptor signaling pathway (Figure 6D).

The TANlncSig associates with tumor-associated neutrophils

In accordance with previously reported expression levels of the immune cell specific marker genes, cibersort (https://cibersort.stanford.edu/) was further used to evaluate the levels of immune infiltration of 22 immune subpopulations in high-risk and low-risk patient groups. *t*-test was performed to determine the difference in lymphocyte infiltration levels between the two groups. As shown in Figures 7A,B, in both LUSC and LUAD, high-risk patients were significantly enriched in 12 immune subpopulations, while low-risk patients were enriched in 10 immune subpopulations. Additionally, mononuclear immune cells, including neutrophils, were found to infiltrate significantly more in the high-risk patient group than in several other groups. Single-cell sequencing data of LUSC and LUAD downloaded from GSE127465, cell type notes downloaded from TISCH (http://tisch.comp-genomics.org/). The homologous expression levels of LINC01272 of the TANlncSig in neutrophil cell lines differed significantly from those of malignant cell lines according to a subsequent analysis of neutrophil cell lines (Figures 7C,D). This indicates that these

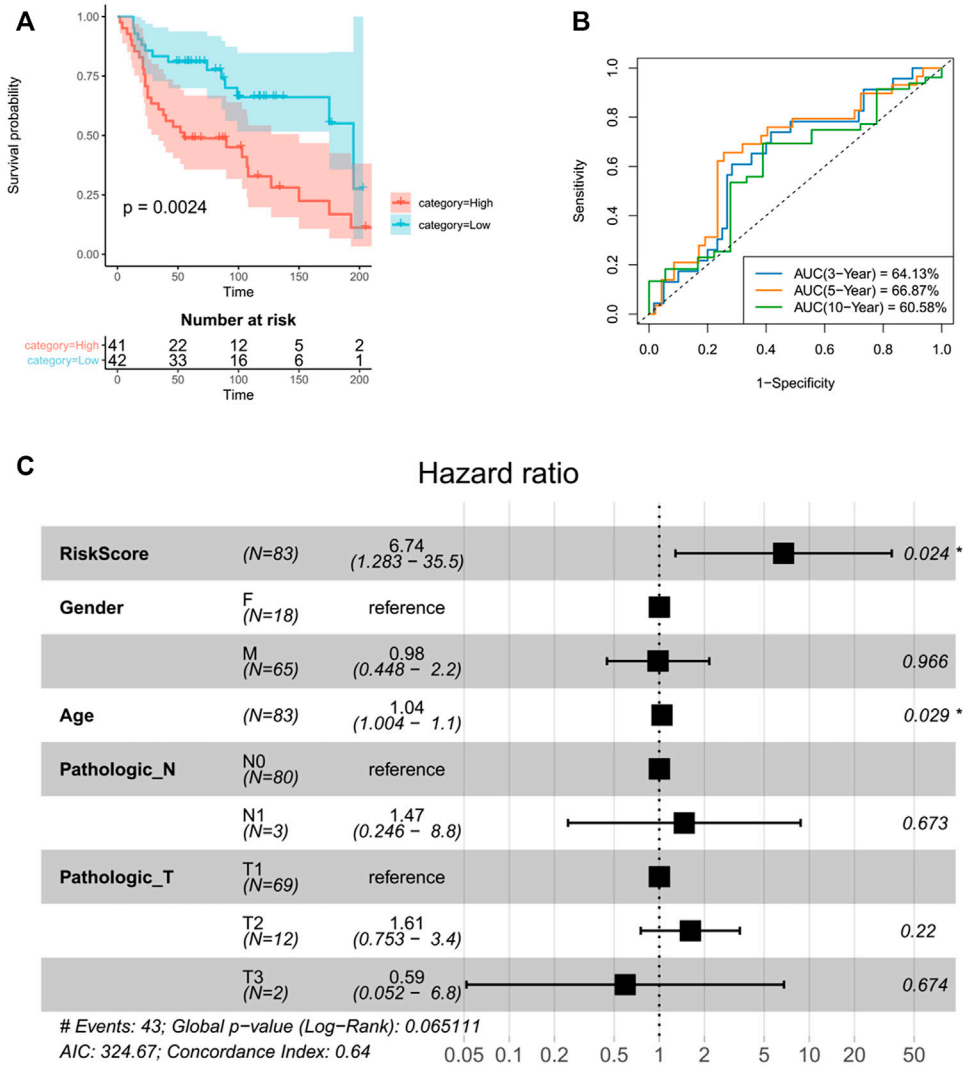


FIGURE 8
The TANlncSig was independently validated in the GSE30219 dataset **(A)**. Kaplan–Meier survival curves of OS were plotted between high- and low-risk groups stratified by the TANlncSig. **(B)**Time ROC curve of luad patients the GSE30219 dataset. **(C)** Visualization of the HRs from a multivariate Cox analysis of the TANlncSig and clinicopathological factors in GSE30219.

lncRNAs are expressed differently in neutrophils compared with malignant cells. In the above study, we found that the TANlncSig was not only associated with patient prognosis but also as a TAN indicator.

TANlncSig was validated over several independent datasets using a microarray platform for prognostic value

TANlncSig was further validated in independent datasets by the microarray platform in order to verify versatility and robustness of TANlncSig. The Affymetrix HG-U133 Plus 2.0 platform was used to analyze 83 LUAD patients from the GSE30219 dataset. As demonstrated again, TANlncSig can distinguish between patients who have high and low survival risk. A total of 83 patients were stratified into 41 high-risk patients and 42 low-risk patients in the GSE30219 dataset. Furthermore, patients in the high-risk group had a marginally poorer outcome than those in the low-risk group (p = 0.0024, ≤0.01; log-rank test) (Figure 8A). The AUC of ROC curve at 3, 5, and 10 years were 64.13, 66.87, and 60.58% respectively (Figure 8B). The results show that TANlncsig can accurately predict the 5-year overall survival of patients, indicating that TANlncsig has good efficacy and certain stability. In order to investigate whether TANlncSig is an independent prognostic factor, a multivariate Cox regression analysis was conducted in

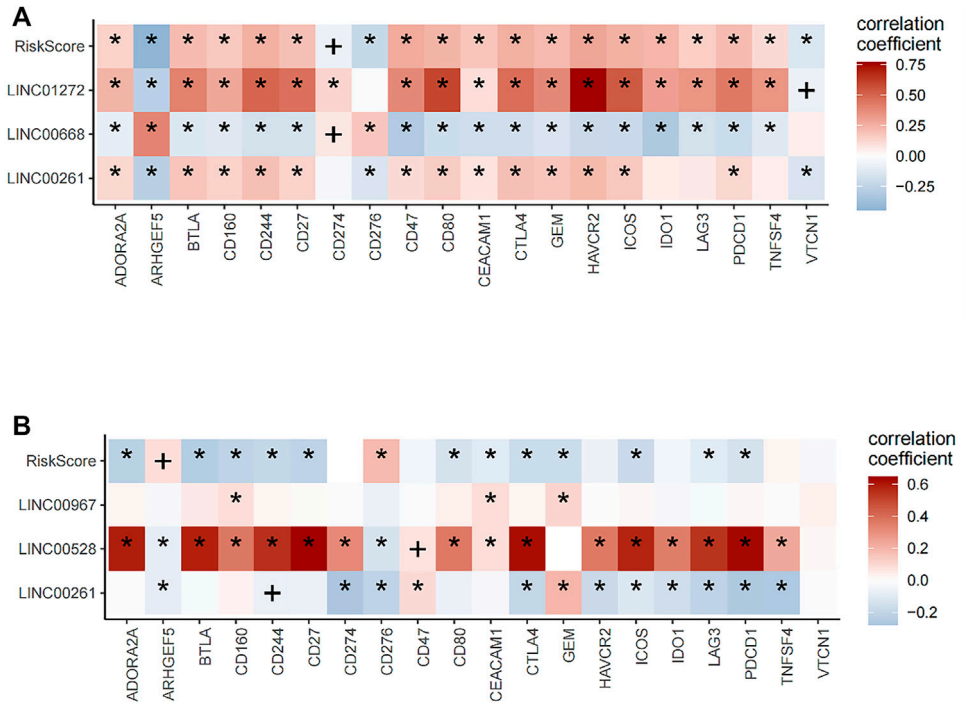


FIGURE 9
Correlation analysis of lncRNA and riskScore with expression levels of immune checkpoint blockade (ICB). **(A)** In lung squamous cell carcinoma, risk score, LINC01272, and LINC00261 positively correlated with the expression of most ICBs, while LINC00668 had negative correlations with the expression of most ICBs. **(B)** In lung adenocarcinoma, risk score had a negative correlation with the expression of most ICBs, while LINC00528 positively correlated with the expression of most ICBs.

patient cohorts. In the independent GSE30219 dataset, the TANlncSig still maintained a significant correlation with OS in the multivariate analysis (HR = 6.74, 95% CI 1.283-35.5, $p = 0.024$, ≤0.01). Thus, these results demonstrate that the TANlncSig helps predict OS independently of other conventional clinical factors (Figure 8C).

Significance of TANlncSig as a marker of immunotherapy

Next, prognostic lncRNAs and risk score were correlated with immune checkpoint molecules expression in LUSC and LUAD patients. In LUSC, risk score, LINC01272, and LINC00261 positively correlated with the expression of most ICBs, while LINC00668 had negative correlations with the expression of most ICBs (Figure 9A). In LUAD, risk score had negative correlations with expression levels of most ICBs, while LINC00528 positively correlated with expression levels of most ICBs (Figure 9B). The expressions of risk score were divided into high and low groups and combined according to the median. The combination was used to analyze the prognosis of immunotherapy for non-small cell lung cancer. In lung squamous cell carcinoma, the combination of CEACAM1, TNFSF4, gem, CD47, vtcn1 and risk score can well stratify the prognosis of patients. In lung adenocarcinoma, all ICB molecules combined with risk score can well predict the prognosis of patients. These results suggest that risk score can be used as an index to predict the response of patients to immunotherapy.

Discussion

In the peripheral blood, neutrophils are the most abundant white blood cells (Dinh et al., 2020). They have a central role in human non-specific immunity. Previous studies suggest that neutrophils inhibit tumors by secreting cytokines and producing reactive oxygen species (Vaughan and Walsh, 2005; Mishalian et al., 2013; Coffelt et al., 2015; Ponzetta et al., 2019). However, other studies indicate that neutrophils in the tumor microenvironment (TME) promote tumorigenesis. Cytokines and chemokines production by invasive neutrophils might affect the recruitment and activation of inflammatory cells in the TME, create an immunosuppressive microenvironment that is conducive for tumorigenesis, regulate tumor growth,

metastasis and angiogenesis, and influence patient prognosis. Traditional methods for quantifying tumor immune cells infiltration based on histology or immunohistochemistry may have bias and variabilities (Yoshihara et al., 2013; Gibney et al., 2016; Spranger and Gajewski, 2018; Zhang et al., 2020; Sanchez-Pino et al., 2021). More recently, RNA-seq analyses have shown that lncRNAs exhibit a better degree of cell type specificity, relative to protein-coding genes in immune cells, highlighting their potential as subpopulation-specific immune cells molecular markers (Huang et al., 2018; Chen et al., 2019; Zhou et al., 2021).

Here, we used a machine learning-based computational framework to identify lncRNA features for evaluating TANs and explored their clinical significance using a combination of lncRNA, immune, and clinical spectrum analyses. The computational framework was used on the TCGA discovery dataset of NSCLC to identify a lncRNA signature (TANlncSig) comprised of 17 lncRNAs obtained from a list of neutrophil-specific lncRNAs using machine learning. Functional enrichment analysis of TANlncSig-related mRNAs showed that TANlncSig is highly correlated with cancer markers of immune response and sustained proliferative signals. Recent experimental evidence on some TANlncSig components is consistent with functional annotations using bioinformatics. It appears that Mir-1303, which is upregulated in tumor tissues, acts as a sponge for LINC01272 and negatively correlates with its expression. A reduction in LINC01272 expression in tissues and cells of NSCLC patients may serve as an independent prognostic marker. LINC01272 overexpression may inhibit NSCLC cells proliferation, migration, and invasion by inhibiting MI-1303 (Zhang and Zhou, 2021). LINC00261 downregulation in gastric cancer is associated with poor prognosis. Ectopic LINC00261 expression disrupts cell migration and invasion, inhibiting metastasis *in vitro* as well as *in vivo*. LINC00261 downregulation promotes cell migration and invasion *in vitro*. LINC00261 overexpression influences epithelial-mesenchymal transition (EMT) through the regulation of E-cadherin, Vimentin and N-cadherin (Liu et al., 2020; Zhai et al., 2021). LINC00668 expression is significantly upregulated *via* STAT3 signaling in NSCLC tissues as well as cell lines. Clinical studies show that upregulated LINC00668 correlates with histological grade, advanced TNM stage, and lymph node metastasis. Additionally, multivariate analyses established that LINC00668 as an independent marker of overall survival (OS) in patients with NSCLC. LINC00668 downregulation inhibits proliferation, migration, and invasion of NSCLC cells and promotes apoptosis. Mechanistically, LINC00668 is a direct target of miR-193a, leading to down-regulation in the expression of its target gene KLF7. STAT3-initiated LINC00668 promotes NSCLC progression by upregulating KLF7 *via* sponging Mir-193a. Therefore, it may serve as a prognostic marker and therapeutic target for NSCLC (An et al., 2019). From the perspective of lncRNA, TANlncSig seems to be a transcriptional marker as a potentially measurable indicator of neutrophil activity and prognosis.

To further assess TANlncSig's role in clinical risk stratification, we evaluated its relationship with survival in patients with NSCLC. When applied to the TCGA RNAseq patient dataset, TANlncSig significantly correlated with patient survival. In TANlncSig, three lung squamous cell carcinoma, neutrophil-specific lncRNAs (LINC01272, LINC00261, and LINC00668) were markedly associated with prognostic outcomes. In lung adenocarcinoma, three neutrophil-specific lncRNAs (LINC00528, LINC00967, and LINC00261) significantly correlated with prognosis. In squamous cell carcinoma and lung adenocarcinoma, correlation analysis of individual lncRNAs and risk score with clinical features (TNM staging) revealed that risk score varied significantly with tumor stage. After adjusting for traditional clinical factors, TANlncSig was verified to be an independent prognostic marker for differentiating between poor and good survival outcomes across patient datasets.

Immune checkpoint inhibitors (ICIs) have emerged as effective lung cancer immunotherapies (Suresh et al., 2018; Iams et al., 2020). Some of the drugs acting on the immune checkpoints, CTLA4 and PD-1/PD-L1, have excellent performance against various tumors. Although significant breakthroughs have been made on CTLA4 and PD-1/PD-L1 antibodies, single-drug effective rates are only about 20%, and they benefit a limited proportion of patients (Magiera-Mularz et al., 2017; Lingel and Brunner-Weinzierl, 2019; Rotte, 2019; Yang and Hu, 2019; Liu and Zheng, 2020). The limited efficacy is attributable to the immune system's complexity. Indeed, immune cells, cytokines, and immune adjuvants in the TME interact with each other, limiting the effects of drugs on individual targets. Thus, drugs that target different links and aspects of tumor immunity are needed to enhance immunotherapy outcomes. Up to 29 immunoglobulin superfamily members and 26 members of the tumor necrosis factor receptor superfamily are expressed on T-cell surfaces alone, and there have been preclinical or clinical studies on related immune targets and drugs. Specific immune checkpoints include lymphocyte activating gene 3 (LAG-3), T-cell immunoglobulin mucin 3 (TIM-3), and V region Ig inhibitor (VISTA). Non-specific immune checkpoints include human killer cell immunoglobulin like receptor (KIR), indoleamine 2, 3-dioxidase (IDO), and CD47, these novel immune checkpoint molecules are expected to provide hints for clinical and basic research (Manser et al., 2015; Munn and Mellor, 2016; Burugu et al., 2018; Huang et al., 2020; Logtenberg et al., 2020). VISTA, (B7-H5, PD-1H) is an immunomodulatory receptor that inhibits T-cell response. VISTA is overexpressed on CD11b myeloid cells (e.g., macrophages, monocytes, neutrophils, and dendritic cells)

and it is found that in humans and mice at a lower level in primitive $CD4^+$ and $CD8^+$ T-cells as well as Tregs. With two potential protein kinase C binding sites and proline residues acting as docking sites in its cytoplasmic tail domain, VISTA can serve as both a receptor and a ligand (Huang et al., 2020; Mutsaers et al., 2021). OX40 (TNFRSF4) has been found to be expressed in activated NK cells, T-cells, NKT cells, as well as neutrophils, and acts as an auxiliary costimulatory immune checkpoint (Curti et al., 2013; Aspeslagh et al., 2016; Buchan et al., 2018). Combining immune checkpoint genes and TANlncSig showed combined prognostic effects on patient survival, in line with previous findings that immunomotor interactions between neutrophilic infiltration and expression levels of checkpoint genes affect the outcome of cancer patients and immunotherapy may also be associated with this condition. In combination with earlier findings, it appears that TANlncSig is correlated with immunosuppressive phenotypes and could predict ICI response. Together, these results indicate that TANlncSig can complement and/or add information to existing immune checkpoint genetic markers.

Due to few gene mutations, lung squamous cell carcinoma is less selective than adenocarcinoma with regards to treatment options, and its survival time (about 1 year) is shorter than that of adenocarcinoma (Travis et al., 2021). Thus, novel, effective advanced lung squamous cell carcinoma treatments are needed to improve patient outcomes. The emergence of immune checkpoint inhibitors in recent years has markedly improved treatment options for advanced lung squamous cell carcinoma patients. Immune checkpoint inhibitors have substantially changed advanced lung squamous cell carcinoma treatment, leading to a shift from retro line immunotherapy to front-line treatment options. Originally approved as second-line treatment after platinum-based dual therapy, palivizumab is now recommended as a single-agent first-line treatment or in combination with chemotherapy. Although treatments targeting the immune checkpoints PD-1 and CTLA4 are successful in many cancers, not all patients benefit from them. Our findings indicate that the combination of CEACAM1, TNFSF4, GEM, CD47, VTCN1, and TANlncSig in squamous cell carcinoma can effectively stratify patients by prognosis, highlighting these immune checkpoint receptors as potential therapeutic targets against advanced lung cancer.

Conclusion

In conclusion, we used a machine learning-based computational framework to identify lncRNA features of TANs (TANlncSig) *via* comprehensive analyses of lncRNA, immune, as well as clinical features. TANlncSig revealed a substantial and repeatable correlation with outcomes, even after adjustments of clinical covariates. Analysis of correlation between prognostic lncRNAs and risk score with the expression of immune checkpoint molecules demonstrated that TANlncSig can predict immunotherapy. The study is the first to define lncRNA characteristics of tumor-associated neutrophils, highlighting the importance of lncRNAs in immune responses and the potential for more precise and personalized treatment cancer immunotherapy.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

ZT, QW, PC, and HG designed the experiments; ZT, QW, and JS performed the experiments; PC, HG, and YP prepared figures; ZT, QW, CL, and CZ was responsible for statistical analysis and provided helpful suggestions; ZT and QW wrote the manuscript. All authors read and approved the final version of the article.

Funding

This work was Funded by the National Natural Science Foundation of China (No. 81871865); Science and Technology Commission of Shanghai Municipality (No. 19411950300); Shanghai Innovative Collaboration Project (No. 2020CXJQ02); Natural Science Foundation of Tianjin (No. 19JCQNJC12300); “Dream Tutor” New Person Cultivation Program of Shanghai Pulmonary Hospital (No. fkxr 1907).

Conflict of interest

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

Publisher’s note

References

An, Y. X., Shang, Y. J., Xu, Z. W., Zhang, Q. C., Wang, Z., Xuan, W. X., et al. (2019). STAT3-induced long noncoding RNA LINC00668 promotes migration and invasion of non-small cell lung cancer via the miR-193a/KLF7 axis. *Biomed. Pharmacother.* 116, 109023. doi:10.1016/j.biopha.2019.109023

Aspeslagh, S., Postel-Vinay, S., Rusakiewicz, S., Soria, J. C., Zitvogel, L., and Marabelle, A. (2016). Rationale for anti-OX40 cancer immunotherapy. *Eur. J. Cancer* 52, 50–66. doi:10.1016/j.ejca.2015.08.021

Atianand, M. K., Caffrey, D. R., and Fitzgerald, K. A. (2017). Immunobiology of long noncoding RNAs. *Annu. Rev. Immunol.* 35, 177–198. doi:10.1146/annurev-immunol-041015-055459

Bao, S., Zhao, H., Yuan, J., Fan, D., Zhang, Z., Su, J., et al. (2020). Computational identification of mutator-derived lncRNA signatures of genome instability for improving the clinical outcome of cancers: A case study in breast cancer. *Brief. Bioinform.* 21 (5), 1742–1755. doi:10.1093/bib/bbz118

Buchan, S. L., Rogel, A., and Al-Shamkhani, A. (2018). The immunobiology of CD27 and OX40 and their potential as targets for cancer immunotherapy. *Blood* 131 (1), 39–48. doi:10.1182/blood-2017-07-741025

Burugu, S., Dancsok, A. R., and Nielsen, T. O. (2018). Emerging targets in cancer immunotherapy. *Semin. Cancer Biol.* 52 (2), 39–52. doi:10.1016/j.semcancer.2017.10.001

Butturini, E., Carcereri de Prati, A., Boriero, D., and Mariotto, S. (2019). Tumor dormancy and interplay with hypoxic tumor microenvironment. *Int. J. Mol. Sci.* 20, 4305. doi:10.3390/ijms20174305

Chen, F., Chen, J., Yang, L., Liu, J., Zhang, X., Zhang, Y., et al. (2019). Extracellular vesicle-packaged HIF-1α-stabilizing lncRNA from tumour-associated macrophages regulates aerobic glycolysis of breast cancer cells. *Nat. Cell Biol.* 21 (4), 498–510. doi:10.1038/s41556-019-0299-0

Chen, R., Manochakian, R., James, L., Azzouqa, A. G., Shi, H., Zhang, Y., et al. (2020). Emerging therapeutic agents for advanced non-small cell lung cancer. *J. Hematol. Oncol.* 13 (1), 58. doi:10.1186/s13045-020-00881-7

Chen, Y. G., Satpathy, A. T., and Chang, H. Y. (2017). Gene regulation in the immune system by long noncoding RNAs. *Nat. Immunol.* 18 (9), 962–972. doi:10.1038/ni.3771

Coffelt, S. B., Kersten, K., Doornebal, C. W., Weiden, J., Vrijland, K., Hau, C. S., et al. (2015). IL-17-producing γδ T cells and neutrophils conspire to promote breast cancer metastasis. *Nature* 522 (7556), 345–348. doi:10.1038/nature14282

Curti, B. D., Kovacsovics-Bankowski, M., Morris, N., Walker, E., Chisholm, L., Floyd, K., et al. (2013). OX40 is a potent immune-stimulating target in late-stage cancer patients. *Cancer Res.* 73 (24), 7189–7198. doi:10.1158/0008-5472.CAN-12-4174

Das, S., and Johnson, D. B. (2019). Immune-related adverse events and anti-tumor efficacy of immune checkpoint inhibitors. *J. Immunother. Cancer* 7 (1), 306. doi:10.1186/s40425-019-0805-8

Dinh, H. Q., Eggert, T., Meyer, M. A., Zhu, Y. P., Olingy, C. E., Llewellyn, R., et al. (2020). Coexpression of CD71 and CD117 identifies an early unipotent neutrophil progenitor population in human bone marrow. *Immunity* 53 (2), 319–334. doi:10.1016/j.immuni.2020.07.017

Dolladille, C., Ederhy, S., Sassier, M., Cautela, J., Thuny, F., Cohen, A. A., et al. (2020). Immune checkpoint inhibitor rechallenge after immune-related adverse events in patients with cancer. *JAMA Oncol.* 6 (6), 865–871. doi:10.1001/jamaoncol.2020.0726

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: New players in cell differentiation and development. *Nat. Rev. Genet.* 15 (1), 7–21. doi:10.1038/nrg3606

Ficarra, V., Galfano, A., Mancini, M., Martignoni, G., and Artibani, W. (2007). TNM staging system for renal-cell carcinoma: Current status and future perspectives. *Lancet. Oncol.* 8 (6), 554–558. doi:10.1016/S1470-2045(07)70173-0

Galluzzi, L., Humeau, J., Buqué, A., Zitvogel, L., and Kroemer, G. (2020). Immunostimulation with chemotherapy in the era of immune checkpoint inhibitors. *Nat. Rev. Clin. Oncol.* 17 (12), 725–741. doi:10.1038/s41571-020-0413-z

Gibney, G. T., Weiner, L. M., and Atkins, M. B. (2016). Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *Lancet. Oncol.* 17 (12), e542–e551. doi:10.1016/S1470-2045(16)30406-5

Hu, G., Tang, Q., Sharma, S., Yu, F., Escobar, T. M., Muljo, S. A., et al. (2013). Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nat. Immunol.* 14, 1190–1198. doi:10.1038/ni.2712

Huang, D., Chen, J., Yang, L., Ouyang, Q., Li, J., Lao, L., et al. (2018). NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death. *Nat. Immunol.* 19 (10), 1112–1125. doi:10.1038/s41590-018-0207-y

Huang, X., Zhang, X., Li, E., Zhang, G., Wang, X., Tang, T., et al. (2020). Vista: An immune regulatory protein checking tumor and immune cells in cancer immunotherapy. *J. Hematol. Oncol.* 13 (1), 83. doi:10.1186/s13045-020-00917-y

Iams, W. T., Porter, J., and Horn, L. (2020). Immunotherapeutic approaches for small-cell lung cancer. *Nat. Rev. Clin. Oncol.* 17 (5), 300–312. doi:10.1038/s41571-019-0316-z

Khoja, L., Day, D., Wei-Wu Chen, T., Siu, L. L., and Hansen, A. R. (2017). Tumour- and class-specific patterns of immune-related adverse events of immune checkpoint inhibitors: A systematic review. *Ann. Oncol.* 28 (10), 2377–2385. doi:10.1093/annonc/mdx286

Ko, E. C., Raben, D., and Formenti, S. C. (2018). The integration of radiotherapy with immunotherapy for the treatment of non-small cell lung cancer. *Clin. Cancer Res.* 24 (23), 5792–5806. doi:10.1158/1078-0432.CCR-17-3620

Lecot, P., Sarabi, M., Pereira Abrantes, M., Mussard, J., Koenderman, L., Caux, C., et al. (2019). Neutrophil heterogeneity in cancer: From biology to therapies. *Front. Immunol.* 10, 2155. doi:10.3389/fimmu.2019.02155

Lingel, H., and Brunner-Weinzierl, M. C. (2019). CTLA-4 (CD152): A versatile receptor for immune-based therapy. *Semin. Immunol.* 42, 101298. doi:10.1016/j.smim.2019.101298

Liu, S., Zheng, Y., Zhang, Y., Zhang, J., Xie, F., Guo, S., et al. (2020). Methylation-mediated LINC00261 suppresses pancreatic cancer progression by epigenetically inhibiting c-Myc transcription. *Theranostics* 10 (23), 10634–10651. doi:10.7150/thno.44278

Liu, Y., and Zheng, P. (2020). Preserving the CTLA-4 checkpoint for safer and more effective cancer immunotherapy. *Trends Pharmacol. Sci.* 41 (1), 4–12. doi:10.1016/j.tips.2019.11.003

Logtenberg, M. E. W., Scheeren, F. A., and Schumacher, T. N. (2020). The CD47-sirpα immune checkpoint. *Immunity* 52 (5), 742–752. doi:10.1016/j.immuni.2020.04.011

Magiera-Mularz, K., Skalniak, L., Zak, K. M., Musielak, B., Rudzinska-Szostak, E., Berlicki, L., et al. (2017). Bioactive macrocyclic inhibitors of the PD-1/PD-L1 immune checkpoint. *Angew. Chem. Int. Ed. Engl.* 56 (44), 13732–13735. doi:10.1002/anie.201707707

Manser, A. R., Weinhold, S., and Uhrberg, M. (2015). Human KIR repertoires: Shaped by genetic diversity and evolution. *Immunol. Rev.* 267 (1), 178–196. doi:10.1111/imr.12316

Marchese, F. P., Raimondi, I., and Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.* 18 (1), 206. doi:10.1186/s13059-017-1348-2

Mishalian, I., Bayuh, R., Levy, L., Zolotarov, L., Michaeli, J., and Fridlender, Z. G. (2013). Tumor-associated neutrophils (TAN) develop pro-tumorigenic properties during tumor progression. *Cancer Immunol. Immunother.* 62 (11), 1745–1756. doi:10.1007/s00262-013-1476-9

Moch, H., Artibani, W., Delahunt, B., Ficarra, V., Knuechel, R., Montorsi, F., et al. (2009). Reassessing the current UICC/AJCC TNM staging for renal cell carcinoma. *Eur. Urol.* 56 (4), 636–643. doi:10.1016/j.eururo.2009.06.036

Munn, D. H., and Mellor, A. L. (2016). Ido in the tumor microenvironment: Inflammation, counter-regulation, and tolerance. *Trends Immunol.* 37 (3), 193–207. doi:10.1016/j.it.2016.01.002

Mutsaers, P., Balcioglu, H. E., Kuiper, R., Hammerl, D., Wijers, R., van Duin, M., et al. (2021). V-domain Ig suppressor of T cell activation (VISTA) expression is an independent prognostic factor in multiple myeloma. *Cancers (Basel)* 13 (9), 2219. doi:10.3390/cancers13092219

Nielsen, S. R., Strøbech, J. E., Horton, E. R., Jackstadt, R., Laitala, A., Bravo, M. C., et al. (2021). Suppression of tumor-associated neutrophils by lorlatinib attenuates pancreatic cancer growth and improves treatment with immune checkpoint blockade. *Nat. Commun.* 12 (1), 3414. doi:10.1038/s41467-021-23731-7

Ponzetta, A., Carriero, R., Carnevale, S., Barbagallo, M., Molgora, M., Perucchini, C., et al. (2019). Neutrophils driving unconventional T cells mediate resistance against murine sarcomas and selected human tumors. *Cell* 178 (2), 346–360. doi:10.1016/j.cell.2019.05.047

Ranzani, V., Rossetti, G., Panzeri, I., Arrigoni, A., Bonnal, R. J., Curti, S., et al. (2015). The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nat. Immunol.* 16, 318–325. doi:10.1038/ni.3093

Rinn, J. L., and Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81, 145–166. doi:10.1146/annurev-biochem-051410-092902

Rotte, A. (2019). Combination of CTLA-4 and PD-1 blockers for treatment of cancer. *J. Exp. Clin. Cancer Res.* 38 (1), 255. doi:10.1186/s13046-019-1259-z

Sage, A., Ng, K., Marshall, E., Enfield, K., Stewart, G., Martin, S., et al. (2018). MA24.06 long non-coding rna expression patterns delineate infiltrating immune cells in the lung tumour microenvironment. *J. Thorac. Oncol.* 13, S443–S444. doi:10.1016/j.jtho.2018.08.524

Sanchez-Pino, M. D., Gilmore, L. A., Ochoa, A. C., and Brown, J. C. (2021). Obesity-associated myeloid immunosuppressive cells, key players in cancer risk and response to immunotherapy. *Obes. (Silver Spring)* 29 (6), 944–953. doi:10.1002/oby.23108

Schoenfeld, A. J., Arbour, K. C., Rizvi, H., Iqbal, A. N., Gadgeel, S. M., Girshman, J., et al. (2019). Severe immune-related adverse events are common with sequential PD-(L)1 blockade and osimertinib. *Ann. Oncol.* 30 (5), 839–844. doi:10.1093/annonc/mdz077

Schupp, J., Krebs, F. K., Zimmer, N., Trzeciak, E., Schuppan, D., and Tuettenberg, A. (2019). Targeting myeloid cells in the tumor sustaining microenvironment. *Cell. Immunol.* 343, 103713. doi:10.1016/j.cellimm.2017.10.013

Socinski, M. A. (2014). Update on taxanes in the first-line treatment of advanced non-small-cell lung cancer. *Curr. Oncol.* 21 (5), e691–e703. doi:10.3747/co.21.1997

Spranger, S., and Gajewski, T. F. (2018). Impact of oncogenic pathways on evasion of antitumour immune responses. *Nat. Rev. Cancer* 18 (3), 139–147. doi:10.1038/nrc.2017.117

Suresh, K., Naidoo, J., Lin, C. T., and Danoff, S. (2018). Immune checkpoint immunotherapy for non-small cell lung cancer: Benefits and pulmonary toxicities. *Chest* 154 (6), 1416–1423. doi:10.1016/j.chest.2018.08.1048

Travis, W. D., Dacic, S., Sholl, L. M., WistubaII, Adusumilli, P., Bubendorf, L., et al. (2021). IASLC multidisciplinary recommendations for pathologic assessment of lung cancer resection specimens after neoadjuvant therapy. *J. Thorac. Oncol.* 16 (1), 709–740. doi:10.1016/j.jtho.2020.01.005

Vansteenkiste, J., Wauters, E., Reymen, B., Ackermann, C. J., Peters, S., and De Ruysscher, D. (2019). Current status of immune checkpoint inhibition in early-stage NSCLC. *Ann. Oncol.* 30 (8), 1244–1253. doi:10.1093/annonc/mdz175

Vaughan, J. E., and Walsh, S. W. (2005). Neutrophils from pregnant women produce thromboxane and tumor necrosis factor-alpha in response to linoleic acid and oxidative stress. *Am. J. Obstet. Gynecol.* 193 (3), 830–835. doi:10.1016/j.ajog.2005.01.057

Wang, L., Felts, S. J., Van Keulen, V. P., Scheid, A. D., Block, M. S., Markovic, S. N., et al. (2018). Integrative genomewide analysis of long noncoding RNAs in diverse immune cell types of melanoma patients. *Cancer Res.* 78, 4411–4423. doi:10.1158/0008-5472.CAN-18-0529

Wu, L., and Zhang, X. H. (2020). Tumor-associated neutrophils and macrophages-heterogenous but not chaotic. *Front. Immunol.* 11, 553967. doi:10.3389/fimmu.2020.553967

Yang, J., and Hu, L. (2019). Immunomodulators targeting the PD-1/PD-L1 protein-protein interaction: From antibodies to small molecules. *Med. Res. Rev.* 39 (1), 265–301. doi:10.1002/med.21530

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Yue, P., Harper, T., Bacot, S. M., Chowdhury, M., Lee, S., Akue, A., et al. (2018). BRAF and MEK inhibitors differentially affect nivolumab-induced T cell activation by modulating the TCR and AKT signaling pathways. *Oncoimmunology* 8 (1), e1512456. doi:10.1080/2162402X.2018.1512456

Zhai, S., Xu, Z., Xie, J., Zhang, J., Wang, X., Peng, C., et al. (2021). Epigenetic silencing of LncRNA LINC00261 promotes c-myc-mediated aerobic glycolysis by regulating miR-222-3p/HIPK2/ERK axis and sequestering IGF2BP1. *Oncogene* 40 (2), 277–291. doi:10.1038/s41388-020-01525-3

Zhang, B., Wu, Q., Li, B., Wang, D., Wang, L., and Zhou, Y. L. (2020). m6A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. *Mol. Cancer* 19 (1), 53. doi:10.1186/s12943-020-01170-0

Zhang, S., and Zhou, J. (2021). Low LINC01272 predicts poor prognosis of non-small cell lung cancer and its biological function in tumor cells by inhibiting miR-1303. *Oncol. Lett.* 22 (3), 652. doi:10.3892/ol.2021.12913

Zhao, R., Li, B., Zhang, S., He, Z., Pan, Z., Guo, Q., et al. (2021). The N6-methyladenosine-modified pseudogene HSPA7 correlates with the tumor microenvironment and predicts the response to immune checkpoint therapy in glioblastoma. *Front. Immunol.* 12, 653711. doi:10.3389/fimmu.2021.653711

Zhou, M., Zhang, Z., Bao, S., Hou, P., Yan, C., Su, J., et al. (2021). Computational recognition of lncRNA signature of tumor-infiltrating B lymphocytes with potential implications in prognosis and immunotherapy of bladder cancer. *Brief. Bioinform.* 22 (3), bbaa047. doi:10.1093/bib/bbaa047

Zhou, M., Zhang, Z., Zhao, H., Bao, S., Cheng, L., and Sun, J. (2018). An immune-related six-lncRNA signature to improve prognosis prediction of glioblastoma multiforme. *Mol. Neurobiol.* 55, 3684–3697. doi:10.1007/s12035-017-0572-9

Zhou, M., Zhao, H., Xu, W., Bao, S., Cheng, L., and Sun, J. (2017). Discovery and validation of immune-associated long non-coding RNA biomarkers associated with clinically molecular subtype and prognosis in diffuse large B cell lymphoma. *Mol. Cancer* 16, 16. doi:10.1186/s12943-017-0580-4

Zhu, J., Li, R., Tiselius, E., Roudi, R., Teghararian, O., Suo, C., et al. (2017). Immunotherapy (excluding checkpoint inhibitors) for stage I to III non-small cell lung cancer treated with surgery or radiotherapy with curative intent. *Cochrane Database Syst. Rev.* 12 (12), CD011300. doi:10.1002/14651858.CD011300.pub2

frontiers | Frontiers in Genetics

TYPE Systematic Review
PUBLISHED 10 November 2022
DOI 10.3389/fgene.2022.944278

OPEN ACCESS

EDITED BY
Yilin Zhang,
The University of Chicago, United States

REVIEWED BY
Qi Shengcai,
Shanghai Stomatology Prevention Hospital, China
Xiaoqi Yang,
Huazhong University of Science and Technology, China

*CORRESPONDENCE
Xingyong Li,
gssylxy@163.com
Wenyuan Luo,
cf15104026385@163.com

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 15 May 2022
ACCEPTED 27 October 2022
PUBLISHED 10 November 2022

CITATION
Zhang F, Zhang G, Zhang H, Pu X, Chi F, Zhang D, Xin X, Gao M, Luo W and Li X (2022), HOXA-AS2 may be a potential prognostic biomarker in human cancers: A meta-analysis and bioinformatics analysis.
Front. Genet. 13:944278.
doi: 10.3389/fgene.2022.944278

HOXA-AS2 may be a potential prognostic biomarker in human cancers: A meta-analysis and bioinformatics analysis

Fan Zhang[1,2†], Guangming Zhang[1,2†], Helin Zhang[1,2†], Xingyu Pu[2], Fei Chi[1,2], Dengxiao Zhang[1,2], Xiaoming Xin[1,2], Mingxuan Gao[2], Wenyuan Luo[2*] and Xingyong Li[2*]

[1]The First Clinical Medical College of Gansu University of Chinese Medicine (Gansu Provincial Hospital), Lanzhou, China, [2]Department of Orthopedics, Gansu Provincial Hospital, Lanzhou, China

Background: Dysregulation of long non-coding (lncRNA) has been reported in various solid tumors. HOXA cluster antisense RNA 2 (HOXA-AS2) is a newly identified lncRNA with abnormal expression in several human malignancies. However, its prognostic value remains controversial. This meta-analysis synthesized available data to clarify the association between HOXA-AS2 expression levels and clinical prognosis in multiple cancers.

Methods: Four public databases (Embase, PubMed, Web of Science, The Cochrane Library) were used to identify eligible studies. Hazard ratios (HRs) and odds ratios (ORs) with their 95% confidence intervals (CIs) were combined to assess the correlation of HOXA-AS2 expression with survival outcomes and clinicopathological features of cancer patients. Publication bias was measured using Begg's funnel plot and Egger's regression test, and the stability of the combined results was measured using sensitivity analysis. Additionally, multiple public databases were screened and extracted to validate the results of this meta-analysis.

Results: The study included 20 studies, containing 1331 patients. The meta-analysis showed that the overexpression of HOXA-AS2 was associated with poor overall survival (HR = 2.06, 95% CI 1.58–2.69, $p < 0.001$). In addition, the high expression of HOXA-AS2 could forecast advanced tumor stage (OR = 3.89, 95% CI 2.90–5.21, $p < 0.001$), earlier lymph node metastasis (OR = 3.48, 95% CI 2.29–5.29, $p < 0.001$), larger tumor size (OR = 2.36, 95% CI 1.52–3.66, $p < 0.001$) and earlier distant metastasis (OR = 3.54, 95% CI 2.00–6.28, $p < 0.001$). However, other clinicopathological features, including age (OR = 1.09, 95% CI 0.86–1.38, $p = 0.467$), gender (OR = 0.92, 95% CI 0.72–1.18, $p = 0.496$), depth of invasion (OR = 2.13, 95% CI 0.77–5.90, $p = 0.146$) and differentiation (OR = 1.02, 95% CI 0.65–1.59, $p = 0.945$) were not significantly different from HOXA-AS2 expression.

Conclusion: Our study showed that the overexpression of HOXA-AS2 was related to poor overall survival and clinicopathological features. HOXA-AS2 may serve as a potential prognostic indicator and therapeutic target for tumor treatment.

KEYWORDS
lncRNA, HOXA-AS2, cancers, prognosis, meta-analysis, bioinformatics analysis

Introduction

Cancer is the second greatest cause of death in most parts of the world, and it has become the most common public pathological condition on the planet (Chu et al., 2020). Traditional cancer treatments, such as surgery, adjuvant medical treatment, and actinotherapy have improved dramatically over the last century (Zhang et al., 2019). Despite this, 5-year cancer survival rates remain poor, particularly for patients with advanced tumor stage or

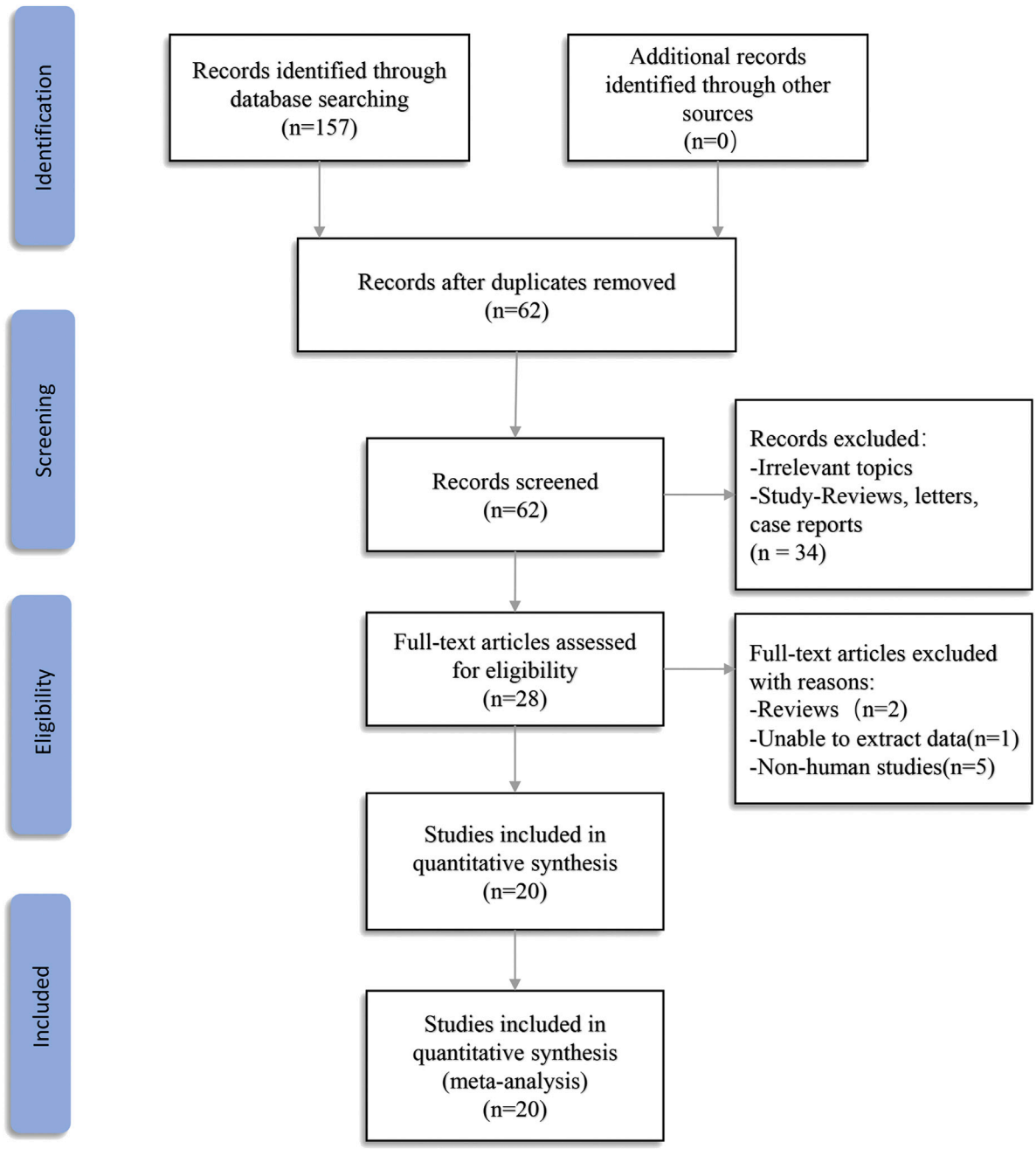


FIGURE 1
Flow diagram of this meta-analysis.

TABLE 1 Characteristics of studies in this meta-analysis.

Study	Year	Country	Cancer type	Sample type	Total Size(n)	Detection Method	Cutoff	Outcome	Multivariate Analysis	HR statistic	NOS score
Chen RH	2021	China	CC	Tissue	27	RT-qPCR	mean	OS	NR	Rep	6
Chen RW	2021	China	OSCC	Tissue	46	RT-qPCR	NR	NR	NR	NR	5
Cui	2019	China	LC	Tissue	80	RT-qPCR	mean	OS	NR	SC	7
Ding	2017	China	CRC	Tissue	69	RT-qPCR	NR	NR	NR	NR	5
Fang	2017	China	BC	Tissue	38	RT-qPCR	NR	OS	NR	SC	6
Jiang	2019	China	TC	Tissue	68	NR	mean	OS	NR	SC	7
Li	2016	China	CRC	Tissue	30	RT-qPCR	NR	OS	NR	SC	6
Li	2017	China	LC	Tissue	103	RT-qPCR	median	OS	Yes	SC	8
Liu	2019	China	LC	Tissue	52	RT-qPCR	median	OS	NR	SC	7
Lu	2020	China	HCC	Tissue	106	RT-qPCR	median	OS	Yes	Rep	8
Qu	2020	China	AML	Blood	108	RT-qPCR	median	OS	Yes	Rep	8
Wang F	2016	China	HCC	Tissue	112	RT-qPCR	NR	OS	NR	SC	6
Wang Y	2018	China	OS	Tissue	66	RT-qPCR	NR	NR	NR	NR	5
Wang L	2019	China	OSA	Tissue	27	RT-qPCR	mean	OS	NR	SC	6
Wang F	2019	China	Bladder cancer	tissue	80	RT-qPCR	NR	NR	NR	NR	5
Wu	2019	China	LGG	tissue	50	RT-qPCR	NR	NR	NR	NR	5
Xia	2018	China	TC	tissue	128	RT-qPCR	mean	NR	NR	NR	6
Xiao	2020	China	PCa	tissue	68	RT-qPCR	mean	OS	NR	SC	7
Xie	2015	China	GC	tissue	55	RT-qPCR	median	OS	NR	SC	7
Zhang	2018	China	HCC	tissue	58	RT-qPCR	NR	NR	NR	NR	5

HR, hazard ratio; GC, gastric cancer; CRC, colorectal cancer; HCC, hepatocellular carcinoma; PCa, Prostate cancer; CC, cervical cancer; OSA, osteosarcoma; LGG, lower-grade glioma; TC, thyroid Cancer; AML, acute myeloid leukemia; OSCC, oral squamous cell carcinoma; LC, lung cancer; BC, breast cancer; NR, no report; OS, overall survival; PFS, progression-free survival; Rep, report; SC, survival curve; RT-qPCR, real-time quantitative polymerase chain reaction.

metastasis (Li et al., 2019). One of the most significant causes is the lack of a good biomarker for detecting cancer early and predicting the clinical outcome of cancer patients (Ye et al., 2019). The significance of biomarkers in cancer has garnered increased attention in recent years across various fields, and they are thought to play critical roles in effectively screening or diagnosing cancer (Tang et al., 2020).

Long noncoding RNAs (lncRNAs) are RNA molecules that are longer than 200 nucleotides and cannot code for proteins (Zhou et al., 2019). A huge number of lncRNAs are produced during the active transcription of the human genome (Morlando and Fatica, 2018). One of the functions of lncRNAs *in vivo* is as tumor suppressors or oncogenes (Xu et al., 2021). Increasing evidence suggest that lncRNAs play a synergistic role in tumorigenesis or tumor suppression and that aberrant lncRNA expression is linked to cell proliferation, growth, and metastasis (Wang Y. et al., 2020). The development of RNA-targeted therapies has presented possibility of lncRNA-guided cancer therapy (Bhan et al., 2017). The inhibition of lncRNA function by RNA depletion and the removal of lncRNA exons encoding essential functional domains using splice-switching oligonucleotides may be the mechanism for targeting lncRNAs for cancer therapy (Kole et al., 2012; Schmitt and Chang, 2016). Therefore, functional lncRNA can be used as a biomarker for cancer diagnosis and for predicting treatment outcome and patient prognosis (Wang J. et al., 2020).

LncRNA HOXA cluster antisense RNA 2 (HOXA-AS2) is located on chromosome 7p15.2, a 1048-bp lncRNA, between the HOXA3 and HOXA4 genes of the HOXA cluster (Liu et al., 2019). Previous studies found that HOXA-AS2 was up-regulated in certain cancers. The increased expression of HOXA-AS2 typically predicts poor prognosis for patients with several cancers including cervical cancer (CC) (Chen and He, 2021), oral squamous cell carcinoma (OSCC) (Chen et al., 2021), lung cancer (LC) (Li and Jiang, 2017; Cui et al., 2019; Liu et al., 2019), colorectal cancer (CRC) (Li et al., 2016; Ding et al., 2017), breast cancer (BC) (Fang et al., 2017), thyroid cancer (TC) (Xia et al., 2018; Jiang et al., 2019), hepatocellular carcinoma (HCC) (Wang et al., 2016; Zhang et al., 2018; Lu et al., 2020), acute myeloid leukemia (AML) (Qu et al., 2020), bladder cancer (Wang

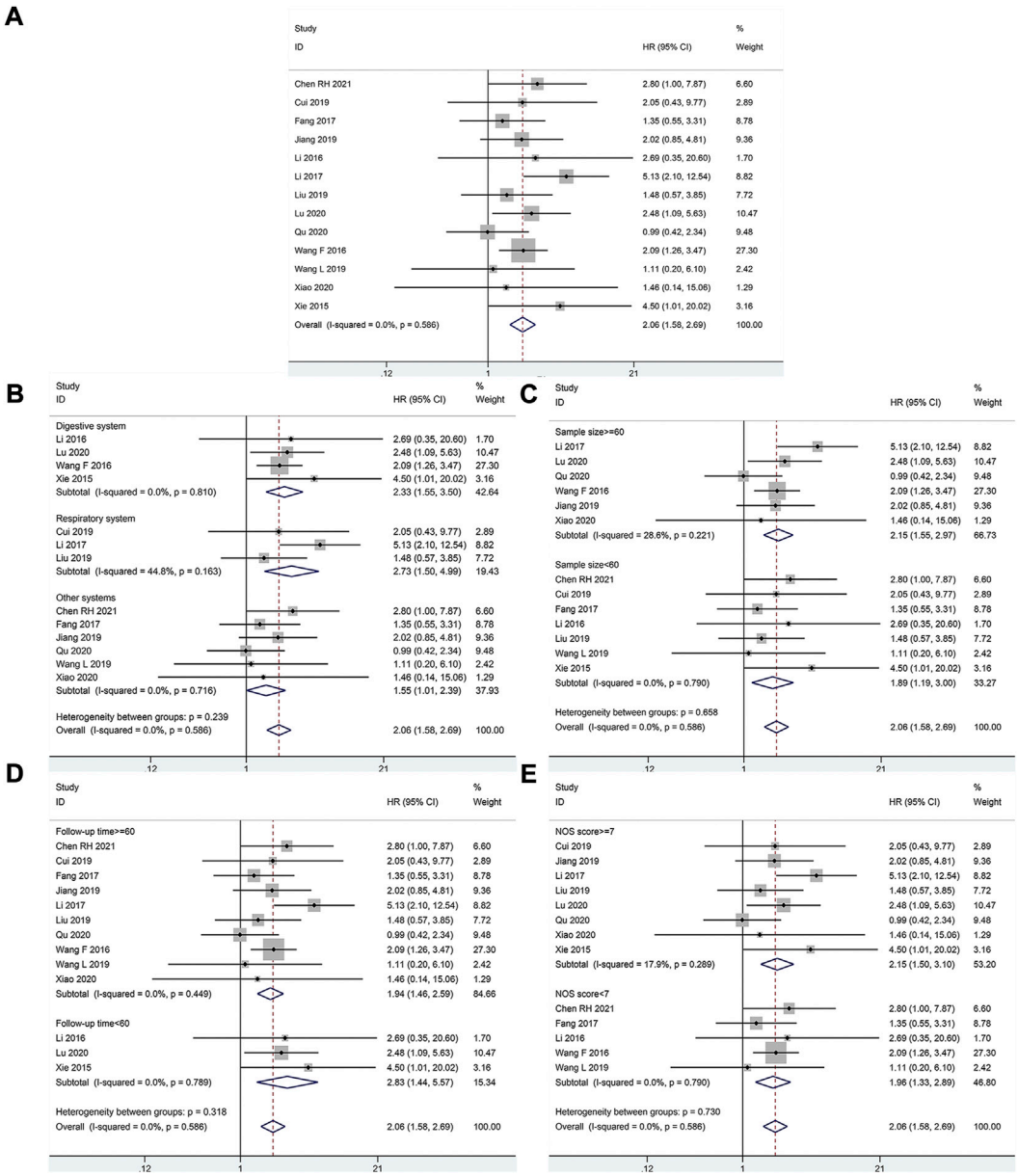


FIGURE 2
Relationship between HOXA-AS2 expression and overall survival. **(A)** Forest plots for association of HOXA-AS2 expression with overall survival. **(B)** Subgroup analysis stratified by cancer type. **(C)** Subgroup analysis stratified by sample size. **(D)** Subgroup analysis stratified by follow-up time. **(E)** Subgroup analysis stratified by NOS score.

F. et al., 2019), osteosarcoma (OSA) (Wang et al., 2018; Wang L. et al., 2019), lower-grade glioma (LGG) (Wu et al., 2019), prostate cancer (PCa) (Xiao and Song, 2020), gastric cancer (GC) (Xie et al., 2015). A high level of HOXA-AS2 expression is associated with poor overall survival (OS) and clinicopathological characteristics such as differentiation, tumor node metastasis (TNM) stage, lymph node metastasis (LNM). However, it is not clear the prognostic value of HOXA-AS2, as most of the published studies were based on a small group of patients. We explored the prognostic value of HOXA-AS2 in pan-cancer for the first time using meta-analysis. Furthermore, we further validated and explored the prognostic value of HOXA-AS2 in multiple databases through bioinformatics analysis, and explored HOXA-AS2-related genes and potential pathways. Also, the role of HOXA-AS2 in tumor immunity was investigated to

TABLE 2 Subgroup meta-analysis of pooled HRs for OS.

Stratified analysis	Studies (n)	Number of patients	Pooled HR (95% CI)	*P*-value	Heterogeneity		
					I^2, %	*P*-value	Model
Cancer type							
Digestive system	4	303	2.33(1.55–3.50)	<0.001	0.0	0.810	Fixed
Respiratory system	3	195	2.73(1.50–4.99)	0.001	44.8	0.163	Fixed
Other systems	6	336	1.55(1.01–2.39)	0.046	0.0	0.716	Fixed
Sample size							
≥60	6	565	2.15(1.55–2.97)	<0.001	28.6	0.221	Fixed
<60	7	269	1.89(1.19–3.00)	0.007	0.0	0.790	Fixed
Follow-up time (month)							
≥60	10	643	1.94(1.46–2.59)	<0.001	0.0	0.449	Fixed
<60	3	191	2.83(1.44–5.57)	<0.001	0.0	0.789	Fixed
NOS score							
≥7	8	600	2.15(1.50–3.10)	<0.001	17.9	0.289	Fixed
<7	5	234	1.96(1.33–2.89)	0.001	20.6	0.790	Fixed

CI, Confidence interval; HR, Hazard ratio.

identify the potential of HOXA-AS2 as a novel tumor marker and therapeutic target.

Materials and methods

Registration

The study was registered in the International Platform of Registered Systematic Review and Meta-Analysis Protocols (the registration number is: CRD42021292257). Because the present study was a systematic review and meta-analysis, Institutional Review Board (IRB) approval was not required.

Search strategy

Quality meta-analysis guidelines were followed to search for and find related papers in the Embase, PubMed, Web of Science, and The Cochrane Library. Key terms include the following: "HOXA-AS2" "long noncoding RNA HOXA-AS2" "lncRNA HOXA-AS2" "HOXA cluster antisense RNA 2" "HOXA3as" "neoplasm" "cancer" "malignancy" "neoplasia" "melanoma" "tumor" "sarcoma" "carcinoma" or "adenoma". These terms were used to maximize the likelihood of finding a relevant article. The literature search included articles revealed as of 15 November 2021. A manual search of the reference lists of the retrieved literature was performed to confirm the eligible

TABLE 3 Association of HOXA-AS2 expression with clinicopathological features.

Clinicopathological parameters	Patients (n)	Or (95%CI)	*P* Value	Heterogeneity (I^2, *P*)	Model
Age (elderly vs. nonelderly)	1181	1.09(0.86–1.38)	0.467	0.0%, 0.677	Fixed
Gender (male vs. female)	1185	0.92(0.72–1.18)	0.496	10.1%, 0.338	Fixed
Tumor stage (III + IV vs. I + II)	551	3.89(2.90–5.21)	<0.001	0.0%, 0.507	Fixed
Lymph node metastasis (positive vs. negative)	433	3.48(2.29–5.29)	<0.001	0.0%, 0.526	Fixed
Tumor size (big vs. small)	829	2.36(1.52–3.66)	<0.001	54.4%, 0.015	Random
Differentiation (poor vs. well)	341	1.02(0.65–1.59)	0.945	0.0%, 0.460	Fixed
Depth of invasion (III + IV vs. I + II)	148	2.13(0.77–5.90)	0.146	57.0%, 0.127	Random
Distant metastasis (Yes vs. No)	277	3.54(2.00–6.28)	<0.001	0.0%, 0.870	Fixed

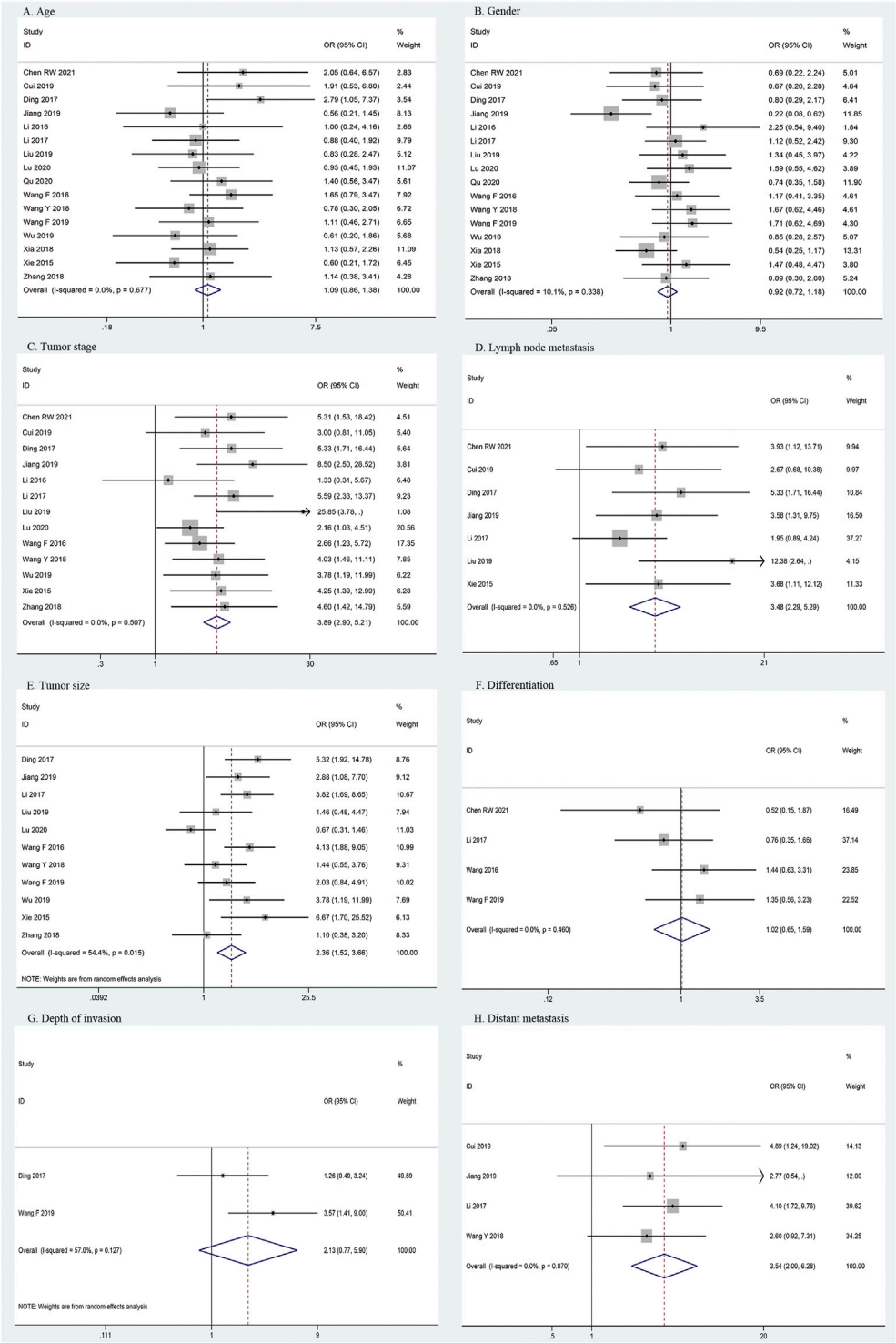


FIGURE 3
Forest plots for association of HOXA-AS2 expression with clinicopathological features. **(A)** Age. **(B)** Gender. **(C)** Tumor stage. **(D)** Lymph node metastasis. **(E)** Tumor size. **(F)** Differentiation. **(G)** Depth of invasion. **(H)** Distant metastasis.

studies included. Any conflicts between the inclusion and exclusion clauses were resolved through group discussion.

Participants, interventions, and comparators

Studies that complied with the following criteria were eventually included: The inclusion criteria were: (a) the use of real-time quantitative polymerase chain reaction (RT-qPCR) analysis to determine the expression of HOXA-AS2 in neoplastic tissues; (b) patients diagnosed with cancer, and the study described a link between HOXA-AS2 and survival data or clinicopathology; (c) the patients were divided into two groups according to the expression level of HOXA-AS2, and (d) the quantitative hazard ratios (HRs) of OS could be extracted from the text or survival curve. The exclusion criteria were: (a) studies not related to tumors or HOXA-AS2; (b) duplicate publications; (c) reviews, conference abstracts, or case reports, and (d) studies that lacked relevant data.

Data extraction

Two researchers extracted information from each study, and any disagreement was resolved by discussing it with a third author. We obtained the following data and information from every study: (a) first author, (b) publication year, (c) country of origin, (d) cancer type, (e) number of samples, (f) HOXA-AS2 expression detection technique, (g) cut-off value, (h) sample size with high and low HOXA-AS2 expression, (i) HRs and 95% confidence intervals (CIs) for OS, (j) clinicopathologic parameters, and (k) follow-up times. OS data was directly obtained or extracted from the Kaplan-Meier (KM) curves using Engauge Digitizer version 4.1 software and the HRs and 95% CIs were computed.

Quality assessment

Two reviewers extracted information individually based on the inclusion and exclusion criteria. Some disagreements were resolved in consultation with a third reviewer. The quality of the studies was assessed using the Newcastle–Ottawa scale (NOS). The scale uses nine elements to judge a study, and a score of one is satisfied for an exact item. Total scores range from 0 to 9. A NOS score of ≥ 7 represents high-quality analysis results.

Statistical analysis

All statistical analyses were conducted using Stata software (version 12.0). The correlation of HOXA-AS2 expression with survival and clinicopathological features of tumor patients was assessed using HRs and odds ratios (ORs) with their 95% CIs, respectively. The chi-squared test and I^2 statistic were preferred to determine the heterogeneity between studies. If there is strong heterogeneity ($P_Q < 0.1$, $I^2 > 50\%$), we considered the random-effect model was applied, and the fixed-effect model was applied otherwise. All results are shown as Forest plots. Egger's test and Begg's funnel plot were used to evaluate publication bias, and sensitivity analysis was conducted to evaluate the robustness of the results.

Public data and tools

HOXA-AS2 expression levels in tumors and normal tissues of different solid tumors were analyzed by the Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn) online database (based on TCGA and GTEx databases) (cutoff, $p < 0.01$). The survival outcomes were then verified by plotting the correlation between HOXA-AS2 expression and OS as a KM curve. Moreover, we further explored the prognostic value of HOXA-AS2 in various cancers using the Biomarker Exploration of Solid Tumors (BEST, https://rookieutopia.com) online tool. Additionally, we further explored the correlation between HOXA-AS2 and drug response through the CellMiner database (Reinhold et al., 2012) using the R-package "readxl", "impute" and "limma" options.

Correlation of HOXA-AS2 expression with tumor immunity

Initially, we analyzed the relationship between HOXA-AS2 expression and the level of immune cell infiltration (ICI) in various cancers based on the R packages "ggExtra", "ggpubr" and "ggplot2" via the CIBERSORT tool. Next, stromal and immune scores were calculated for each tumor sample using the ESTIMATE algorithm. The correlation between HOXA-AS2 expression and tumor microenvironment (TME) was assessed by the R package "ESTIMATE" and "limma". In addition, the relation of HOXA-AS2 expression and tumor mutational load (TMB), tumor microsatellite instability (MSI) and immune checkpoint genes were further evaluated. TMB scores were calculated by Perl scripts, and MSI scores were determined from TCGA database mutation data. The results were visualized using the R package "RColorBrewer" and "reshape2".

Analysis of HOXA-AS2-related genes and construction of signaling pathway network

To further investigate the value of HOXA-AS2, we obtained related genes from the MEM-Multi Experiment Matrix database. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were carried out. The top findings with a p-value of less than 0.05 were deemed significant. Finally, we used Cytoscape software to create a signal pathway network.

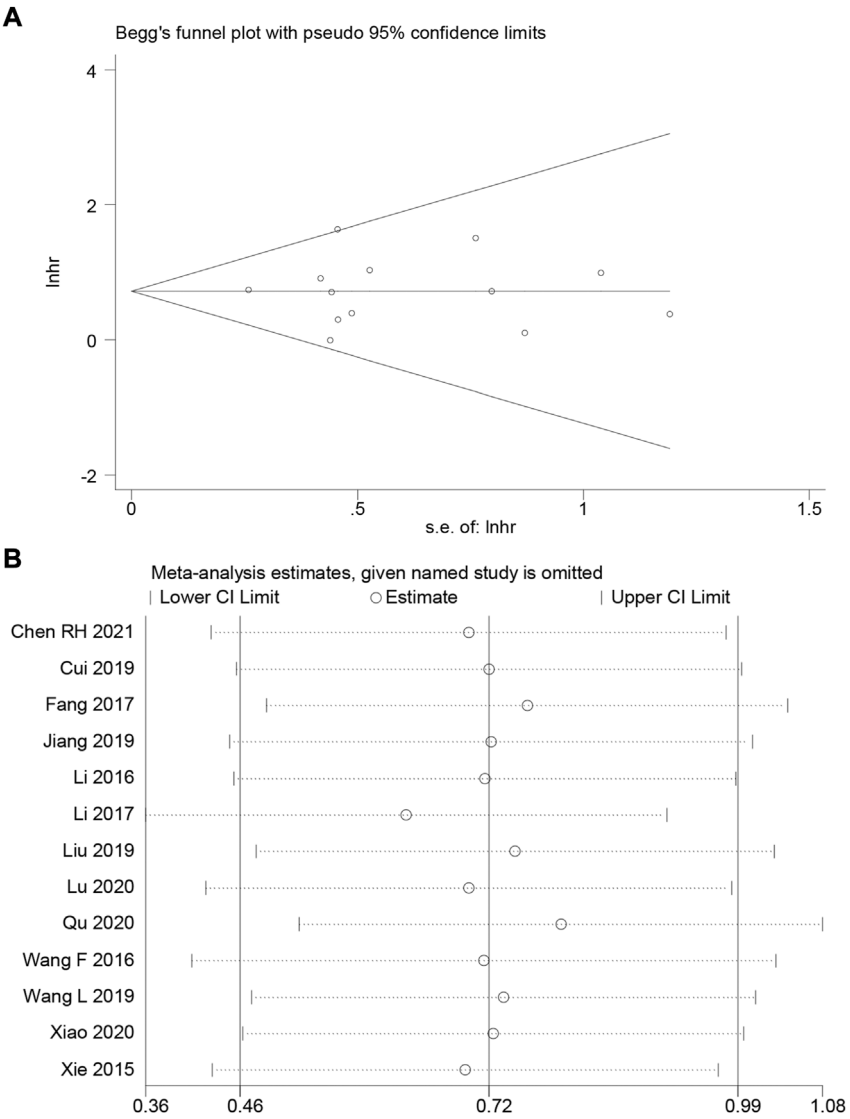


FIGURE 4
(A) Begg's funnel plot of HOXA-AS2 for overall survival. **(B)** A sensitivity analysis of pooled HR for overall survival.

Results

Identification of articles

A total of 157 records were found in four electronic databases (Embase = 57, PubMed = 47, Web of Science = 53, The Cochrane Library = 0). Ninety-five duplicate articles were deleted using Endnote X9 software. After screening the titles and abstracts, 34 articles were excluded because they were not associated with the review topic or as a result of reviews, meta-analysis, letters, or expert opinions. Hence, a full-text examination was conducted for 28 articles. One article was excluded because we were unable to extract data. Two papers were eliminated because they were reviews, and five articles were eliminated because they were cell-based studies. In the end, 20 articles were included in the final meta-analysis (Figure 1).

Characteristics of the included articles

All selected articles were published between 2015 and 2021 and included 1331 patients, all of whom were from China. The smallest sample size was 27, and the largest sample size was 128. Among the twenty studies, one focused on LGG (Wu et al., 2019), one on CC (Chen and He, 2021),

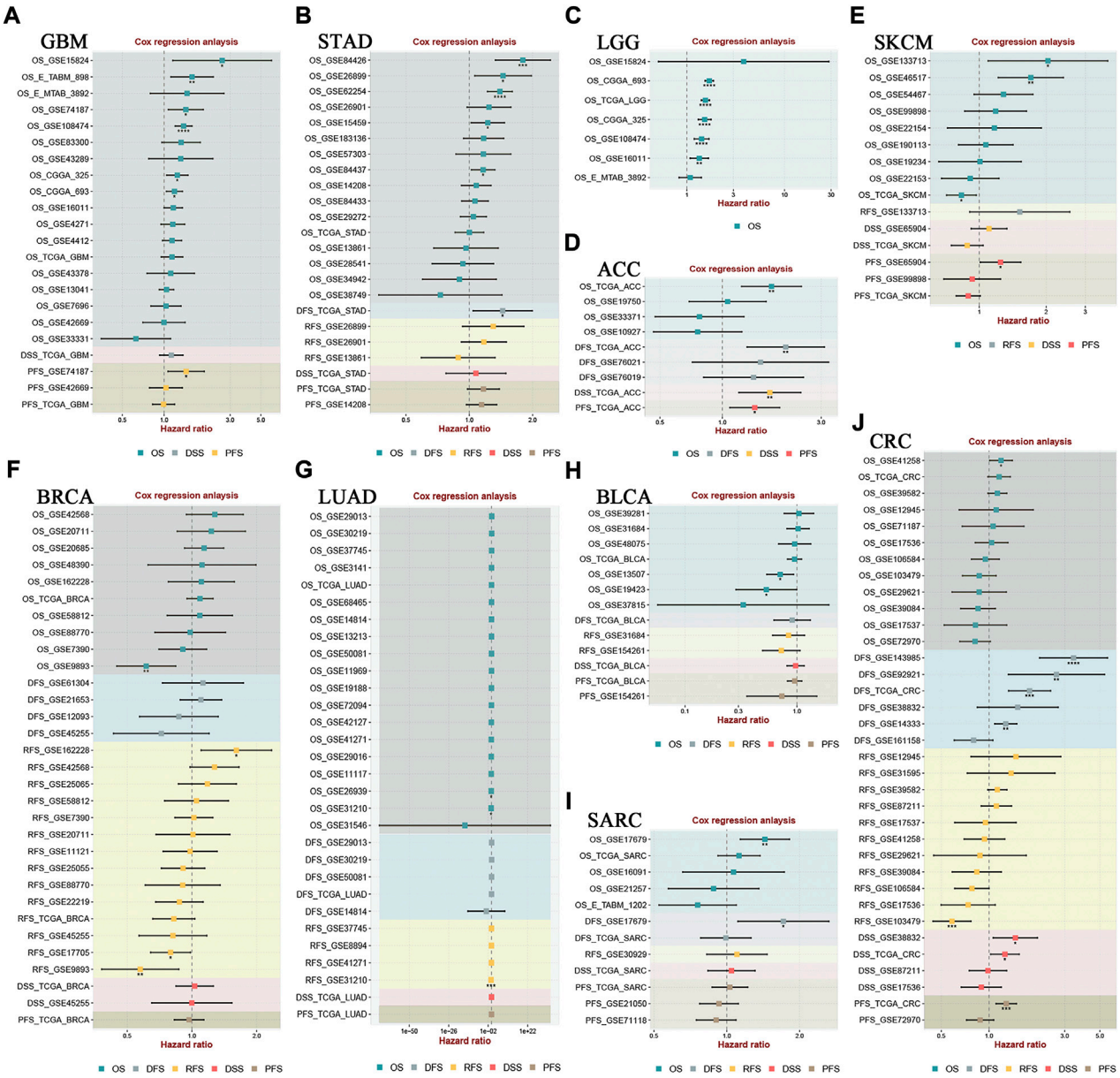


FIGURE 5
The correlation between HOXA-AS2 expression and survival in different tumor types was analyzed in BEST. Cox regression analysis of **(A)** GBM, **(B)** STAD, **(C)** LGG, **(D)** ACC, **(E)** SKCM, **(F)** BRCA, **(G)** LUAD, **(H)** BLCA, **(I)** SARC, **(J)** CRC. $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$.

one on OSCC (Chen et al., 2021), three on LC (Li and Jiang, 2017; Cui et al., 2019; Liu et al., 2019), two on CRC (Li et al., 2016; Ding et al., 2017), one on BC (Fang et al., 2017), two on TC (Xia et al., 2018; Jiang et al., 2019), three on HCC (Wang et al., 2016; Zhang et al., 2018; Lu et al., 2020), two on OSA (Wang et al., 2018; Wang L. et al., 2019), one on PCa (Xiao and Song, 2020), one on GC (Xie et al., 2015), one on bladder cancer (Wang F. et al., 2019), and one on AML (Qu et al., 2020). The expression of the indicated genes in cancer tissues was measured by RT-qPCR. All eligible studies were dichotomized into low and high HOXA-AS2 expression groups based on a cut-off value. The follow-up time ranged from 25 to 120 months. All included studies were cohort studies, 65% (13/20) of which reported OS rates. The main characteristics of the eligible studies are shown in Table 1.

Association of HOXA-AS2 with OS

A total of 834 patients in 13 studies reported a link between HOXA-AS2 expression and OS. Since there was no significant heterogeneity between the studies, a fixed-effects model was used to calculate the HR and 95% CI. The pooled HR was 2.06, which indicated that HOXA-AS2 overexpression predicted poor OS in these patients with neoplasms (Figure 2A). Furthermore, KM survival analysis was applied to determine OS in different subgroups of patients according to tumor type (digestive system, respiratory system, others) (Figure 2B), sample size (n ≥ 60 or n < 60) (Figure 2C), follow-up time (≥60 months or < 60 months) (Figure 2D), and NOS score (NOS scores ≥ 7 or < 7) (Figure 2E). As depicted in Table 2, higher HOXA-AS2 expression levels were significantly associated with worse OS.

Association between HOXA-AS2 and clinicopathologic parameters

Correlations between HOXA-AS2 expression and the clinicopathological features of the patients are shown in Table 3. The meta-analysis results showed that higher HOXA-AS2 expression levels tended to be significantly associated with advanced tumor stage (OR = 3.89, 95% CI 2.90–5.21, $p < 0.001$) (Figure 3C), earlier LNM (OR = 3.48, 95% CI 2.29–5.29, $p < 0.001$) (Figure 3D), larger tumor size (OR = 2.36, 95% CI 1.52–3.66, $p < 0.001$) (Figure 3E) and earlier distant metastasis (OR = 3.54, 95% CI 2.00–6.28, $p < 0.001$) (Figure 3H). However, age (OR = 1.09, 95% CI 0.86–1.38, $p = 0.467$) (Figure 3A), gender (OR = 0.92, 95% CI 0.72–1.18, $p = 0.496$) (Figure 3B), depth of invasion (OR = 2.13, 95% CI 0.77–5.90, $p = 0.146$) (Figure 3G), and differentiation (OR = 1.02, 95% CI 0.65–1.59, $p = 0.945$) (Figure 3F), had no significant link with increased HOXA-AS2 expression levels.

Publication bias and sensitivity analysis

We employ Begg's funnel plot and Egger's regression test to identify publication bias of OS. The shape of Begg's funnel was essentially symmetrical, with no visible asymmetry (Figure 4A), and Egger's regression analysis did not reveal the presence of publication bias (Pr > |t| = 0.971). We ran a sensitivity analysis by eliminating one qualified study to analyze the influence of a single study on the result. According to the analysis, the results were not significantly influenced (Figure 4B). This verifies the reliability of the meta-analysis conclusions.

Validation of HOXA-AS2 expression in public databases

We used the TCGA dataset to analyze the degree of HOXA AS2 expression in various tumors to further corroborate our results. HOXA-AS2 was aberrantly expressed in glioblastoma multiforme (GMB), acute myeloid leukemia (LAML), pancreatic adenocarcinoma (PAAD), and thymoma (THYM), compared to normal controls (Supplementary Figure S1A). A violin plot revealed that the degree of HOXA-AS2 expression in human cancer was highly related to the clinical stage (Supplementary Figure S1B). We used GEPIA to create survival graphs by combining HOXA-AS2 expression data with the OS data of patients with malignancies in the entire TCGA dataset, which included 9491patients separated into high (4741) and low (4750) groups of HOXA-AS2 expression based on median levels. The results showed that increased HOXA-AS2 expression predicted poor OS, confirming the meta-analysis results (Supplementary Figure S1C). Additionally, we explored the link of HOXA-AS2 expression and tumor prognosis using Cox regression model through the BEST online tool. The findings revealed that there was a significant correlation between HOXA-AS2 expression and the prognosis of GBM, stomach adenocarcinoma (STAD), LGG, adrenocortical carcinoma (ACC), skin cutaneous melanoma (SKCM), breast invasive carcinoma (BRCA), lung adenocarcinoma (LUAD), bladder urothelial carcinoma (BLCA), sarcoma (SARC), and CRC in at least two datasets (Figure 5).

HOXA-AS2 and drug response

To further explore the significance of HOXA-AS2 in guiding cancer treatment, we analyzed the relationship between HOXA-AS2 and drug response. The results revealed that patients with high HOXA-AS2 expression had a better drug response to XL−147, Cpd−401, cordycepin, fenretinide, estramustine, and arsenic trioxide. In contrast, AS−703569, ENMD−2076, SB−1317, benzaldehyde (BEN), staurosporine, aminoflavone, amonafide, midostaurin, sapitinib, and KW−2449 had a better drug response in patients with low HOXA-AS2 expression (Figure 6).

Correlation analysis of HOXA-AS2 expression with tumor immunity

Correlation analysis between HOXA-AS2 expression and ICI levels identified remarkable correlations between HOXA-AS2 expression and ICI levels in BRCA ($n = 7$), kidney renal papillary cell carcinoma (KIRP) ($n = 6$), kidney renal clear cell carcinoma (KIRC) (n = 5), head and Neck squamous cell carcinoma (HNSC) ($n = 4$), STAD ($n = 4$), thyroid carcinoma (THCA) ($n = 3$), testicular germ cell tumors (TGCT) (n = 3), BLCA ($n =$

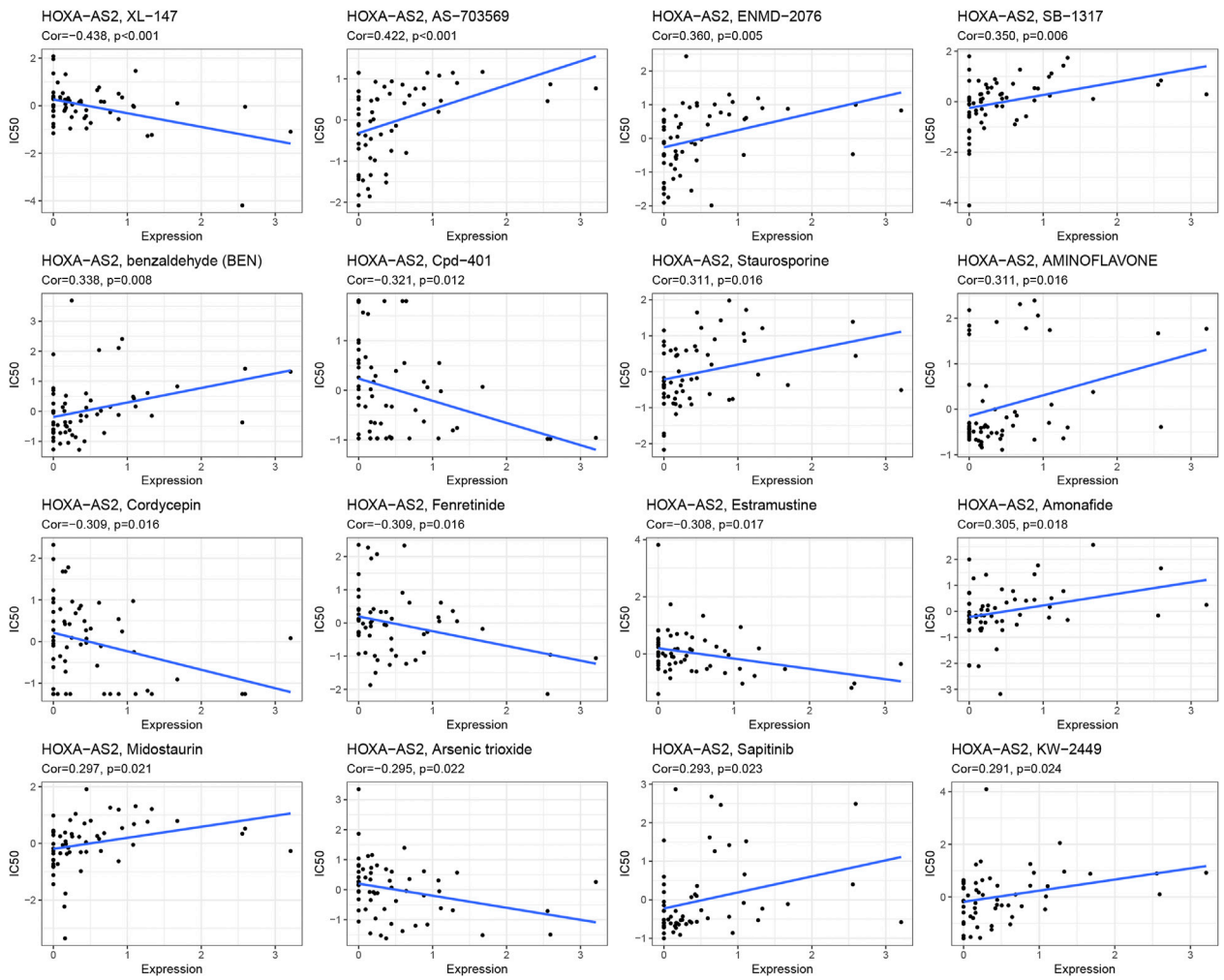


FIGURE 6
An illustration of the association between HOXA-AS2 expression and expected medication response.

3), LGG (n = 3), THYM (n = 2), ovarian serous cystadenocarcinoma (OV) (*n* = 2), esophageal carcinoma (ESCA) (n = 2), prostate adenocarcinoma (PRAD) (*n* = 1), mesothelioma (MESO) (*n* = 1), lung squamous cell carcinoma (LUSC) (n = 1), LAML (n = 1), and cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC) (n=1). Detailed information on the subpopulations of infiltrating immune cells in various cancer types is illustrated in Figure 7. HOXA-AS2 expression was negatively correlated with the levels of infiltrating M0 macrophages in ESCA, NHSC, CESC, THCA, BRCA, LUSC, and BLCA (Figure 7A). Similarly, HOXA-AS2 expression was negatively associated with the levels of infiltrating neutrophils in STAD, HNSC, KIRC, and BRCA (Figure 7B). In regard to monocytes, their infiltration levels were negatively correlated with the HOXA-AS2 expression in LGG (Figure 7C). HOXA-AS2 expression was also negatively related to the levels of infiltrating M1 macrophages in BLCA, and KIRP (Figure 7D). Moreover, HOXA-AS2 expression was negatively correlated with the levels of infiltrating M2 macrophages in BRCA, but positively associated with TGCT (Figure 7E). HOXA-AS2 expression was positively associated with the levels of infiltrating naive B cells in STAD, OV, and BRCA, but positively associated with KIRP and TGCT (Figure 7F). HOXA-AS2 expression was negatively correlated with the infiltrating levels of activated CD4 memory T cells in BLCA, KIRC, KIRP, STAD, TGCT, and THYM (Figure 7G). Furthermore, HOXA-AS2 expression presented a positive relationship with the levels of infiltrating plasma cells in BRCA (Figure 7H). The levels of infiltrating CD8 T cells were positively associated with HOXA-AS2 expression in PRAD, HNSC, and BRCA, but negatively related in KIRP (Figure 7I). A positive association with infiltrating follicular helper T cells

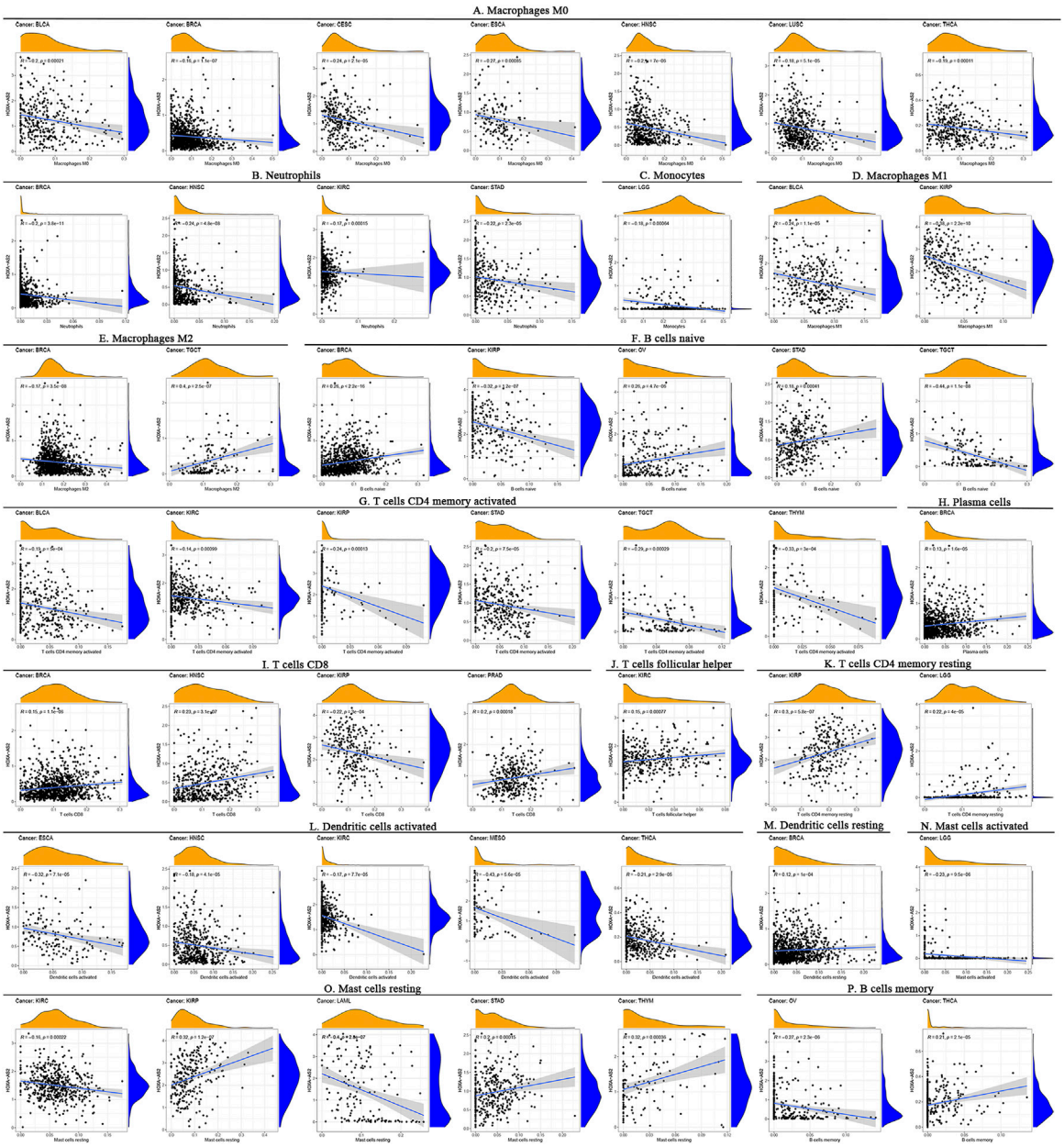


FIGURE 7
Correlation between HOXA-AS2 gene expression and the level of immune cell infiltration in pan-cancerous tissues. HOXA-AS2 expression significantly correlated with infiltrating levels of M0 macrophages in BLCA, BRCA, CESC, ESCA, NHSC, LUSC, and THCA **(A)**, neutrophils in BRCA, HNSC, KIRC, and STAD **(B)**, monocytes in LGG **(C)**, M1 macrophages in BLCA and KIRP **(D)**, M2 macrophages in BRCA and TGCT **(E)**, naive B cells in BRCA, KIRP, OV, STAD, and TGCT **(F)**, CD4 memory T cells in BLCA, KIRC, KIRP, STAD, TGCT, and THYM **(G)**, plasma cells in BRCA **(H)**, CD8 T cells in BRCA, HNSC, KIRP, and PRAD **(I)**, follicular helper T cells in KIRC **(J)**, resting CD4 memory T cells in KIRP and LGG **(K)**, activated dendritic cells in ESCA, HNSC, KIRC, MESO, and THCA **(L)**, resting DCs in BRCA **(M)**, activated mast cells in LGG **(N)**, resting mast cells in KIRC, KIRP, LAML, STAD, and THYM **(O)**, memory B cells in OV and THCA **(P)**.

was identified in KIRC (Figure 7J). The levels of infiltrating resting CD4 memory T cells were positively associated with HOXA-AS2 expression in KIRP, and LGG (Figure 7K). HOXA-AS2 expression was negatively associated with the levels of infiltrating activated dendritic cells in MESO, HNSC, KIRC, ESCA, and THCA (Figure 7L). Moreover, HOXA-AS2 expression was positively associated with the levels of infiltrating resting DCs in BRCA (Figure 7M). The levels of infiltrating

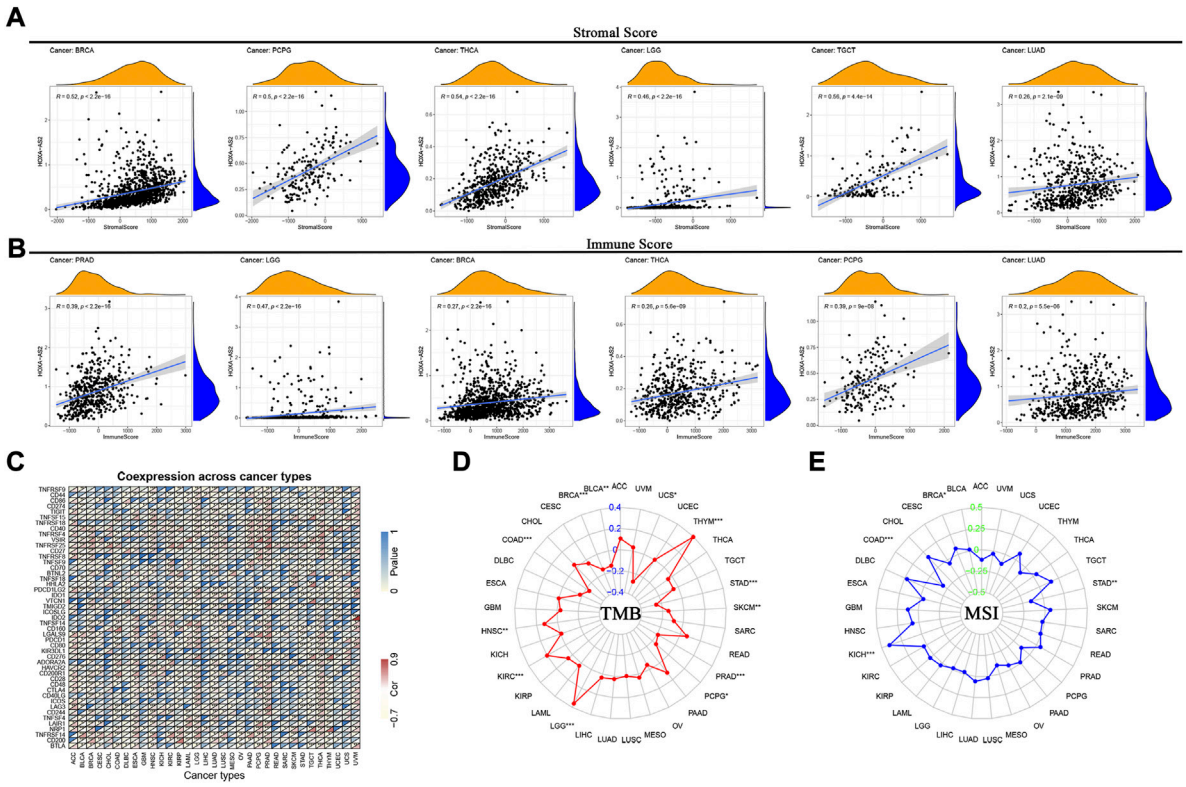


FIGURE 8
Correlations between HOXA-AS2 expression and immunity, including stromal score, immune score, immune checkpoint genes, TMB and MSI in cancers. **(A)** Correlation of HOXA-AS2 expression with the stromal score in pan-cancer. **(B)** Correlation of HOXA-AS2 expression with the immune score in pan-cancer. **(C)** Correlation of HOXA-AS2 expression and immune checkpoint genes. **(D)** The radar chart displays the correlation of TMB with HOXA-AS2 expression. The red curve indicates the correlation coefficient, and the blue value indicates the range. **(E)** The radar chart displays the correlation of MSI with HOXA-AS2 expression. The blue curve indicates the correlation coefficient, and the green value indicates the range. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

activated mast cells were negatively associated with HOXA-AS2 expression in LGG (Figure 7N). In contrast, the levels of infiltrating resting mast cells were positively correlated with HOXA-AS2 expression in THYM, STAD, and KIRP, but negatively related in KIRC and LAML (Figure 7O). The levels of infiltrating memory B cells were positively associated with HOXA-AS2 expression in THCA, but negatively associated in OV (Figure 7P).

Furthermore, to explore its relationship with the TME, we analyzed the association of HOXA-AS2 expression with stromal and immune scores. Our findings showed that HOXA-AS2 expression correlated with the stromal scores of 13 cancers, the top 6 tumors were BRCA, pheochromocytoma and paraganglioma (PCPG), THCA, LGG, TGCT, and LUAD (Figure 8A); and with the immune scores of 11 cancers, the top 6 tumors were PRAD, LGG, BRCA, THCA, PCPG, and LUAD (Figure 8B).

Moreover, the correlation of HOXA-AS2 expression with immune checkpoint genes showed that CD44, CD40, VSIR, LGALS9, TNFRSF14 and TNFRSF25 were significantly associated with HOXA-AS2 expression in several cancers, especially in LGG, PRAD, LUAD, THCA, and HNSC (Figure 8C). We evaluated the relation between HOXA-AS2 expression and TMB/MSI as well, and found that there was a significant positive correlation between its expression and TMB in LGG, THYM, KIRC, and HNSC, while there was a significant negative correlation in STAD, BRCA, colon adenocarcinoma (COAD), PRAD, SKCM, BLCA, uterine carcinosarcoma (UCS), and PCPG (Figure 8D). Also, there was a significant positive correlation between HOXA-AS2 expression and MSI in kidney chromophobe (KICH) and BRCA, while there was a significant negative correlation in COAD and STAD (Figure 8E).

Analysis of HOXA-AS2-related genes

The top 150 genes were screened co-expressed by HOXA-AS2 using the MEM-Multi Experiment Matrix database. HOXA2, HOXA5, and HOXA3 were the top three co-expressed genes ranked by *p*-value and interrelated with HOXA-AS2 expression (Supplementary Figure S2). In

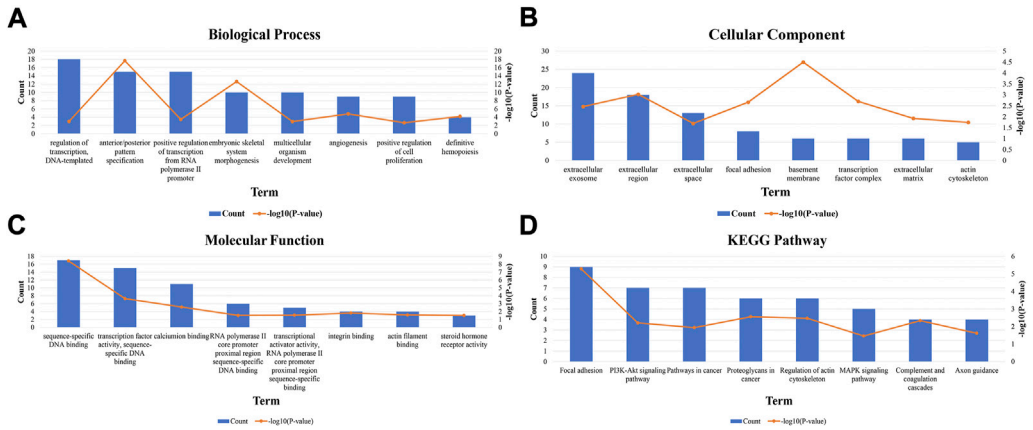


FIGURE 9
GO terms and the KEGG pathway. **(A)** GO enrichment of target genes in biological process ontolagy ($p < 0.05$).**(B)** GO enichment of target genes in cellular component ontology ($p < 0.05$). **(C)** GO enrichment of target genes in molecular function ontology ($p < 0.05$).**(D)** The top 8 pathways related to the differentially expressed genes by the KEGG database analysis. BP, biological process; CC, cellular component GO, gene ontology analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; MF, molecular function.

TABLE 4 Gene ontology analysis of HOXA-AS2-related genes.

GO number	Description	Genes	*P* Value
GO:0009952	anterior/posterior pattern specification	RARG, NR2F2, HOXA10, HOXA9, HOXA3, HOXB4, HOXB3, HOXA2, HOXB2, HOXA7, HOXA6, HOXB7, HOXA5, HOXB6, HOXB5	2.25E-18
GO:0048704	embryonic skeletal system morphogenesis	HOXA3, HOXB4, HOXB3, HOXB2, HOXA7, HOXB7, HOXA6, HOXB6, HOXA5, HOXB5	2.37E-13
GO:0043565	sequence-specific DNA binding	RARG, NR2F2, MEIS2, HOXA10, HOXA9, HOXA3, HOXB4, HOXB3, RARB, HOXB2, HOXA1, HOXA7, HOXA6, HOXB7, HOXA5, HOXB6, HOXA4	3.81E-09
GO:0001525	Angiogenesis	LAMA5, TGFB2, MEIS1, NRP2, ID1, HOXA3, HOXB3, EPHB2, HOXA7	1.56E-05
GO:0005604	basement membrane	LAMA5, CCDC80, COL4A1, NTN4, P3H2, FBLN1	3.18E-05
GO:0060216	definitive hemopoiesis	MEIS1, HOXA9, HOXB4, HOXB3	6.45E-05
GO:0003700	transcription factor activity, sequence-specific DNA binding	RARG, NR2F2, MEIS2, MECOM, ID1, HOXA3, HOXB4, HOXB3, HOXB2, HOXA6, HOXB7, HOXA5, TEAD2, HOXB6, HOXA4	0.000228
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	WWTR1, RARG, MEIS2, CYR61, EGFR, HOXA10, MEIS1, HOXB4, RARB, HOXA2, HOXA7, MET, HOXA5, TEAD2, HOXB5	0.000338
GO:0007435	salivary gland morphogenesis	TGFB2, TWSG1, EGFR	0.000511
GO:0005576	extracellular region	LAMA5, TGFB2, NRP2, C1R, CFI, NTN4, FBLN1, LTBP3, TFPI, CYR61, CYB5D2, BMP1, COL4A1, PDGFC, SERPING1, IGFBP6, EPHB2, MET	0.000943

addition, GO and KEGG pathway analyses were performed to explore the underlying molecular mechanisms (Figure 9, Table 4), and a signaling pathway network was constructed using Cytoscape software (Figure 10).

Discussion

According to global cancer statistics 2020, there were approximately 19.3 million new cancer cases and 10.0 million cancer deaths worldwide (Sung et al., 2021). Despite the variety of treatments available today, cancer has a high rate of recurrence and death, resulting in increased costs and poor patient prognosis (Huang et al., 2020). For most types of cancer, early detection and treatment improve prognoses. LncRNA was recently found to have huge clinical value in the early diagnosis and novel treatment of patients with cancer (Dunn et al., 2010). Numerous studies showed that lncRNAs were involved in a variety of physiological and pathological processes and had

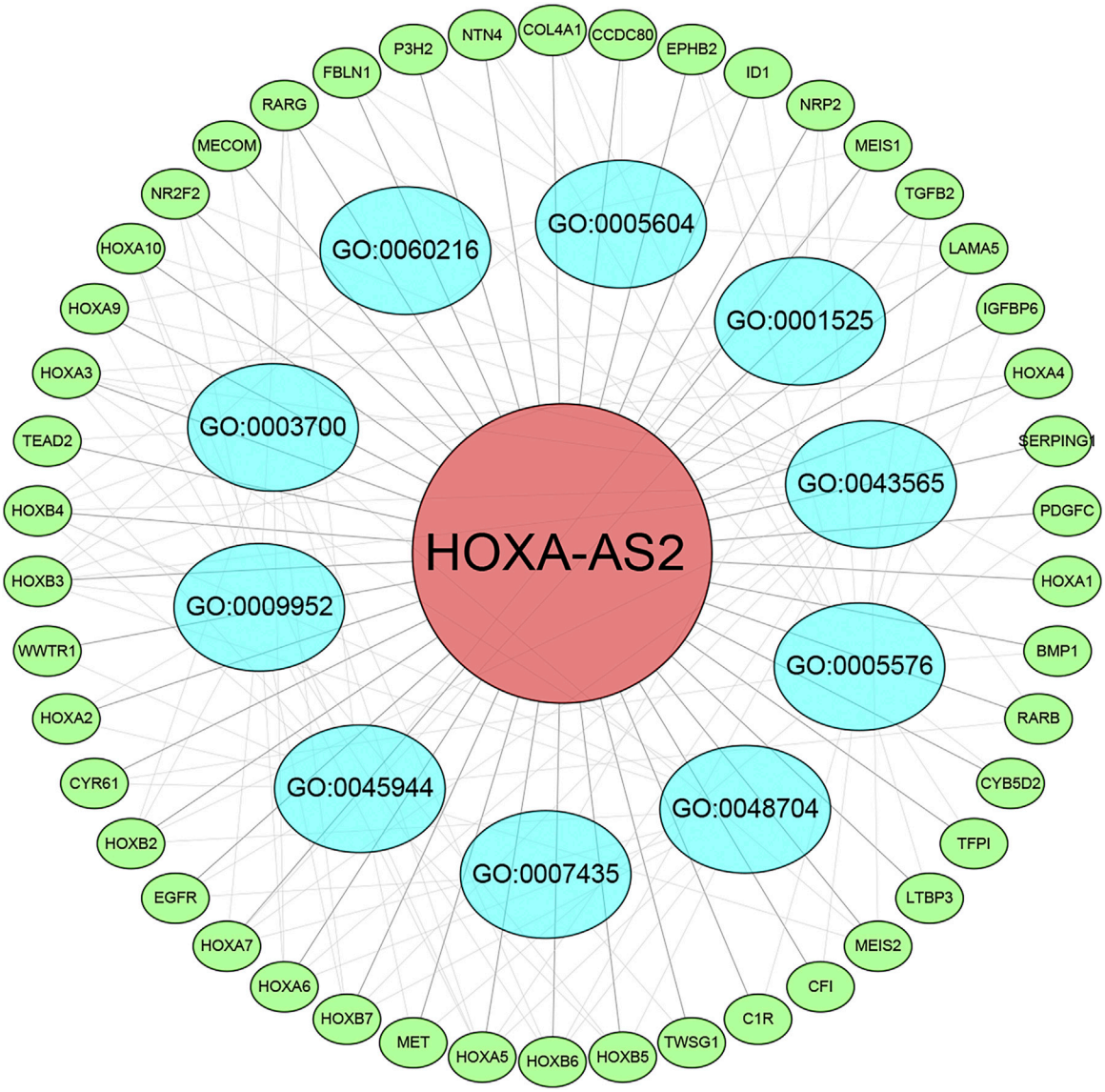


FIGURE 10
Differentially expressed gene interaction network analysis. Green nodes represent target genes and sky blue nodes represent the related pathway. As indicated in red, HOXA-AS2 localized at the center of the network.

important effects on tumorigenesis and tumor growth (Renganathan and Felley-Bosco, 2017). For instance, prostate cancer-associated transcript 6 (PCAT6) was significantly increased in various cancers. The overexpression of PCAT6 was closely correlated with OS, TNM stage, distant metastasis, LNM, tumor size, and the degree of differentiation in cancer patients (Shi et al., 2021), which may be a new cancer-related biomarker. In recent years, numerous studies found that the HOXA-AS2 was overexpressed in a variety of tumor types, such as CC (Chen and He, 2021), OSCC (Chen et al., 2021), PCa (Xiao and Song, 2020) and so on. The function of HOXA-AS2 in many malignancies has yet to be realized. Thus, we conducted the current meta-analysis to assess the predictive and clinical importance of HOXA-AS2 aberrant expression in cancer patients.

Our meta-analysis found that HOXA-AS2 overexpression was related to a lower chance of survival. We also examined the connection between HOXA-AS2 overexpression and several clinicopathological features. HOXA-AS2 overexpression was associated with tumor stage, LNM, tumor size and distant metastasis. However, there were no significant associations

TABLE 5 Summary of HOXA-AS2 functional roles and related genes.

Cancer	Expression	Functional role	Related genes	References
Oral squamous cell Carcinoma	Upregulate	Cell proliferation and migration	miR-567/CDK8	Chen et al. (2021)
Colorectal cancer	Upregulate	Cell proliferation and apoptosis	p21 and KLF2	Li et al. (2016)
				Ding et al. (2017)
Breast cancer	Upregulate	Cell proliferation	miR-520c-3p	Fang et al. (2017)
Non-small cell lung cancer	Upregulate	Cell migration, invasion, proliferation, metastasis	miR-520a-3p	Li and Jiang, (2017)
				Cui et al. (2019)
				Liu et al. (2019)
Acute myeloid leukemia	Upregulate	Cell proliferation, invasion	SOX4/PI3K/AKT	Qu et al. (2020)
Hepatocellular Carcinoma	Upregulate	Cell migration, invasion	p-AKT, MMP-2 and MMP-9/miR-520c-3p/GPC3	Wang et al. (2016)
				Zhang et al. (2018)
				Lu et al. (2020)
Bladder cancer	Upregulate	Cell proliferation, invasion	miR-125b/Smad2	Wang et al. (2019a)
Osteosarcoma	Upregulate	Cell migration and invasion	miR-124–3p/E2F3	Wang et al. (2018)
				Wang et al. (2019b)
Glioma	Upregulate	Cell proliferation and invasion and promoted apoptosis	RND3	Xie et al. (2015)
Thyroid	Upregulate	Cell migration and invasion	miR-520c-3p/S100A4	Xia et al. (2018)
				Jiang et al. (2019)
Prostate cancer	Upregulate	Cell proliferation, migration, invasion and EMT	miR-509–3p/PBX3	Xiao and Song, (2020)
Gastric cancer	Upregulate	Cell proliferation and apoptosis	P21/PLK3/DDIT3	Xie et al. (2015)
Cervical cancer	Upregulate	Cell proliferation migration, invasion	miR-509–3p/BTN3A1	Chen and He, (2021)

with age, gender, differentiation, or depth of invasion. Furthermore, we performed further analysis of the prognostic role of HOXA-AS2 in various cancers using several public databases. Among them, Cox regression analysis indicated that HOXA-AS2 had a better prognostic value in GBM, STAD, LGG, ACC, SKCM, BRCA, LUAD, BLCA, SARC, and CRC. These results suggest that HOXA-AS2 is a predictor of poor prognosis in cancer patients.

Our findings demonstrated a strong correlation between HOXA-AS2 expression and immunity in multiple tumors. Tumor-infiltrating immune cells play an irreplaceable role in tumor development. Previous research has shown that HOXA-AS2 could influence glioma progression by regulating Treg cell proliferation and immune tolerance (Zhong et al., 2022). We found that HOXA-AS2 was associated with multiple infiltrating immune cells in a variety of tumors. However, the regulatory mechanism of HOXA-AS2 on ICI still remains to be further confirmed by abundant experiments. The interaction between TME and tumor cells is decisive for tumor survival and progression. Immune cells and stromal cells are key components of the TME (Xiao and Yu, 2021). We found that HOXA-AS2 expression correlated with immune cell scores in 11 tumors and with stromal cell scores in 13 tumors. It indicates that HOXA-AS2 has an essential role in the TME. TMB and MSI are well-directed for tumor immunotherapy. Our findings indicated that HOXA-AS2 expression correlated with TMB in 12 cancers and with MSI in 4 cancers. In summary, HOXA-AS2 may further influence the prognosis of cancer patients *via* modulation of tumor immunity.

Although HOXA-AS2 was demonstrated to be a major predictive factor for patients with various malignancies in several studies, the basic principle of how HOXA-AS2 caused cancer remains unknown. According to the results of this study, the overexpression of HOXA-AS2 can significantly aid cancer growth and metastasis. In contrast, inhibited HOXA-AS2 expression, significantly reduced cell proliferation, migration, and invasion, as well as the carcinogenesis process. In PCa, HOXA-AS2 exhibited a negative connection with miR-509–3p. The inhibition of HOXA-AS2 prevented PCa cells from proliferating and migrating (Xiao and Song, 2020). This suggested that HOXA-AS2 could be used as a therapeutic target to treat PCa. In addition, by decreasing miR-520c-3p expression, HOXA-AS2 enhanced the growth and spread of HCC (Wang et al., 2016). HOXA-AS2 was elevated in OSCC tissues and increased OSCC cell proliferation by sponging miR-567/CDK8 (Chen et al., 2021). Table 5 highlights the association between HOXA-AS2 and malignancies to investigate functionally associated genes.

We used the MEM-Multi Experiment Matrix database to predict target genes and perform the signaling pathway analysis of HOXA-AS2 to further investigate its value. HOXA2, HOXA5, and HOXA3,

all of which play important roles in cancer, were strongly associated with HOXA-AS2 expression in our study. Following that, we conducted GO analysis, which indicated that the sequence-specific DNA binding, extracellular exosomes, and angiogenesis of HOXA-AS2 were all significantly related. HOXA AS2 was highly connected to cancer-associated pathways in KEGG analysis.

Nevertheless, this meta-analysis had several limitations. First, some HRs and the corresponding 95% CIs were extracted from KM curves. Second, the qualifying studies were all performed in China, so it is unclear whether the results can be generalized to patients in other countries. To address this limitation, we validated the correlation between HOXA-AS2 expression and prognosis of cancer patients in public databases. Third, the included studies were inconsistent in dividing expression according to cut-off values. Additionally, only a few trials were included, and some cancer types had very low sample sizes. Thus, more clinical investigations should be conducted to assess the potential prognostic role of HOXA-AS2 expression in cancer types that were not included.

Conclusion

In summary, this meta-analysis found that HOXA-AS2 overexpression was linked to the poor prognosis of cancer patients and could be used as a new prognostic biomarker and therapeutic target for various malignancies. The predictive usefulness of HOXA-AS2 in tumors has to be confirmed in more studies, including other cancer types.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding authors.

Author contributions

XL and WL conceived this study. FZ, MG, and XP were responsible for the collection, extraction, and analysis of the data. FZ, GZ, and HZ was responsible for writing the paper. DZ, XX, and FC performed the quality evaluation and completed data analysis. XL and WL reviewed the paper. All authors read and approved the final manuscript.

Funding

This work was financially supported by Natural Science Foundation of Gansu Province (No. 20JR10RA365); Key R and D Program of Gansu Province-Social Development (No. 20YF8FA092).

Acknowledgments

We would like to thank all the researchers and study participants for their contributions.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.944278/full#supplementary-material

References

Bhan, A., Soleimani, M., and Mandal, S. S. (2017). Long noncoding RNA and cancer: A new paradigm. *Cancer Res.* 77 (15), 3965–3981. doi:10.1158/0008-5472.Can-16-2634

Chen, R., and He, P. (2021). Long noncoding RNA HOXA-AS2 accelerates cervical cancer by the miR-509-3p/BTN3A1 axis. *J. Pharm. Pharmacol.* 73 (10), 1387–1396. doi:10.1093/jpp/rgab090

Chen, R., Wang, X., Zhou, S., and Zeng, Z. (2021). LncRNA HOXA-AS2 promotes tumor progression by suppressing miR-567 expression in oral squamous cell carcinoma. *Cancer Manag. Res.* 13, 5443–5455. doi:10.2147/cmar.S305946

Chu, H. Y., Chen, Y. J., Hsu, C. J., Liu, Y. W., Chiou, J. F., Lu, L. S., et al. (2020). Physical cues in the microenvironment regulate stemness-dependent homing of breast cancer cells. *Cancers (Basel)* 12 (8), E2176. doi:10.3390/cancers12082176

Cui, T. J., Lin, G. S., Dai, Y. M., Zheng, J. P., Chen, Z., Chen, Q., et al. (2019). LncRNA HOXA-AS2 regulates microRNA-216a-5p to promote malignant progression of non-small cell lung cancer. *Eur. Rev. Med. Pharmacol. Sci.* 23 (3), 264–273. doi:10.26355/eurrev_201908_18656

Ding, J., Xie, M., Lian, Y., Zhu, Y., Peng, P., Wang, J., et al. (2017). Long noncoding RNA HOXA-AS2 represses P21 and KLF2 expression transcription by

binding with EZH2, LSD1 in colorectal cancer. *Oncogenesis* 6 (1), e288. doi:10.1038/oncsis.2016.84

Dunn, B. K., Wagner, P. D., Anderson, D., and Greenwald, P. (2010). Molecular markers for early detection. *Semin. Oncol.* 37 (3), 224–242. doi:10.1053/j.seminoncol.2010.05.007

Fang, Y., Wang, J., Wu, F., Song, Y., Zhao, S., and Zhang, Q. (2017). Long noncoding RNA HOXA-AS2 promotes proliferation and invasion of breast cancer by acting as a miR-520c-3p sponge. *Oncotarget* 8 (28), 46090–46103. doi:10.18632/oncotarget.17552

Huang, S., Yang, J., Fong, S., and Zhao, Q. (2020). Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. *Cancer Lett.* 471, 61–71. doi:10.1016/j.canlet.2019.12.007

Jiang, L., Wu, Z., Meng, X., Chu, X., Huang, H., and Xu, C. (2019). LncRNA HOXA-AS2 facilitates tumorigenesis and progression of papillary thyroid cancer by modulating the miR-15a-5p/HOXA3 Axis. *Hum. Gene Ther.* 30 (5), 618–631. doi:10.1089/hum.2018.109

Kole, R., Krainer, A. R., and Altman, S. (2012). RNA therapeutics: Beyond RNA interference and antisense oligonucleotides. *Nat. Rev. Drug Discov.* 11 (2), 125–140. doi:10.1038/nrd3625

Li, Q., Dai, Y., Wang, F., and Hou, S. (2016). Differentially expressed long noncoding RNAs and the prognostic potential in colorectal cancer. *Neoplasma* 63 (6), 977–983. doi:10.4149/neo_2016_617

Li, W., Li, J., Mu, H., Guo, M., and Deng, H. (2019). MiR-503 suppresses cell proliferation and invasion of gastric cancer by targeting HMGA2 and inactivating WNT signaling pathway. *Cancer Cell. Int.* 19, 164. doi:10.1186/s12935-019-0875-1

Li, Y., and Jiang, H. (2017). Up-regulation of long non-coding RNA HOXA-AS2 in non-small cell lung cancer is associated with worse survival outcome. *Int. J. Clin. Exp. Pathol.* 10 (9), 9690–9696.

Liu, Y., Lin, X., Zhou, S., Zhang, P., Shao, G., and Yang, Z. (2019). Long noncoding RNA HOXA-AS2 promotes non-small cell lung cancer progression by regulating miR-520a-3p. *Biosci. Rep.* 39 (5), BSR20190283. doi:10.1042/bsr20190283

Lu, Q., Gao, J., Tang, S., Li, Z., Wang, X., Deng, C., et al. (2020). Integrated RNA sequencing and single-cell mass cytometry reveal a novel role of LncRNA HOXA-AS2 in tumorigenesis and stemness of hepatocellular carcinoma. *Onco. Targets. Ther.* 13, 10901–10916. doi:10.2147/OTT.S272717

Morlando, M., and Fatica, A. (2018). Alteration of epigenetic regulation by long noncoding RNAs in cancer. *Int. J. Mol. Sci.* 19 (2), E570. doi:10.3390/ijms19020570

Qu, Y., Wang, Y., Wang, P., Lin, N., Yan, X., and Li, Y. (2020). Overexpression of long noncoding RNA HOXA-AS2 predicts an adverse prognosis and promotes tumorigenesis via SOX4/PI3K/AKT pathway in acute myeloid leukemia. *Cell. Biol. Int.* 44 (8), 1745–1759. doi:10.1002/cbin.11370

Reinhold, W. C., Sunshine, M., Liu, H., Varma, S., Kohn, K. W., Morris, J., et al. (2012). CellMiner: A web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res.* 72 (14), 3499–3511. doi:10.1158/0008-5472.Can-12-1370

Renganathan, A., and Felley-Bosco, E. (2017). Long noncoding RNAs in cancer and therapeutic potential. *Adv. Exp. Med. Biol.* 1008, 199–222. doi:10.1007/978-981-10-5203-3_7

Schmitt, A. M., and Chang, H. Y. (2016). Long noncoding RNAs in cancer pathways. *Cancer Cell.* 29 (4), 452–463. doi:10.1016/j.ccell.2016.03.010

Shi, S. B., Cheng, Q. H., Gong, S. Y., Lu, T. T., Guo, S. F., Song, S. M., et al. (2021). PCAT6 may be a new prognostic biomarker in various cancers: A meta-analysis and bioinformatics analysis. *Cancer Cell. Int.* 21 (1), 370. doi:10.1186/s12935-021-02079-4

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Tang, L., Chen, Y. L., Tang, X., Wei, D., Xu, X. Y., and Yan, F. (2020). Long noncoding RNA DCST1-AS1 promotes cell proliferation and metastasis in triple-negative breast cancer by forming a positive regulatory loop with miR-873-5p and MYC. *J. Cancer* 11 (2), 311–323. doi:10.7150/jca.33982

Wang, F., Wu, D., Chen, J., Chen, S., He, F., Fu, H., et al. (2019a). Long non-coding RNA HOXA-AS2 promotes the migration, invasion and stemness of bladder cancer via regulating miR-125b/Smad2 axis. *Exp. Cell. Res.* 375 (1), 1–10. doi:10.1016/j.yexcr.2018.11.005

Wang, F., Yang, H., Deng, Z., Su, Y., Fang, Q., and Yin, Z. (2016). HOX antisense lincRNA HOXA-AS2 promotes tumorigenesis of hepatocellular carcinoma. *Cell. Physiol. biochem.* 40 (1-2), 287–296. doi:10.1159/000452545

Wang, J., Wang, W., Tang, Q., Lu, L., Luo, Z., Li, W., et al. (2020a). Long non-coding RNA lnc-GNAT1-1 suppresses liver cancer progression via modulation of epithelial-mesenchymal transition. *Front. Genet.* 11, 1029. doi:10.3389/fgene.2020.01029

Wang, L., Wang, L., and Zhang, X. (2019b). Knockdown of lncRNA HOXA-AS2 inhibits viability, migration and invasion of osteosarcoma cells by miR-124-3p/E2F3. *Onco. Targets. Ther.* 12, 10851–10861. doi:10.2147/ott.S220072

Wang, Y., Yang, C., Liu, X., Zheng, J., Zhang, F., Wang, D., et al. (2020b). Transcription factor AP-4 (TFAP4)-upstream ORF coding 66 aa inhibits the malignant behaviors of glioma cells by suppressing the TFAP4/long noncoding RNA 00520/microRNA-520f-3p feedback loop. *Cancer Sci.* 111 (3), 891–906. doi:10.1111/cas.14308

Wang, Y., Zhang, R., Cheng, G., Xu, R., and Han, X. (2018). Long non-coding RNA HOXA-AS2 promotes migration and invasion by acting as a ceRNA of miR-520c-3p in osteosarcoma cells. *Cell. Cycle* 17 (13), 1637–1648. doi:10.1080/15384101.2018.1489174

Wu, L., Zhu, X., Song, Z., Chen, D., Guo, M., Liang, J., et al. (2019). Long non-coding RNA HOXA-AS2 enhances the malignant biological behaviors in glioma by epigenetically regulating RND3 expression. *Onco. Targets. Ther.* 12, 9407–9419. doi:10.2147/ott.S225678

Xia, F., Chen, Y., Jiang, B., Du, X., Peng, Y., Wang, W., et al. (2018). Long noncoding RNA HOXA-AS2 promotes papillary thyroid cancer progression by regulating miR-520c-3p/S100A4 pathway. *Cell. Physiol. biochem.* 50 (5), 1659–1672. doi:10.1159/000494786

Xiao, S., and Song, B. (2020). LncRNA HOXA-AS2 promotes the progression of prostate cancer via targeting miR-509-3p/PBX3 axis. *Biosci. Rep.* 40 (8), BSR20193287. doi:10.1042/bsr20193287

Xiao, Y., and Yu, D. (2021). Tumor microenvironment as a therapeutic target in cancer. *Pharmacol. Ther.* 221, 107753. doi:10.1016/j.pharmthera.2020.107753

Xie, M., Sun, M., Zhu, Y. N., Xia, R., Liu, Y. W., Ding, J., et al. (2015). Long noncoding RNA HOXA-AS2 promotes gastric cancer proliferation by epigenetically silencing P21/PLK3/DDIT3 expression. *Oncotarget* 6 (32), 33587–33601. doi:10.18632/oncotarget.5599

Xu, X., Zhong, Z., Shao, Y., and Yi, Y. (2021). Prognostic value of MEG3 and its correlation with immune infiltrates in gliomas. *Front. Genet.* 12, 679097. doi:10.3389/fgene.2021.679097

Ye, J., Sun, H., Feng, Z., Zhang, Q., Xia, Y., Ji, Y., et al. (2019). Prognostic significance of LncRNA GHET1 expression in various cancers: A systematic review and meta-analysis. *Biosci. Rep.* 39 (10), BSR20190608. doi:10.1042/bsr20190608

Zhang, C., Ren, X., He, J., Wang, W., Tu, C., and Li, Z. (2019). The prognostic value of long noncoding RNA SNHG16 on clinical outcomes in human cancers: A systematic review and meta-analysis. *Cancer Cell. Int.* 19, 261. doi:10.1186/s12935-019-0971-2

Zhang, Y., Xu, J., Zhang, S., An, J., Zhang, J., Huang, J., et al. (2018). HOXA-AS2 promotes proliferation and induces epithelial-mesenchymal transition via the miR-520c-3p/GPC3 Axis in hepatocellular carcinoma. *Cell. Physiol. biochem.* 50 (6), 2124–2138. doi:10.1159/000495056

Zhong, C., Tao, B., Li, X., Xiang, W., Peng, L., Peng, T., et al. (2022). HOXA-AS2 contributes to regulatory T cell proliferation and immune tolerance in glioma through the miR-302a/KDM2A/JAG1 axis. *Cell. Death Dis.* 13 (2), 160. doi:10.1038/s41419-021-04471-4

Zhou, Q., Hou, Z., Zuo, S., Zhou, X., Feng, Y., Sun, Y., et al. (2019). LUCAT1 promotes colorectal cancer tumorigenesis by targeting the ribosomal protein L40-MDM2-p53 pathway through binding with UBA52. *Cancer Sci.* 110 (4), 1194–1207. doi:10.1111/cas.13951

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 16 November 2022
DOI 10.3389/fgene.2022.961278

Check for updates

OPEN ACCESS

EDITED BY
Peter Hart,
Roosevelt University College of Pharmacy, United States

REVIEWED BY
Jiani Ma,
China University of Mining and Technology, China
Shijia Zhai,
Lanzhou University Second Hospital, China

*CORRESPONDENCE
Xiangyu Wang,
wang_xy123@126.com

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 04 June 2022
ACCEPTED 02 November 2022
PUBLISHED 16 November 2022

CITATION
Wu S, Ballah AK, Che W and Wang X (2022), M^{7}G-related LncRNAs: A comprehensive analysis of the prognosis and immunity in glioma. *Front. Genet.* 13:961278. doi: 10.3389/fgene.2022.961278

M^{7}G-related LncRNAs: A comprehensive analysis of the prognosis and immunity in glioma

Shuaishuai Wu, Augustine K. Ballah, Wenqiang Che and Xiangyu Wang*

First Affiliated Hospital, Jinan University, Department of Neurosurgery, Guangzhou, China

Today, numerous international researchers have demonstrated that N^7-methylguanosine (m^{7}G) related long non-coding RNAs (m^{7}G-related lncRNAs) are closely linked to the happenings and developments of various human beings' cancers. However, the connection between m^{7}G-related lncRNAs and glioma prognosis has not been investigated. We did this study to look for new potential biomarkers and construct an m^{7}G-related lncRNA prognostic signature for glioma. We identified those lncRNAs associated with DEGs from glioma tissue sequences as m^{7}G-related lncRNAs. First, we used Pearson's correlation analysis to identify 28 DEGs by glioma and normal brain tissue gene sequences and predicated 657 m^{7}G-related lncRNAs. Then, eight lncRNAs associated with prognosis were obtained and used to construct the m^{7}G risk score model by lasso and Cox regression analysis methods. Furthermore, we used Kaplan-Meier analysis, time-dependent ROC, principal component analysis, clinical variables, independent prognostic analysis, nomograms, calibration curves, and expression levels of lncRNAs to determine the model's accuracy. Importantly, we validated the model with external and internal validation methods and found it has strong predictive power. Finally, we performed functional enrichment analysis (GSEA, aaGSEA enrichment analyses) and analyzed immune checkpoints, associated pathways, and drug sensitivity based on predictors. In conclusion, we successfully constructed the formula of m^{7}G-related lncRNAs with powerful predictive functions. Our study provides instructional value for analyzing glioma pathogenesis and offers potential research targets for glioma treatment and scientific research.

KEYWORDS

glioma, lncRNA, m 7 G, prognosis model, immune analysis, therapy

Introduction

Glioma is one of the most common aggressive and fatal primary tumors in the central nervous system, accounting for approximately 30% of cases (Mousavi et al., 2022). They are graded by the World Health Organization (WHO) as I to IV with increasing malignancy based on the histopathological characteristics of the tumor (Ostrom et al., 2019). Although genetic and molecular testing has brought advances in disease diagnosis, surgery, radiotherapy, and other comprehensive treatments have brought hope to patients; their prognosis is still poor. It is getting more severe economic pressure and burdening patients, their families, and society (Frances et al., 2022; Haddad et al., 2022). Thus, there is an urgent need to detect glioma-related biomarkers in our clinical care for early identification and diagnosis and to investigate new therapeutic approaches.

Although long non-coding RNA (lncRNA) is non-coding RNA that cannot be translated into protein RNA molecules, several reports have demonstrated that lncRNA regulates tumorigenesis and development (Yang et al., 2016; Chen et al., 2021). For example, *LINC01503* promotes the cancer stem cell properties of glial cells by reducing the degradation of GLI2 (Wei et al., 2022). The lncRNA *HOXA-AS2* can enhance the expression of *KDM2A*/*JAG1*, which can contribute to Treg cell proliferation and immune tolerance in gliomas and promote tumor development (Zhong et al., 2022). The downregulation of lncRNA *TTTY15*, which targets miR-4500, could regulate the proliferation and apoptosis of A172 glioma cells (Wang et al., 2022). LncRNA *IRAIN* overexpression inhibits glioma progression and temozolomide resistance by suppressing the PI3 K-related signaling pathway (Guo et al., 2022). LncRNA *KB-1460A1.5* suppresses glioma development through the miR-130a-3p feedback loop (Xu et al., 2022). Despite some progress in previous studies, few biomarkers have been studied for lncRNA prognosis to differentiate patients. Therefore, we investigated the prognostic role of m^7G-related lncRNAs in glioma by identifying m^7G-related DEGs in glioma in order to be able to find more useful biomarkers for glioma.

N^7-methylguanosine (m^7G) refers to the methylation of guanosine at the N^7 position. m^7G RNA modification is one of the most common posttranscriptional modifications; it is widely distributed in the 5′hat region of tRNA, rRNA, and eukaryotic mRNA and plays an essential role in gene expression, protein synthesis and transcriptional stability (Pei and Shuman, 2002; Jaffrey, 2014; Song et al., 2020). M^7G can regulate the secondary structure of RNA or protein-RNA interaction through electrostatic and spatial effects (Furuichi, 2015). Current studies have demonstrated that almost every stage of the life cycle can be adjusted by m^7G modifications, such as transcription (Pei and Shuman, 2002), mRNA splicing (Jiang et al., 2018), nuclear output (Lewis and Izaurralde, 1997), and translation (Marchand et al., 2018). The mutation of m^7G methyltransferase is related to many diseases. Mutations, knockouts, and overexpression of m^7G-related genes, such as WD repeat domain 4 (*WDR4*), lead to microcephalic primordial dwarfism (Sauna and Kimchi-Sarfaty, 2011), Nervous system damage (Lin et al., 2018), and impairment of learning and memory abilities (Pereira et al., 2009). Furthermore, *METTL1* is an author of m^7G, essential for suppressing lung cancer cell migration through m^7G editing on RAS and MYC driver genes (Balzeau et al., 2017; Pandolfini et al., 2019). Also, overexpression of mettl1 and bad prognosis of patients with liver cancer is associated with the downregulation of tumor suppressors in hepatocellular carcinoma (Barbieri et al., 2017; Tian et al., 2019). The tRNA N^7-methylguanosine modification mediated by *METTL1*/*WDR4* promotes the development of squamous cell carcinoma (Chen et al., 2022). Furthermore, METTL1-m7G-EGFR/EFEMP1 axis is a precise mechanism for bladder cancer development (Ying et al., 2021). Therefore, if we want a further biological understanding of the interaction between lncRNA and cancer, we must study m^7G modifications and explore new prognostic and therapeutic markers. In this study, we constructed a formula based on m^7G prognosis-related lncRNAs; and verified their outstanding performance in prognosis prediction. The lncRNAs associated with glioma prognosis were also identified, which may provide potential research directions for analyzing glioma's pathogenesis and clinical treatment.

Materials and methods

Patients and datasets

We downloaded glioma data (GBM and LGG) and normal brain tissue RNA transcriptome data from the Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) website (698 glioma samples and 1152 normal human brain samples, respectively). Validation data were available from the China Glioma Genome Atlas (CGGA,1018 glioma samples). Meanwhile, clinical information of glioma patients was downloaded from the TCGA and CGGA databases, and patients without follow-up data or an overall survival <30 days were excluded. Since the data in this study were obtained from public databases, ethics committee approval was not required according to the relevant regulations of the databases.

Identify the expression of m^7G-related genes

First, we obtained 3 genes from the published article about m^7G (Tomikawa, 2018; Pandolfini et al., 2019; Teng et al., 2021). Then we searched for three biological pathways related to m^7G in

GSEA and extracted genes involved in each pathway. After removing duplicate genes and summarizing the above genes, we obtained 29 genes. Then, we used Wilcoxon's method (*FDR* < 0.05, Log_2FC >1) to screen the genes with significant differences in the expression level between glioma and normal tissues based on these 29 genes. After deleting the genes with no significant differences, the remaining ones are m^7G-related differentially expressed genes (m^7G-related DEGs), and named them m^7G-related genes (Supplementary Material S1). Expression of m^7G-related DEGs samples were visualized using vioplot. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis implemented in R.

Establishment of the risk signature

First, we performed co-expression analysis of 28 m^7G-related genes and lncRNAs in the TCGA and GTEx glioma and normal brain tissue datasets, identifying 657 m^7G-related lncRNAs (Pearson correlation coefficients >0.4, $p < 0.001$, Supplementary Material S2). Secondly, the prognostic relationship of m^7G-related lncRNAs was assessed by univariate Cox regression (Supplementary Material S3). In the univariate analysis, the m^7G-related lncRNAs with $p < 0.01$ (539 lncRNAs) were included in the least absolute shrinkage and selection operator (Lasso) regression. The results derived from Lasso regression were then incorporated into a multivariate Cox model to derive eight prognostic m^7G-related lncRNAs and create the risk scores (RS)formula:

$$\text{risk score} = \sum_{i=1}^{n} \text{coef m7GLnc}Sigi \times \text{EXP m7GLnc}Sigi$$

The "coef m7GLncsigi" in this "risk scores" formula represents the coefficient value, which is the regression coefficient of the 8 prognostic lncRNAs derived from the multifactorial regression analysis. The "EXP m7GLncSigi" in the formula represents the expression levels of the 8 m7G-related lncRNAs. By using the RS formula, we can get the risk value of each patient. And after getting the risk value of all patients, we can find out the median risk value of the patients. According to the median value, we can determine the level of risk of the patients.

Validation of the risk scoring model

Kaplan-Meier (K-M) analysis, time-dependent ROC, principal component analysis (PCA), independent prognostic analysis, nomogram (1-, 3-, and 5-year), calibration curve and the expression level of lncRNAs are used to determine the accuracy of the model. In the CGGA validation sample, we applied the same intermediate values to assess the validity and reliability of our RS formula using the same way as above. We also use the same approach to randomly divide the TCGA data into two groups for internal validation.

Functional annotation of prognostic m^7G-related LncRNAs

We divided the patients into high-risk and low-risk groups based on the median risk score. GSEA (version 4.1.0, ($p < 0.05$ and FDR <0.25))was used for functional enrichment analysis (Subramanian et al., 2005). The infiltrating fraction of 16 immune cells and the activity of 13 immune-related pathways were measured by ssGSEA (Rooney et al., 2015). We also explored the relationship between risk scores and immune checkpoints in both risk groups (Yao et al., 2021).

Drug sensitivity correlation analysis

To find more drugs for the treatment of glioma, we focused on evaluating and predicting immune-related drugs. According to the online tool Cancer Drug Sensitivity Genomics, the IC50 of different drugs on glioma samples was predicted using the R package 'pRRophetic'. The main use of 'pRRophetic' is to predict phenotypes from gene expression data (to predict clinical outcomes using Cancer Genome Project CGP cell line data), to predict drug sensitivity in external cell lines (CCLE) and also for clinical data prediction.

Statistical analysis

This study used R software (version 4.1.2) and GSEA software for statistical analysis. Wilcoxon test was used to identify the expression levels of m7G-related DEGs in cancer and normal tissues. Survival curves were generated using the Kaplan-Meier method and compared using the log-rank test. Cox regression and Lasso regression were used to evaluate the prognostic influences of m7G-related lncRNA features.

Results

Differential expression and enrichment analysis of m^7G-related genes

After analysis, we found that 28 m7G-related genes were significantly differentially expressed between glioma and normal tissues (Figure 1A). Specifically, *NUDT11*, *IFIT5*, *GEMIN5*, *METTL1*, *CYFIP1*, *NCBP1*, *WDR4*, *NUDT10*, *EIF3D*, *LARP1*, *DCP2*, *DCPS*, *AGO2*, *NCBP2*, *EIF4G3* and *LSM1* were highly expressed in tumor samples ($p < 0.001$). *NSUN2*, *NUDT4*, *EIF4E2*, *SNUPN*, *NCBP3*, *NUDT3*, *EIF4E3*, *EIF4E*, *NCBP2L*, *EIF4E1B*, *EIF4A1* and *NUDT4B* were lowly expressed in tumor samples ($p < 0.001$). The expression levels of *NUDT16* were not significantly different ($p = 0.517$) (Figure 1B). In addition, to further understand the intrinsic association between the 28 m^7G-related genes, we also performed a correlation analysis. The results showed that the positive correlation between *GEMIN5* and *NCBP1* was the most significant, and the negative correlation between *GEMIN5* and *EIF4A1* was the most significant (Figure 1C). The above results

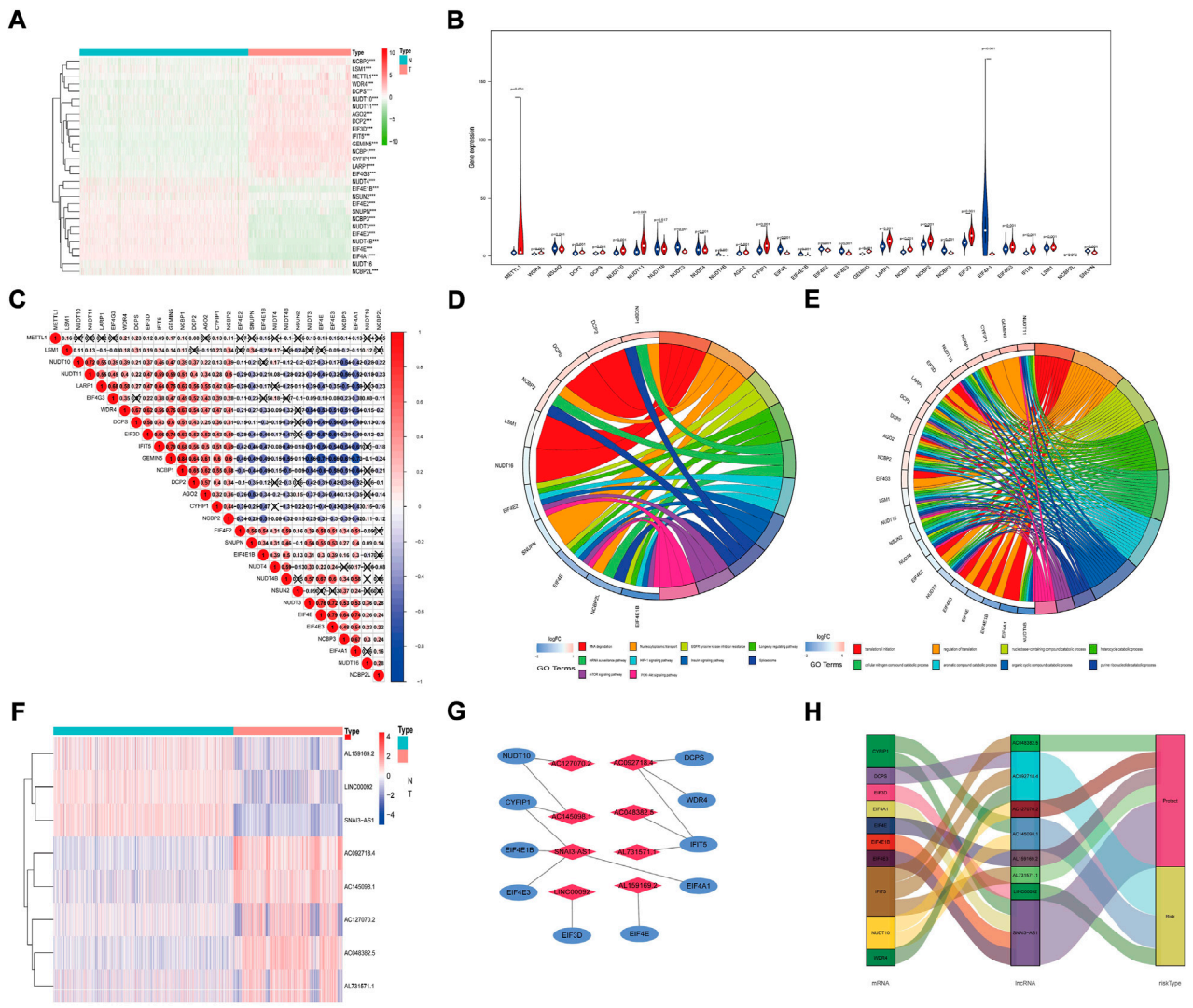


FIGURE 1
Expression of m^{7}G-related mRNAs and prognostic m^{7}G-Related lncRNAs. **(A)** Heat Map shows the expression levels of eight m^{7}G-related mRNAs. **(B)** The violplot shows the differentially m^{7}G-related mRNAs. Blue represents normal sample, and red represents the glioma sample. **(C)** Spearman correlation analysis of m^{7}G-related mRNAs. **(D)** KEGG circle diagram of m^{7}G-related DEGs. **(E)** GO circle diagram of m^{7}G-related DEGs. **(F)** The expression levels of eight prognostic m^{7}G-Related lncRNAs. **(G)** The co-expression network of prognostic m^{7}G-related lncRNAs. **(H)** Sankey diagram of prognostic m^{7}G-Related lncRNAs. lncRNAs, long non-coding RNAs; N, normal; T, tumor.

suggest some interaction between m^{7}G-related genes in glioma. Then, KEGG pathway analysis showed that m^{7}G-related DEGs were mainly enriched in RNA degradation, Nucleocytoplasmic transport, *EGFR* tyrosine kinase inhibitor resistance, Longevity regulating pathway, mRNA surveillance pathway, *HIF-1* signalling pathway, Insulin signalling pathway, Spliceosome, mTOR signalling pathway, *HIF-1* signalling pathway and PI3K-Akt signalling pathway (Figure 1D). GO analysis showed that DEGs were mainly enriched in the regulation of translation, nucleobase-containing compound catabolic process, heterocycle catabolic process, *Etc.* (Figure 1E).

Screening prognostic m^{7}G-related LncRNAs

We identified 658 lncRNAs associated with m^{7}G-related genes. Univariate Cox regression analysis showed that 540 lncRNAs were linked to patient prognosis. One hundred thirty-two were considered risk lncRNAs with HR > 1, while 408 were protective lncRNAs with HR < 1. After Lasso regression, 28 m7G-associated lncRNAs were identified. Finally, multivariate Cox regression identified 8 lncRNAs with the best prognostic correlation (*AC048382.5*, *AC127070.2*,

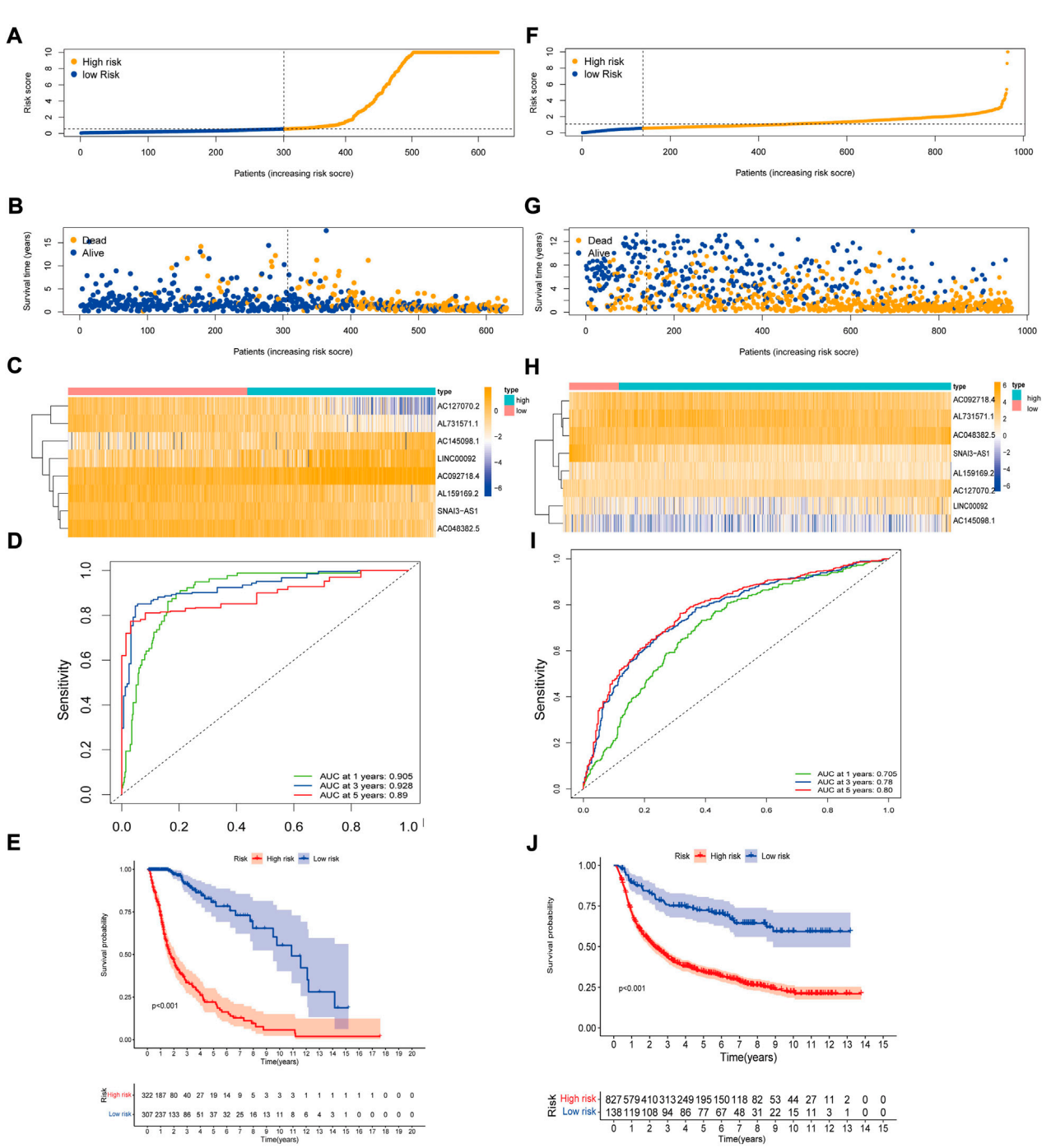


FIGURE 2
Construction and validation of the 8 prognostic m^{7}G-related lncRNAs signature for survival prediction. **(A)** Distribution of RS in TCGA. **(B)** Survival time and status of patients in TCGA. **(C)** Heatmap of prognostic m^{7}G-related lncRNAs of RS in TCGA. **(D)** ROC curve for TCGA. **(E)** KM curve for TCGA. **(F)**Distribution of RS in CGGA. **(G)**Survival time and status of patients in CGGA. **(H)**Heatmap of m^{7}G-related lncRNAs of RS in CGGA. **(I)**ROC curve for CGGA. **(J)** KM curve for CGGA.

AL159169.2, *AL731571.1*, *SNAI3-AS1*, *AC092718.4*, *AC145098.1*, *LINC00092*) (Supplementary Material S4). The expression levels of the eight prognostic m^{7}G-related lncRNAs are shown (Figure 1F). We used the Cytoscape and 'galluvial' R packages to visualize the lncRNAs. The co-expression network contained 14 lncRNA-mRNA pairs (Figure 1G, R2>0.4, $p < 0.001$). *SNAI3-AS1* was co-expressed with four related genes (*EIF4A1*, *EIF4E3*, *EIF4E1B*, and *CYFIP1*), AC092718.4 was co-expressed with three

related genes (*IFIT5*, *DCPS*, and *WDR4)*, and *AC145098.1* was co-expressed with two related genes (*CYFIP1* and *NUDT10*), *AC127070.2* co-expressed with *NUDT10*, *AC048382.5* and *AL731571.1* both co-expressed with *IFIT5*, *LINC00092* co-expressed with *EIF3D* and *AL159169.2* co-expressed with *EIF4E*. *AC048382.5*, *AC127070.2*, *AL159169.2*, *AL731571.1*, and *SNAI3-AS1* were protective factors, while *AC092718.4*, *AC145098.1*, and *LINC00092* were risk factors (Figure 1H).

Development and validation of prognostic models

Based on the above eight lncRNAs, we constructed a prognostic model and calculated the risk score for each patient using the risk score model. The risk score formula worked as follows: risk score =(0.620302782 × AC092718.4 expression) + (0.492232265 × LINC00092 expression) + (0.724211508 × AC145098.1 expression) + (-0.922536934 × SNAI3-AS1 expression) + (-0.922536934 × AC048382.5 expression) + (-0.846208391 × AC127070.2 expression) + (-0.924348861 × AL731571.1 expression) + (-0.807182397 × AL159169.2 expression). After obtaining a risk score for each patient, the patients were divided into two groups based on the median risk score: a high-risk group and a low-risk group (Figure 2A). We found that more and more patients died as the risk score increased (Figure 2B). Figure 2C showed eight prognostic m7G-related lncRNAs involved in two groups by heat map. The ROC curve area showed the excellent predictive capability of the model based on eight survival-related lncRNAs. In the TCGA data, the AUC values were 0.905, 0.928, and 0.89 at 1, 3, and 5 years, respectively (Figure 2D). According to KM analysis, patients with high RS had worse survival rates than those with low RS (Figure 2E).

Using the same cut-off from the TCGA data for the CGGA validation data, it was possible to distinguish the high-risk group from the low-risk group. However, the number of patients in the low-risk group was significantly lower (Figure 2F). CGGA patients showed that high-risk patients are positively associated with poor prognosis (Figure 2G). The expression of prognostic m^7G-related lncRNAs in CGGA resembled that in TCGA samples (Figure 2H). In the CGGA sample, the AUC values were0.705, 0.78, and 0.80 at 1, 3, and 5 years, respectively (Figure 2I). KM analysis performed on CGGA data showed the same results as TCGA data ($p < 0.001$, Figure 2J). The validation results in the two validation datasets of TCGA also demonstrate the excellent predictive power of the model (Supplementary Material S5).

Validation of PCA analysis and the expression of prognostic LncRNAs

The distribution of patients based on whole genes, m^7G-related genes, m^7G-related lncRNAs, and prognostic m^7G-related lncRNAs was visualized using PCA plots. The results showed that m^7G survival-associated lncRNA showed the best results (Figures 3A–D). High- and low-risk patients can be distributed in different quadrants according to the RS of prognostic m^7G-related lncRNAs.

We evaluated the expression levels of m^7G-related lncRNAs in the TCGA dataset. We found that all genes differed significantly in different grades (Figure 3E), and all but one of the genes had similar trends across stages (Figure 3F). In TCGA and CGGA datasets, the same trend of gene expression was shown with increasing tumor grade.

Validation of the correlations between clinical variables and risk score

Using TCGA data, we analyzed the correlations between these clinical variables and the eight lncRNAs risk scores. The risk scores were correlated with age, survival status, and tumor stage; *AC048382.5* was associated with age, survival status, and stage; *AC127070.2* was correlated with survival status, sex, and stage; *AC145098.1* was correlated with survival status and stage; *AL159169.2* was correlated with age, survival status, and stage; *AL731571.1* was correlated with age, survival status, gender, and staging; and *LINC00092* was associated with age and survival status. (Figure 4). The above results showed that our screened m^7G-related lncRNAs had the excellent predictive ability.

Development and validation of nomogram

In TCGA and CGGA data, we analyzed the independent prognostic factors of glioma patients by Cox regression. Univariate and multivariate Cox regression analyses showed that risk score was an independent predictor (HR = 1.253, 95% CI: 1.192–1.317, $p < 0.001$; HR = 1.127, 95% CI:1.096–1.160, $p < 0.001$) of OS in glioma patients (Figure 5A, B,D,E). We constructed a column line plot containing clinicopathological variables and risk scores to facilitate clinical work (Figure 5C,F). The calibration curves showed good agreement between actual OS and predicted survival rates (Figures 5J–L).

Functional annotation of m^7G-related lncRNAs

We used GSEA to investigate further the differences between the two subgroups for eight m^7G-related lncRNAs. In KEGG analysis, the main added functions were systemic lupus-erythematosus, n-glycan-synthesis, and glutathione-metabolism. Decreased functions were wnt-signalling-pathway, taste-transduction, and terpenoid-backbone-biosynthesis (Figure 6A). Most of these pathways are mainly responsible

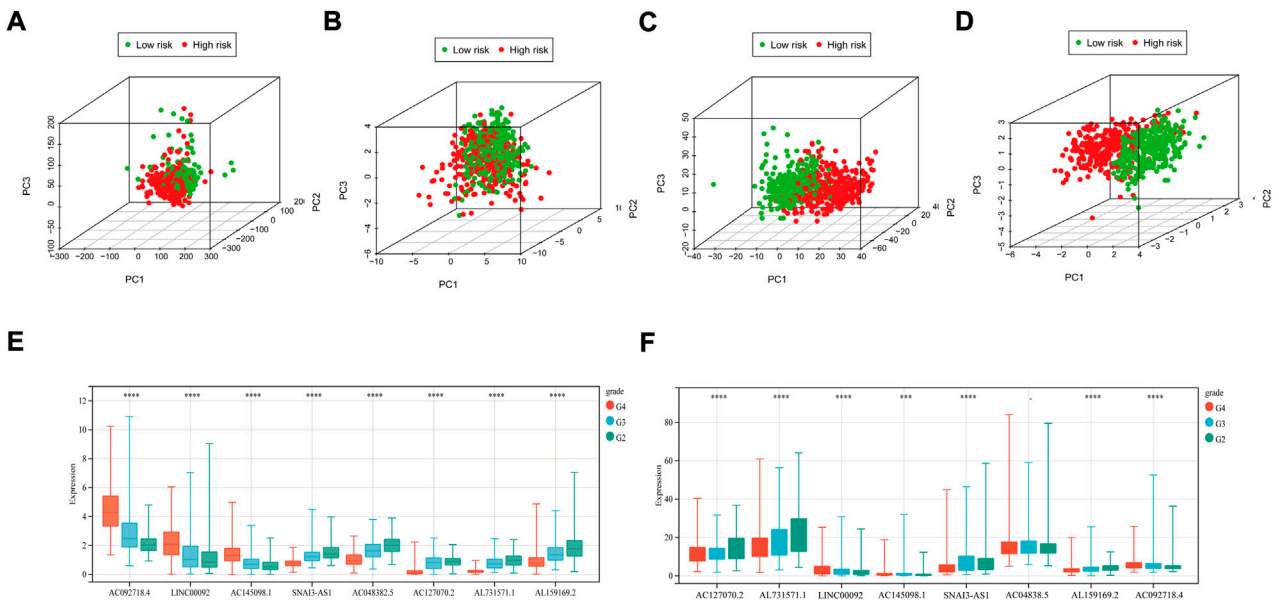


FIGURE 3
PCA maps of the TCGA glioma dataset show the distribution of patients based on the **(A)** whole genome; **(B)** m^7G-related gene sets; **(C)** m^7G-related lncRNAs; and **(D)** prognostic m^7G-related lncRNAs. Red: high-risk; Green dots: low-risk. Expression profile of 8 prognostic m^7G-related lncRNAs with different glioma grades. **(E)**m^7G-related lncRNAs expression with different glioma grades in TCGA datasets. **(F)** prognostic m^7G-related lncRNAs expression with different glioma grades in CGGA datasets. (G2: WHO II, G3: WHO III, G4: WHO IV). $*p < 0.05$, $**p < 0.01$ and $***p < 0.001$.

for immune-related diseases and metabolic pathways. So, this suggests that poor prognosis in high-risk patients is likely to be closely related to tumor immune-related pathways.

Investigation of immune-related pathways

We quantified the enrichment scores of ssGSEA by measuring the immune cell subpopulations and related pathways to investigate further the correlation between risk scores and immune cells and functions. In the high-risk group, we found a significant rise in most cells (B cells, $CD8^+$ T cells, DCs, Tregs, *etc.*) (Figure 6B). T-cell-co-inhibition, APC-co-stimulation, CCR, T-cell-co-stimulation, and type I IFN response were higher in the high-risk group than in the low-risk group (Figure 6C). The above results suggest that the high-risk group's immune function was more active. We also compared the analysis of differences in immune checkpoint expression between the two groups because of the importance of checkpoint-based immunotherapy (Figure 6D).

Correlation of predictive features between drug sensitivities

We also analyzed the correlation between predictive characteristics and tumor immune-related drugs. The results found lower IC50 of Cisplatin, Etoposide, and Rapamycin in the high-risk group and higher IC50 of Lenalidomide and PAC-1 in the high-risk group. (Figures 6E–I), which helps to explore individualized treatment regimens suitable for high-risk patients.

Discussion

Glioma is a common brain tumor, accounting for 78% of primary malignant brain tumors in the brain, and its overall prognosis has been poor. Therefore, exploring the early diagnosis of glioma and accurately predicting the prognostic markers is of crucial clinical significance (Linzey et al., 2019). Many studies have shown the critical role of m^7G in cancer development, mainly focusing on the regulation of tumor cell genesis and progression, but few investigations on cancer prognosis (Orellana et al., 2021; Rong et al., 2021; Xia et al., 2021). Several studies have recently emerged by constructing mRNA and lncRNA predictive signatures associated with glioma autophagy, pyrogenesis, m^6A, and ferrogenesis can be used to predict the prognosis of glioma patients (Maimaiti et al., 2022) (Zhou et al., 2021a) (Guan et al., 2021) (Shi et al., 2022). However, the study of prognostic m7G-related lncRNAs in glioma has not been reported. Therefore, we purpose to investigate the prognostic role of m^7G-related lncRNA in glioma and provide a new approach for the future clinical treatment of glioma.

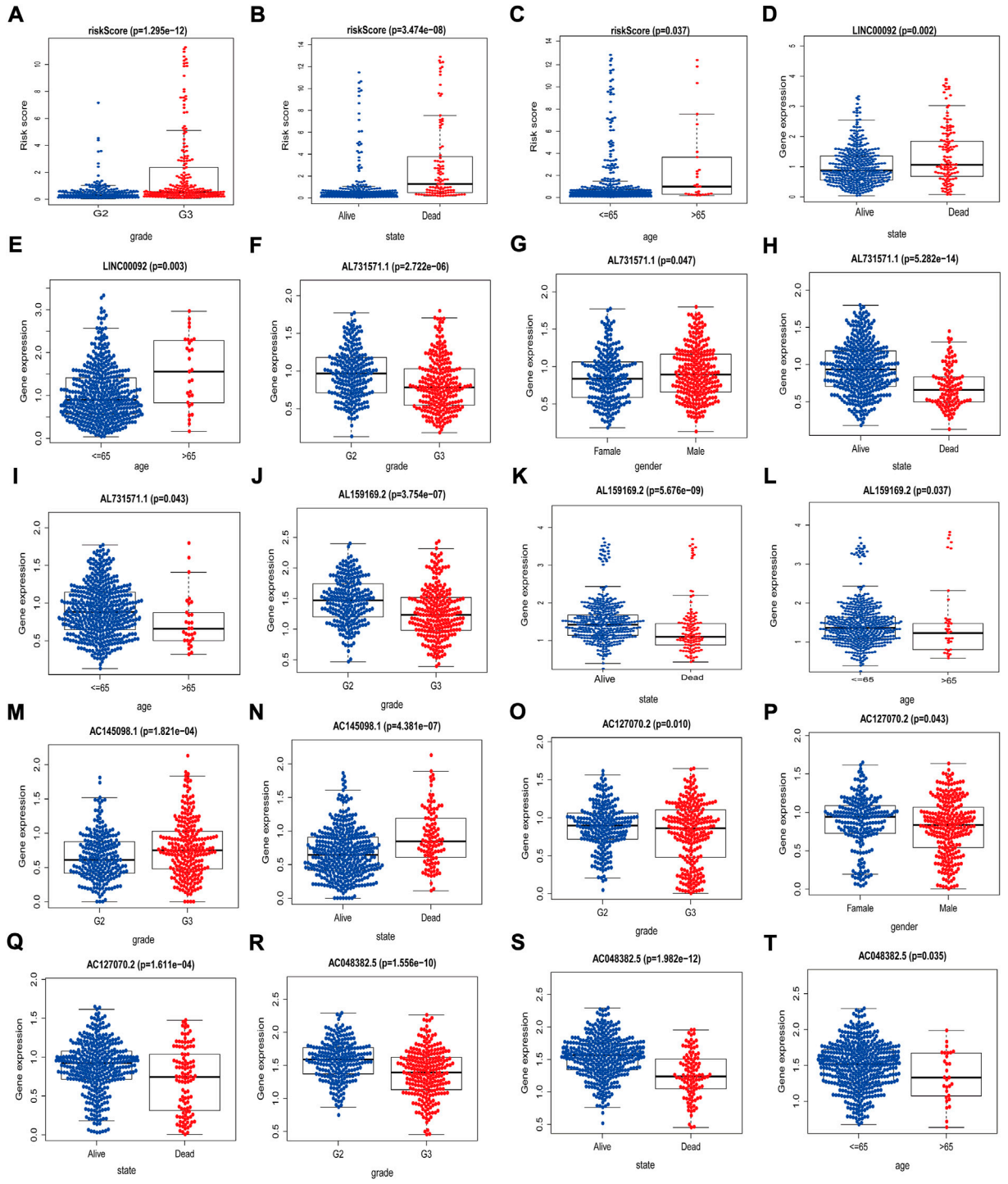


FIGURE 4
Associations between risk scores/related lncRNAs and clinical features. **(A–C)** Association between risk score and gender, state, and age. **(D–E)** Association between *LINC00092* expression level and state, age. **(F–I)** Association between *AL731571.1* expression level and grade, gender, state and age. **(J–L)** Association between *AL159169.2* expression level and grade, state and age. **(M–N)** Association between *AC145098.1* expression level and grade and state. **(O–Q)** Association between *AL127070.2* expression level and grade, gender and state. **(R–T)** Association between *AC048382.5* expression level and grade, state and age.

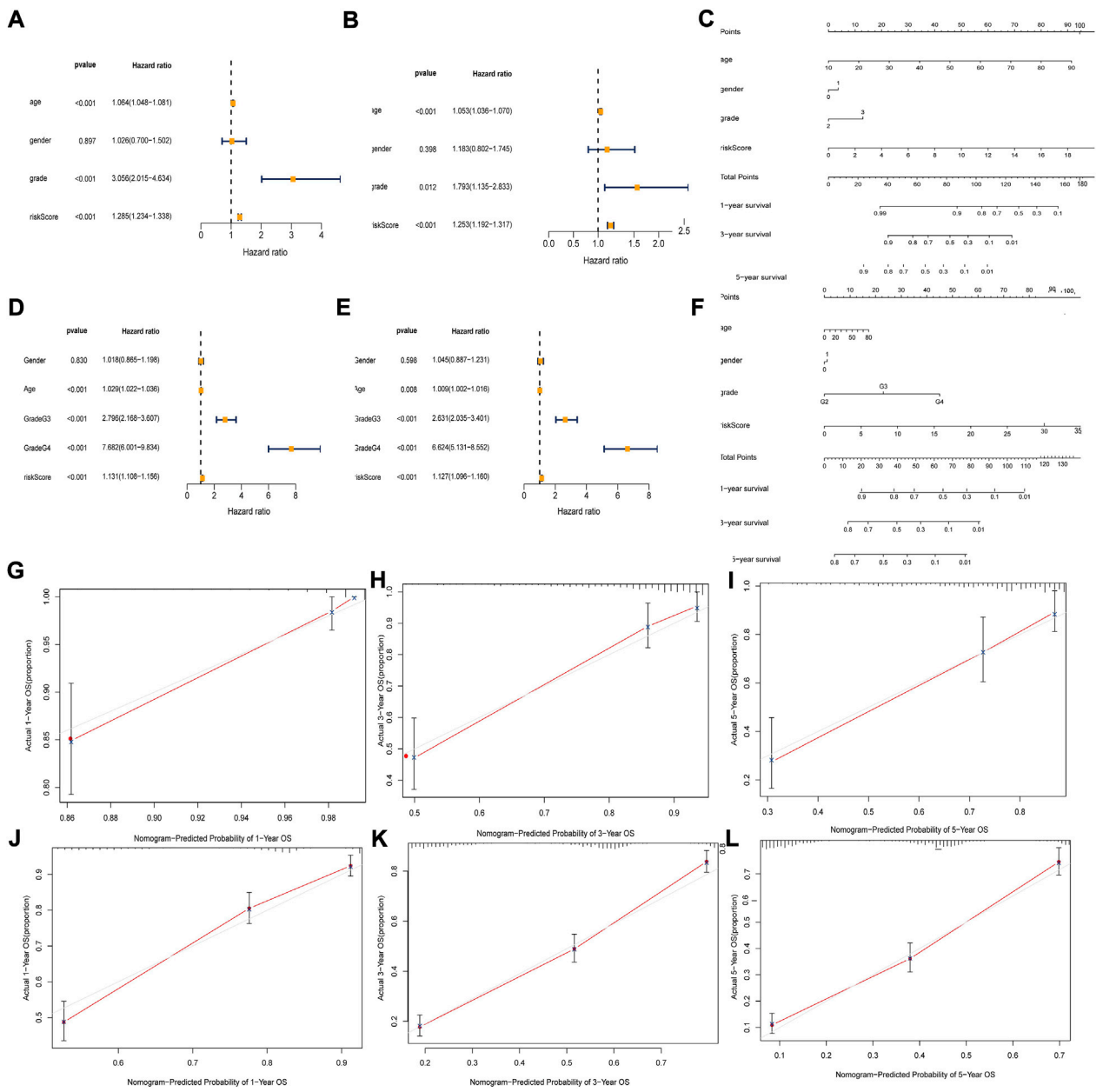


FIGURE 5
Independent prognosis analysis of risk score. **(A and D)** Univariate COX Forest plot of the risk score in TCGA and CGGA. **(B and E)** Multivariate COX Forest plot of the risk score in TCGA and CGGA. **(C and F)** Nomogram based on prognostic features in TCGA and CGGA. **(G–I)** Calibration plots of the nomogram for predicting the probability of OS at 1, 3, and 5 years in the TCGA. **(J–L)** Calibration plots of the nomogram for predicting the probability of OS at 1, 3, and 5 years in the CGGA.

This study first obtained 28 DEGs associated with N^7-methylguanosine. Then, KEGG analysis showed that DEGs were mainly enriched in RNA degradation, nucleocytoplasmic transport, mRNA surveillance pathway, HIF-1, mTOR, and HIF1-PI3K-Akt signaling pathway. GO analysis showed that DEGs were primarily enriched in the translational initiation activity, regulation of translation, RNA 7–methylguanosine, *etc.* Existing studies have modified mRNA by adding an m^7G 5′ cap to protect mRNA from premature degradation (Kasprzyk and Jemielity, 2021). EGFR plays a crucial role in the METTL1-m7G axis in bladder cancer (Ying et al., 2021). Upregulated WDR4 expression increases m^7G methylation levels in hepatocellular carcinoma (Xia et al., 2021). Hickey et al. reported that m^7G-MP, the cap analog, is a potent and specific inhibitor of eukaryotic translation (Hickey et al.,

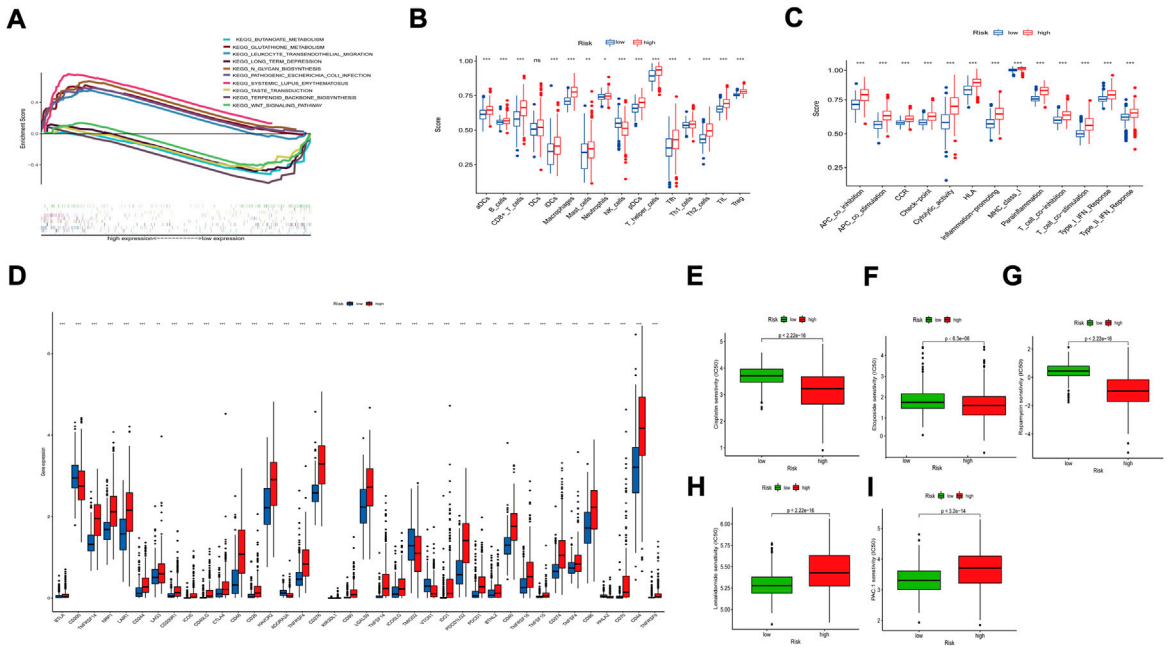


FIGURE 6
Functional enrichment analysis of 8 prognostic m^{7}G-related lncRNAs. **(A)** KEGG analysis of 8 prognostic m^{7}G-related lncRNAs. **(B)** The infiltration levels of 16 immune cells. **(C)** The correlation between the predictive signature and 13 immune-related functions. **(D)** Expression of immune checkpoints. aDCs, activated dendritic cells; iDCs, immature dendritic cells; NK, natural killer; pDCs, plasmacytoid dendritic cells; Tfh, T follicular helper; Th1, T helper type 1; Th2, T helper type 2; TIL, tumor-infiltrating lymphocyte; Treg, T regulatory cell; APC, antigen-presenting cell; CCR, chemokine receptor; HLA, human leukocyte antigen; MHC, major histocompatibility complex; IFN, interferon. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ns, non-significant. Comparison of treatment drugs sensitivity between high- and low-risk groups. **(E–I)** IC50 of Cisplatin, Etoposide, Rapamycin, Lenalidomide, PAC.1 in high and low-risk groups. IC50, half-maximal inhibitory concentration.

1976). The above results suggest that m^{7}G-related genes maybe participate in cancer development through various pathways such as transcription and translation. However, further studies are needed to explore the function of m^{7}G-related genes in glioma.

In addition, there are pieces of evidence that lncRNAs play an essential part in cancer (Ho et al., 2022; Liang et al., 2022; Zhang et al., 2022). *SNAI3-AS1*, an m^{7}G prognosis-associated lncRNA, is an important tumor modifier in hepatocellular carcinoma tumor progression (Li et al., 2020). Recently, it has been reported that autophagy-related lncRNA features can accurately predict the prognosis of glioma patients (Maimaiti et al., 2022). Ferroptosis-associated lncRNAs can also predict the prognosis of glioma patients (Shi et al., 2022). Therefore, it is important to identify the predictive value of m^{7}G-related lncRNAs in glioma patients and could provide potential directions for future experimental studies of m^{7}G and clinical studies of glioma. In this study, we identified 8 prognostic m^{7}G-related lncRNAs (*AC048382.5*, *AC127070.2*, *AL159169.2*, *AL731571.1*, *SNAI3-AS1*, *AC092718.4*, *AC145098.1*, *LINC00092*) for establishing prognostic model. We also found mRNAs (*EIF4A1*, *EIF4E3*, *EIF4E1B*, *CYFIP1*, *DCPS*, *WDR4*, *NUDT10*, *IFIT5*, *EIF3D*, *EIF4E*) were significantly co-expressed with these lncRNAs. Among them, eIF4E binds the 7-methyl-GTP portion of the 5′ cap structure of cytoplasmic mRNA and plays a part in translation initiation and regulation (Merrick and Pavitt, 2018). Additional studies have found that DCPS acts on m^{7}G through mRNA decay (Ng et al., 2015). *WDR4* undergoes a malignant transformation of cells through overexpression of m^{7}G (Orellana et al., 2021). *EIF4* acts as a cap-binding protein to enhance m^{7}G cap stabilization of transcripts and plays an important role in malignancy through upregulation (Culjkovic-Kraljacic et al., 2020). In conclusion, the above reports provide evidence for our related studies on N^7-methylguanosine. In analyzing two databases with the same median, we found that the number of deaths increased as the risk score increased. The 5-year AUC values (AUC = 0.89, AUC = 0.80) in both TCGA and CGGA data demonstrated the success of the model construction in predicting the prognosis of glioma patients. Furthermore, eight lncRNAs expression in different grades of glioma, the correlation between risk scores and clinical characteristics also increases their predictive power.

Then, GSEA shows that the high-risk group mainly enriched systemic lupus-erythematosus, n-glycan-biosynthesis, glutathione-synthesis, and leukocyte-transendothelial migration. N^7-methyladenosine, a common methylation modification of RNA, plays an essential role in autoimmune diseases like RA and SLE (Agris et al., 1992; Zhou et al., 2021b). N-glycan plays a significant part in breast and oral cancers (Hirano and Furukawa, 2022; Wu et al.,

2022). Glutathione affects tumor progression by altering oxidative stress sensitivity in astrocytic tumors (Moreira Franco et al., 2021). Increased expression of lymphocyte-specific protein 1 (LSP1) will cause leukocyte migration and inhibition of the immune microenvironment in GBM (Cao et al., 2020). The above results suggest that the occurrence and development of gliomas are also most likely to be closely related to immune-related pathways. The ssGSEA results showed a significant rise in most cells (macrophages, $CD8^+$ T cells, mast cells, Tregs, *etc.*) in the high-risk group. Some of the above findings have been confirmed by studies. For example, $CD8^+$T-cell infiltration is associated with poor prognosis in patients with BC (Hou et al., 2020; Liu et al., 2020). High infiltration of tumour-associated macrophages was associated with low-grade glioma and thyroid cancer (Ryder et al., 2008; Li et al., 2022). The number of mast cells was positively linked to poor prognosis in patients with prostate cancer (Zhang et al., 2020).

The degree of MC infiltration in mice and human gliomas is proportional to the malignancy of the tumor (Polajeva et al., 2011; Polajeva et al., 2014). The ratio of high neutrophils to lymphocytes predicts a poorer OS in BC patients (Tan, 2017). Pathological grading of gliomas is positively correlated with infiltrating neutrophils (Khan et al., 2020). Increased infiltration of Tregs indicates a poor prognosis in patients with hepatocellular carcinoma (Tu et al., 2016). lncRNA *HOXA-AS2* promotes Treg proliferation and immune tolerance through the miR-302A/*KDM2A* axis to promote glioma progression and poor prognosis (Zhong et al., 2022). Increased Treg and MDSC in mouse gliomas can lead to a decrease in overall survival (Zhai et al., 2021). We found higher HLA and type I IFN response scores in the high-risk group, except for increased tumor immune cell infiltration. Thus, decreased antitumor immunity in high-risk groups may be responsible for poor prognosis. We found significant differences in immune checkpoint expression between the high-risk and low-risk groups. We also studied the sensitivity of immune-related drugs among patients and found that high-risk patients may be sensitive to Cisplatin, Etoposide, and Rapamycin and resistant to Lenalidomide, PAC-1. This implies that high-risk groups may benefit from treatment with multiple immune-related drugs. We hope the above study provides a basis for precise, individualized treatment of glioma patients.

However, our study has some limitations. In the first place, we only used CGGA and CGGA database data for verification and still required external data to test the applicability of predicted signatures. Next, the mechanism of action of m^7G-related lncRNAs in glioma needs to be further validated experimentally.

Conclusion

We successfully built a formula for m^7G-related lncRNAs with powerful predictive functions and screened lncRNAs with prognostic values. These studies add some instructional value to glioma etiopathogenesis and clinical treatment analysis. And these m^7G-related lncRNAs may become new biomarkers and are expected to provide new ideas for glioma therapeutic approaches.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

SW had the initial idea for this study, performed the statistical analysis, and wrote the manuscript; AB and WC and XW revised the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors would like to thank the TCGA, GTEx, CGGA databases, and GSEA websites for data availability.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.961278/full#supplementary-material

References

Agris, P. F., Kovacs, S. A., Boak, A. M., and Chen, J. (1992). Spatial localization of distinct rheumatic disease-associated epitopes and the RNA "cap" of the U1 snRNP particle. *Immunol. Invest.* 21 (2), 111–122. doi:10.3109/08820139209066183

Balzeau, J., Menezes, M. R., Cao, S., and Hagan, J. P. (2017). The LIN28/let-7 pathway in cancer. *Front. Genet.* 8, 31–16. doi:10.3389/fgene.2017.00031

Barbieri, I., Tzelepis, K., Pandolfini, L., Shi, J., Millan-Zambrano, G., Robson, S. C., et al. (2017). Promoter-bound METTL3 maintains myeloid leukaemia by m6A-dependent translation control. *Nature* 552 (7683), 126–131. doi:10.1038/nature24678

Cao, J-Y., Guo, Q., Guan, G-F., Zhu, C., Zou, C-Y., Zhang, L-Y., et al. (2020). Elevated lymphocyte specific protein 1 expression is involved in the regulation of leukocyte migration and immunosuppressive microenvironment in glioblastoma. *Aging* 12 (2), 1656–1684. doi:10.18632/aging.102706

Chen, J., Li, K., Chen, J., Wang, X., Ling, R., Cheng, M., et al. (2022). Aberrant translation regulated by METTL1/WDR4-mediated tRNA N7-methylguanosine modification drives head and neck squamous cell carcinoma progression. *Cancer Commun.* 42, 223–244. doi:10.1002/cac2.12273

Chen, M., Nie, Z., Li, Y., Gao, Y., Wen, X., Cao, H., et al. (2021). A new ferroptosis-related lncRNA signature predicts the prognosis of bladder cancer patients. *Front. Cell Dev. Biol.* 9, 699804. doi:10.3389/fcell.2021.699804

Culjkovic-Kraljacic, B., Skrabanek, L., Revuelta, M. V., Gasiorek, J., Cowling, V. H., Cerchietti, L., et al. (2020). The eukaryotic translation initiation factor eIF4E elevates steady-state m(7)G capping of coding and non-coding transcripts. *Proc. Natl. Acad. Sci. U. S. A.* 117 (43), 26773–26783. doi:10.1073/pnas.2002360117

Frances, S. M., Velikova, G., Klein, M., Short, S. C., Murray, L., Wright, J. M., et al. (2022). Long-term impact of adult WHO grade II or III gliomas on health-related quality of life: A systematic review. *Neurooncol. Pract.* 9 (1), 3–17. doi:10.1093/nop/npab062

Furuichi, Y. (2015). Discovery of m(7)G-cap in eukaryotic mRNAs. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 91 (8), 394–409. doi:10.2183/pjab.91.394

Guan, S., He, Y., Su, Y., and Zhou, L. (2021). A risk signature consisting of eight m^6A methylation regulators predicts the prognosis of glioma. *Cell. Mol. Neurobiol.* 42, 2733–2743. doi:10.1007/s10571-021-01135-x

Guo, A., Lin, R., Zheng, S., Fang, G., Lin, Z., Zhuang, Z., et al. (2022). Overexpression of lncRNA IRAIN restrains the progression and Temozolomide resistance of glioma via repressing IGF-1R-PI3K-NF-κB signalling pathway. *Histol. Histopathol.* 37 (6), 543–554. doi:10.14670/HH-18-425

Haddad, A. F., Young, J. S., Oh, J. Y., Okada, H., Aghi, M. K., Pereira, K. M., et al. (2022). Clinical characteristics and outcomes of null-cell versus silent gonadotroph adenomas in a series of 1166 pituitary adenomas from a single institution. *Neurosurg. Focus* 52 (2), E13. doi:10.3171/2020.3.FOCUS20114

Hickey, E. D., Weber, L. A., and Baglioni, C. (1976). Inhibition of initiation of protein synthesis by 7-methylguanosine-5'-monophosphate. *Proc. Natl. Acad. Sci. U. S. A.* 73 (1), 19–23. doi:10.1073/pnas.73.1.19

Hirano, K., and Furukawa, K. (2022). Biosynthesis and biological significances of LacdiNAc group on N- and O-glycans in human cancer cells. *Biomolecules* 12 (2), 195. doi:10.3390/biom12020195

Ho, K-H., Shih, C-M., Liu, A-J., and Chen, K-C. (2022). Hypoxia-inducible lncRNA MIR210HG interacting with OCT1 is involved in glioblastoma multiforme malignancy. *Cancer Sci.* 113 (2), 540–552. doi:10.1111/cas.15240

Hou, W., Xue, M., Shi, J., Yang, M., Zhong, W., Fan, X., et al. (2020). PD-1 topographically defines distinct T cell subpopulations in urothelial cell carcinoma of the bladder and predicts patient survival. *Urol. Oncol.* 38 (8), e1–e685. doi:10.1016/j.urolonc.2020.04.009

Jaffrey, S. R. (2014). An expanding universe of mRNA modifications. *Nat. Struct. Mol. Biol.* 21 (11), 945–946. doi:10.1038/nsmb.2911

Jiang, S., Xie, Y., He, Z., Zhang, Y., Zhao, Y., Chen, L., et al. (2018). m6ASNP: a tool for annotating genetic variants by m(6)A function. *Gigascience* 7 (5). doi:10.1093/gigascience/giy035

Kasprzyk, R., and Jemielity, J. (2021). Enzymatic assays to explore viral mRNA capping machinery. *Chembiochem.* 22 (23), 3236–3253. doi:10.1002/cbic.202100291

Khan, S., Mittal, S., McGee, K., Alfaro-Munoz, K. D., Majd, N., Balasubramaniyan, V., et al. (2020). Role of neutrophils and myeloid-derived suppressor cells in glioma progression and treatment resistance. *Int. J. Mol. Sci.* 21 (6), E1954. doi:10.3390/ijms21061954

Lewis, J. D., and Izaurralde, E. (1997). The role of the cap structure in RNA processing and nuclear export. *Eur. J. Biochem.* 247 (2), 461–469. doi:10.1111/j.1432-1033.1997.00461.x

Li, G., Li, L., Li, Y., Qian, Z., Wu, F., He, Y., et al. (2022). An MRI radiomics approach to predict survival and tumour-infiltrating macrophages in gliomas. *Brain.* 145, 1151–1161. doi:10.1093/brain/awab340

Li, Y., Guo, D., Lu, G., Mohiuddin Chowdhury, A. T. M., Zhang, D., Ren, M., et al. (2020). LncRNA SNAI3-AS1 promotes PEG10-mediated proliferation and metastasis via decoying of miR-27a-3p and miR-34a-5p in hepatocellular carcinoma. *Cell Death Dis.* 11 (8), 685. doi:10.1038/s41419-020-02840-z

Liang, J., Liu, C., Xu, D., Xie, K., and Li, A. (2022). LncRNA NEAT1 facilitates glioma progression via stabilizing PGK1. *J. Transl. Med.* 20 (1), 80. doi:10.1186/s12967-022-03273-2

Lin, S., Liu, Q., Lelyveld, V. S., Choe, J., Szostak, J. W., and Gregory, R. I. (2018). Mettl1/Wdr4-Mediated m(7)G tRNA methylome is required for normal mRNA translation and embryonic stem cell self-renewal and differentiation. *Mol. Cell* 71 (2), 244–255. doi:10.1016/j.molcel.2018.06.001

Linzey, J. R., Sivakumar, W., Johnson, J. N., Ivan, M. E., Haider, A. S., Philips, C. A., et al. (2019). Young neurosurgeons committee of the American association of neurological surgeons: Training ground for future leaders in organized neurosurgery in the United States of America. *World Neurosurg.* 123, 59–63. doi:10.1016/j.wneu.2018.11.206

Liu, Z., Zhou, Q., Wang, Z., Zhang, H., Zeng, H., Huang, Q., et al. (2020). Intratumoral TIGIT(+)CD8(+)T-cell infiltration determines poor prognosis and immune evasion in patients with muscle-invasive bladder cancer. *J. Immunother. Cancer* 8 (2), e000978. doi:10.1136/jitc-2020-000978

Maimaiti, A., Tuerhong, M., Wang, Y., Aisha, M., Jiang, L., Wang, X., et al. (2022). An innovative prognostic model based on autophagy-related long non-coding RNA signature for low-grade glioma. *Mol. Cell. Biochem.* 477, 1417–1438. doi:10.1007/s11010-022-04368-6

Marchand, V., Ayadi, L., Ernst, F. G. M., Herder, J., Bourguignon-Igel, V., Galvani, A., et al. (2018). AlkAniline-Seq: Profiling of m(7)G and m(3)C RNA modifications at single nucleotide resolution. *Angew. Chem. Int. Ed. Engl.* 57 (51), 16785–16790. doi:10.1002/anie.201810946

Merrick, W. C., and Pavitt, G. D. (2018). Protein synthesis initiation in eukaryotic cells. *Cold Spring Harb. Perspect. Biol.* 10 (12), a033092. doi:10.1101/cshperspect.a033092

Moreira Franco, Y. E., Alves, M. J., Uno, M., Moretti, I. F., Trombetta-Lima, M., de Siqueira Santos, S., et al. (2021). Glutaminolysis dynamics during astrocytoma progression correlates with tumor aggressiveness. *Cancer Metab.* 9 (1), 18. doi:10.1186/s40170-021-00255-8

Mousavi, S. M., Derakhshan, M., Baharloii, F., Dashti, F., Mirazimi, S. M. A., Mahjoubin-Tehran, M., et al. (2022). Non-coding RNAs, and glioblastoma: Insight into their roles in metastasis. *Mol. Ther. Oncolytics* 24, 262–287. doi:10.1016/j.omto.2021.12.015

Ng, C. K., Shboul, M., Taverniti, V., Bonnard, C., Lee, H., Eskin, A., et al. (2015). Loss of the scavenger mRNA decapping enzyme DCPS causes syndromic intellectual disability with neuromuscular defects. *Hum. Mol. Genet.* 24 (11), 3163–3171. doi:10.1093/hmg/ddv067

Orellana, E. A., Liu, Q., Yankova, E., Pirouz, M., De Braekeleer, E., Zhang, W., et al. (2021). METTL1-mediated m(7)G modification of Arg-TCT tRNA drives oncogenic transformation. *Mol. Cell* 81 (16), 3323–3338.e14. doi:10.1016/j.molcel.2021.06.031

Ostrom, Q. T., Cioffi, G., Gittleman, H., Patil, N., Waite, K., Kruchko, C., et al. (2019). CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2012-2016. *Neuro. Oncol.* 21, V1–V100. doi:10.1093/neuonc/noz150

Pandolfini, L., Barbieri, I., Bannister, A. J., Hendrick, A., Andrews, B., Webster, N., et al. (2019). METTL1 promotes let-7 MicroRNA processing via m7G methylation. *Mol. Cell* 74 (6), 1278–1290. doi:10.1016/j.molcel.2019.03.040

Pei, Y., and Shuman, S. (2002). Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J. Biol. Chem.* 277 (22), 19639–19648. doi:10.1074/jbc.M200015200

Pereira, P. L., Magnol, L., Sahun, I., Brault, V., Duchon, A., Prandini, P., et al. (2009). A new mouse model for the trisomy of the Abcg1-U2af1 region reveals the complexity of the combinatorial genetic code of down syndrome. *Hum. Mol. Genet.* 18 (24), 4756–4769. doi:10.1093/hmg/ddp438

Polajeva, J., Bergstrom, T., Edqvist, P-H., Lundequist, A., Sjosten, A., Nilsson, G., et al. (2014). Glioma-derived macrophage migration inhibitory factor (MIF) promotes mast cell recruitment in a STAT5-dependent manner. *Mol. Oncol.* 8 (1), 50–58. doi:10.1016/j.molonc.2013.09.002

Polajeva, J., Sjosten, A. M., Lager, N., Kastemar, M., Waern, I., Alafuzoff, I., et al. (2011). Mast cell accumulation in glioblastoma with a potential role for stem cell

factor and chemokine CXCL12. *Plos One* 6 (9), e25222. doi:10.1371/journal.pone.0025222

Rong, D., Sun, G., Wu, F., Cheng, Y., Sun, G., Jiang, W., et al. (2021). Epigenetics: Roles and therapeutic implications of non-coding RNA modifications in human cancers. *Mol. Ther. Nucleic Acids* 25, 67–82. doi:10.1016/j.omtn.2021.04.021

Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160 (1-2), 48–61. doi:10.1016/j.cell.2014.12.033

Ryder, M., Ghossein, R. A., Ricarte-Filho, J. C. M., Knauf, J. A., and Fagin, J. A. (2008). Increased density of tumor-associated macrophages is associated with decreased survival in advanced thyroid cancer. *Endocr. Relat. Cancer* 15 (4), 1069–1074. doi:10.1677/ERC-08-0036

Sauna, Z. E., and Kimchi-Sarfaty, C. (2011). Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.* 12 (10), 683–691. doi:10.1038/nrg3051

Shi, J., Lai, D., Zuo, X., Liu, D., Chen, B., Zheng, Y., et al. (2022). Identification of ferroptosis-related biomarkers for prognosis and immunotherapy in patients with glioma. *Front. Cell Dev. Biol.* 10, 817643. doi:10.3389/fcell.2022.817643

Song, B., Tang, Y., Chen, K., Wei, Z., Rong, R., Lu, Z., et al. (2020). m7GHub: deciphering the location, regulation, and pathogenesis of internal mRNA N7-methylguanosine (m(7)G) sites in human. *Bioinformatics* 36 (11), 3528–3536. doi:10.1093/bioinformatics/btaa178

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Tan, Y. G. (2017). High neutrophil-to-lymphocyte ratio predicts worse overall survival in patients with advanced/metastatic urothelial bladder cancer. *Bju Int.* 119, 17–18.

Teng, P-C., Liang, Y., Yarmishyn, A. A., Hsiao, Y-J., Lin, T-Y., Lin, T-W., et al. (2021). RNA modifications and epigenetics in modulation of lung cancer and pulmonary diseases. *Int. J. Mol. Sci.* 22 (19), 10592. doi:10.3390/ijms221910592

Tian, Q-H., Zhang, M-F., Zeng, J-S., Luo, R-G., Wen, Y., Chen, J., et al. (2019). METTL1 overexpression is correlated with poor prognosis and promotes hepatocellular carcinoma via PTEN. *J. Mol. Med.* 97 (11), 1535–1545. doi:10.1007/s00109-019-01830-9

Tomikawa, C. (2018). 7-Methylguanosine modifications in transfer RNA (tRNA). *Int. J. Mol. Sci.* 19 (12), E4080. doi:10.3390/ijms19124080

Tu, J-F., Ding, Y-H., Ying, X-H., Wu, F-Z., Zhou, X-M., Zhang, D-K., et al. (2016). Regulatory T cells, especially ICOS+ FOXP3(+) regulatory T cells, are increased in the hepatocellular carcinoma microenvironment and predict reduced survival. *Sci. Rep.* 6, 35056. doi:10.1038/srep35056

Wang, Y., Du, B., Ma, H., and Zhang, Z. (2022). Down-regulation of lncRNA TTTY15 targeting miR-4500 to inhibit the biological characteristics of A172 glioma cells. *Zhonghua Yi Xue Yi Chuan Xue za Zhi* 39 (2), 171–175. doi:10.3760/cma.j.cn511374-20201116-00801

Wei, P., Jiang, J., Xiao, M., Zeng, M., Liu, X., Zhao, B., et al. (2022). The transcript ENST00000444125 of lncRNA LINC01503 promotes cancer stem cell properties of glioblastoma cells via reducing FBXW1 mediated GLI2 degradation. *Exp. Cell Res.* 412 (1), 113009. doi:10.1016/j.yexcr.2022.113009

Wu, Y., Liu, Y., Shang, Z., Liu, X., Xu, Y., and Liu, W. (2022). N-Glycomic profiling reveals dysregulated glycans related to oral cancer using MALDI-MS. *Anal. Bioanal. Chem.* 414 (5), 1881–1890. doi:10.1007/s00216-021-03822-6

Xia, P., Zhang, H., Xu, K., Jiang, X., Gao, M., Wang, G., et al. (2021). MYC-targeted WDR4 promotes proliferation, metastasis, and sorafenib resistance by inducing CCNB1 translation in hepatocellular carcinoma. *Cell Death Dis.* 12 (7), 691. doi:10.1038/s41419-021-03973-5

Xu, L., Wu, Q., Yan, H., Shu, C., Fan, W., Tong, X., et al. (2022). Long non-coding RNA KB-1460A1.5 inhibits glioma tumorigenesis via miR-130a-3p/TSC1/mTOR/YY1 feedback loop. *Cancer Lett.* 525, 33–45. doi:10.1016/j.canlet.2021.10.033

Yang, D., Sun, L., Li, Z., and Gao, P. (2016). "Noncoding RNAs in regulation of cancer metabolic reprogramming," in *Long and short non-coding RNAs in cancer biology*. *Advances in experimental medicine and biology*. Editor E. Song, 191–215.

Yao, J., Chen, X., Liu, X., Li, R., Zhou, X., and Qu, Y. (2021). Characterization of a ferroptosis and iron-metabolism related lncRNA signature in lung adenocarcinoma. *Cancer Cell Int.* 21 (1), 340. doi:10.1186/s12935-021-02027-2

Ying, X., Liu, B., Yuan, Z., Huang, Y., Chen, C., Jiang, X., et al. (2021). METTL1-m(7) G-EGFR/EFEMP1 axis promotes the bladder cancer development. *Clin. Transl. Med.* 11 (12), e675. doi:10.1002/ctm2.675

Zhai, L., Bell, A., Ladomersky, E., Lauing, K., Bollu, L., Nuygen, B., et al. (2021). Tumor cell ido enhances immune suppression and decreases survival independent of tryptophan metabolism in glioblastoma. *Neuro-Oncology* 23, 94. doi:10.1093/neuonc/noab196.371

Zhang, E., Dai, F., Mao, Y., He, W., Liu, F., Ma, W., et al. (2020). Differences of the immune cell landscape between normal and tumor tissue in human prostate. *Clin. Transl. Oncol.* 22 (3), 344–350. doi:10.1007/s12094-019-02128-5

Zhang, Y., Zhang, Y., Wang, S., Cao, B., Hu, D., Jia, J., et al. (2022). LINC00467 facilitates the proliferation, migration and invasion of glioma via promoting the expression of inositol hexakisphosphate kinase 2 by binding to miR-339-3p. *Bioengineered* 13 (2), 3370–3382. doi:10.1080/21655979.2021.2018098

Zhong, C., Tao, B., Li, X., Xiang, W., Peng, L., Peng, T., et al. (2022). HOXA-AS2 contributes to regulatory T cell proliferation and immune tolerance in glioma through the miR-302a/KDM2A/JAG1 axis. *Cell Death Dis.* 13 (2), 160. doi:10.1038/s41419-021-04471-4

Zhou, W., Wang, X., Chang, J., Cheng, C., and Miao, C. (2021). The molecular structure and biological functions of RNA methylation, with special emphasis on the roles of RNA methylation in autoimmune diseases. *Crit. Rev. Clin. Lab. Sci.*59 (3), 203–218. doi:10.1080/10408363.2021.2002256

Zhou, Z., Wei, J., Lu, B., Jiang, W., Bao, Y., Li, L., et al. (2021). Comprehensive characterization of pyroptosis patterns with implications in prognosis and immunotherapy in low-grade gliomas. *Front. Genet.* 12, 763807. doi:10.3389/fgene.2021.763807

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 26 January 2023
DOI 10.3389/fgene.2023.1062060

OPEN ACCESS

EDITED BY
Yilin Zhang,
The University of Chicago, United States

REVIEWED BY
Yinhao Chen,
Nantong University, China
Yonghua Wang,
The Affiliated Hospital of Qingdao University, China

*CORRESPONDENCE
Bin Fu,
urofbin@163.com
Xiaoqiang Liu,
shaw177@163.com

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

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 05 October 2022
ACCEPTED 16 January 2023
PUBLISHED 26 January 2023

CITATION
Li S, Jiang M, Yang L, Zheng F, Liu J, Situ X, Liu X, Weipeng L and Fu B (2023), Identification of platinum resistance-related gene signature for prognosis and immune analysis in bladder cancer. *Front. Genet.* 14:1062060. doi: 10.3389/fgene.2023.1062060

Identification of platinum resistance-related gene signature for prognosis and immune analysis in bladder cancer

Sheng Li[1,2†], Ming Jiang[1,2†], Lin Yang[1,2†], Fucun Zheng[1,2], Jiahao Liu[1,2], Xiong Situ[1,2], Xiaoqiang Liu[1,2]*, Liu Weipeng[1,2] and Bin Fu[1,2]*

[1]Department of Urology, Nanchang, China, [2]The First Affiliated Hospital of Nanchang University, Nanchang, China

Purpose: Currently, there is limited knowledge about platinum resistance-related long non-coding RNAs (lncRNAs) in bladder cancer. We aim to identify platinum resistance-related lncRNAs and construct a risk model for accurate prognostic prediction of bladder cancer.

Methods: Transcriptomic and clinical data were extracted from The Cancer Genome Atlas (TCGA) database, and platinum resistance-related genes were obtained from HGSOC-Platinum. The platinum resistance-related lncRNAs were obtained by the Spearman correlation analysis. Then, we constructed a risk score model through Cox regression analysis and the LASSO algorithm. The model was verified by analyzing the median risk score, Kaplan-Meier curve, receiver operating characteristic (ROC) curve, and heatmap. We also developed a nomogram and examined the relationship between the risk score model, immune landscape, and drug sensitivity. Lastly, we assessed the differential expression of PRR-lncRNAs in the cisplatin-resistant bladder cancer cell line and the normal bladder cancer cell line using qRT-PCR.

Results: We developed and validated an eight-platinum resistance-related lncRNA risk model for bladder cancer. The risk model showed independent prognostic significance in univariate and multivariate Cox analyses. Based on multivariate analysis, we developed a nomogram. The modified model is both good predictive and clinically relevant after evaluation. Furthermore, immune-related and drug-sensitivity analyses also showed significant differential expression between high and low-risk groups. The qRT-PCR demonstrated that most of the lncRNAs were upregulated in cisplatin-resistance cancerous tissues than in control tissues.

Conclusion: We have developed a predictive model based on eight platinum resistance-related lncRNAs, which could add meaningful information to clinical decision-making.

KEYWORDS

bladder cancer, platinum resistance, long non-coding RNA, os, prognosis model

Introduction

Bladder cancer (BLCA) is the world's 10th most commonly diagnosed cancer (Babjuk et al., 2022). Although nearly 75% of bladder cancers are non-muscle-invasive (NMIBC), 45%–50% of patients with NMIBC will experience recurrence, and 6%–40% will progress (Slovacek et al., 2021). Patients with NMIBC are prone to develop muscle-invasive bladder

cancer (MIBC) after repetition and have a high risk of metastasis and poor prognosis, with a few surviving for more than 5 years (Chou et al., 2016; Malmstrom et al., 2017). Approximately 50% of MIBC patients eventually develop the disease at distant sites because of disseminated micrometastases, even after undergoing radical cystectomy and pelvic lymph node dissection (Patel et al., 2020). Hence, identifying specific tumor factors and providing new biomarkers are necessary to accurately diagnose, treat, and predict bladder cancer's outcome.

Platinum-based chemotherapy drugs are one of the most commonly used drugs for treating various tumors, especially for the systemic management of muscle-invasive and advanced bladder cancer (Ghosh, 2019). Initially, sensitive tumors, frequently observed in cancers, eventually develop chemoresistance. Unfortunately, the development of platinum resistance results in significant tumor recurrence and decreased overall patient survival (Hu et al., 2018). Non-coding RNAs that are longer than 200 nucleotides are long non-coding RNAs (LncRNAs). Since the development of high-throughput sequencing in recent years, many non-coding genes have been discovered to regulate the occurrence, development, metastasis, and chemotherapy resistance in cancers (Gao et al., 2020; Liu et al., 2020; Wu et al., 2020; Li et al., 2021a; Lu et al., 2021). It also significantly impacts bladder cancer, such as lncRNA KCNQ1OT1 facilitates the progression by targeting MiR-218–5p/HS3ST3B1 (Li et al., 2021b), and lncRNA CASC11 promotes cancer cell proliferation in bladder cancer through miRNA-150 (Luo et al., 2019).

Nevertheless, the role and prognostic value of platinum resistance-related (PRR) lncRNAs in BLCA have yet to be expounded. Consequently, we investigated the correlation between bladder cancer and PRR lncRNAs. As well as functional enrichment analysis of PRR lncRNAs, we analyzed immune cell infiltration, immune checkpoints, tumor mutational burden (TMB), immunotherapy, and drug sensitivity between high- and low-risk patients. Besides, we used a nomogram to visualize the overall survival of BLCA patients. It is hoped that new biomarkers can be provided for the personalized treatment of BLCA patients.

Methods

Data download and processing

The Cancer Genome Atlas (TCGA) database was accessed to obtain RNA sequencing data, tumor mutational burden (TMB) data, and related clinical information on bladder cancer patients. Transcriptome FPKM data was extracted using Strawberry Perl for further analysis. Genes expressing less than one in more than half of the samples were deleted. Moreover, clinically incomplete samples were excluded from the follow-up clinical correlation analysis. The results of comprehensive immunogenomic analyses of bladder cancer were obtained from The Cancer Immunome Database (TCIA, https://www.tcia.at/home). Platinum resistance-related genes were downloaded from HGSOC-Platinum (http://ptrc-ddr.cptac-data-view.org). Using the limma package in R software, a differential expression matrix for platinum resistance-related genes (PRR) was created. The criteria for differential expression analysis were | log 2 (fold change) | >1 and a false discovery rate (FDR) < 0.01.

Identification of platinum resistance-related (PRR) LncRNAs

Spearman correlation coefficients were calculated based on differential expression PRR genes and lncRNA expression profiles to recognize platinum resistance-related lncRNAs (|R2 | >0.45 and $p < 0.05$).

Construction of platinum resistance-related prognostic signature and GSEA

Firstly, univariate Cox regression analysis was utilized to evaluate the prognostic value of PRR lncRNAs. When the p-value was lower than 0.01, it was incorporated into the LASSO regression analysis. Then, based on the above results, we developed the platinum resistance-related prognostic model. Platinum resistance-related prognostic scores for each patient were calculated as follows: Risk score = (Coef (lncRNA1) * expression lncRNA1) + (Coef (lncRNA2) * expression lncRNA2) +. . . . . .+ (Coef (lncRNA n) * expression lncRNA n). Eventually, due to the median risk score, patients were divided into low- and high-risk groups. The Kaplan–Meier curve was generated with the log-rank test to compare the two groups' overall survival (OS). To evaluate the predictive performance of the signature, we used the 'timeROC' R package to generate a receiver operating characteristic curve (ROC). A heat map was used to show the difference in platinum resistance-related lncRNA expression profiles between the high/low-risk groups. We randomly split the entire cohort into a 1:1 train and a test set for internal validation to assess the risk model feasibility. Validation cohorts were calculated using the same formula as the total cohort, and the same validation method was applied. We used the Gene Set Enrichment Analysis (GSEA) to examine the molecular mechanisms underlying low- and high-risk groups. p values less than 0.05 were considered statistically significant.

Building and validating a nomogram

Univariate Cox and multivariate Cox regression analyses were used to identifying potential prognostic factors for the risk model and clinical features. Then, we constructed a nomogram by incorporating the meaningful variables ($p< 0.05$). Clinicians can easily use the nomogram to assess 1-, 3-, and 5 year overall survival in bladder cancer patients. The receiver operating characteristic (ROC) and calibration plots were calculated to estimate the discriminative accuracy of the nomogram. All of these will be validated on training and test sets.

Comprehensive analysis of the relationship between the risk model and tumor microenvironment and immunity

The ESTIMATE algorithm was used to assess immune infiltration in bladder cancer patients (Supplementary Table S1). The difference in immune cell infiltration between the high-risk and low-risk groups of patients was evaluated using TIMER, CIBERSORT, CIBERSORT-ABS, QUANTISEQ, MCP-counter, XCELL, and EPIC algorithms. In addition, the potential immune checkpoint was acquired from previous literature. We detected the expression levels of immune checkpoint-related genes between the two groups. Furthermore, we used TCIA data to predict the relationship between platinum

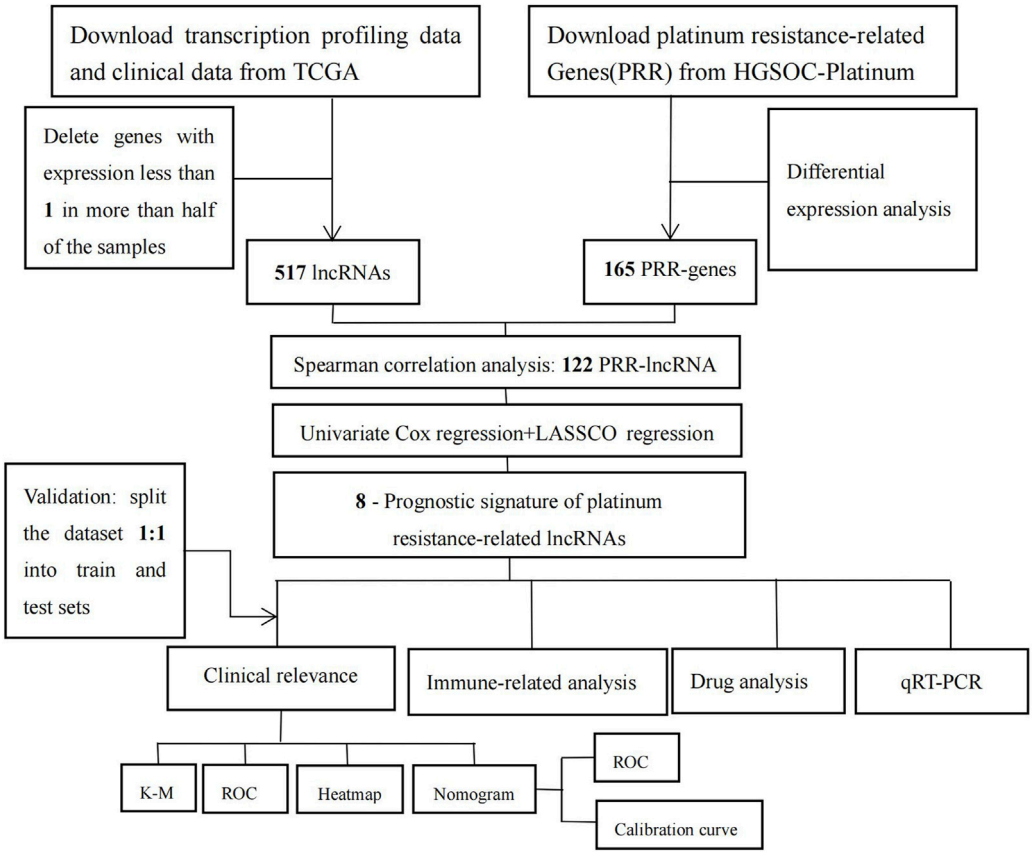


FIGURE 1
Flow-process diagram of the study.

resistance-related prognostic scores and immunotherapy sensitivity. TMB between the two groups was also analyzed. Box plots were generated to visualize the differences.

Drug sensitivity analysis

We use the "pRRophetic" package in R software to predict the drug's half-maximal inhibitory concentration (IC50) value between the high-risk and low-risk groups. Moreover, we considered p values less than 0.05 to be statistically significant. Box plots were generated to visualize the differences.

Cell culture and qRT-PCR

Human BC cells of T24 were purchased from the Cell Bank of Culture Collection of the Chinese Academy of Sciences, Shanghai Institute of Cell Biology (Shanghai, China). The exponential growth phase T24 cells were selected, and 200 μg/mL was chosen as the initial drug concentration according to the pre-experiment. The same concentration was repeated three times, each 2 days. Continue using the previous concentration of cisplatin for 2 days after passage, and then gradually increase the concentration. If the cell condition is not good, replace the medium without cisplatin. When the cell condition is normal, continue to add medicine. A cisplatin-resistant bladder cancer cell line, T24-CDDP, was established after cisplatin continued for 10 months. The T24-CDDP cell lines were validated by Cell Counting Kit-8 (CCK-8) assay, and GraphPad Prism9 was used to plot the cell IC50. All cells are cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco) and at 37°C in 5% CO2. Invitrogen TRIzol reagent was used for total RNA extraction and the Takara PrimeScript RT reagent Kit for cDNA synthesis. Real-time quantitative PCR was performed using SYBR Green (Roche, Switzerland). Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as an endogenous reference. At least three replicates of each reaction were performed. Supplementary Table S2 shows the primer sequences.

Result

Basic information

Figure 1 shows the flowchart of our study. We obtained gene expression profiles of 431 bladder tumor patient samples, including 412 tumors and 19 adjacent normals, from the TCGA database. Samples with incomplete clinical information were removed. Supplementary Table S3 contains the clinical data for the remaining 372 tumor samples. Then we randomly split the entire cohort into a 1:1 train and a test set for internal validation. Data on 412 bladder cancers containing information on immunotherapy were downloaded from the TCIA database (Supplementary Table S4).

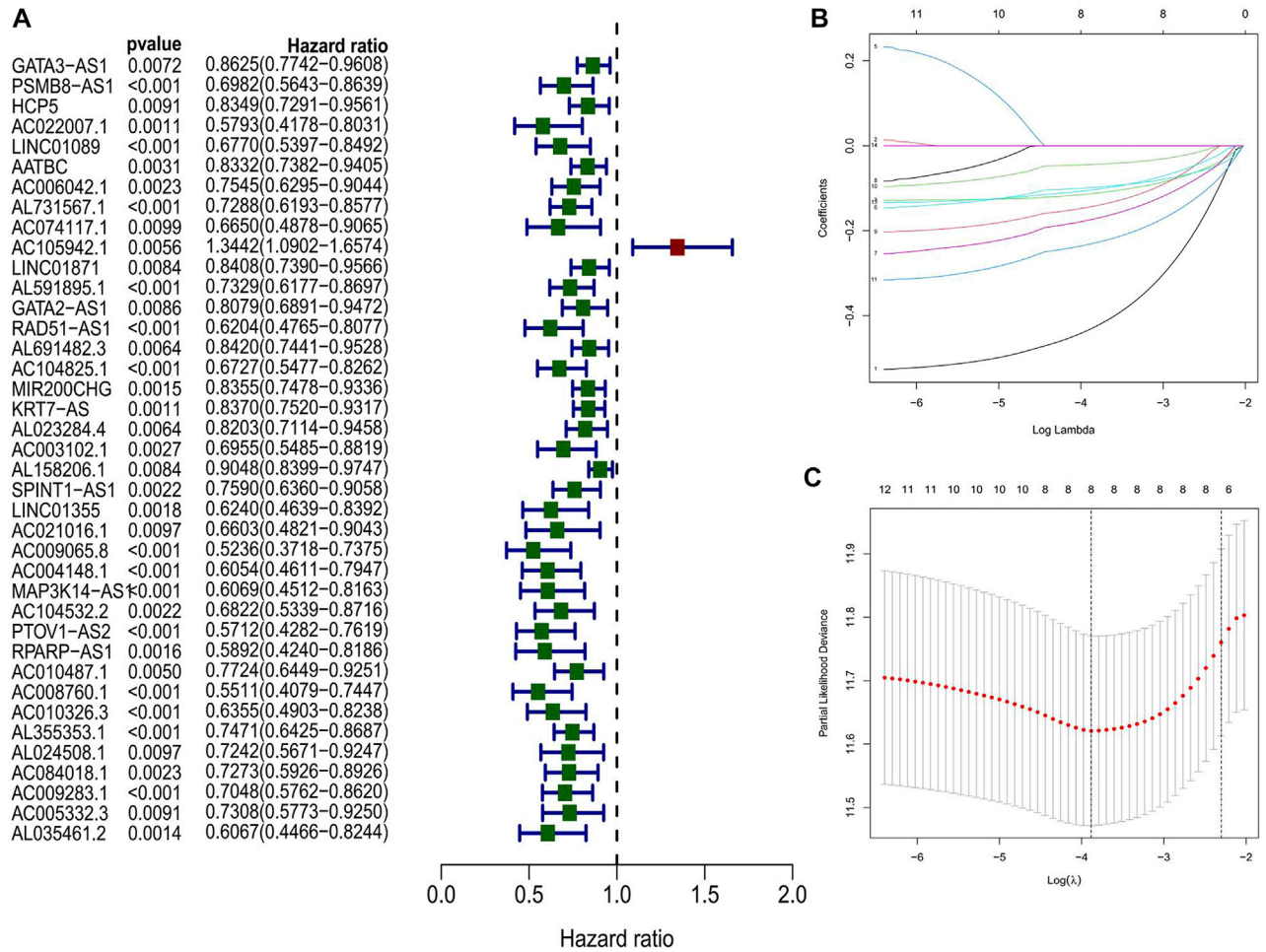


FIGURE 2
(A) Identification of prognostic PRR by univariate Cox regression analysis in the whole group **(B, C)** Lasso regression analysis in the entire group.

Differentially expressed (DE) platinum resistance genes and LncRNA

Supplementary Table S5 records that 936 platinum resistance genes were extracted from HGSOC-Platinum. A comparison of bladder cancer tissues with normal tissues identified 165 DE genes (78 were upregulated and 87 were downregulated). Supplementary Figure S1 shows the heatmap evaluation of DE genes. In Supplementary Figure S2, a volcano map represents the distribution of all DE genes according to log10FDR and log2FC. Throughout the gene expression profiles, 16882 lncRNAs were identified. Five hundred eleven lncRNAs remained after deleting genes with an expression of less than one in more than half of the samples (Supplementary Table S6). Then, we identified 122 platinum resistance-related lncRNAs by correlation Spearson analysis, as shown in Supplementary Table S7.

Construction and validation of a platinum resistance-related lncRNA risk model

Using univariate Cox regression, 39 platinum resistance-associated lncRNAs were identified. Other than AC105942.1, which was a high-risk prognostic LncRNA, all others were low-risk ($p < 0.01$, Figure 2A). In the LASSO regression analysis, eight platinum resistance-related lncRNAs were associated with prognostic factors in bladder cancer (BCa) patients (Supplementary Table S8). It was verified through cross-validation that the LASSO regression analysis optimal value was the right one (Figures 2B, C). The formula of the risk score was as follows: Risk score = (−0.431424975893598*PSMB8-AS1) + (−0.1130343821 6125*AL731567.1) + (−0.0984074363105057*AC104825.1) + (−0.173932427517578*AC009065.8) + (−0.1434206 81656449* MAP3K14-AS1) + (−0.0447654488946425*PTOV1-AS2) + (−0.231285006432146*AC008760.1) + (−0.10488485275485*AL35 5353.1). This risk model divided patients into high-risk and low-risk groups based on the median risk score. The Kaplan-Meier survival analysis showed that low-risk BCa patients had a significantly better overall survival than patients at high risk (Figure 3A). Moreover, based on the risk model, the scatterplot demonstrated a correlation between survival time and risk score for BCa patients. There was a correlation between patients' risk scores and their mortality from bladder cancer. The higher the score, the greater the risk (Figure 3D). As shown in the heat map (Figure 3G), these eight-platinum resistance-related lncRNAs were highly expressed as protective factors in the low-risk group. Lastly, overall survival AUCs of 1-, 3-, and 5 years were 0.709,

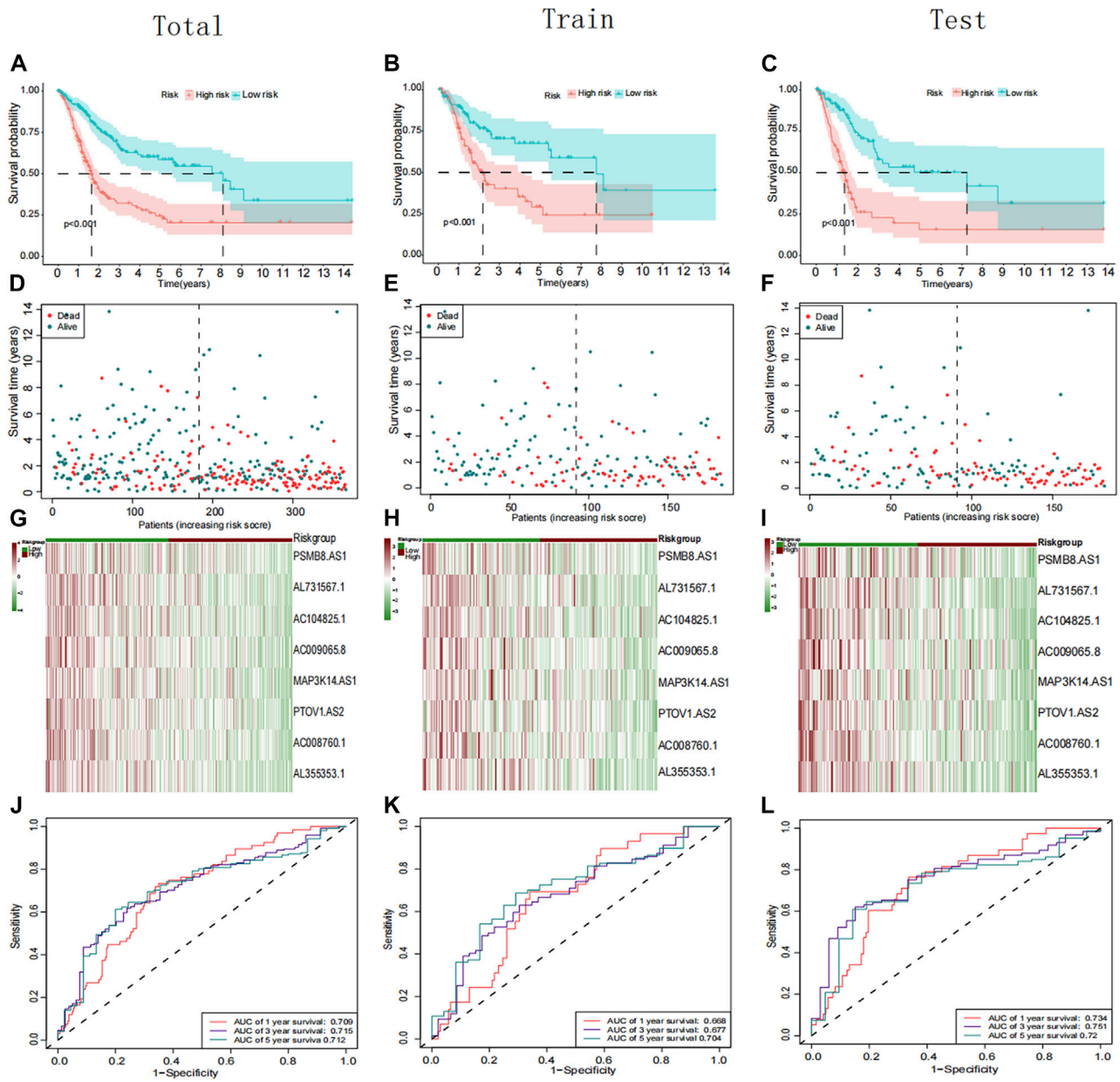


FIGURE 3
Prognostic analysis of the PRR lncRNAs signature in the total, training cohort, and testing group. **(A–C)** Kaplan–Meier curve of the patient in the whole, training cohort, and testing groups. **(D–F)** The rank of calculated risk scores in the total, training cohort, and testing groups. **(G–I)** Heatmap showed the differences of 8 PRR lncRNAs in the whole, training cohort, and testing groups. **(J–L)** Time-independent receiver operating characteristic (ROC) analysis in the total, training cohort, and testing groups.

0.715, and 0.712, respectively (Figure 3J). In the training and testing groups, we validated the risk model. These two groups used the same methods to identify high-risk and low-risk patients. Figure 3B and Figure 3C illustrate the relationship between risk scores and survival. The prognostics between the different risk patients in the training and testing groups were shown in Figure 3E and Figure 3F. There was a significant decrease in the overall survival of the high-risk group compared with the low-risk group. The heat maps were consistent across the entire group (Figures 3H, I). Figures 3K, L showed that both training and testing groups achieved ideal AUC values.

Construction and assessment, a new type of nomogram

A multivariate and univariate Cox analysis of clinical variables, including age, grade, stage, T stage, and risk scores, revealed that the risk model was the most significant prognostic factor (Figures 4A, B). Then, according to the critical variables in the multiple regression analysis ($p< 0.05$), a prognostic nomogram of bladder cancer patients was established (Figure 4C), which could be used to predict the 1-, 3-, and 5 year OS rates of patients. In the entire cohort, the AUC of values for 1-, 3-, and 5 year OS were 0.783, 0.765, and 0.760,

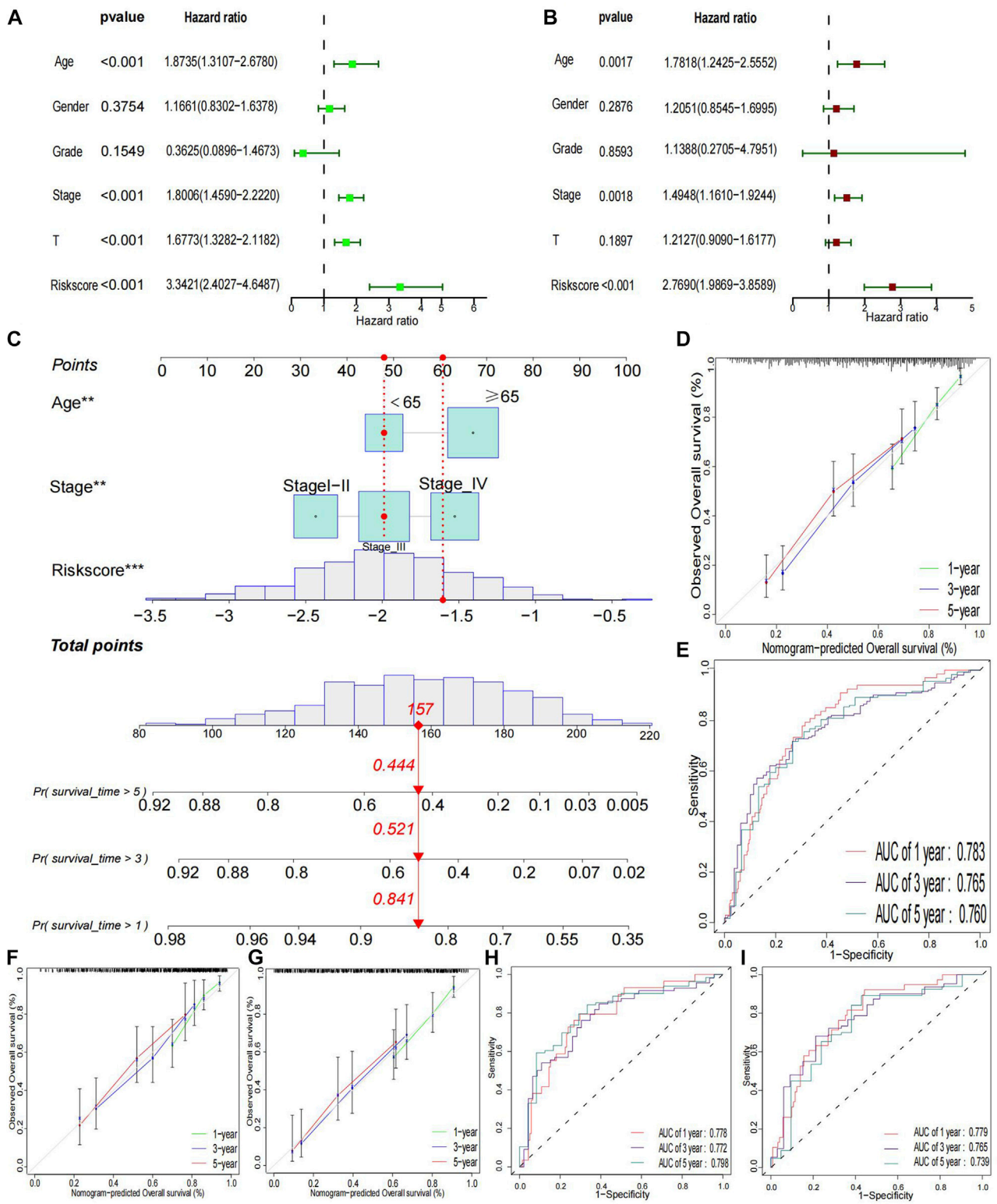


FIGURE 4
Clinical relevance Analysis and Validation **(A)** Univariate prognostic analysis. **(B)** Multivariate prognostic analysis. **(C)** Constructed a nomogram in whole groups. **(D, E)** Calibration curves and ROC for 1-, 3-, and 5 year in the entire group. **(F–I)** Calibration curves and ROC for 1-, 3-, and 5-year in training and testing groups, respectively.

respectively (Figure 4E). Calibration plots for 1-, 3-, and 5 years were generated to verify our model across the entire cohort. All calibration plots fall near the 45-degree diagonal line (Figure 4D). The AUC values and calibration plots of the training set and test set show that the nomogram has good discriminative power (Figures 4F–I).

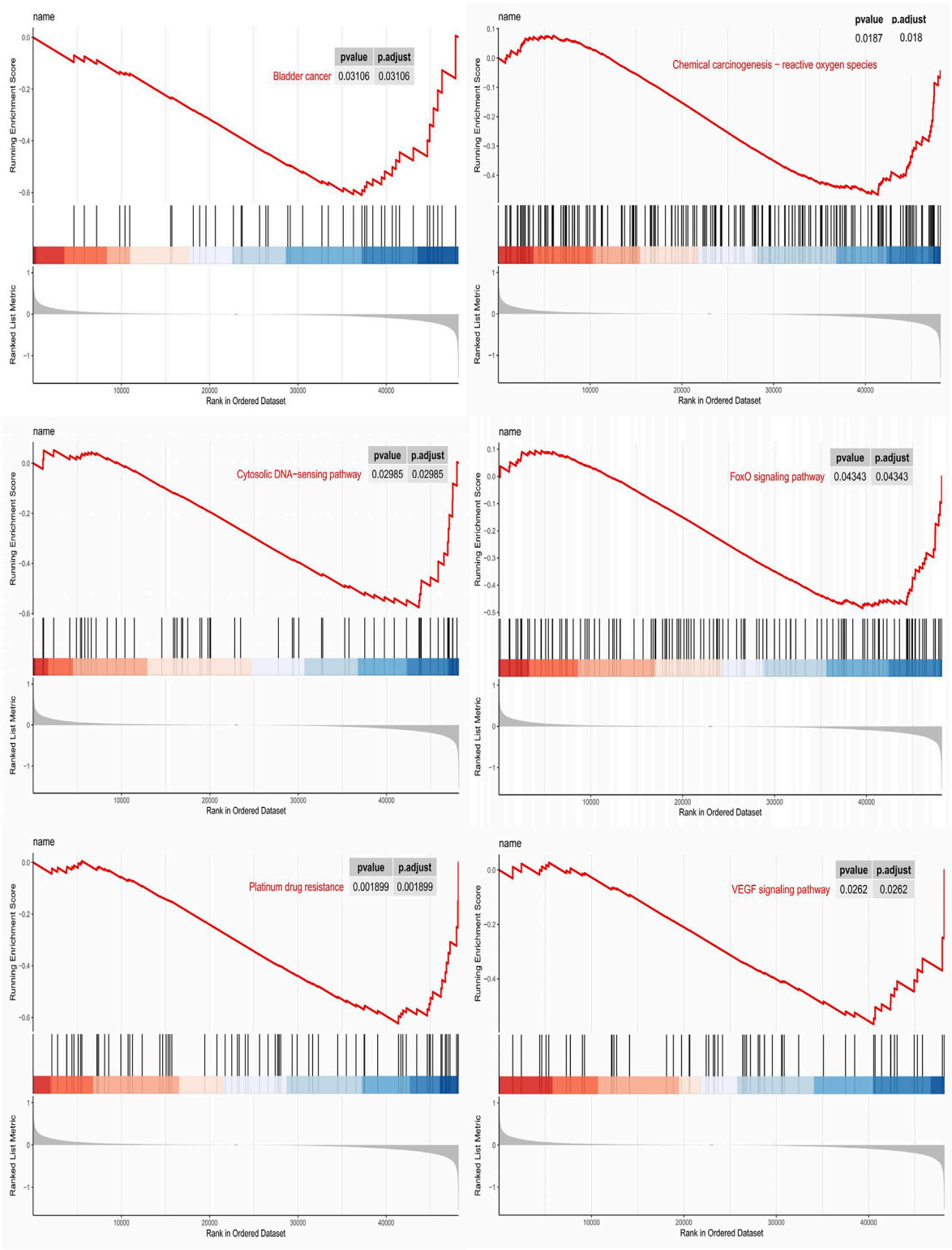


FIGURE 5
GSEA analysis.

GSEA analysis of platinum resistance-related (PRR) LncRNAs

GSEA analysis was performed to elucidate the biological function of PRR-based signatures further. GSEA revealed that PRR lncRNA prognostic models mainly regulated cancer- and platinum-related pathways, such as Bladder cancer, Cytosolic DNA–sensing pathway, VEGF signaling pathway, FoxO signaling pathway, Chemical carcinogenesis–reactive oxygen species, and Platinum drug resistance (Figure 5).

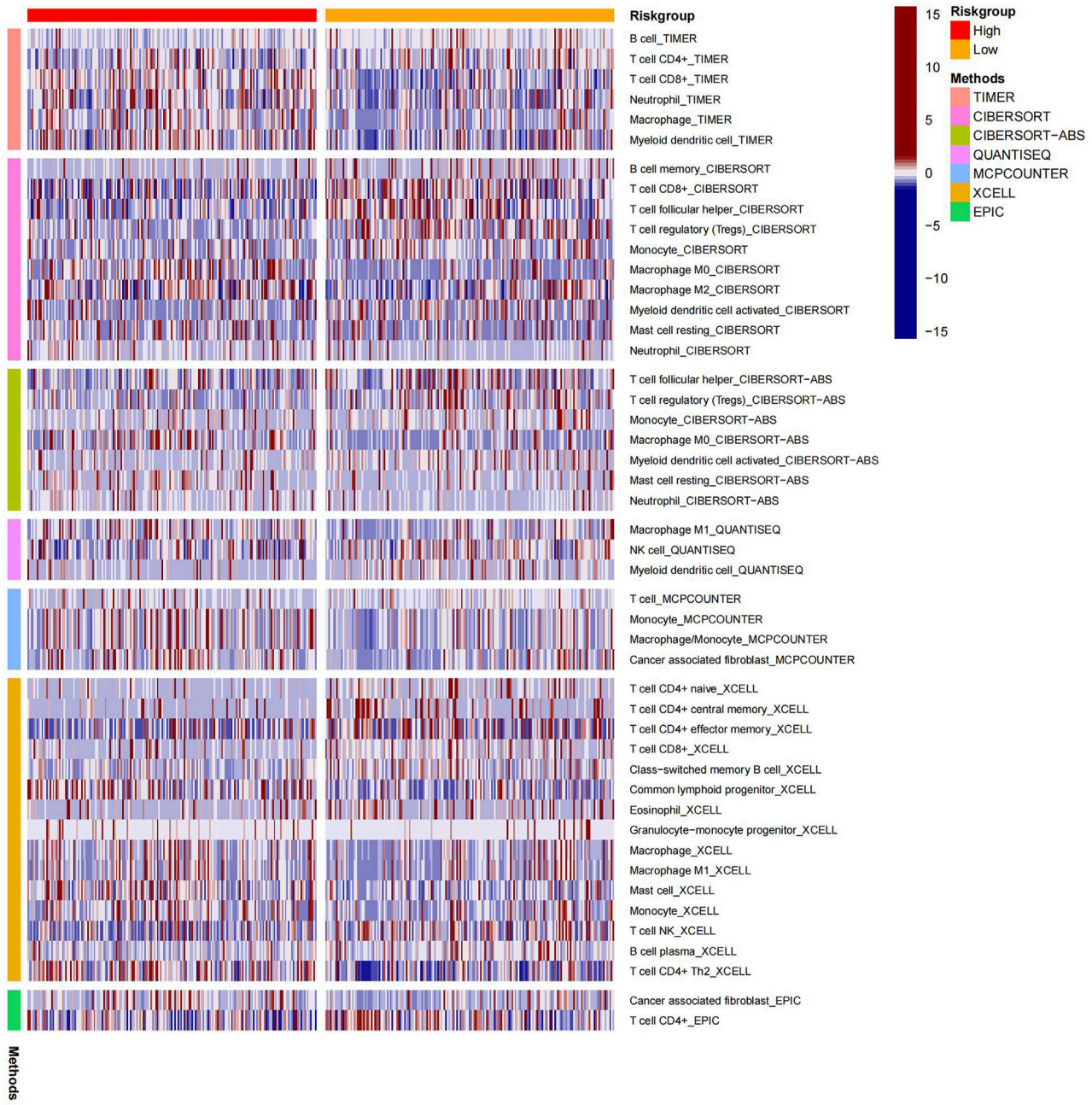


FIGURE 6
Heat map of immune cell infiltration in high and low-risk groups.

Immune-related analysis of BLCA patients using the risk model

The heatmap displayed the relationship between the risk model and immune infiltration (Figure 6). In the low-risk group, $CD4^+$ T-cell, $CD8^+$ T-cell, and regulatory cells infiltrated more than in the high-risk group. At the same time, macrophages and monocytes were more prevalent in high-risk populations. Furthermore, based on immune checkpoint analysis, representative immune checkpoint-related genes, such as PDCD1LG2, CD44, CD47, CD276, PVR, and TNFSF9, were remarkably upregulated when compared with low-risk group samples (Figure 7). Comparison of somatic mutations in patients with high and low-risk scores and visualization of the top 20 genes with the highest mutation frequency (Figures 8A, B). There was no significant difference in TMB between the high-risk and low-risk groups. By analysis, we found that patients with low-risk scores were more sensitive to immunotherapy, whether they were CTLA4+ or PD-1+ or both positive (Figures 8C–F).

Drug sensitive and qRT-PCR

A further investigation was conducted to assess the sensitivity difference of drugs in two groups of patients with

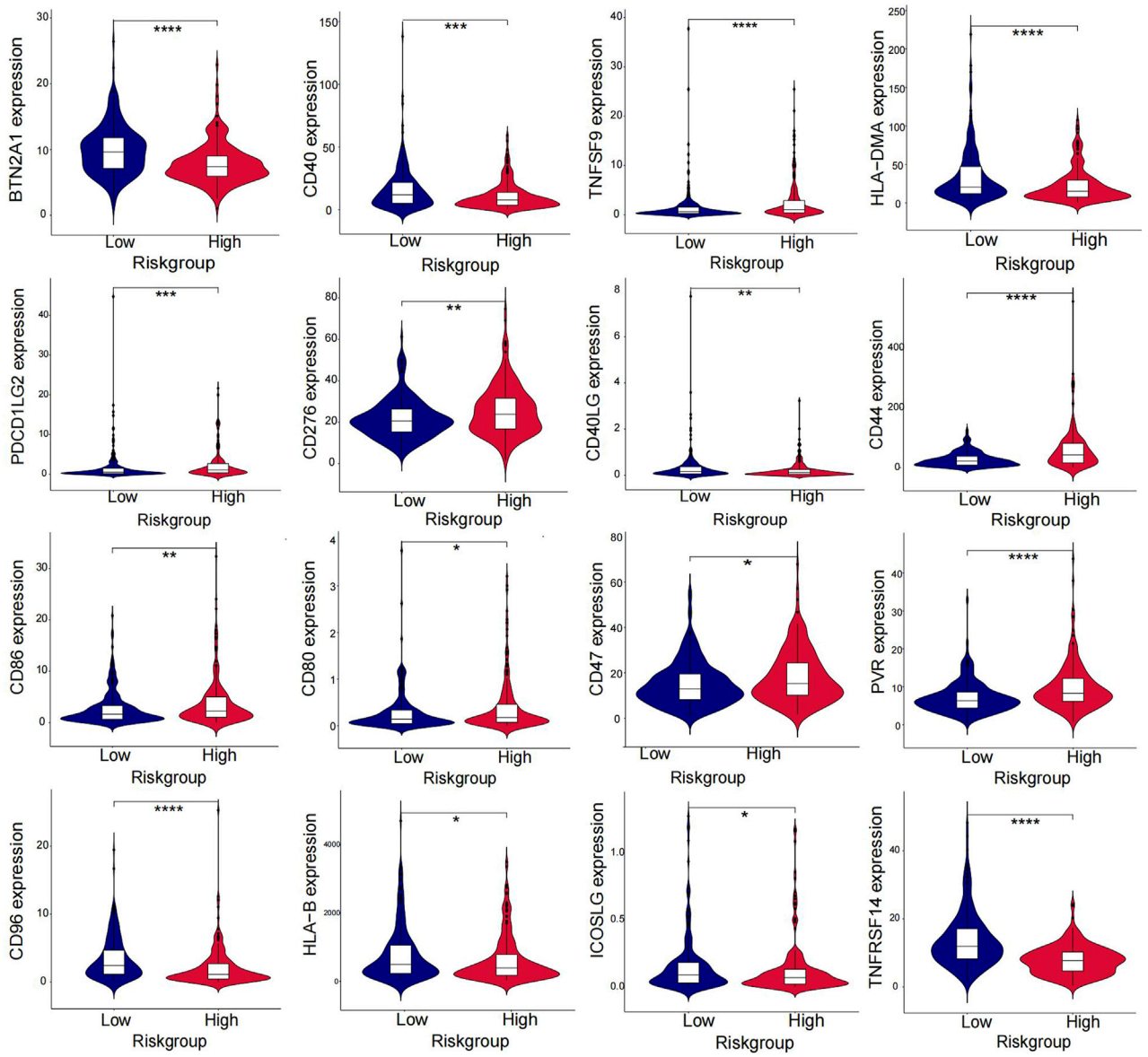


FIGURE 7
Differences in immune checkpoints between high and low-risk groups.

bladder cancer to improve the therapeutic outcome. The analysis results indicated that IC50 values of drugs, including BIBW2992, Erlotinib, Gefitinib, and Lapatinib, were higher in high-risk patients than those of low risk. While IC50 values of drugs containing Cisplatin, Gemcitabine, Mitomycin C, Methotrexate, Vinblastine, Vinorelbine, Doxorubicin, Docetaxel, Thapsigargin, and Pazopanib were much higher in the low-risk patients than those of the high-risk (Figure 9). In Figures 10A, B, we can see that the cisplatin IC50 for T24-CDDP was significantly higher than that of T24, implying the thriving culture of our drug-resistant cells. As shown in Figure 10C, AC008760.1, PTOV1-AS2, AL355353.1, AC104825.1, and MAP3K14-AS1 were more highly expressed in cisplatin-resistance T24 cells than normal T24 cells.

Discussion

In recent years, many studies have focused on the role lncRNAs play in bladder cancer (BLCA). Lia et al. developed and validated an eight-pyroptosis-related lncRNA prognostic model for BLCA (Lia et al., 2022). Luo et al. found that lncRNA RP11-89 facilitates tumorigenesis and ferroptosis resistance in BLCA (Luo et al., 2021). Tong et al. constructed a prognostic epithelial-mesenchymal transition-related lncRNA risk model in BLCA (Tong et al., 2021). Hu et al. discussed the roles and mechanisms of lncRNAs in cisplatin chemoresistance, including changes in cellular uptake or efflux of a drug, apoptosis, autophagy, related signaling pathways, and so on 7). However, studies on the prognosis of platinum resistance-associated (PRR) lncRNAs in BLCA are still limited. Accordingly, we explored the relationship between PRR lncRNAs and the prognosis of BLCA.

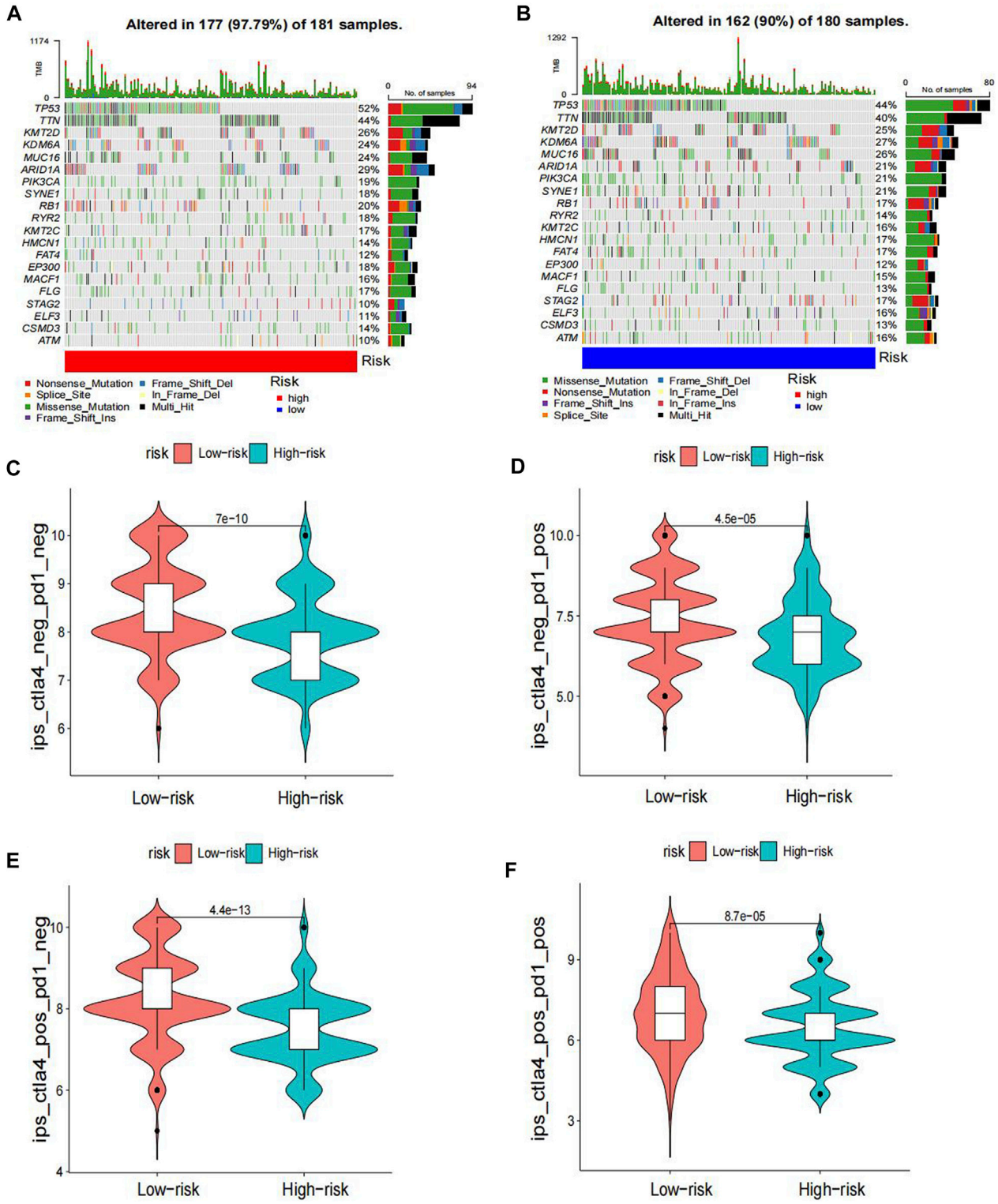


FIGURE 8
Differences in TMB and immunotherapy sensitivity between the two groups.

Based on eight PRR lncRNAs, we established a BLCA risk prognosis model, and median risk scores categorized patients into high and low-risk groups. Kaplan-Meier survival, heatmap, and ROC analyses have shown the good predictive ability of our risk model. Moreover, the immune-related and drug-sensitivity analysis also showed significant differences between high- and low-risk groups. These identified PRR lncRNAs were protection factors: PSMB8-AS1, AL731567.1, AC104825.1, AC009065.8, MAP3K14-AS1, AL355353.1,

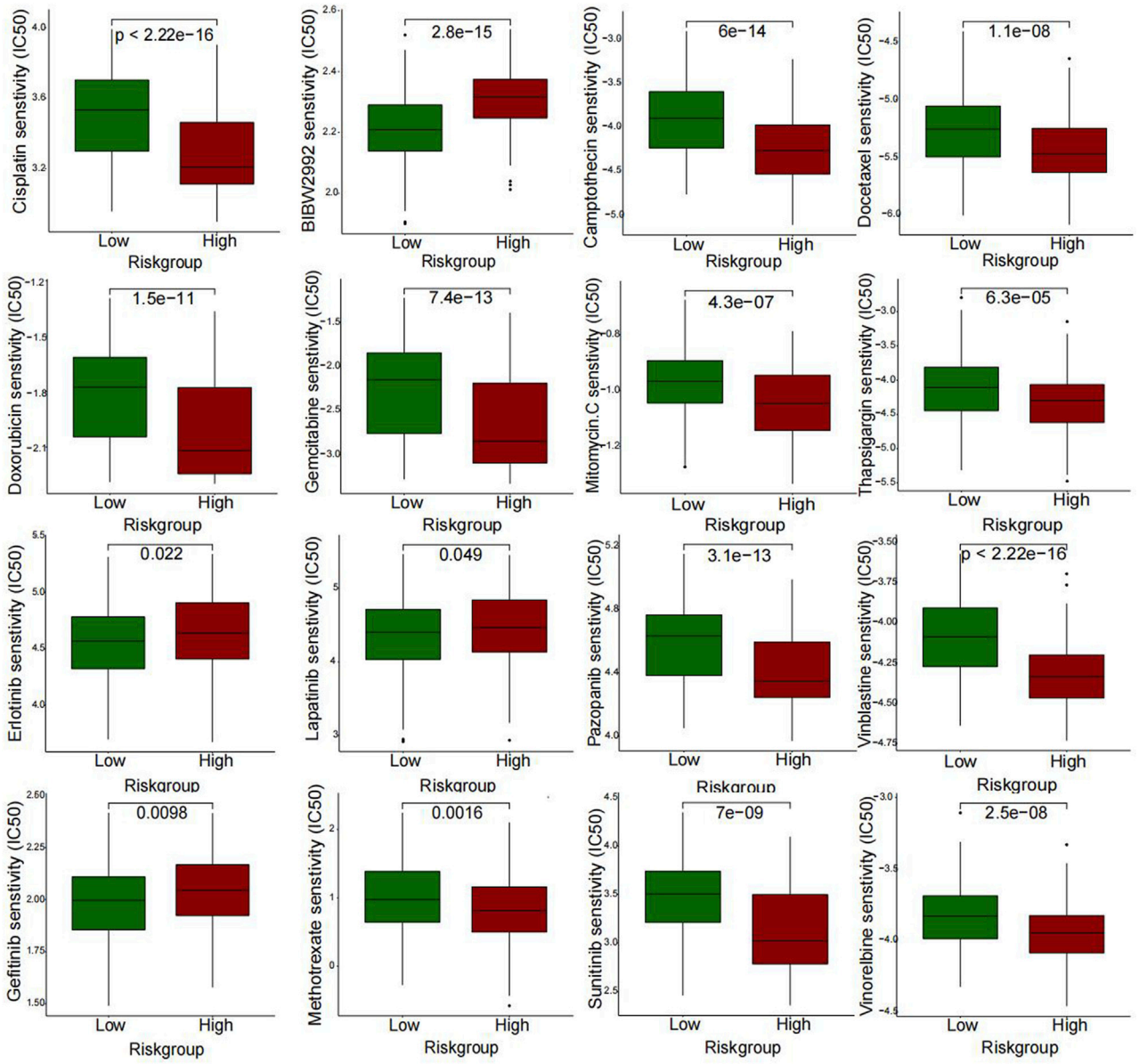


FIGURE 9
Differences in drug sensitivity between the two groups.

AC008760.1, and PTOV1-AS2. When protective factors are expressed at a higher level, the prognosis for BLCA patients is better.

Recent studies have found LncRNA PSMB8-AS1 to be a prognostic marker and a protective factor in BLCA (Tong et al., 2021; Mo et al., 2022). Zhang et al. proposed that PSMB8-AS1 promotes pancreatic cancer progression by regulating the miR-382–3p/STAT1/PD-L1 axis. Thus, it is worthwhile to explore PSMB8-AS1's mechanism of action in bladder cancer (Zhang et al., 2020). MAP3K14-AS1 was recognized as a highly prevalent and specific methylated locus in colorectal cancer, which can be used to monitor tumor burden dynamics in liquid biopsy under different therapeutic regimens (Barault et al., 2018). Kuang et al. revealed that the necroptosis-related lncRNAs MAP3K14-AS1 and AL731567.1 were considered protective effectors in BLCA (XiaYu et al., 2022). AL355353.1 was found to be associated with glycometabolism in BLCA and affected prognosis (Tang et al., 2022). Liu et al. revealed that PTOV1-AS2 might affect the prognosis of pancreatic cancer through TP53-associated signature (Liu et al., 2021). LncRNA AC008760.1 was identified as expressed lower in bladder urothelial carcinoma cells than in normal urothelial cells (Li et al., 2022), which was consistent with our findings. Moreover, the knockdown of AC008760.1 can significantly promote the proliferation and migration of bladder cancer cells. Furthermore, AC009065.8 and AC104825.1 in BLCA are rarely reported in research, and thus, the specific mechanism is also worthy of further investigation (Chen et al., 2020).

We further compared several clinical variables to assess our risk model's predictive ability. Three independent prognostic factors were identified: age, stage, and risk score. As previously reported, age and stage are prominent risk factors for multiple tumors, including bladder

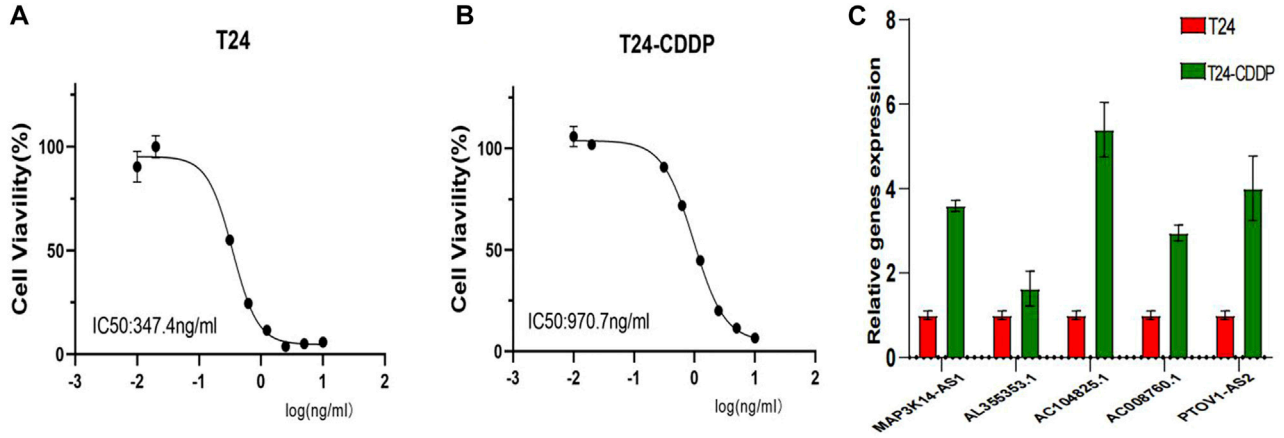


FIGURE 10
(A, B) Cisplatin IC50 for T24-CDDP and T24. **(C)** The expression of AC008760.1, PTOV1-AS2, AL355353.1, AC104825.1, and MAP3K14-AS1 in T24 and T24-CDDP cell lines.

cancer (Hu et al., 2022; Lu et al., 2022). Further comparison showed that the model's prediction performance is superior to age and stage, demonstrating its high predictive power. To increase the clinical applicability of the model, we used a nomogram to visualize the survival probability of bladder cancer patients. In the training and testing sets, both ROC and calibration curves showed the good predictive ability of the nomogram.

Immunotherapy has shown promising results in the management of BLCA. In our study, high-risk and low-risk groups were compared regarding the immune checkpoint. Figure 7 displays that TNFSF9, PDCL1LG2, PVR, CD44, CD86, CD80, CD47, and CD276 in the immune checkpoint were expressed in the high-risk group, while BTN2A1, CD40, CD40LG, HLA−DMA, HLA−B, CD96, ICOSLG, and TNFRSF14 were mainly expressed in the low-risk group. For the former, except for TNFSF9, almost all other genes in the immune checkpoint were reported in bladder cancer and were associated with poor outcomes (Kiss et al., 2019; Kucan Brlic et al., 2019; Hu et al., 2020; Yang et al., 2020; Yan et al., 2021; Harland et al., 2022). It is thought that tumors with more mutated genes tend to produce more mutant RNAs and proteins that are more easily recognized by the immune system and respond well to immunotherapy. Thus, we also analyzed the difference in TMB in the two risk groups. Although there was no significant difference in TMB between the high-risk and low-risk groups, we found that the low-risk group gained more immunophenoscores, which can be used to predict response to immune checkpoint inhibitors. The results showed that patients with low-risk scores were more sensitive to immunotherapy, whether they were CTLA4+ or PD-1-. Therefore, combined with our risk model, we found that immunotherapy could be a good option for bladder cancer patients with platinum resistance.

Many studies show that immune infiltration correlates with prognosis (Hatogai and Sweis, 2020; Zheng et al., 2020). Consequently, the rate of immune cell infiltration between different risk groups was calculated. Comparing the low-risk group with the high-risk group, we found that $CD8^+$ T-cell and regulatory cells were significantly increased. Induced tumor cell death is the primary function of $CD8^+$ cells (Henning et al., 2018). Moreover, the numbers of macrophages and monocytes have risen notably in the high-risk group, which are generally involved in defending against external attacks (Xia et al., 2020). Due to this, we considered that platinum resistance-related lncRNA is closely related to immune infiltration in bladder cancer.

Drug sensitivity analysis showed that high-risk groups were more sensitive to cisplatin because of their relatively low expression of platinum-resistance-associated LncRNAs, which further adds to the reliability of our findings. Moreover, drug analysis results also showed that IC50 of BIBW2992, erlotinib, gefitinib, and lapatinib were lower in low-risk patients, implying that drug-resistant patients were more sensitive to those drugs. T24 bladder cancer cells are inhibited in proliferation and invasion by BIBW2992/Afatinib (Tang et al., 2015). A primary mechanism of gefitinib is that it interferes with the metabolic functions of tumor cells and inhibits EGFR signaling in a meaningful manner (Peng et al., 2016). One recent study suggests lapatinib as a first-line option for treating muscle-invasive urothelial carcinoma in dogs (Maeda et al., 2022). These four drugs all belong to EGFR family inhibitors, which are expected to play a significant role in future bladder cancer treatments. Although we did not find suitable primer sequences for the rest three lncRNAs, it may be because their base sequences are long or technical problems. For the most, we evaluated the expression level of most lncRNAs in our signature. The expression trend followed the bioinformatic prediction.

However, some areas still need to be addressed in this study. Firstly, this is a retrospective study using TCGA datasets. Retrospective studies may have selection and information bias. For example, in light of the small sample size, stage I was grouped with stage II. Secondly, external validation needed to be improved as other databases lacked lncRNA expression profiles or overall survival data. Finally, although we experimentally validated the differential expression of PPR lncRNAs in platinum-resistant bladder cancer cells, the underlying mechanisms of how the detected platinum-resistance-related lncRNAs impact the prognosis of bladder cancer require further study by basic experiments.

Conclusion

Based on eight PRR lncRNAs, we constructed a prognosis model for BLCA patients. As well as providing prognostic information and immune analysis, our risk model can give a new direction for chemotherapy or targeted therapy for BLCA patients.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding authors.

Author contributions

XL, LW, and BF contributed to the conception of the study. SL, MJ, LY, FZ, and JL contributed significantly to the analysis and manuscript preparation.

Funding

This study was supported by the Jiangxi Provincial "Double Thousand Plan" Fund Project (Grant No. jxsq2019201027), Key Project of Natural Science Foundation of Jiangxi Province (20212ACB206013), and Youth Project of Natural Science Foundation of Jiangxi Province (20212BAB216037).

Acknowledgments

The TCGA and TCIA database provides cancer patient data, significantly facilitating clinical research.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1062060/full#supplementary-material

References

Babjuk, M., Burger, M., Capoun, O., Cohen, D., Comperat, E. M., Dominguez Escrig, J. L., et al. (2022). European association of urology guidelines on non-muscle-invasive bladder cancer (Ta, T1, and carcinoma *in situ*). *Eur. Urol.* 81 (1), 75–94. doi:10.1016/j.eururo.2021.08.010

Barault, L., Amatu, A., Siravegna, G., Ponzetti, A., Moran, S., Cassingena, A., et al. (2018). Discovery of methylated circulating DNA biomarkers for comprehensive non-invasive monitoring of treatment response in metastatic colorectal cancer. *Gut* 67 (11), 1995–2005. doi:10.1136/gutjnl-2016-313372

Chen, Q., Hu, L., Huang, D., Chen, K., Qiu, X., and Qiu, B. (2020). Six-lncRNA immune prognostic signature for cervical cancer. *Front. Genet.* 11, 533628. doi:10.3389/fgene.2020.533628

Chou, R., Selph, S. S., Buckley, D. I., Gustafson, K. S., Griffin, J. C., Grusing, S. E., et al. (2016). Treatment of muscle-invasive bladder cancer: A systematic review. *Cancer* 122 (6), 842–851. doi:10.1002/cncr.29843

Gao, N., Li, Y., Li, J., Gao, Z., Yang, Z., Li, Y., et al. (2020). Long non-coding RNAs: The regulatory mechanisms, research strategies, and future directions in cancers. *Front. Oncol.* 10, 598817. doi:10.3389/fonc.2020.598817

Ghosh, S. (2019). Cisplatin: The first metal-based anticancer drug. *Bioorg Chem.* 88, 102925. doi:10.1016/j.bioorg.2019.102925

Harland, N., Maurer, F. B., Abruzzese, T., Bock, C., Montes-Mojarro, I. A., Fend, F., et al. (2022). Elevated expression of the immune checkpoint ligand CD276 (B7-H3) in urothelial carcinoma cell lines correlates negatively with the cell proliferation. *Int. J. Mol. Sci.* 23 (9), 4969. doi:10.3390/ijms23094969)

Hatogai, K., and Sweis, R. F. (2020). The tumor microenvironment of bladder cancer. *Adv. Exp. Med. Biol.* 1296, 275–290. doi:10.1007/978-3-030-59038-3_17

Henning, A. N., Roychoudhuri, R., and Restifo, N. P. (2018). Epigenetic control of CD8+ T cell differentiation. *Nat. Rev. Immunol.* 18 (5), 340–356. doi:10.1038/nri.2017.146

Hu, J., Lai, C., Shen, Z., Yu, H., Lin, J., Xie, W., et al. (2022). A prognostic model of bladder cancer based on metabolism-related long non-coding RNAs. *Front. Oncol.* 12, 833763. doi:10.3389/fonc.2022.833763

Hu, Y., Zhang, Y., Gao, J., Lian, X., and Wang, Y. (2020). The clinicopathological and prognostic value of CD44 expression in bladder cancer: A study based on meta-analysis and TCGA data. *Bioengineered* 11 (1), 572–581. doi:10.1080/21655979.2020.1765500

Hu, Y., Zhu, Q. N., Deng, J. L., Li, Z. X., Wang, G., and Zhu, Y. S. (2018). Emerging role of long non-coding RNAs in cisplatin resistance. *Once Targets Ther.* 11, 3185–3194. doi:10.2147/OTT.S158104

Kiss, B., van den Berg, N. S., Ertsey, R., McKenna, K., Mach, K. E., Zhang, C. A., et al. (2019). CD47-Targeted near-infrared photoimmunotherapy for human bladder cancer. *Clin. Cancer Res.* 25 (12), 3561–3571. doi:10.1158/1078-0432.CCR-18-3267

Kucan Brlic, P., Lenac Rovis, T., Cinamon, G., Tsukerman, P., Mandelboim, O., and Jonjic, S. (2019). Targeting PVR (CD155) and its receptors in anti-tumor therapy. *Cell Mol. Immunol.* 16 (1), 40–52. doi:10.1038/s41423-018-0168-y

Li, X., Zhang, C., Peng, X., Li, Y., Chen, G., Gou, X., et al. (2022). A novel risk score model based on five angiogenesis-related long non-coding RNAs for bladder urothelial carcinoma. *Cancer Cell Int.* 22 (1), 157. doi:10.1186/s12935-022-02575-1

Li, Y., Shi, B., Dong, F., Zhu, X., Liu, B., and Liu, Y. (2021). LncRNA KCNQ1OT1 facilitates the progression of bladder cancer by targeting MiR-218-5p/HS3ST3B1. *Cancer Gene Ther.* 28 (3-4), 212–220. doi:10.1038/s41417-020-00211-6

Li, Z., Meng, X., Wu, P., Zha, C., Han, B., Li, L., et al. (2021). Glioblastoma cell-derived lncRNA-containing exosomes induce microglia to produce complement C5, promoting chemotherapy resistance. *Cancer Immunol. Res.* 9 (12), 1383–1399. doi:10.1158/2326-6066.CIR-21-0258

Lia, T., Shao, Y., Regmi, P., and Li, X. (2022). Development and validation of pyroptosis-related lncRNAs prediction model for bladder cancer. *Biosci. Rep.* 42 (1). doi:10.1042/BSR20212253)

Liu, X., Chen, B., Chen, J., and Sun, S. (2021). A novel tp53-associated nomogram to predict the overall survival in patients with pancreatic cancer. *BMC Cancer* 21 (1), 335. doi:10.1186/s12885-021-08066-2

Liu, Y., Cheng, G., Huang, Z., Bao, L., Liu, J., Wang, C., et al. (2020). Long noncoding RNA SNHG12 promotes tumour progression and sunitinib resistance by upregulating CDCA3 in renal cell carcinoma. *Cell Death Dis.* 11 (7), 515. doi:10.1038/s41419-020-2713-8

Lu, L., Liu, L. P., Zhao, Q. Q., Gui, R., and Zhao, Q. Y. (2021). Identification of a ferroptosis-related LncRNA signature as a novel prognosis model for lung adenocarcinoma. *Front. Oncol.* 11, 675545. doi:10.3389/fonc.2021.675545

Lu, Z., Tang, F., Li, Z., Lai, Y., Lu, Z., Zhang, J., et al. (2022). Prognosis risk model based on pyroptosis-related lncRNAs for bladder cancer. *Dis. Markers* 2022, 7931393. doi:10.1155/2022/7931393

Luo, H., Xu, C., Le, W., Ge, B., and Wang, T. (2019). lncRNA CASC11 promotes cancer cell proliferation in bladder cancer through miRNA-150. *J. Cell Biochem.* 120 (8), 13487–13493. doi:10.1002/jcb.28622

Luo, W., Wang, J., Xu, W., Ma, C., Wan, F., Huang, Y., et al. (2021). LncRNA RP11-89 facilitates tumorigenesis and ferroptosis resistance through PROM2-activated iron export by sponging miR-129-5p in bladder cancer. *Cell Death Dis.* 12 (11), 1043. doi:10.1038/s41419-021-04296-1)

Maeda, S., Sakai, K., Kaji, K., Iio, A., Nakazawa, M., Motegi, T., et al. (2022). Lapatinib as first-line treatment for muscle-invasive urothelial carcinoma in dogs. *Sci. Rep.* 12 (1), 4. doi:10.1038/s41598-021-04229-0

Malmstrom, P. U., Agrawal, S., Blackberg, M., Bostrom, P. J., Malavaud, B., Zaak, D., et al. (2017). Non-muscle-invasive bladder cancer: A vision for the future. *Scand. J. Urol.* 51 (2), 87–94. doi:10.1080/21681805.2017.1283359

Mo, X., Hu, D., Li, Y., Nai, A., Ma, F., Bashir, S., et al. (2022). A novel pyroptosis-related prognostic lncRNAs signature, tumor immune microenvironment and the associated regulation axes in bladder cancer. *Front. Genet.* 13, 936305. doi:10.3389/fgene.2022.936305

Patel, V. G., Oh, W. K., and Galsky, M. D. (2020). Treatment of muscle-invasive and advanced bladder cancer in 2020. *CA Cancer J. Clin.* 70 (5), 404–423. doi:10.3322/caac.21631

Peng, M., Huang, Y., Tao, T., Peng, C. Y., Su, Q., Xu, W., et al. (2016). Metformin and gefitinib cooperate to inhibit bladder cancer growth via both AMPK and EGFR pathways joining at Akt and Erk. *Sci. Rep.* 6, 28611. doi:10.1038/srep28611

Slovacek, H., Zhuo, J., and Taylor, J. M. (2021). Approaches to non-muscle-invasive bladder cancer. *Curr. Oncol. Rep.* 23 (9), 105. doi:10.1007/s11912-021-01091-1

Tang, D., Li, Y., Tang, Y., Zheng, H., Luo, W., Li, Y., et al. (2022). Recognition of glycometabolism-associated lncRNAs as prognosis markers for bladder cancer by an innovative prediction model. *Front. Genet.* 13, 918705. doi:10.3389/fgene.2022.918705

Tang, Y., Zhang, X., Qi, F., Chen, M., Li, Y., Liu, L., et al. (2015). Afatinib inhibits proliferation and invasion and promotes apoptosis of the T24 bladder cancer cell line. *Exp. Ther. Med.* 9 (5), 1851–1856. doi:10.3892/etm.2015.2314

Tong, H., Li, T., Gao, S., Yin, H., Cao, H., and He, W. (2021). An epithelial-mesenchymal transition-related long noncoding RNA signature correlates with the prognosis and progression in patients with bladder cancer. *Biosci. Rep.* 41 (1). doi:10.1042/BSR20203944)

Wu, H., Liu, B., Chen, Z., Li, G., and Zhang, Z. (2020). MSC-induced lncRNA HCP5 drove fatty acid oxidation through miR-3619-5p/AMPK/PGC1α/CEBPB axis to promote stemness and chemo-resistance of gastric cancer. *Cell Death Dis.* 11 (4), 233. doi:10.1038/s41419-020-2426-z

Xia, Y., He, X. T., Xu, X. Y., Tian, B. M., An, Y., and Chen, F. M. (2020). Exosomes derived from M0, M1 and M2 macrophages exert distinct influences on the proliferation and differentiation of mesenchymal stem cells. *PeerJ* 8, e8970. doi:10.7717/peerj.8970

XiaYu, K., GuoGang, C., CanXuan, L., and WeiBin, X. (2022). Comprehensive characterization of necroptosis-related lncRNAs in bladder cancer identifies a novel signature for prognosis prediction. *Dis. Markers* 2022, 2360299. doi:10.1155/2022/2360299

Yan, X., Du, G. W., Chen, Z., Liu, T. Z., and Li, S. (2021). CD86 molecule might Be a novel immune-related prognostic biomarker for patients with bladder cancer by bioinformatics and experimental assays. *Front. Oncol.* 11, 679851. doi:10.3389/fonc.2021.679851

Yang, Y., Wang, X., Bai, Y., Feng, D., Li, A., Tang, Y., et al. (2020). Programmed death-ligand 2 (PD-L2) expression in bladder cancer. *Urol. Oncol.* 38 (6), e9–e603. doi:10.1016/j.urolonc.2020.01.001

Zhang, H., Zhu, C., He, Z., Chen, S., Li, L., and Sun, C. (2020). LncRNA PSMB8-AS1 contributes to pancreatic cancer progression via modulating miR-382-3p/STAT1/PD-L1 axis. *J. Exp. Clin. Cancer Res.* 39 (1), 179. doi:10.1186/s13046-020-01687-8

Zheng, Z., Mao, S., Zhang, W., Liu, J., Li, C., Wang, R., et al. (2020). Dysregulation of the immune microenvironment contributes to malignant progression and has prognostic value in bladder cancer. *Front. Oncol.* 10, 542492. doi:10.3389/fonc.2020.542492

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 23 February 2023
DOI 10.3389/fgene.2023.1007135

OPEN ACCESS

EDITED BY
Yilin Zhang,
The University of Chicago, United States

REVIEWED BY
Silke Cameron,
University Medical Center Göttingen,
Germany
Maria Elena Turba,
Genefast srl, Italy

*CORRESPONDENCE
Haibo Qiu,
qiuhb@sysucc.org.cn

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

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 30 July 2022
ACCEPTED 10 February 2023
PUBLISHED 23 February 2023

CITATION
Guo J, Feng S, Yu H, Ou B, Jiang D, Zhuang W, Ding C, Chen X, Zhang M, Ling Y, Zeng Y and Qiu H (2023), Transcriptomic study of gastrointestinal stromal tumors with liver metastasis. *Front. Genet.* 14:1007135. doi: 10.3389/fgene.2023.1007135

Transcriptomic study of gastrointestinal stromal tumors with liver metastasis

Jianrong Guo[1†], Shoucheng Feng[1†], Hong Yu[1], Biyi Ou[1], Dan Jiang[1], Wei Zhuang[2], Chao Ding[1], Xiaojiang Chen[1], Miaoquan Zhang[1], Yudong Ling[1], Yi Zeng[1] and Haibo Qiu[1*]

[1]Department of Gastric Surgery, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-Sen University Cancer Center, Guangzhou, China, [2]Department of Pharmacy, Women and Children's Hospital, School of Medicine, Xiamen University, Xiamen, Fujian, China

Introduction: GIST (gastrointestinal stromal tumor) is the most prominent mesenchymal neoplasms of the gastrointestinal tract, and liver is the most common metastasis site for GIST. The molecular mechanism leading to liver metastasis of GIST is currently unclear.

Methods: With the goal of revealing the underlying mechanism, we performed whole-genome gene expression profiling on 18 pairs of RNA samples comprised of GIST tissues (with liver metastasis) and corresponding non-tumor tissues. After identifying differentially expressed gene, functional annotation and signal pathway analyses were conducted. GSE13861, datasets that compare GIST (without liver metastasis) with adjacent tissues, served as a comparison.

Results: A total of 492 up-regulated genes and 629 down-regulated genes were identified as differentially expressed genes between liver metastasis tissues and non-tumor tissues. We characterized expression patterns of DEGs identified from our cohort and GSE13861 that show signatures of enrichment for functionality. In subsequent gene set enrichment analysis, differentially expressed genes were mainly enriched in Epithelial Mesenchymal Transition in both datasets. 493 genes were overlapped among our whole-genome gene expression profiling results and GSE13861, consisting 188 up-regulated genes and 305 down-regulated genes. By using CytoHubba plugin of Cytoscape, CDH1, CD34, KIT, PROM1, SOX9, FGF2, CD24, ALDH1A1, JAG1 and NES were identified as top ten hub genes in tumorigenesis and liver metastasis of GIST. higher expression levels of FGF2, JAG1, CD34, ALDH1A1 and the lower expression level of CDH1 were respectively associated with unfavorable overall survival. Meanwhile higher expression levels of CD34, FGF2, KIT, JAG1, ALDH1A were correlated with worse disease-free survival.

Discussion: The present study may help to provide candidate pathways and targets for treatment of GIST and prevention methods to liver metastasis.

KEYWORDS

gastrointestinal stromal tumor, epithelial mesenchymal transition, liver metastasis, tumorigenesis, differentially expressed genes (DEG)

1 Introduction

GIST (gastrointestinal stromal tumor) is the most prominent mesenchymal neoplasms of the gastrointestinal tract, and their prevalence is on the rise (Corless et al., 2011). Activating mutations in the receptor tyrosine kinase encoding genes KIT (KIT proto-oncogene, receptor tyrosine kinase) or PDGFRA (platelet -derived growth factor receptor alpha) are extensively seen in GISTs (Serrano and George, 2020). These mutations cause constitutive activation of KIT or PDGFRA-mediated ligand independent activation and signaling (Joensuu et al., 2013). GISTs can appear everywhere in the gastrointestinal tract, although they're most prevalent in the stomach (50%–60%) and small intestine (30%–35%), with the colon and rectum (5%) and oesophagus (1%) (Joensuu et al., 2012). Liver metastasis (LM) from GIST is very common, and a primary tumor is diagnosed simultaneously in 15%–50% of cases. Furthermore, after excision of a high-risk GIST, up to 40%–80% of individuals may emerge with liver metastasis over a period of about 2 years (Ng et al., 1992; DeMatteo et al., 2000; DeMatteo et al., 2009). However, the mechanisms of GIST invasion and acquisition of the potential to metastasize are still unknown. Acquiring a better knowledge of the molecular process behind liver metastasis of GIST is crucial, as it might result in new anticancer treatment targets and greatly contribute to advances in diagnostic approaches.

Gene chip, also known as gene profile, is a gene detection method that has been used for over a decade. Gene chips can instantly identify all of the genes' expression information within the same sample time-point, making them ideal for detecting differentially expressed genes (DEGs) (Wang, 2000). Therefore, we collected GIST tissues of patients with liver metastasis and corresponding non-tumor tissues (stomach and intestinal tissue) yielding sufficient RNA for gene expression profiling. Meanwhile we also downloaded mRNA microarray data from the Gene Expression Omnibus (GEO) and jointly analyzed our gene expression profiling data with online data for identifying differentially expressed genes which may play an important role in tumorigenesis and liver metastasis of GIST. Gene Ontology (GO) annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were applied to further provide an overview of the function of the screened DEGs. Then a protein-protein interaction (PPI) network was constructed to determine the hub genes and survival analyses of the screened hub genes were carried out using Gene Expression Profiling Interactive Analysis (GEPIA).

In this study, we first performed gene chip detection on GIST tumor sample and peri cancerous tissues of 18 GIST patients with liver metastases, obtained microarray dataset, and obtained organized microarray dataset of GIST with no liver metastasis and paracancer tissues from the GEO database. Differentially expressed genes were analyzed separately, and the enrichment of DEGs in the two datasets were analyzed. The STRING website and Cytoscape software were used to find out the key genes that promote the tumorigenesis and liver metastasis of GIST. Finally, we explored the potential of these key genes as prognostic markers of gastrointestinal tumors using Kaplan–Meier Survival analyses. This study helps us better understand the molecular mechanism of GSIT tumorigenesis and liver metastasis.

2 Materials and methods

2.1 Clinical samples

GIST tissues of patients with liver metastasis and corresponding non-tumor tissue (stomach and intestinal tissue) samples were obtained from Sun Yat-sen University Cancer Center under protocols approved by the institutional review board at Sun Yat-sen University Cancer Center. Written informed consent was obtained from all patients enrolled in the study. All experiments using clinical samples were carried out in accordance with the approved guidelines.

2.2 Microarray analysis

All samples were frozen in liquid nitrogen at −80°C. The total RNA of samples was extracted by TRIZOL method, and the total RNA was examined by NanoDrop 2000 and Agilent Bioanalyzer 2100. The qualified sample goes into the chip experiment. The standards of quality control are: Thermo NanoDrop 2000:1.7 < A260/A280 < 2.2; Agilent 2100 Bioanalyzer: RIN ≥ 7.0 and 28S/18S > 0.7. Affymetrix GeneChip Human Primeview array (Affymetrix, Santa Clara, CA, United States) was used to analyze global expression pattern of 28,869 well-annotated genes. RNA samples were amplified and labeled using the 3′IVT Expression Kit and GeneChip WT Terminal Labeling and Control Kit from Affymetrix. Affymetrix's GeneChip Fluidics Station 450 was used to carry out the normal washing treatment after the samples were hybridized at 45°C for 16 h. The arrays were then scanned using the GeneChip Scanner 7G procedure. Quantile normalization of gene expression was performed using the normalizeBetweenArrays function in limma.

We also downloaded the following gene expression profiles from the GEO: GSE13861 (including six GIST and 19 surrounding normal fresh frozen tissues) (Cho et al., 2011) for further analysis.

2.3 DEG identification

R language limma package was used to identify DEGs in our cohort and GSE13861 separately. The log-fold change (FC) in expression and adjusted *p*-values (adj. P) were determined. The adj. P using the Benjamini–Hochberg method with default values were applied to correct the potential false-positive results. DEGs were defined as genes that satisfied the specified cutoff criterion of adj. $p > 0.05$ and $|\log FC| > 2.0$. The Venn diagram online tool was used to look at the intersecting genes. In order to illustrate the volcano plot of DEGs, visual hierarchical cluster analysis was also carried out.

2.4 GO annotation and KEGG pathway enrichment analyses of DEGs

To reveal the functions of DEGs, GO annotation and KEGG pathway enrichment analyses were conducted. Biological process (BP), cellular component (CC), and molecular function (MF) were

TABLE 1 Details of mutations, clinical features for the 18 GIST patients with liver metastasis.

No.	Age	Sex	Site	Size (cm)	Mitotic index	Grade	Metastasis	Mutation
1	48	M	small intestine	4	200	high	liver	K11
2	71	F	stomach	7.8	50	high	liver	K11
3	58	M	small intestine	6	90	high	liver	K11
4	63	F	stomach	10.3	10	high	liver	K11
5	59	M	small intestine	8	55	high	liver	K11
6	57	M	stomach	7	>5	high	liver	K11
7	23	M	stomach	4.3	15	high	liver	K11
8	54	F	small intestine	4.7	4	low	liver	K11
9	57	F	stomach	3.9	>30	high	liver	K11
10	50	F	stomach	2.5	6	medium	liver	K11
11	60	M	stomach	3.7	6	medium	liver	K11
12	48	F	stomach	4.5	15	high	liver	K11
13	33	F	stomach	6	4	medium	liver	K11
14	57	F	stomach	2.3	<3	low	liver	K11
15	59	F	stomach	4.9	14	high	liver	K11
16	59	M	stomach	5	9	medium	liver	K11
17	68	F	stomach	2.6	20	high	liver	K11
18	52	M	stomach	7.5	>10	high	liver	K11

the three categories that made up the GO terms. Statistical significance was determined to be adj. $p < 0.05$. Resulting p-values are adjusted for multiple testing using the "Benjamini–Hochberg" method.

2.5 Gene set enrichment analysis (GSEA)

To find out the different mechanisms between GIST with liver metastasis and GIST without metastasis, GSEA (Version: 3.0; http://software.broadinstitute.org/gsea/index.jsp) was performed (Subramanian et al., 2005). The threshold was set at $p < 0.05$.

2.6 Construction of PPI network and screening of hub genes

A database called Search Tool for the Retrieval of Interacting Genes (STRING) is used to study the functional protein association networks (Szklarczyk et al., 2017). The filtered DEGs had already been added to the STRING database. All PPI pairs with a cumulative score greater than 0.4 were retrieved. High-degree nodes seem to be essential for maintaining the network's overall stability. The degree of all nodes was calculated by Cytoscape (v3.6.1) plugin cytoHubba using the MCC algorithm (Chin et al., 2014), in this experiment, the genes with the top 10 highest MCC score values were considered as hub genes.

2.7 Kaplan–meier survival analyses of the hub genes

Survival analysis of hub genes was based on Kaplan–Meier Survival analyses, using GEPIA (http://gepia.cancer-pku.cn/) tool. According to the expression of each hub gene, the cancer patients were divided into low or high expression group based on the median mRNA expression of hub genes, at statistical significance of $p < 0.05$.

3 Result

3.1 Characteristics of GIST patients with liver metastasis in our cohort

Our cohort consisting of 18 paired GIST tissues of patients with liver metastasis (LM) and corresponding non-tumor tissue (NT) samples. Details of mutations, clinical features for the 18 GIST patients with liver metastasis are presented in Table 1. Eight of the 18 patients were male and 10 were female. The youngest patient was 23 and the oldest was 71. Four of the 18 GISTs are small-intestine GISTs, and the remaining 14 are stomach GISTs. All patients presented with liver metastases. And all of the patients harbored a single non-synonymous mutation in KIT (Kit exon 11). The tumor size, mitotic index and location of primary tumors are demonstrated in Table 1.

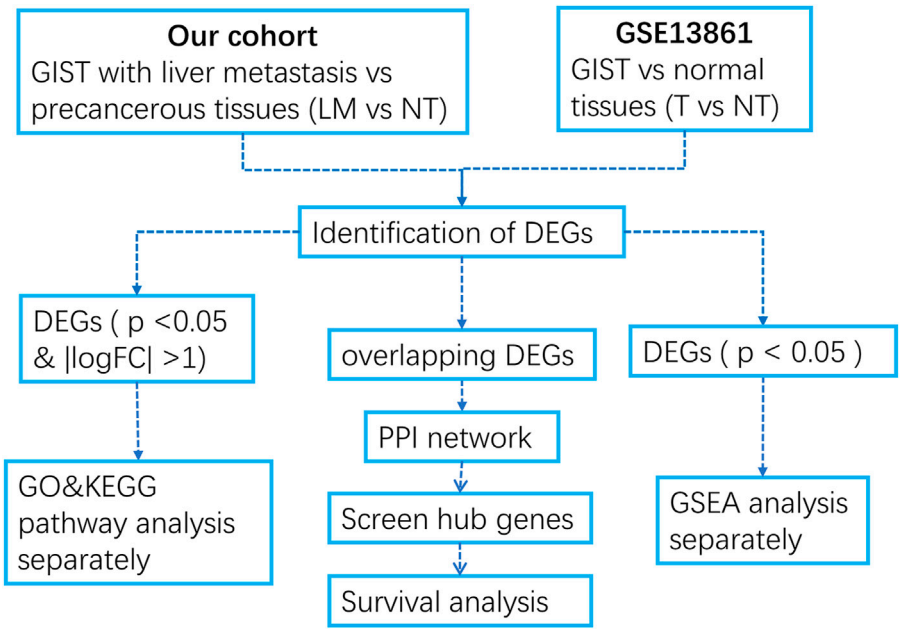


FIGURE 1
Flow diagram of the data collection and method implementation in this work.

3.2 Identification of differentially expressed genes (DEGs)

We developed a flow diagram to show our process (Figure 1). To characterize the tumor biology of GIST with liver metastasis, we performed whole-genome gene expression profiling in 18 pairs of RNA samples comprised of GIST with LM and NT tissues. 1121 genes were found to differentially express between LM and adjacent tissues, including 492 upregulated genes and 629 downregulated genes (Supplementary Data Sheet S1). Volcano map of DEGs was shown in Figure 2A. Subsequently, heatmap of DEGs was created, in which the mRNA expression profiles of LM and NT resulted in obviously separate clusters (Figure 2B). Principle Component Analysis (PCA) and hierarchical cluster analysis results were demonstrated in Figures 2C, D. GSE13861 (including 6 GIST and 19 surrounding normal fresh frozen tissues) is a dataset that compare GIST without liver metastasis with adjacent tissues, which serves as a comparison. DEGs in GSE13861 were calculated according to the criteria of $p < 0.05$ and |logFC| >2.0. 924 genes were found to differentially express between GIST and adjacent tissues, including 313 upregulated genes and 611 downregulated genes (Supplementary Data Sheet S2). Volcano map of DEGs is shown in Supplementary Figure S1A. Hierarchical clustering heatmap of DEGs was shown in Supplementary Figures S1B, C Shows PCA results of GSE13861. Hierarchical cluster analysis was visualized and important details were demonstrated in Supplementary Figure S1D.

3.3 GO and KEGG analysis of DEGs reveal the different enrich patterns of GIST with LM and GIST without LM

To characterize the biological mechanism of GIST liver metastasis, gene enrichment analysis including Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were conducted. DEGs acquired from the two datasets were subjected to enrichment separately. For GO biological process (BP), DEGs in our cohort were mainly enriched in *cell junction assembly*, *cell-substrate adhesion* and *urogenital development*, while DEGs in GSE13861 were mainly enriched in *extracellular matrix organization*, *extracellular structure organization* and *external encapsulating structure organization*. In terms of cellular component (CC), DEGs in our cohort were mainly enriched in *collagen-containing extracellular matrix*, *cell-cell junction* and *apical part of cell*. The CC enrichment results of GSE13861 were very similar to our cohort. For GO molecular function (MF), results were also similar between these two cohorts (Figures 3A, B). We further explored the function significance of these DEGs using KEGG pathway analysis. DEGs in our cohort were mainly enriched in *PI3K-Akt signaling pathway* and *Tight junction*, while DEGs in GSE13861 were mainly enriched in *Fluid shear stress and atherosclerosis* and *Metabolism of xenobiotics by cytochrome P450* (Figures 3C, D). Changes in gene expression in *PI3K-Akt signaling pathway* and *Tight junction* signaling pathways in our cohort are depicted in detail in Figures 4A, B.

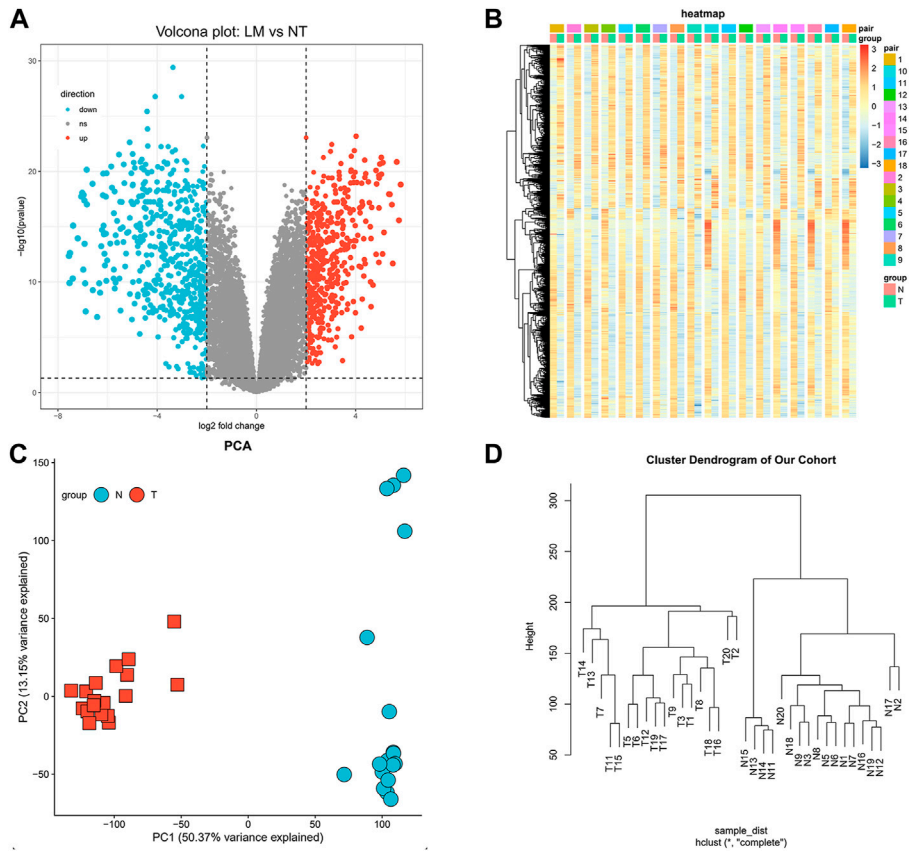


FIGURE 2
Identification of differentially expressed genes. **(A)** Volcano map of differentially expressed genes (Upregulated genes in red, downregulated genes in blue). **(B)** Hierarchical clustering heatmap of DEGs screened on the basis of FC > 2.0 and a corrected p-value < 0.05. **(C)** Shows PCA results of our cohort. **(D)** Visual hierarchical cluster analysis.

3.4 Gene set enrichment analysis reveal the differences between GIST with LM and GIST without LM

GSEA was performed to identify the gene sets that were statistically different between the normal controls and GIST group (Taking $p < 0.05$ as the boundary value). The results illustrated that *Epithelial mesenchymal transition* (EMT) was the most significantly upregulated pathway in both cohorts (Figures 5A–C, E). DEGs in our cohort were also positively correlated and significantly enriched in *IL2 Stat5 Signaling* (Figures 5A, D, NES = 1.767 & P.adj <0.001). While in GSE13861, *IL2 Stat5 Signaling* was not in the top10-enriched pathways (Figure 5B).

3.5 PPI network construction and hub genes selection and analysis

To identify those genes which play significant roles in both tumorigenesis and liver metastasis of GIST, GSE13861 dataset containing GIST primary tumor tissues (PT) and corresponding non-tumor tissues (NT) was co-analyzed. The Venn diagram (Figure 7A) illustrated a total of 493 genes overlapped among our microarray results and GSE13861, consisting 188 upregulated genes and 305 downregulated genes (Supplementary Data Sheet S3). Using the STRING and Cytoscape databases, a PPI network of potential interactions between overlapping genes was constructed (Figure 6). The hub genes were selected from the PPI network using the MCC algorithm of CytoHubba plugin. According to the MCC scores, the top ten highest-scored genes included CDH1, CD34, KIT, PROM1, SOX9, FGF2, CD24, ALDH1A1, JAG1, and NES (Figure 7B and Supplementary Data Sheet S4). The abbreviations, names, and functions of these genes are displayed in Table 2. The function of these hub genes was analyzed by Metascape, in which as expected, these genes were mainly enriched in pathways in cell-cell adhesion (Figure 7C).

3.6 Validation and prognostic value of hub genes

Among above mentioned 10 hub genes, the expressions of CD34, KIT, PROM1, NES, and FGF2 respectively were higher in GIST (with LM) tissues (Figures 8A–E) compared to NT tissues (p-values all <0.001). Meanwhile reverse trend was found for the

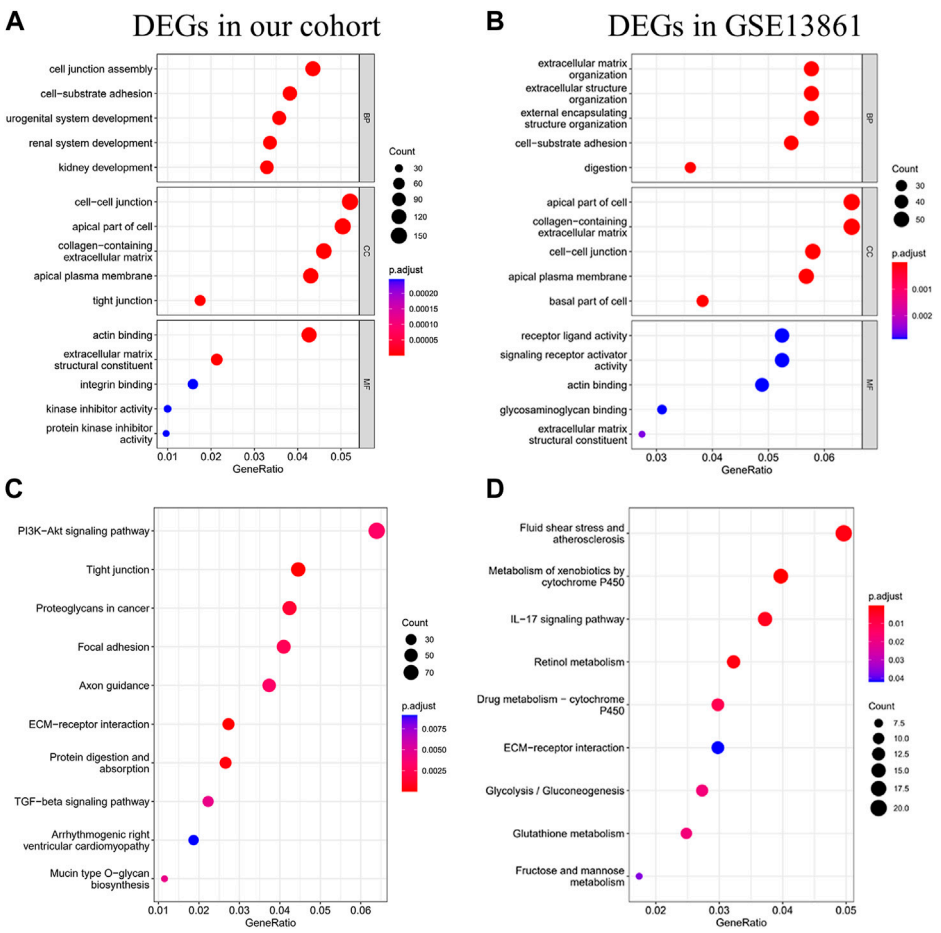


FIGURE 3
GO and KEGG analysis of DEGs. GO analysis **(A)**, and KEGG analysis **(C)** of DEGs in our cohort. GO analysis **(B)**, and KEGG analysis **(D)** of DEGs in GSE13861 dataset.

expressions of the rest hub genes CDH1, SOX9, CD24, ALDH1A1, and JAG1 (Figures 8E–J, *p*-values all <0.001). These results are nearly identical to the findings from the GES13861 dataset (Figures 8K–T). Prognostic significance of hub genes was investigated in several types of gastrointestinal tumors including stomach adenocarcinoma, colon adenocarcinoma, esophageal carcinoma and rectal adenocarcinoma by the GEPIA database. The Kaplan-Meier analyses suggested that higher expression levels of FGF2, JAG1, CD34, and ALDH1A1 and the lower expression level of CDH1 were respectively associated with worse overall survival (OS) (Figure 9). Meanwhile higher expression levels of CD34, FGF2, KIT, JAG1, and ALDH1A were correlated with worse disease-free survival (DFS) (Figure 10).

4 Discussion

During the past decade, GIST has become the prominent focus of molecularly targeted therapy for solid tumors (Poveda et al., 2017; Hemming et al., 2018). GIST are more prevalent than previously thought, according to population-based studies (Corless and Heinrich, 2008). The incidence of GIST was found to be 14.5 per million population, with the highest frequency being observed in older individuals and there was no gender difference (Gold and DeMatteo, 2006). The hallmarks of cancer consist of six biological traits: sustaining proliferative signaling, evading growth suppressors, evasion of apoptosis, limitless replicative potential, inducing angiogenesis, and ability to invade and metastasize (Hanahan and Weinberg, 2011). It is worth noting that the last characteristic, invasion and metastasis is vital for progressive nature of cancer. Many malignancies favor certain organs as metastatic sites, including the lungs, bone marrow, and liver. Liver metastases are a major cause of death in patients with colorectal cancer. The liver environment, which includes ECM and stromal cells, may encourage metastatic colonization. Metastatic colorectal cancer cell lines responded more favorably to ECM derived from primary rat hepatocytes than to ECM from fetal rat fibroblast cultures (Zvibel et al., 1998). The D6.1A tetraspanin, a cell-surface organizer, interacted with the 64 integrin and enhanced liver colonization by pancreatic cancer cells injected intraperitoneally (Herlevsen et al., 2003).

Patients with GIST have a high risk of recurrence (about 55–72 percent) and a dismal survival rate due to malignant cells preferentially metastasizing to liver tissue (DeMatteo et al., 2000; Bayraktar et al., 2010). Cho et al. discovered that Compared to KIT

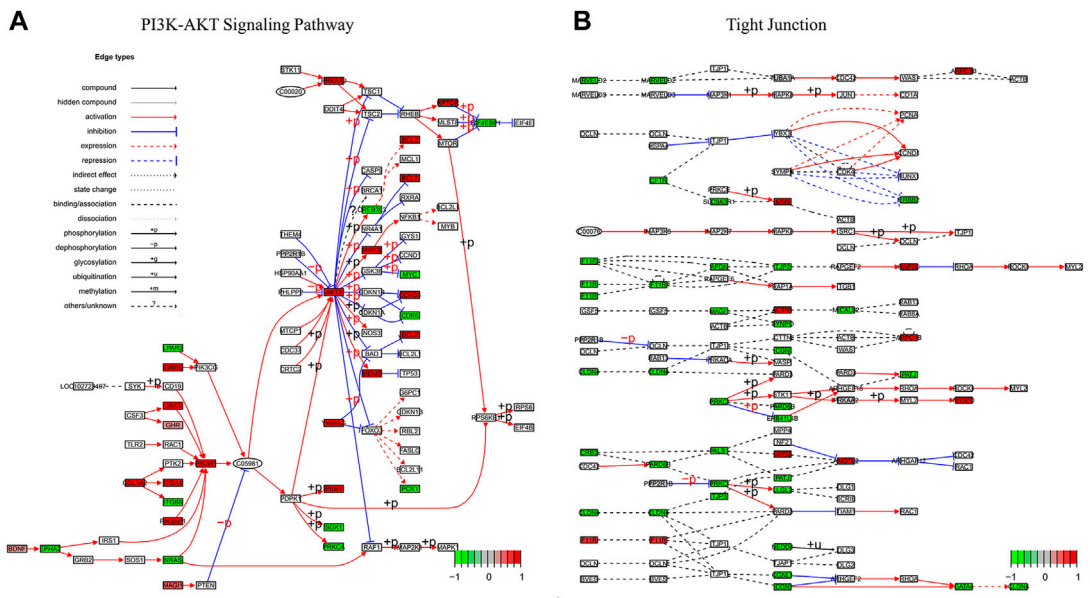


FIGURE 4
Pathview map of **(A)** PI3K-AKT Signaling Pathway (map 04,151) and **(B)** Tight Junction (map04530) using data of our cohort. Upregulated genes in red, downregulated genes in green.

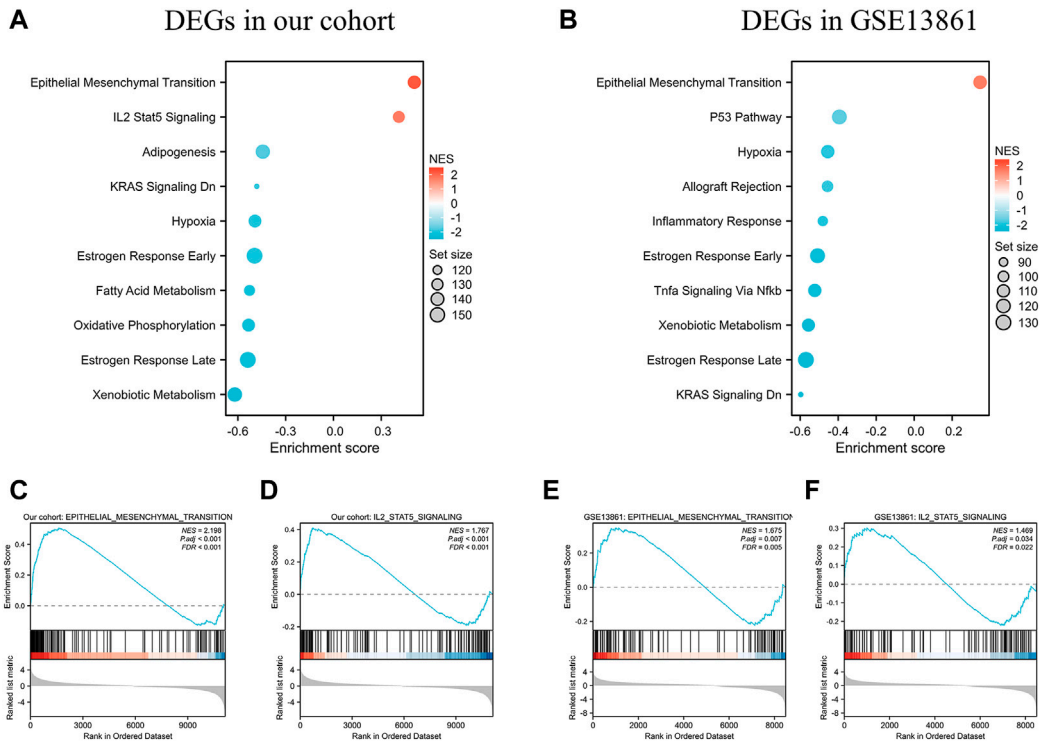


FIGURE 5
GSEA analysis of DEGs in the data sets. **(A)** The top 10 enriched KEGG items for the DEGs in our cohort, and **(B)** DEGs in GSE13861 dataset. Taking $p < 0.05$ as the boundary value. Significant enrichment of the Epithelial Mesenchymal Transition **(C)** and IL-2 STAT5 Signaling **(D)** with DEGs in our cohort. Significant enrichment of the Epithelial Mesenchymal Transition **(E)** and IL-2 STAT5 Signaling **(F)** with DEGs in GSE13861.

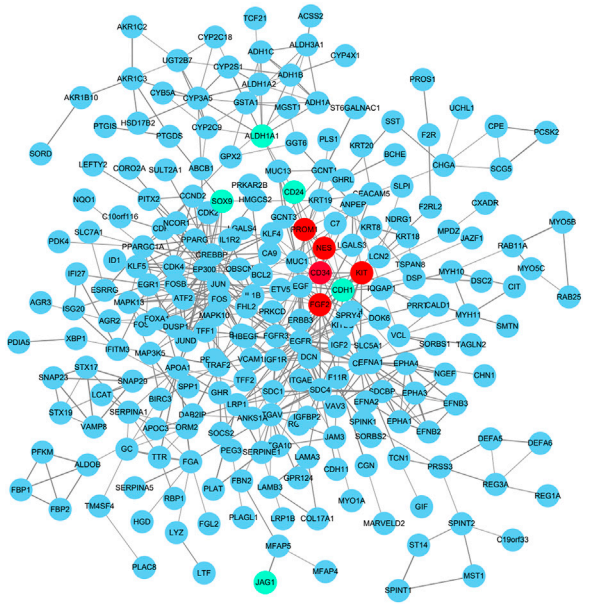


FIGURE 6
Protein–protein interaction network of 188 upregulated genes and 305 downregulated genes were analyzed using Cytoscape software. The edges between 2 nodes represent the gene-gene interactions. Upregulated hub genes in red, downregulated hub genes in teal.

mutation-negative GISTs, KIT mutation-positive GISTs had more frequent liver metastases and worse mortality (Cho et al., 2006). Wang et al. reported that the KIT exon 11,557-558 deletion upregulates CXCR4 by increasing ETV1 binding to the CXCR4 promoter in GIST cells, which in turn encourages liver metastasis (Wang et al., 2016). As such, to better understand GIST biological behavior and inform the development of treatment strategies, it is critical to identify the significant genes that regulate the liver metastasis of GIST. Advances in bioinformatics have been conducive to identify molecular targets that indicate the progression of GIST (Amirnasr et al., 2019; Ohshima et al., 2019).

In this study, a total of 492 upregulated genes and 629 downregulated genes were identified in GIST with LM compared to corresponding NT. Function annotation based on GO and KEGG analyses demonstrated that DEGs were mainly enriched in *cell junction assembly*, *tight junction*, *actin binding* and *PI3K-Akt signaling pathway*. GSEA results indicated that *IL-2 STAT5 Signaling* may be a vital pathway which promotes liver metastasis of GIST. Meanwhile, *EMT signal pathway* is the most significant and positive enriched pathway in both our cohort and GSE13861, which indicated that EMT may play a significant role in tumorigenesis and liver metastasis of GIST. Furthermore, to identify genes which play essential roles in both tumorigenesis and liver metastasis of GIST, our data and GSE13861 dataset were co-analyzed. A totally of 493 genes overlapped among our microarray results and GSE13861, including 188 upregulated genes and 305 downregulated genes. Then a PPI network of putative interactions between overlapping genes was

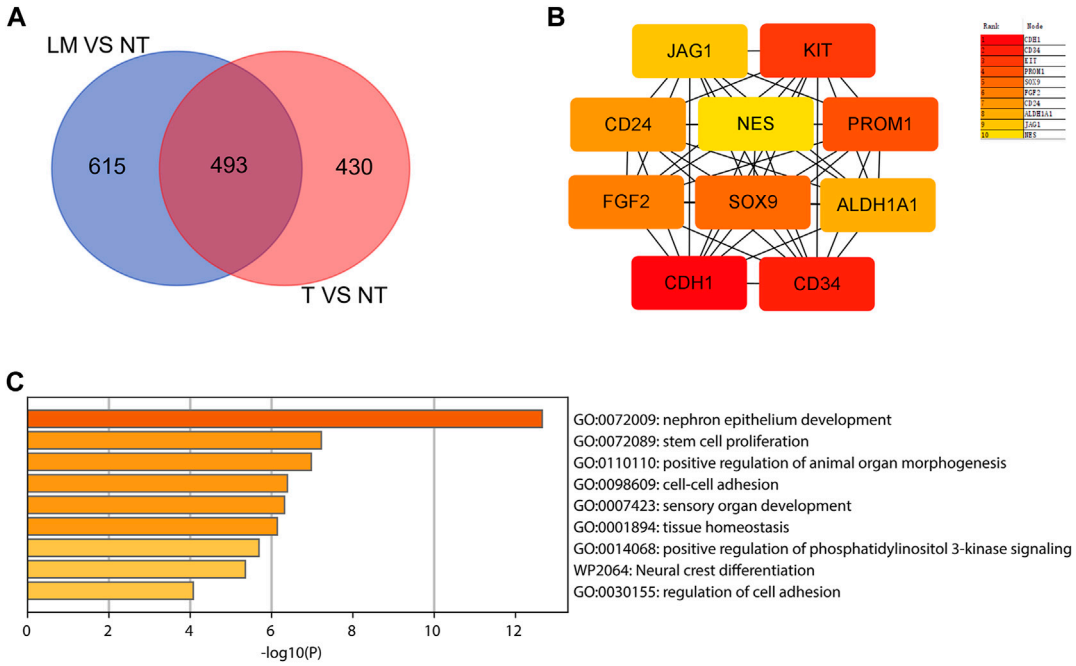


FIGURE 7
Hub genes selection and analysis. **(A)** Venn diagram shows the 493 overlapping DEGs. **(B)** The top 10 hub genes in the PPI network were screened by Cytoscape plugin cytoHubba. The 10 identified hub genes such as CDH1, CD34, KIT, PROM1, SOX9, FGF2, CD24, ALDH1A1, JAG1, NES are displayed from red (high degree value) to yellow (low degree value). **(C)** GO and KEGG pathway enrichment analysis of the 10 hub genes.

TABLE 2 Details of hub genes.

Gene symbol	Degree	Full name	Gene function
CDH1	63	Epithelial cadherin	Loss of CDH1 is thought to contribute to progression in cancer by increasing proliferation, invasion, and/or metastasis
CD34	40	CD34	CD34 is a cell surface glycoprotein and function as a cell-cell adhesion factor.
KIT	39	KIT proto-oncogene receptor tyrosine kinase	Mutations in this gene are associated with gastrointestinal stromal tumors, mast cell disease, acute myelogenous leukemia, and piebaldism.
PROM1	37	prominin-1	PROM1 is often expressed on adult stem cells, where it is thought to function in maintaining stem cell properties by suppressing differentiation.
SOX9	37	SRY-box transcription factor 9	SOX-9 plays a pivotal role in male sexual development; by working with Sf1, SOX-9 can produce AMH in Sertoli cells to inhibit the creation of a female reproductive system.
FGF2	36	fibroblast growth factor 2	FGF2 is involved in a variety of biological processes, including cell growth, morphogenesis, tissue repair, tumor growth and invasion.
CD24	34	CD24	CD24 is overexpressed in many cancers and some cancer stem cells and is associated with the development, invasion, and metastasis of cancer cells.
ALDH1A1	32	aldehyde dehydrogenase 1 family member A1	High ALDH1A1 activity is closely related to stemness phenotype of several tumors, possibly contributing to cancer progression and diffusion in the body.
JAG1	31	jagged canonical Notch ligand 1	JAG1/Notch signaling cascades activate a number of oncogenic factors that regulate cellular functions such as proliferation, metastasis, drug-resistance, and angiogenesis.
NES	30	Nestin	Nestin may be a marker for newly synthesized tumor vessels and a therapeutic target for tumor angiogenesis.

created using the STRING and Cytoscape databases and hub genes were selected from the PPI network using the MCC algorithm of CytoHubba plugin. According to the MCC scores, the top ten highest-scored genes were CDH1, CD34, KIT, PROM1, SOX9, FGF2, CD24, ALDH1A1, JAG1, and NES.

The phosphatidylinositol PI3K/AKT/mTOR pathway is a critical survival pathway for cell proliferation, apoptosis, autophagy and translation in neoplasms (Patel, 2013). Constitutive autophosphorylation of RTKs has an impact on the activation of the PI3K/AKT/mTOR pathway (Vara et al., 2004; Fruman and Rommel, 2014). In several preclinical and early-stage clinical trials PI3K/AKT/mTOR signaling inhibition has been considered as a promising targeted therapy strategy for GISTs (Duan et al., 2020). Our results suggest that, unlike GIST, liver-metastatic GIST has more genes enriched in the *PI3K-Akt signaling pathway*. We hypothesized that *PI3K-Akt signaling pathway* is an important pathway to promote liver metastasis of GIST. It can be used as a target to prevent and treat liver metastasis of GIST.

Tight junction is the most talked-about structure in epithelial and endothelial cells because they control permeability (Jiang et al., 1999; Tsukita et al., 1999). It is an area where neighboring cells' plasma membranes make a sequence of connections that appear to totally obstruct the extracellular space, forming an intercellular barrier and intramembrane diffusion fence (Wong and Gumbiner, 1997). The majority of malignancies are characterized by abnormal growth control, tissue architecture loss, and loss of differentiation. The feature that cancer cells' mutual adhesiveness is much less than that of normal cells is a key characteristic of cancer cells (Martin and Jiang, 2009). Reduced cell-cell interaction leads cancer cells to rebel against the social order, resulting in the breakdown of overall tissue architecture, a morphological hallmark of malignancy. The loss cell-cell junction and tight junction are changes associated with cancer progression to an invasive, metastatic state (Thomson et al., 2011).

The cytokine interleukin-2 (IL-2) was first discovered in 1976 as a T cell growth factor (Morgan et al., 1976). While IL-2 has been shown to activate several STAT family members, including STAT1, STAT3, and STAT5, STAT5 is the predominant IL-2 signaling molecule (Hou et al., 1995; Lin et al., 1995). Indeed, IL-2 has also been shown to signal *via* the Mitogen Activated Protein Kinase (MAPK) pathway, *via* extracellular signal-regulated kinase (ERK), as well as the PI3K pathway (González-García et al., 1997; Liao et al., 2013; Ross and Cantrell, 2018). In this study, we identified *IL-2 STAT5 Signaling* is the second and positively enriched pathway using GSEA in DEGs in our cohort, while in GSE13861, *IL-2 Stat5 Signaling* was not in the top10-enriched pathways. This result indicates that *IL-2 STAT5 Signaling* may be a vital pathway which promotes liver metastasis of GIST.

The extracellular matrix (ECM) performs many functions in addition to its structural role; as a major component of the cellular microenvironment it influences cell behaviors such as proliferation, adhesion and migration, and regulates cell differentiation and death (Hynes, 2009). Abnormal ECM dynamics can result in uncontrolled cell proliferation and invasion, failure of cell death, and loss of cell differentiation, which can lead in congenital abnormalities and pathological processes such as tissue fibrosis and cancer. As the ECM's significance in tumor progression becomes more evident, cancer treatment strategies have started to focus on specific ECM components in an effort to reduce metastasis (Walker et al., 2018; Paolillo and Schinelli, 2019; Girigoswami et al., 2021).

Epithelial mesenchymal transition (EMT) is a crucial developmental process that triggers the transdifferentiation of polarized epithelial cells into mesenchymal cells during tumor invasion and metastasis (Kalluri and Weinberg, 2009; Polyak and Weinberg, 2009). Cancer cells acquire invasive and metastatic characteristics with activation of EMT, which facilitates effective

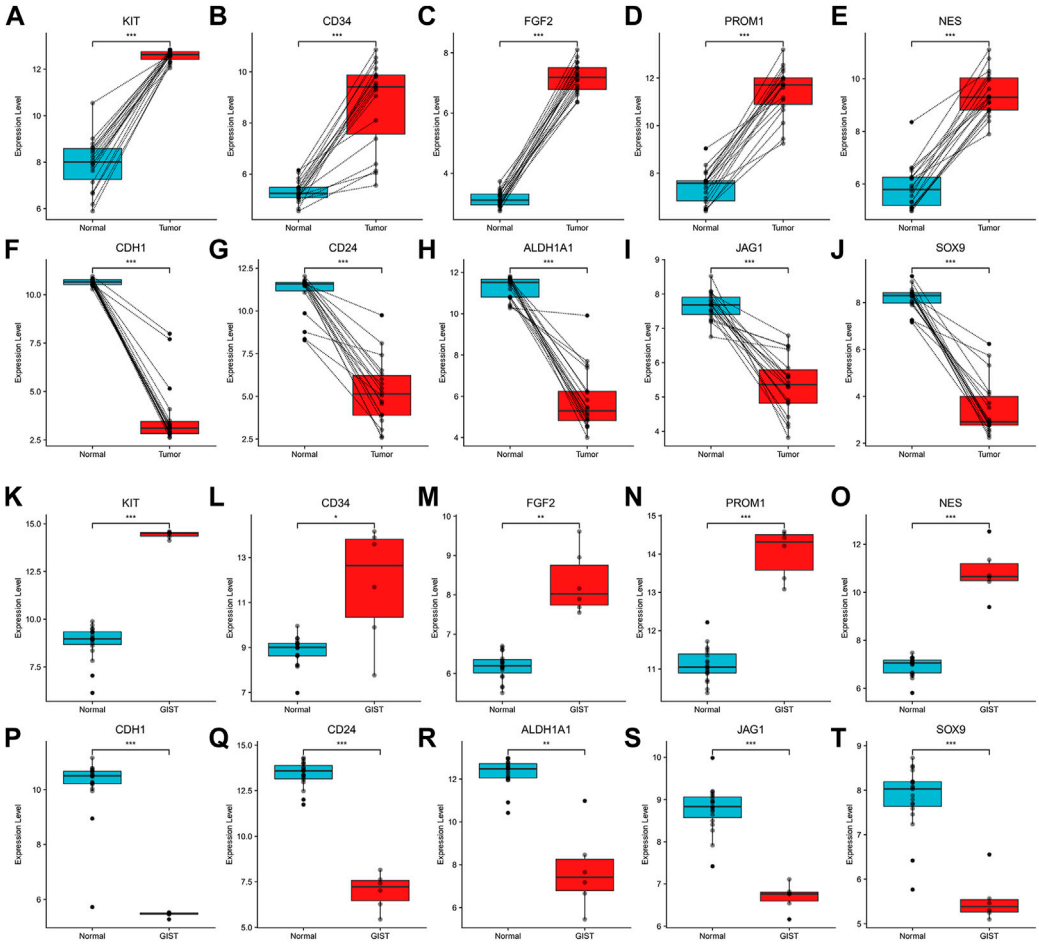


FIGURE 8
Expression of hub genes. **(A–J)**. Expression of KIT, CD34, FGF2, PROM1, NES, CDH1, CD24, ALDH1A1, JAG1 and SOX9 in Our cohort and **(K–T)** GSE13861. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

colonization of distal target organs (Tsai and Yang, 2013). In line with previous study, we found that EMT signal pathway enriched in GIST tissues of patients with liver metastasis compared to corresponding pericancerous tissues, which indicated that EMT may play a significant role in liver metastasis of GIST.

E-cadherin (also known as cadherin-1 or CDH1), a protein belonging to the cadherin family, is possibly one of the most potent and extensively researched regulators of adhesion. Together with associated Catenins, E-cadherin is essential for regulating cell adhesion, signaling and transcription in cancers and controlling metastatic progression (Jiang and Mansel, 2000). Alteration in cell adhesion molecules (CAMs), such as E-cadherin affect the processes of cell-cell adhesion and cell-matrix adhesion and subsequently their metastatic potential. It also regulates the cell cycle regulators p27kip1 and p57kip2, which are essential for cell-cell contact inhibition in healthy tissue but are lost or disrupted in cancer cells, primarily due to the loss of E-cadherin in cancer cells (Croix et al., 1998; Cavallaro and Christofori, 2004a; Migita et al., 2008). Therefore, decreased cell-cell adhesion not only increases the potential for metastatic dissemination of cancer cells, but also encourages unchecked cell proliferation through the absence of contact inhibition (Cavallaro and Christofori, 2004b). Indeed, studies has shown a correlation between reduced E-cadherin and α-catenin expression with increased tumor cell invasiveness (Zschiesche et al., 1997). Sheng Liu et al. demonstrated that reduced E-cadherin expression was correlated with distant metastasis of GIST and E-cadherin was thus considered as risk factor for GIST metastasis. In our study, E-cadherin had been identified as the top hub gene and to be involved in the process of tumorigenesis and liver metastasis of GIST. The results of our study demonstrated decreased expression levels of E-cadherin were associated with unfavorable OS in gastrointestinal tumors. Therefore, we believe that it mediates the liver metastasis of GIST and can be used as a target for the treatment of metastatic GIST.

ETV1, a transcription factor from the ETS family, is a master regulator of the normal lineage specification and development of the ICCs which are the precursors to GIST (Chi et al., 2010). Hao-Chen Wang et al. reported that upregulating ETV1 expression induced CXCR4 expression, which promoted liver metastasis of GIST (Wang et al., 2016). We compared ETV1 expression in our cohort and found that ETV1 are upregulated in GIST tissues of patients with liver metastasis compared with corresponding non-tumor tissue (Supplementary Figures S2A, C). Our result supports ETV1's

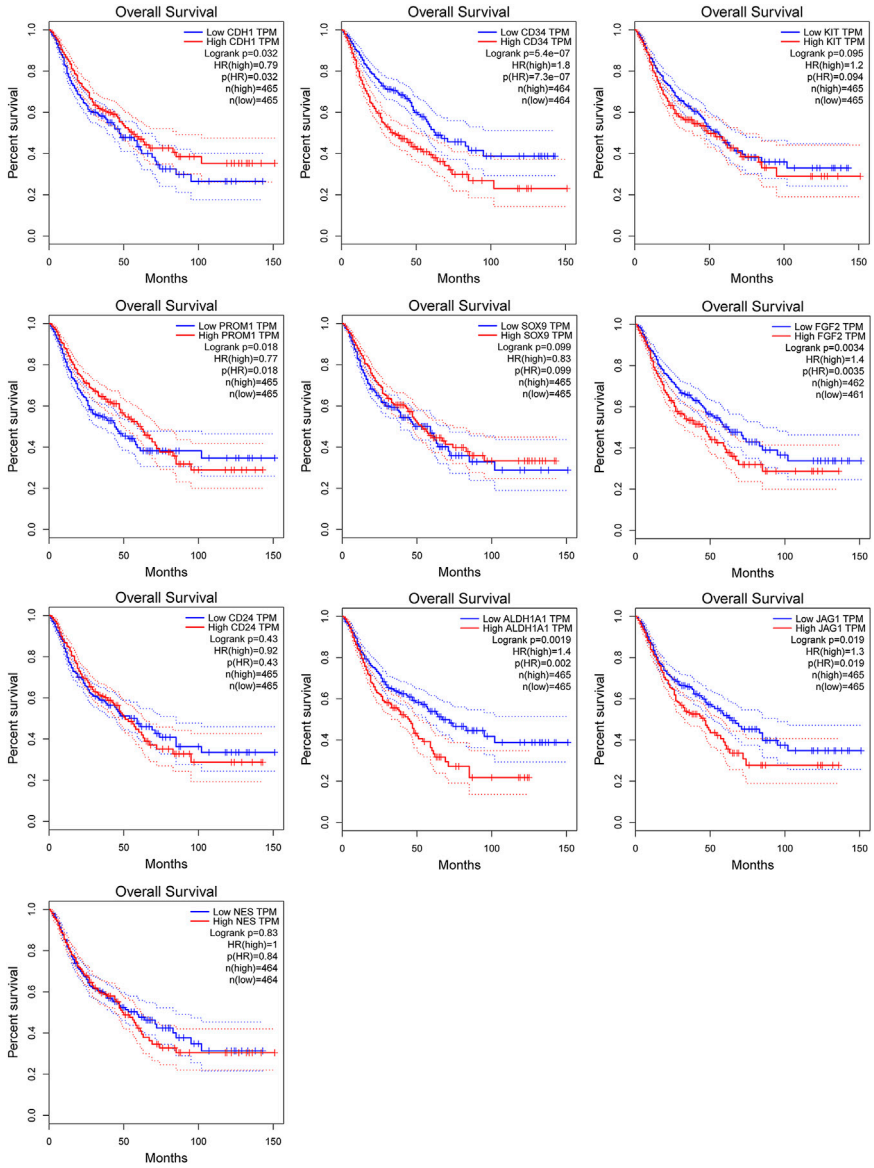


FIGURE 9
Kaplan-Meier curves of hub genes expression and overall survival in gastrointestinal tumors. Data are presented as the hazard ratio with a 95% confidence interval. Log-rank $p < 0.05$ was regarded as statistically significant.

stimulative role in liver metastasis of GIST. Besides, it has been demonstrated that ETV4 expression impacted Wnt/catenin signaling and was correlated to an aggressive phenotype in GIST (Zeng et al., 2017). However, our results showed no significant difference in ETV1 expression levels in GIST compared to the adjacent tissues in both our cohort and GSE13861 (Supplementary Figures S2B, D). Further research in this area is needed.

The major limitation of the present study is that Tumor transcriptome programs are rather diverse, both within tumor cells due to somatic genetic changes and within tumor microenvironments due to extensive infiltration of the stroma and other cell types in the tumor. An average gene expression profile from microarray can mask the real signals causing the liver metastasis of GIST from a rare cell population or cell type. Besides, it has been indicated that long non-coding RNAs (LncRNAs) participate in certain pro-metastatic stages, such as the epithelial mesenchymal transition, invasion and migration, and organotrophic colonization, and they also have an impact on the metastatic tumor microenvironment (Amirnasr et al., 2020; Liu et al., 2021). The gene chips we used in current study only contain probes for protein-coding mRNAs but not LncRNAs. Thus, further researches should be conducted to elucidate the potential function of LncRNAs in liver metastasis of GIST. Moreover, a direct comparison of liver metastases and primary sites of GIST maybe a better study protocol. But, on one hand, liver metastases from GIST patients are difficult to obtain because they are usually treated by ablation. On the other hand, we think that the transcription level of GIST with liver metastasis has already changed before metastasis, the potential role of these genes in promoting liver metastasis cannot be ignored. This information is lost if direct compare liver metastases samples and primary lesions. It would be better if we collected GIST

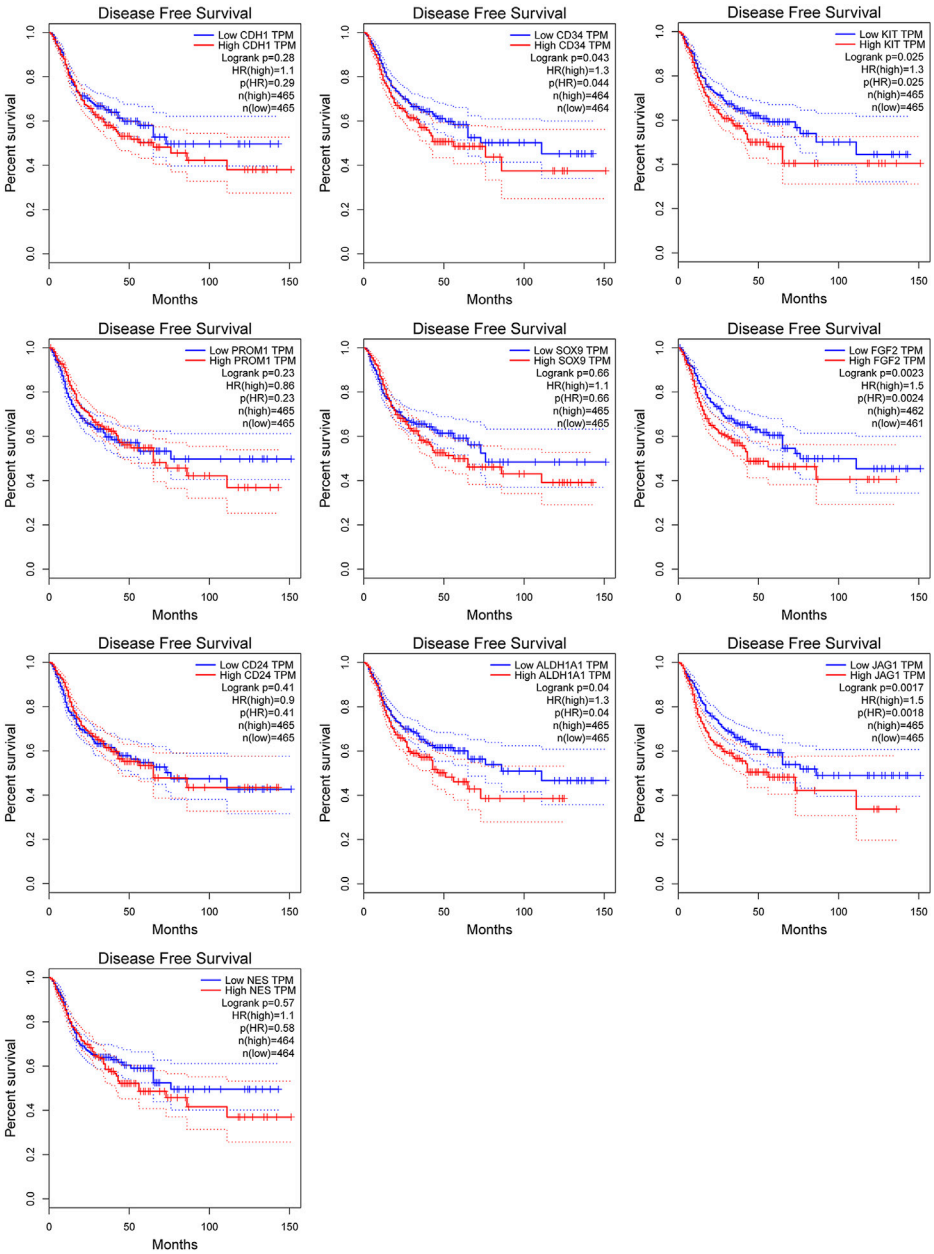


FIGURE 10
Kaplan-Meier curves of hub genes expression and disease-free survival in gastrointestinal tumors. Data are presented as the hazard ratio with a 95% confidence interval. Log-rank $p < 0.05$ was regarded as statistically significant.

specimens without liver metastasis and adjacent tissues at the same time. This reduces batch effects compared to using data from GEO databases for comparison. Furthermore, there is currently no public database contains both prognostic and gene sequencing data of GIST. And, our cohort contained too few cases (only 18 patients) to survival analysis. So, we can only retreat to the next best, using TCGA database for survival analysis. Whether these hub genes in GIST have prognostic value remains to be further confirmed.

In summary, through analyzing data of self-made whole-genome gene expression profiling and GEO dataset, we identified those signal pathways and hub genes that played significant roles in the tumorigenesis and liver metastasis of GIST. Further studies with larger sample sizes should be carried out to validate the present findings. Additionally, experimental evidence is warranted to investigate the functional roles of the identified hub genes in the liver metastasis of GIST. We sincerely hope that this present study will contribute to the discovery of therapeutic target for liver metastatic GIST.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author.

Author contributions

HQ: Conceptualization; Supervision; Project administration. JG: Investigation; Formal analysis; Writing—review and editing. SF: Investigation; Writing–original Draft. HY, BO, and DJ: Data collecting. CD, XC, and MZ: Writing–Review and Editing. YL and YZ: Data Curation. WZ: Collecting of clinical samples.

Funding

This work was supported by NSFC grants (NSFC 81602061).

Acknowledgments

The authors would like to thank their patients for their participation in the study.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1007135/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Identification of differentially expressed genes of GSE13861 dataset. **(A)** Volcano map of differentially expressed genes (Upregulated genes in red, downregulated genes in blue). **(B)** Hierarchical clustering heatmap of DEGs screened on the basis of FC >2.0 and a corrected P value <0.05. **(C)** Shows PCA results of our cohort. **(D)** Visual hierarchical cluster analysis.

SUPPLEMENTARY FIGURE S2
Expression of ETV1 and ETV4. **(A, B)** Expression of ETV1 and ETV4 in GIST (with liver metastasis) tissues and correspond non-tumor tissues in our cohort. **(C, D)**. Expression of ETV1 and ETV4 in primary GIST tissues and correspond non-tumor tissues in GSE13861 dataset.

References

Amirnasr, A., Gits, C. M. M., van Kuijk, P. F., Smid, M., Vriends, A. L. M., Rutkowski, P., et al. (2019). Molecular comparison of imatinib-naïve and resistant gastrointestinal stromal tumors: Differentially expressed microRNAs and mRNAs. *Cancers (Basel)* 11, 882. doi:10.3390/cancers11060882

Amirnasr, A., Sleijfer, S., and Wiemer, E. A. C. (2020). Non-coding RNAs, a novel paradigm for the management of gastrointestinal stromal tumors. *Int. J. Mol. Sci.* 21, 6975. doi:10.3390/ijms21186975

Bayraktar, U. D., Bayraktar, S., and Rocha-Lima, C. M. (2010). Molecular basis and management of gastrointestinal stromal tumors. *World J. Gastroenterol.* 16, 2726–2734. doi:10.3748/wjg.v16.i22.2726

Cavallaro, U., and Christofori, G. (2004). Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. *Nat. Rev. Cancer* 4, 118–132. doi:10.1038/nrc1276

Cavallaro, U., and Christofori, G. (2004). Multitasking in tumor progression: Signaling functions of cell adhesion molecules. *Ann. N. Y. Acad. Sci.* 1014, 58–66. doi:10.1196/annals.1294.006

Chi, P., Chen, Y., Zhang, L., Guo, X., Wongvipat, J., Shamu, T., et al. (2010). ETV1 is a lineage survival factor that cooperates with KIT in gastrointestinal stromal tumours. *Nature* 467, 849–853. doi:10.1038/nature09409

Chin, C.-H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8, S11. doi:10.1186/1752-0509-8-S4-S11

Cho, S., Kitadai, Y., Yoshida, S., Tanaka, S., Yoshihara, M., Yoshida, K., et al. (2006). Deletion of the KIT gene is associated with liver metastasis and poor prognosis in patients with gastrointestinal stromal tumor in the stomach. *Int. J. Oncol.* 28, 1361–1367. doi:10.3892/ijo.28.6.1361

Cho, J. Y., Lim, J. Y., Cheong, J. H., Park, Y. Y., Yoon, S. L., Kim, S. M., et al. (2011). Gene expression signature–based prognostic risk score in gastric cancer. *Clin. Cancer Res.* 17, 1850–1857. doi:10.1158/1078-0432.CCR-10-2180

Corless, C. L., and Heinrich, M. C. (2008). Molecular pathobiology of gastrointestinal stromal sarcomas. *Annu. Rev. Pathol. Mech. Dis.* 3, 557–586. doi:10.1146/annurev.pathmechdis.3.121806.151538

Corless, C. L., Barnett, C. M., and Heinrich, M. C. (2011). Gastrointestinal stromal tumours: Origin and molecular oncology. *Nat. Rev. Cancer* 11, 865–878. doi:10.1038/nrc3143

Croix, B., St, Rak, J. W., Florenes, V. A., Slingerland, J. M., and Kerbel, R. S. (1998). E-Cadherin–dependent growth suppression is mediated by the cyclin-dependent kinase inhibitor p27KIP1. *J. Cell Biol.* 142, 557–571. doi:10.1083/jcb.142.2.557

DeMatteo, R. P., Lewis, J. J., Leung, D., Mudan, S. S., Woodruff, J. M., and Brennan, M. F. (2000). Two hundred gastrointestinal stromal tumors: Recurrence patterns and prognostic factors for survival. *Ann. Surg.* 231, 51–58. doi:10.1097/00000658-200001000-00008

DeMatteo, R. P., Ballman, K. V., Antonescu, C. R., Maki, R. G., Pisters, P. W. T., Demetri, G. D., et al. (2009). Adjuvant imatinib mesylate after resection of localised, primary gastrointestinal stromal tumour: A randomised, double-blind, placebo-controlled trial. *Lancet* 373, 1097–1104. doi:10.1016/S0140-6736(09)60500-6

Duan, Y., Haybaeck, J., and Yang, Z. (2020). Therapeutic potential of PI3K/AKT/mTOR pathway in gastrointestinal stromal tumors: Rationale and progress. *Cancers (Basel)* 12, 2972. doi:10.3390/cancers12102972

Fruman, D. A., and Rommel, C. (2014). PI3K and cancer: Lessons, challenges and opportunities. *Nat. Rev. Drug Discov.* 13, 140–156. doi:10.1038/nrd4204

Girigoswami, K., Saini, D., and Girigoswami, A. (2021). Extracellular matrix remodeling and development of cancer. *Stem Cell Rev Rep* 17, 739–747. doi:10.1007/s12015-020-10070-1

Gold, J. S., and DeMatteo, R. P. (2006). Combined surgical and molecular therapy: The gastrointestinal stromal tumor model. *Ann. Surg.* 244, 176–184. doi:10.1097/01.sla.0000218080.94145.cf

González-García, A., Mérida, I., Martinez-A, , C., and Carrera, A. C. (1997). Intermediate affinity interleukin-2 receptor mediates survival via a phosphatidylinositol 3-Kinase-dependent pathway. *J. Biol. Chem.* 272, 10220–10226. doi:10.1074/jbc.272.15.10220

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell* 144, 646–674. doi:10.1016/j.cell.2011.02.013

Hemming, M. L., Heinrich, M. C., Bauer, S., and George, S. (2018). Translational insights into gastrointestinal stromal tumor and current clinical advances. *Ann. Oncol.* 29, 2037–2045. doi:10.1093/annonc/mdy309

Herlevsen, M., Schmidt, D.-S., Miyazaki, K., and Zöller, M. (2003). The association of the tetraspanin D6.1A with the alpha6beta4 integrin supports cell motility and liver metastasis formation. *J. Cell Sci.* 116, 4373–4390. doi:10.1242/jcs.00760

Hou, J., Schindler, U., Henzel, W. J., Wong, S. C., and McKnight, S. L. (1995). Identification and purification of human stat proteins activated in response to interleukin-2. *Immunity* 2, 321–329. doi:10.1016/1074-7613(95)90140-x

Hynes, R. O. (2009). The extracellular matrix: Not just pretty fibrils. *Science* 326, 1216–1219. doi:10.1126/science.1176009

Jiang, W. G., and Mansel, R. E. (2000). E-cadherin complex and its abnormalities in human breast cancer. *Surg. Oncol.* 9, 151–171. doi:10.1016/s0960-7404(01)00010-x

Jiang, W. G., Martin, T. A., Matsumoto, K., Nakamura, T., and Mansel, R. E. (1999). Hepatocyte growth factor/scatter factor decreases the expression of occludin and transendothelial resistance (TER) and increases paracellular permeability in human vascular endothelial cells. *J. Cell. Physiol.* 181, 319–329. doi:10.1002/(SICI)1097-4652(199911)181:2<319::AID-JCP14>3.0.CO;2-S

Joensuu, H., Vehtari, A., Riihimaki, J., Nishida, T., Steigen, S. E., Brabec, P., et al. (2012). Risk of recurrence of gastrointestinal stromal tumour after surgery: An analysis of pooled population-based cohorts. *Lancet Oncol.* 13, 265–274. doi:10.1016/S1470-2045(11)70299-6

Joensuu, H., Hohenberger, P., and Corless, C. L. (2013). Gastrointestinal stromal tumour. *Lancet* 382, 973–983. doi:10.1016/S0140-6736(13)60106-3

Kalluri, R., and Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* 119, 1420–1428. doi:10.1172/JCI39104

Liao, W., Lin, J.-X., and Leonard, W. J. (2013). Interleukin-2 at the crossroads of effector responses, tolerance, and immunotherapy. *Immunity* 38, 13–25. doi:10.1016/j.immuni.2013.01.004

Lin, J.-X., Migone, T. S., Tsang, M., Friedmann, M., Weatherbee, J. A., Zhou, L., et al. (1995). The role of shared receptor motifs and common stat proteins in the generation of cytokine pleiotropy and redundancy by IL-2, IL-4, IL-7, IL-13, and IL-15. *Immunity* 2, 331–339. doi:10.1016/1074-7613(95)90141-8

Liu, S. J., Dang, H. X., Lim, D. A., Feng, F. Y., and Maher, C. A. (2021). Long noncoding RNAs in cancer metastasis. *Nat. Rev. Cancer* 21, 446–460. doi:10.1038/s41568-021-00353-1

Martin, T. A., and Jiang, W. G. (2009). Loss of tight junction barrier function and its role in cancer metastasis. *Biochim. Biophys. Acta (BBA) - Biomembr.* 1788, 872–891. doi:10.1016/j.bbamem.2008.11.005

Migita, T., Oda, Y., Masuda, K., Hirata, A., Kuwano, M., Naito, S., et al. (2008). Inverse relationship between E-cadherin and p27Kip1 expression in renal cell carcinoma. *Int. J. Oncol.* 33, 41–47. doi:10.3892/ijo.33.1.41

Morgan, D. A., Ruscetti, F. W., and Gallo, R. (1976). Selective *in vitro* growth of T lymphocytes from normal human bone marrows. *Science* 193, 1007–1008. doi:10.1126/science.181845

Ng, E. H., Pollock, R. E., Munsell, M. F., Atkinson, E. N., and Romsdahl, M. M. (1992). Prognostic factors influencing survival in gastrointestinal leiomyosarcomas. Implications for surgical management and staging. *Ann. Surg.* 215, 68–77. doi:10.1097/00000658-199201000-00010

Ohshima, K., Fujiya, K., Nagashima, T., Ohnami, S., Hatakeyama, K., Urakami, K., et al. (2019). Driver gene alterations and activated signaling pathways toward malignant progression of gastrointestinal stromal tumors. *Cancer Sci.* 110, 3821–3833. doi:10.1111/cas.14202

Paolillo, M., and Schinelli, S. (2019). Extracellular matrix alterations in metastatic processes. *Int. J. Mol. Sci.* 20, 4947. doi:10.3390/ijms20194947

Patel, S. (2013). Exploring novel therapeutic targets in GIST: Focus on the PI3K/Akt/mTOR pathway. *Curr. Oncol. Rep.* 15, 386–395. doi:10.1007/s11912-013-0316-6

Polyak, K., and Weinberg, R. A. (2009). Transitions between epithelial and mesenchymal states: Acquisition of malignant and stem cell traits. *Nat. Rev. Cancer* 9, 265–273. doi:10.1038/nrc2620

Poveda, A., Garcia Del Muro, X., Lopez-Guerrero, J. A., Cubedo, R., Martinez, V., Romero, I., et al. (2017). GEIS guidelines for gastrointestinal sarcomas (GIST). *Cancer Treat. Rev.* 55, 107–119. doi:10.1016/j.ctrv.2016.11.011

Ross, S. H., and Cantrell, D. A. (2018). Signaling and function of interleukin-2 in T lymphocytes. *Annu. Rev. Immunol.* 36, 411–433. doi:10.1146/annurev-immunol-042617-053352

Serrano, C., and George, S. (2020). Gastrointestinal stromal tumor: Challenges and opportunities for a new decade. *Clin. Cancer Res.* 26, 5078–5085. doi:10.1158/1078-0432.CCR-20-1706

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi:10.1073/pnas.0506580102

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: Quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368. doi:10.1093/nar/gkw937

Thomson, S., Petti, F., Sujka-Kwok, I., Mercado, P., Bean, J., Monaghan, M., et al. (2011). A systems view of epithelial–mesenchymal transition signaling states. *Clin. Exp. Metastasis* 28, 137–155. doi:10.1007/s10585-010-9367-3

Tsai, J. H., and Yang, J. (2013). Epithelial–mesenchymal plasticity in carcinoma metastasis. *Genes Dev.* 27, 2192–2206. doi:10.1101/gad.225334.113

Tsukita, S., Furuse, M., Tsukita, S., and Furuse, M. (1999). Occludin and claudins in tight-junction strands: Leading or supporting players? *Trends Cell Biol.* 9, 268–273. doi:10.1016/s0962-8924(99)01578-0

Vara, J. Á. F., Casado, E., de Castro, J., Cejas, P., Belda-Iniesta, C., and Gonzalez-Baron, M. (2004). PI3K/Akt signalling pathway and cancer. *Cancer Treat. Rev.* 30, 193–204. doi:10.1016/j.ctrv.2003.07.007

Walker, C., Mojares, E., and del Río Hernández, A. (2018). Role of extracellular matrix in development and cancer progression. *Int. J. Mol. Sci.* 19, 3028. doi:10.3390/ijms19103028

Wang, H.-C., Li, T. Y., Chao, Y. J., Hou, Y. C., Hsueh, Y. S., Hsu, K. H., et al. (2016). KIT exon 11 codons 557–558 deletion mutation promotes liver metastasis through the CXCL12/CXCR4 Axis in gastrointestinal stromal tumors. *Clin. Cancer Res.* 22, 3477–3487. doi:10.1158/1078-0432.CCR-15-2748

Wang, J. (2000). From DNA biosensors to gene chips. *Nucleic Acids Res.* 28, 3011–3016. doi:10.1093/nar/28.16.3011

Wong, V., and Gumbiner, B. M. (1997). A synthetic peptide corresponding to the extracellular domain of occludin perturbs the tight junction permeability barrier. *J. Cell Biol.* 136, 399–409. doi:10.1083/jcb.136.2.399

Zeng, S., Seifert, A. M., Zhang, J. Q., Kim, T. S., Bowler, T. G., Cavnar, M. J., et al. (2017). ETV4 collaborates with Wnt/β-catenin signaling to alter cell cycle activity and promote tumor aggressiveness in gastrointestinal stromal tumor. *Oncotarget* 8, 114195–114209. doi:10.18632/oncotarget.23173

Zschiesche, W., Schonborn, I., Behrens, J., Herrenknecht, K., Hartveit, F., Lilleng, P., et al. (1997). Expression of E-cadherin and catenins in invasive mammary carcinomas. *Anticancer Res.* 17, 561–567.

Zvibel, I., Brill, S., Halpern, Z., and Papa, M. (1998). Hepatocyte extracellular matrix modulates expression of growth factors and growth factor receptors in human colon cancer cells. *Exp. Cell Res.* 245, 123–131. doi:10.1006/excr.1998.4229

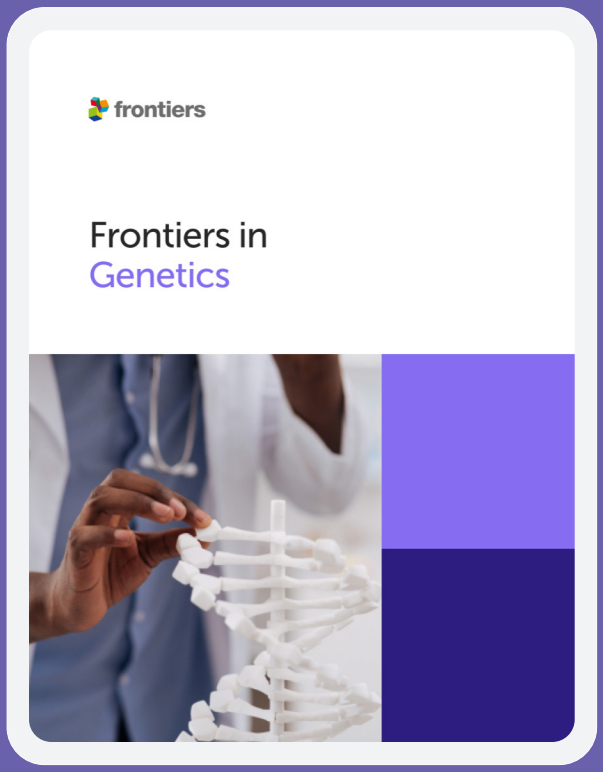
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact