Am J Transl Res 2024;16(6):2212-2232
www.ajtr.org /ISSN:1943-8141/AJTR0154322

Original Article

Analysis of the potential biological significance of glycosylation in triple-negative breast cancer on patient prognosis

Han Zhou[1], Zhiwei Wang[1], Jun Guo[2,3], Zihui Zhu[4], Gang Sun[1,2]

[1]Department of Breast and Thyroid Surgery, The Affiliated Cancer Hospital of Xinjiang Medical University, Urumqi 830011, Xinjiang, China; [2]Key Laboratory of Oncology of Xinjiang Uyghur Autonomous Region, Urumqi 830011, Xinjiang, China; [3]Department of Cancer Research Institute, Affiliated Cancer Hospital of Xinjiang Medical University, Urumqi 830011, Xinjiang, China; [4]Department of Breast Surgery, The Affiliated Cancer Hospital of Xinjiang Medical University, Urumqi 830011, Xinjiang, China

Received November 8, 2023; Accepted May 6, 2024; Epub June 15, 2024; Published June 30, 2024

Abstract: Background: Breast cancer is the most common malignancy in women, with its prognosis varying greatly according to its subtype. Triple-negative breast cancer (TNBC) has the worst prognosis among all subtypes. Glycosylation is a critical factor influencing the prognosis of patients with TNBC. Our aim is to develop a tumor prognosis model by analyzing genes related to glycosylation to predict patient outcomes. Methods: The dataset used in this study was downloaded from the Cancer Genome Atlas Program (TCGA) database, and predictive genes were identified through Cox one-way regression analysis. The model genes with the highest risk scores among the 18 samples were obtained by lasso regression analysis to establish the model. We analyzed the pathways affecting the progression of TNBC and discovered key genes for subsequent research. Results: Our model was constructed using data from TCGA database and validated through Kaplan-Meier curve analysis and Receiver Operating Characteristic (ROC) curve assessment. Our analysis revealed that a high expression of tumor-related chemokines in the high-risk group may be associated with poor tumor prognosis. Furthermore, we conducted a random survival forest analysis and identified two significant genes, namely DPM2 and PINK1, which have been selected for further investigation. Conclusion: The prognostic analysis model, developed based on the glycosylation genes in TNBC, exhibits excellent validation efficacy. This model is valuable for the prognostic analysis of patients with TNBC.

Keywords: TNBC, prognostic model, DPM2, PINK1, immune infiltration, glycosylation

Introduction

Breast cancer remains one of the most prevalent cancers worldwide, with triple-negative breast cancer (TNBC), one of the most prevalent pathological subtypes, representing a significant subtype in current clinical research, and having a high incidence in current clinical studies. TNBC is characterized as a heterogeneous tumor distinguished by the absence of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) expression, linking it to a more aggressive clinical outcome and limited treatment options [1, 2]. Notably, the trend in China shows that TNBC tends to affect younger and premenopausal women more frequently, and the lack of treatment options often results in a poorer prognosis due to higher malignancy and rapid disease progression.

Glycosylation, the enzymatic process that attaches carbohydrates to proteins or lipids, plays a critical role in protein folding and stability, sub-cellular localization, and the functionality of glycoproteins. Compared to normal cells, cancer cells often exhibit significant changes in glycosylation patterns [3]. These alterations, including impaired synthesis of complex cellular glycans and the production of abnormal structures, are closely associated with cancer initiation, progression, metastasis, and chemoresistance. Specific glycosylation processes implicated in cancer include sialylation, fucosylation,

https://doi.org/10.62347/PXAR3644

O-glycan truncation, and alterations in N- and O-glycosylation, which are essential for protein folding regulation and quality control [4]. Changes in tumor cell surface N-oligosaccharides have been linked to enhanced cell proliferation, migration, and invasion [5].

Despite numerous prevention strategies, the prognosis of breast cancer continues to pose a significant global health challenge, affecting individuals, families, communities, and healthcare systems. Our research underscores the critical role of glycosylation in the prognosis of breast cancer. Utilizing open data and experimental validation, we screened glycosylation-related genes that influence the prognosis of TNBC and developed a prognostic model. This model was validated across multiple dimensions using external datasets, including the construction of ROC curve graphs. Furthermore, we explored downstream pathways and differentiated between low-risk and high-risk groups within the model to predict potential target genes. These validations confirm the accuracy and reliability of our model, showcasing it as a valuable tool for guiding the prognosis of TNBC in clinical settings.

Materials and methods

Data

The Cancer Genome Atlas (TCGA) databank, recognized as the most extensive repository for cancer gene information (https://portal.gdc.cancer.gov/), includes data on copy number variation, gene expression, miRNA expression, SNPs, DNA methylation, and other data types. The processed raw mRNA expression data for Triple-Negative Breast Cancer (TNBC) were downloaded, comprising 127 samples. The Series Matrix File for GSE58812, accessible through the NCBI's Gene Expression Omnibus (GEO) database and annotated with the platform GPL570, included data from 107 patients with TNBC, all of whom had complete expression profiles and survival data. Similarly, the Series Matrix File for GSE135565, annotated by GPL570, comprised data from 84 TNBC patients, each with complete survival information and expression profiles. These datasets were analyzed for further investigation.

KEGG and GO functional analysis

The functional annotation of significant genes was performed using the Metascape databank (www.metascape.org) to elucidate the relevant functions of these genes comprehensively. Gene Ontology (GO) pathway analysis was fulfilled for decided genes. Gene Ontology (GO) pathway analysis was conducted for selected genes, with statistical significance set at a minimum overlap of ≥3 and a p-value of ≤0.01.

Model construction and prognosis

Genes associated with glycosylation were carefully selected and utilized for lasso regression to construct a predictive model. A risk score formula was developed for each patient by integrating gene expression data. This study utilized regression coefficients estimated through lasso regression analysis to calculate the risk score, which facilitated the categorization of patients into high-risk and low-risk cohorts. The median risk score served as the demarcation point for group stratification. Kaplan-Meier analysis assessed the survival difference between the two groups, with comparisons made using the log-rank statistical method. Lasso regression and stratified analysis were employed to determine the predictive value of the risk score in patient prognosis. Further more, the "survival-ROC" package was used for ROC curve analysis to evaluate the accuracy of the predictive model.

Analysis of immune cell infiltration

CIBERSORT, a method widely recognized for its ability to analyze and characterize cellular composition using gene expression profiles, serves as a crucial tool for estimating immune cell infiltration. Based on support vector regression and the back-convolution analysis of immune cell subtype expression matrices, the dataset includes 547 biomarkers capable of distinguishing 22 unique phenotypes of human immune cells, such as subsets of B/T-cells, myeloid cells, and plasma cells. The CIBERSORT algorithm analyzed patient data to infer the relative proportions of the 22 immune cell populations infiltrating the tissues. The sum of the estimated scores for immune cell types in each sample equaled 1, facilitating the comparison of differences in immune cell content. Correlation analysis was performed between immune cell composition and gene expression.

GSVA analysis (Gene Set Variance Analysis)

Gene Set Variation Analysis (GSVA) is a non-parametric, unsupervised method utilized to

assess gene set enrichment results from microarrays and transcriptomes. This approach translates changes in individual genes into alterations at the pathway level by evaluating genes of interest comprehensively to determine the biological function of a sample. To reduce redundancy in pathway information, duplicate genes within each gene set were eliminated, specifically removing those identified in two or more pathways. In this study, gene sets (50 hallmark pathways) were downloaded from the Molecular Signatures Database (v7.0). The GSVA algorithm of the R language "GSVA" package was used to score each gene set comprehensively and evaluate potential changes in biological functions across different samples.

GSEA analysis

Participants were stratified into high- and low-risk groups based on the risk assessment model, and signaling pathway differences between the two groups were examined using GSEA. The gene set background was sourced from version 7.0 of the Molecular Signatures Database (MsigDB), which includes annotated genes for subtype pathways. This database was utilized in performing the differential expression analysis of pathway among the subtypes, where significantly enriched gene sets (with an adjusted *p*-value below 0.05) were identified and ranked according to their concordance scores. GSEA is frequently employed in disease typing studies to elucidate biological significance.

Machine learning methods to identify key genes

The Random Survival Forest (RSF) algorithm, analyzed using the Random-Forest-SRC software package, was applied to survival data to screen for candidate genes, providing a ranking of genes related to prognosis (nrep = 1000, representing 1000 Monte Carlo simulations). Our final selections of feature genes were those with relative importance greater than 0.3. The prognostic significance of these identified genes was confirmed by examining the Kaplan-Meier survival curves. Additionally, immunohistochemical images of breast cancer from the Human Protein Atlas (HPA) database were analyzed to assess differences in protein expression levels of key genes.

miRNA network construction

MicroRNAs (miRNAs) are small RNAs capable of regulating gene expression either by inhibiting mRNA translation or promoting mRNA degradation. Further analysis explored the presence of miRNAs in key genes that could regulate the transcription or degradation of detrimental genes. Key gene-associated mi-RNAs were obtained via the miRcode database, and the gene miRNA network was visualized via Cytoscape software.

Analysis of regulatory networks involving critical genes

In our investigation, we utilized the R package "RcisTarget" to predict transcription factors, with all analyses being motif-based. The Normalized Enrichment Score (NES) for motifs depended on the total number of motifs in the database. Annotated files were created by combining motif similarity and gene sequences with motifs identified by the primary data source. The first step in assessing motif overexpression on the gene set involved measuring the Area Under the Curve (AUC) for each pair of motif groups, calculated through the recovery curve of the sequence according to the gene set. The NES for each motif was determined from the AUC distribution across all motifs in the gene set, using rcistarge.hg19.motifdb.cisbpont.500bp as the Gene-motif ranking database.

Drug susceptibility analysis

The genomics database used in this study, the largest of its kind, was the Genomics of Drug Sensitivity in Cancer (GDSC) database (https://www.cancerrxgene.org/). To predict the chemotherapy sensitivity of tumor samples, we utilized the R software package "pRRophetic". The IC50 values for each chemotherapeutic drug were estimated using a regression method, and the accuracy of both the regression and prediction was confirmed through 10-fold cross-validation using the GDSC training set. Default parameter settings were applied, such as using "combat" to correct for batch effects and averaging the expression of duplicated genes.

Statistical analysis

Kaplan-Meier curves were plotted to analyze survival and comparisons were made using log-

rank analysis. Cox proportional hazards modeling was applied for multivariate analysis. All statistical analyses were performed using R software (version 4.3.0). Results were considered to be statistically significant at $P<0.05$, using a two-sided analysis.

Results

Functional enrichment of related genes

The raw TNBC mRNA expression data from the TCGA database were downloaded for processing. To identify signaling pathways associated with glycosylation genes, we acquired a gene set of glycosylation genes from the GeneCards database (https://www.genecards.org/), selecting genes with a relevance score greater than 3. The analysis revealed that these genes are enriched in pathways including protein O-linked glycosylation, protein N-linked glycosylation, and general protein glycosylation (**Figure 1A**, **1B**).

Acquisition of prognosis-related genes and construction of prediction models

Cox univariate regression was employed to identify prognostic genes within the glycosylation gene set in TNBC. A total of 59 prognosis-related genes were identified and visualized (*p*-value <0.05) (**Table 1**; **Figure 1C**). As depicted in the figure, some of the identified prognostic genes act as risk factors for tumorigenesis and development. To further identify key genes within the prognostic gene set, we applied the lasso regression feature selection algorithm, specifically targeting genes characteristic of triple-negative breast cancer. Patients were randomly divided into training and validation sets in a 4:1 ratio. Lasso regression analysis (**Figure 2A-C**) yielded the optimal risk score value for each sample for subsequent analysis (Risk score = GLB1L × (-0.383907613279652) + EPCAM × (-0.00681852065508716) + PTGDS × 0.050439855466918 + ITGB3 × 0.0712510002611501 + UGCG × 0.10293855714384 + NANS × 0.115013479069145 + SMPD1 × 0.12110376616419 + B4GALNT2 × 0.12252210876786 + MMRN1 × 0.1294464788656 + DPM2 × 0.25911172011391 + ALG1 × 0.290667457826843 + KIT × 0.30690596763118 + BGN × 0.309893230261671 + TGFB1 × 0.322208987756023 + PINK1 × 0.325114622613627 + MGAT5B × 0.334265742062431 + SERPINE1 × 0.432850897643511 + TMEM165 × 0.625955517391717). Patients were stratified into low- and high-risk groups based on the median risk score. The Kaplan-Meier curve was utilized to compare survival rates between the high-risk and low-risk groups. Survival analysis indicated a statistically significant difference in overall survival between the high-risk and low-risk groups in both training and test datasets, with the high-risk group exhibiting notably lower overall survival (**Figure 2D**, **2E**).

Both internal and external datasets were utilized to validate the efficacy of the prognostic model

Both internal and external datasets were employed to validate the efficacy of the prognostic model. The model's validation began with the internal test set and training set, utilizing Receiver Operating Characteristic (ROC) curves. The results indicate a robust validation efficacy of the model (**Figure 3A**, **3B**). Validation was further extended to an external dataset, wherein survival data for patients with Triple-Negative Breast Cancer (TNBC) were sourced from the Gene Expression Omnibus (GEO) databank (GSE58812, GSE135565). Risk scores were calculated according to the model's formula, categorizing individuals into high-risk or low-risk groups based on the median value of their risk scores and objective accuracy criteria. The difference in survival between these two groups was assessed using Kaplan-Meier curves to evaluate the robustness of the predictive model. Our analysis confirmed that the overall survival of the high-risk group was significantly lower compared to the low-risk group in the GEO external validation set (**Figure 4A**, **4B**). The model's reliability was further validated through ROC curve analysis on an independent dataset, confirming its strong predictive capability in forecasting patient prognosis (**Figure 4C**, **4D**).

Risk score and immunization correlation analysis

Recent studies highlight the pivotal role of the tumor microenvironment in the progression of breast cancer, including TNBC [6]. Changes in the tumor microenvironment often herald shifts towards more severe cancer subtypes with poorer prognoses [7]; thus, it cannot be over-

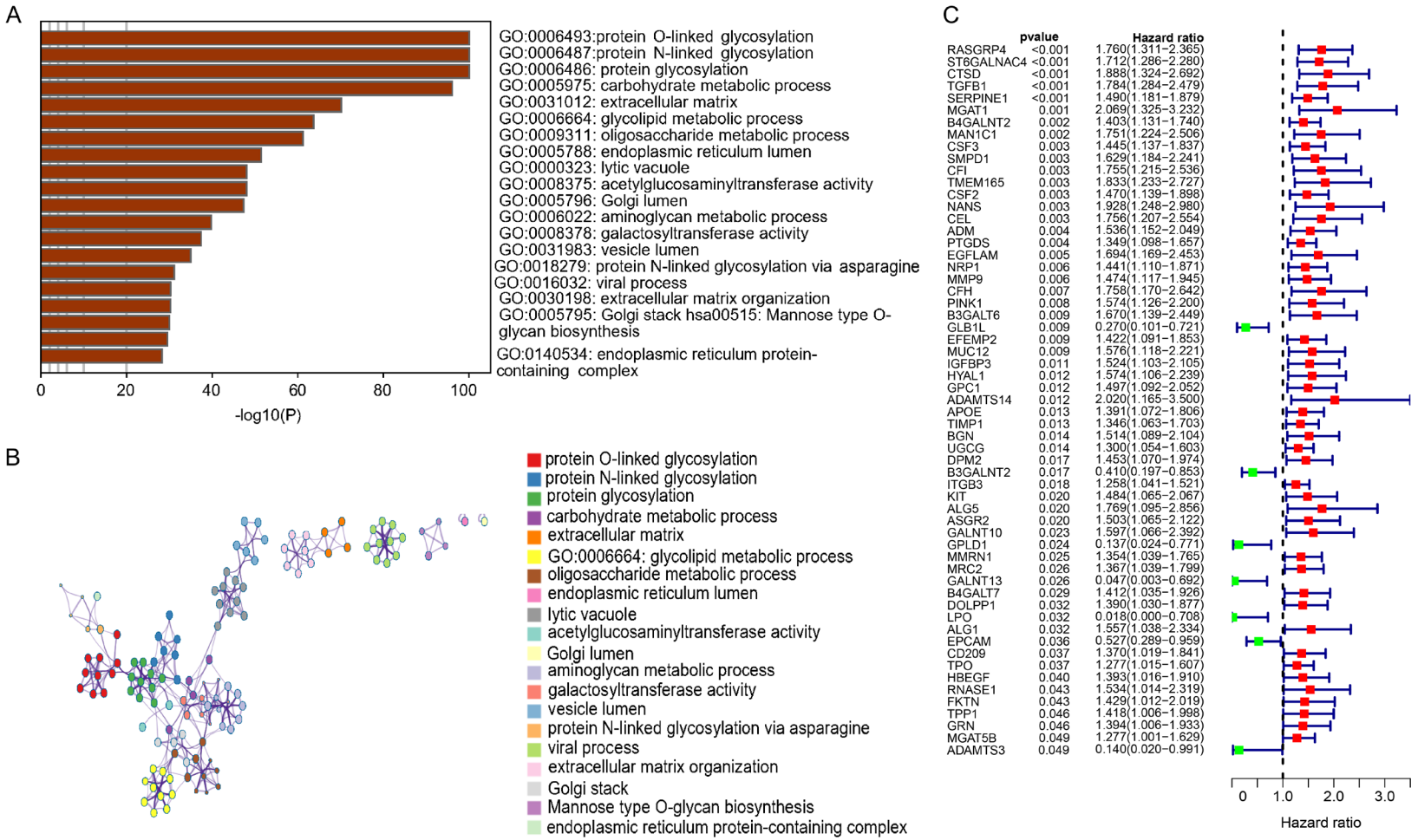

Figure 1. Enrichment of genes related to glycosylation in breast cancer. A. Glycosylation associated genes were obtained from Gene Cards database. B. Pathway enrichment analysis was performed on the selected related genes. C. Prognostic genes associated with modeling were screened by cox analysis.

Table 1. Prognostic genes identified by Cox univariate regression

Gene	HR	z	*p*-value	Lower	Upper
RASGRP4	1.760466688	3.756491138	0.000172312	1.310601129	2.364749193
ST6GALNAC4	1.712180304	3.681529985	0.000231839	1.285917135	2.279743627
CTSD	1.88778118	3.510981722	0.000446455	1.324051595	2.691524858
TGFB1	1.78385391	3.44690807	0.000567041	1.283608545	2.479053904
SERPINE1	1.489668048	3.367170661	0.000759437	1.181238243	1.878631094
MGAT1	2.069193581	3.196547642	0.001390829	1.32485267	3.231726948
B4GALNT2	1.402953496	3.083171995	0.002048067	1.131275205	1.739875941
MAN1C1	1.751252953	3.064388969	0.002181151	1.223783128	2.506070592
CSF3	1.445399468	3.010031698	0.002612204	1.137136869	1.837227936
SMPD1	1.629082978	2.998331765	0.00271462	1.184129356	2.241234319
CFI	1.755159071	2.994361993	0.002750194	1.214503511	2.536496053
TMEM165	1.833214058	2.992232202	0.002769455	1.232548399	2.726605935
CSF2	1.470093729	2.958614693	0.003090252	1.138900109	1.897598881
NANS	1.928472779	2.958365591	0.003092751	1.248116287	2.97969612
CEL	1.755944408	2.944803002	0.003231603	1.207181147	2.554165771
ADM	1.536043883	2.919546152	0.003505415	1.151506072	2.048995544
PTGDS	1.349102052	2.853834349	0.004319505	1.098328582	1.657132825
EGFLAM	1.693652876	2.787861793	0.005305717	1.169372495	2.45299088
NRP1	1.441234801	2.746079809	0.006031209	1.110301016	1.870805953
MMP9	1.474298863	2.74528873	0.006045768	1.117444686	1.945113852
CFH	1.758099479	2.715350024	0.006620576	1.169954279	2.641909888
PINK1	1.573740452	2.653169748	0.007973976	1.125780214	2.19994896
B3GALT6	1.669836854	2.62477294	0.008670677	1.138674067	2.44877371
GLB1L	0.269743994	2.611918184	0.00900358	0.100911865	0.721043281
EFEMP2	1.422013741	2.60499293	0.009187616	1.091090811	1.853304106
MUC12	1.575975296	2.599584251	0.009333676	1.118423969	2.220712539
IGFBP3	1.523911523	2.555808558	0.010594139	1.103199474	2.105064755
HYAL1	1.573553953	2.51814801	0.011797373	1.105711102	2.239348089
GPC1	1.497153517	2.509570186	0.012087819	1.09240593	2.051864233
ADAMTS14	2.019675271	2.505519023	0.012227182	1.165393012	3.50018248
APOE	1.39114832	2.479883198	0.013142543	1.07166468	1.805876116
TIMP1	1.345838864	2.471789551	0.013443863	1.06343312	1.703240395
BGN	1.513622408	2.466694971	0.013636647	1.088882072	2.104041249
UGCG	1.299617172	2.446834486	0.0144117	1.053530147	1.603186012
DPM2	1.453398521	2.394217784	0.016655856	1.070169823	1.973861731
B3GALNT2	0.409857868	2.385668954	0.017048089	0.196965099	0.852859075
ITGB3	1.258077811	2.371781901	0.017702537	1.040668578	1.520906667
KIT	1.483929704	2.334558292	0.019566508	1.065379481	2.066913626
ALG5	1.768650924	2.332022599	0.019699502	1.09524088	2.856107867
ASGR2	1.503283483	2.319418378	0.020372362	1.065214585	2.121507967
GALNT10	1.596722919	2.268891399	0.023274931	1.065789171	2.392146729
GPLD1	0.13700573	2.255291714	0.024115024	0.024351207	0.770827106
MMRN1	1.354281706	2.243004064	0.024896552	1.039010867	1.76521632
MRC2	1.366863415	2.230001339	0.025747354	1.038568987	1.798932588
GALNT13	0.046700766	2.227298915	0.025927303	0.003150391	0.692282885
B4GALT7	1.411941657	2.179011792	0.029330792	1.035286405	1.92563066
DOLPP1	1.390157389	2.150104214	0.031546972	1.029559809	1.87705226

LPO	0.018082749	2.144828813	0.031966537	0.000462101	0.707606758
ALG1	1.556547809	2.139685631	0.032380181	1.037864301	2.334448807
EPCAM	0.526715218	2.098323979	0.035876536	0.289408017	0.958608279
CD209	1.369784817	2.086756963	0.036910109	1.019302532	1.840778753
TPO	1.276957433	2.085047396	0.037064999	1.014774617	1.60687926
HBEGF	1.392744451	2.055796769	0.03980211	1.015562589	1.910012369
RNASE1	1.533697755	2.02725978	0.042635852	1.014298341	2.31906995
FKTN	1.429231123	2.026112138	0.042753295	1.011727973	2.019022561
TPP1	1.417951687	1.995569917	0.045980757	1.006250287	1.998098298
GRN	1.394086614	1.992895009	0.046272945	1.005505093	1.932837038
MGAT5B	1.277340331	1.971783065	0.048634376	1.001468315	1.629206134
ADAMTS3	0.140052138	1.969106595	0.048940852	0.019794443	0.990914543

looked as a crucial factor in tumor disease studies. The tumor microenvironment includes tumor-associated fibroblasts, immune cells, the extracellular matrix, various growth and inflammatory factors, specific physicochemical characteristics, and the cancer cells. It significantly impacts disease diagnosis, patient survival outcomes, and response to treatment. We explored the potential molecular mechanisms by which risk scores influence the progression of triple-negative breast cancer by analyzing the correlation between risk scores and immune infiltration. The comparison revealed significant differences in immune cell content, notably in T cells follicular helper and activated dendritic cells between the low-risk and high-risk groups (**Figure 5A**, **5B**). Furthermore, we analyzed immune regulatory genes, highlighting the expression differences in immune-related chemokines, immunosuppressants, immunostimulatory factors, and immunoreceptors between the high- and low-risk groups (**Figure 6A-D**), including notably high expressions of chemokines such as CCL18, CCL17, and CCL19 in the high-risk group.

Discussion on specific signaling mechanisms related to the prognostic model

We conducted a further investigation into the signaling pathways distinguishing the high- and low-risk models to uncover the molecular mechanisms potentially responsible for the prognostic value of the risk scores on tumor progression. GSVA results indicated that pathways enriched in different patient groups were predominantly associated with IL6 JAK STAT3 SIGNALING, IL2 STAT5 SIGNALING, and HEME METABOLISM (**Figure 7A**). In TNBC, the hyperactivation of IL6 JAK STAT3 SIGNALING may contribute to tumor invasion, progression, and metastasis [8]. The IL2 STAT5 SIGNALING pathway is vital for regulating the immune response and cell proliferation. GSEA results highlighted the involvement of the cGMP-PKG signaling pathway, Oxytocin signaling pathway, and Toll-like receptor signaling pathway (**Figure 7B**). The molecular signaling network linking each of these pathways is illustrated (**Figure 7C**). The cGMP-PKG signaling pathway and Toll-like receptor signaling pathway play a key role in modulating tumor biological behavior. The cGMP-PKG signaling pathway is associated with the proliferation, metastasis, and apoptosis of tumor cells, among other processes [9]. The Toll-like receptor signaling pathway is involved in the immune response and inflammatory processes [10]. The enrichment of these pathways in the high-risk group samples indicates their significant regulatory role in the progression of triple-negative breast cancer, suggesting that perturbations in these signaling pathways between individuals in high- and low-risk groups markedly influence TNBC patient outcomes.

Risk analysis and independent prognostic assessment, along with correlation analysis of various clinical indicators

To enable personalized prognostic assessment and develop prognostic intervention strategies, we divided our sample into two cohorts based on the median risk score distribution: high- and low-risk groups. We utilized regression analysis, presented in a column-line graph format, to demonstrate how various clinical indicators of TNBC and risk score values contributed differ-

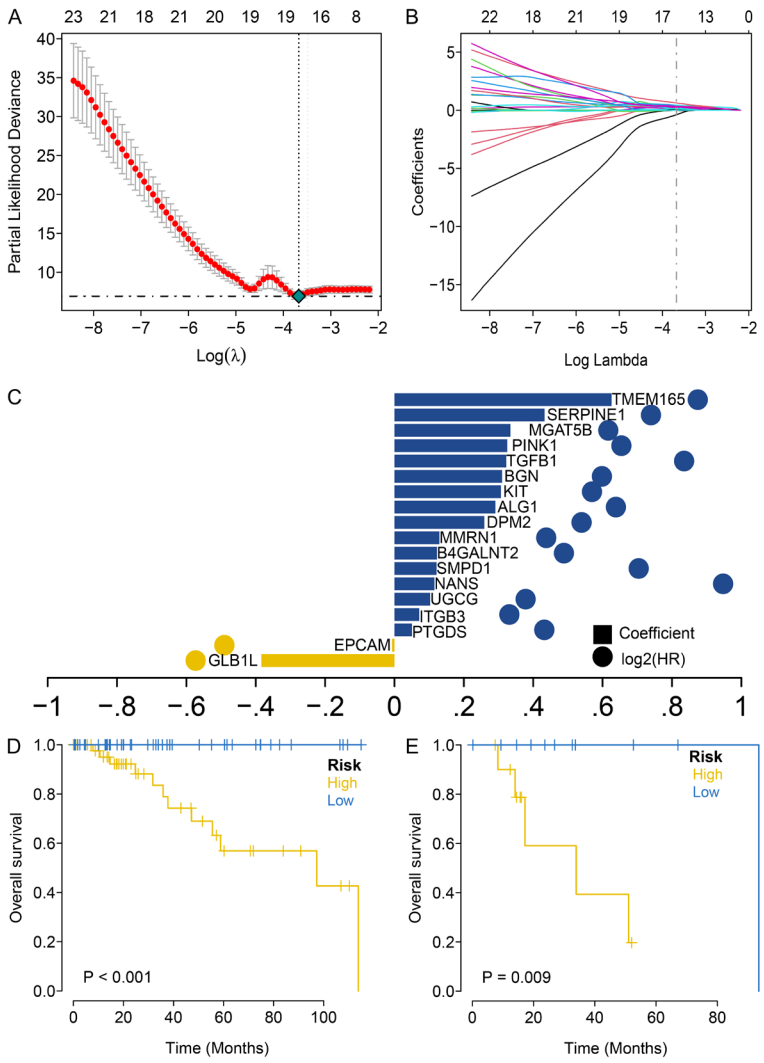

Figure 2. Prognostic model of glycosylation in triple-negative breast cancer (TNBC). A-C. The model-related genes were screened by Lasso analysis. D, E. The K-M curve analysis of high- and low-risk group.

ently to the scoring process across our entire sample. The distribution of risk score values played a significant role in the scoring process at each cancer stage (**Figure 8A**). Furthermore, we performed a prediction analysis of OS at three and five years (**Figure 8B**), discovering that the OS predicted by the nomogram closely aligned with the actual OS of the patients. This finding supports the feasibility of using a nomogram to graphically represent individual patients' prognostic risks. This visualization method allows both physicians and patients to intuitively understand and compare the impact of different factors on prognosis, making risk assessment more interpretable and accessible. Subsequently, through univariate and multivariate analyses, we determined that the risk score was an independent prognostic factor for TNBC patients (**Figure 9A**, **9B**). This implies that the risk score can independently predict a patient's prognostic outcome, either alone or when considering other relevant factors. By analyzing the size of clinical indicator values, we categorized the samples based on their corresponding risk score values and illustrated the outcomes for each clinical indicator grouping in box-and-line plots (**Figure 10A-F**). We observed that the distributions of risk score values in clinical indicators of Fustat, N, and Stage were significantly different between groups (p-value <0.05), suggesting a more general clinical relevance of the model.

Prognostic modeling for tumor immunotherapy and exploration of clinical value

We assessed the potential responsiveness of high- and low-risk cohorts to anti-tumor immunotherapy, finding that the response was poorer in the high-risk group (**Figure 11**). A high Tumor Immune Dysfunction and Exclusion (TIDE) score indicates conditions such as immune escape, abnormal antigen presentation, and abnormal antigen recognition, which may affect the interaction of the immune system with the tumor. Under these conditions, cancer cells can evade the immune system and resist the effects of immune cell-mediated killing, leading to tumor growth, spread, and progression. Therefore, a high TIDE score may correlate with poor prognosis. Additionally, we analyzed the mutation profiles of patients in high- and low-risk groups, noting a significant difference in the proportion of Titin (TTN) allele mutations. Patients in the high-risk group had fewer TTN allele mutations compared to those in the low-risk group (**Figure**

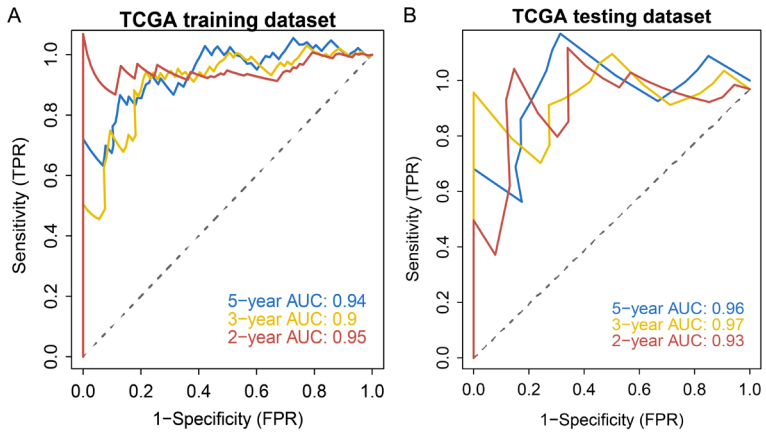

Figure 3. The model was verified by TCGA data set.

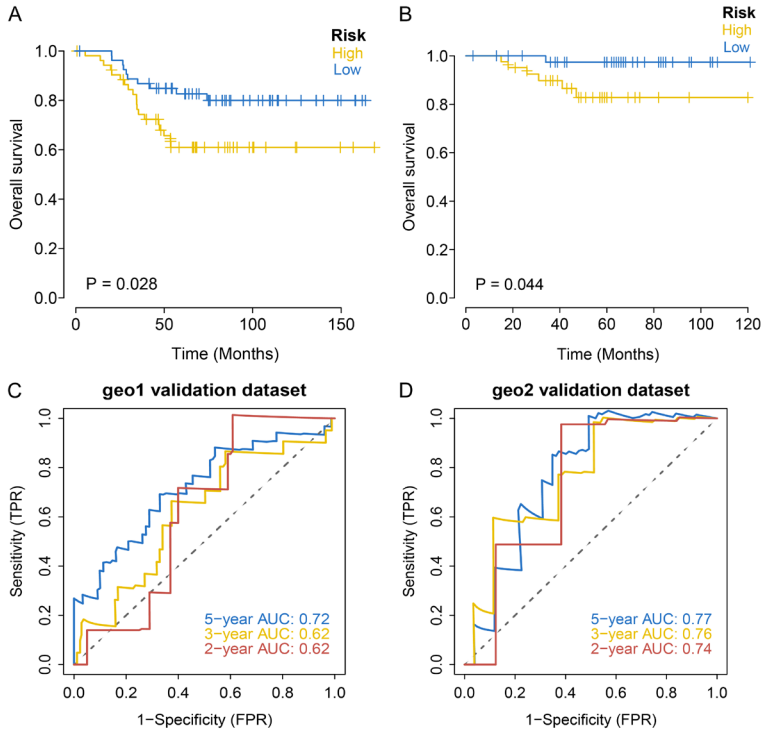

Figure 4. External validation of the data set. A, B. Analysis of Kaplan-Meier survival curve in high-low risk group. C, D. Validation on GEO external datasets.

12A). This suggests that TTN gene expression predicts poorer prognosis and altered metastasis rates in cancer patients [11, 12], further validating the accuracy of our model. To delve into the underlying tumor biology, we explored the relationship between high- and low-risk groups and common immunotherapy-related tumor markers. The comparison of tumor mutation burden (TMB) between high- and low-risk groups is shown in **Figure 12B**. Interestingly, the TMB in the low-risk group was higher than that in the high-risk group. Given the heterogeneity of TNBC, which comprises multiple independent subgroups, this may reflect differential responses to TMB based on subtype specificity.

Machine learning methods to identify key genes

We analyzed the expression levels of 18 genes screened in both high-risk and low-risk groups to identify key genes delineated by our model (**Figure 13A**). These genes underwent Random Survival Forest analysis, with those having a relative importance above 0.3 selected as final markers (**Figure 13B**), highlighting the significance of 4 genes. Survival analysis of these 4 genes revealed statistically significant differences in survival for DPM2 and PINK1 (**Figure 13C**, **13D**). Cox univariate and multivariate analyses demonstrated that these two genes were risk factors (**Figure 13E**, **13F**). This finding was further supported by immunohistochemistry in normal and breast cancer tissues (**Figure 13G**), suggesting these two genes as focal points for future studies.

Key gene-related transcriptional regulation analysis and miRNA network construction

In this study, two pivotal genes were analyzed to uncover shared regulatory mechanisms, including various transcription factors. These transcription factors were identified through cumulative recovery curves (**Figure 14A**), with the analysis indicating a significant motif (MOTIF) cisbp__M5524 and a normalized enrichment score (NES) of 9.37. This suggests a high probability of the key gene PINK1 binding

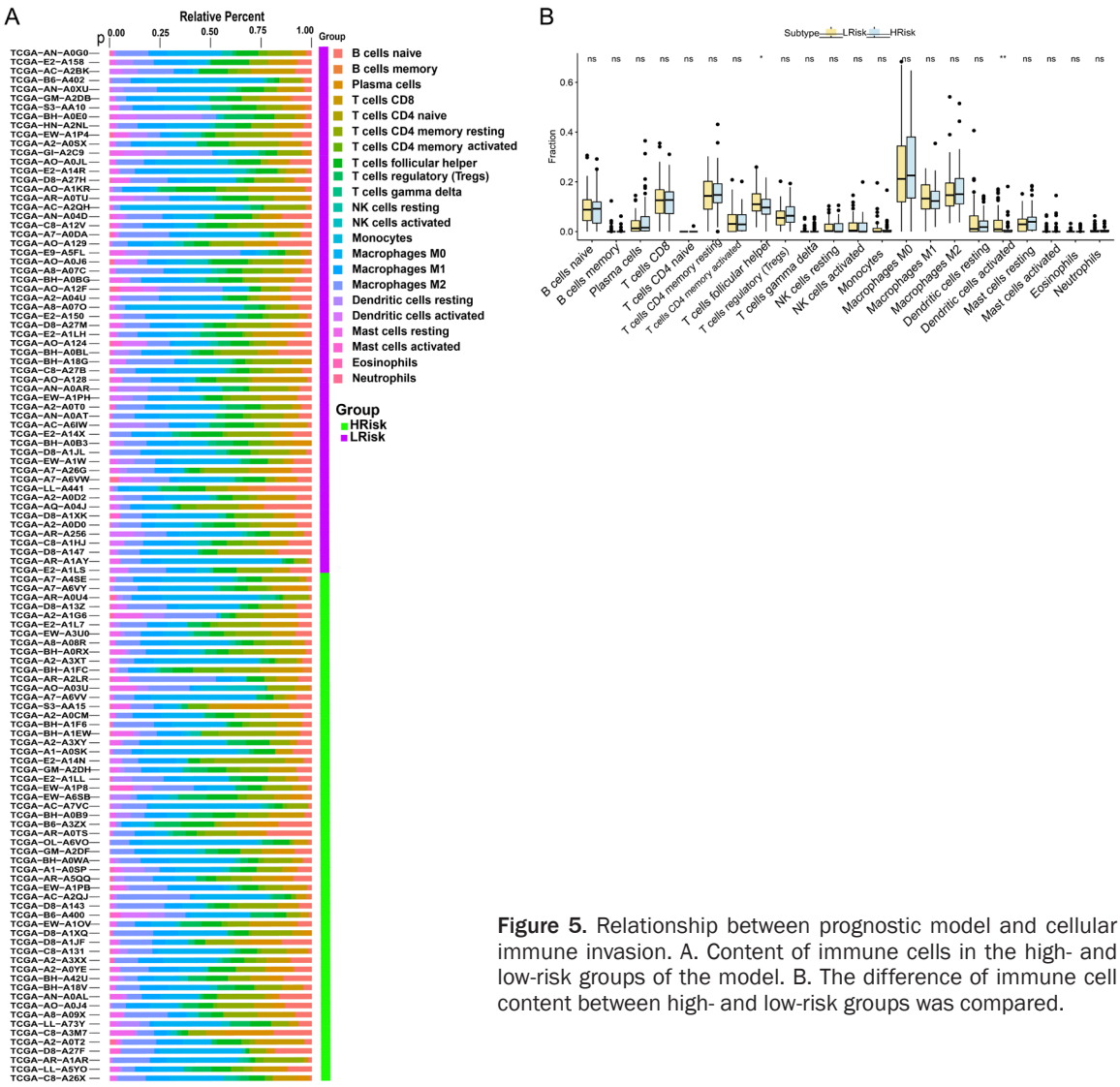

Figure 5. Relationship between prognostic model and cellular immune invasion. A. Content of immune cells in the high- and low-risk groups of the model. B. The difference of immune cell content between high- and low-risk groups was compared.

to the transcription factor HNF4A, setting the direction for our subsequent investigation. Further analysis revealed all enriched motifs of the key genes and their corresponding transcription factors (**Figure 16**). Reverse prediction by the miRcode database identified 24 mRNA-miRNA relationship pairs for DPM2 and 38 pairs for PINK1, involving a single transcription factor. The miRNA network and the transcription factor network related to these key genes were visualized using Cytoscape (**Figure 14B**), providing a comprehensive view of their regulatory landscape.

Clinical drug susceptibility study of prognostic model

For early-stage TNBC, the combination of chemotherapy and surgery is the prevalent treatment approach. After establishing the model, our aim was to predict the sensitivity of tumor samples to chemotherapy using the R package "pRRophetic", based on drug sensitivity data from the GDSC database. Additionally, we sought to explore the correlation between the risk score and the sensitivity to commonly used chemotherapy drugs. The findings revealed a noticeable correlation between the risk score and the sensitivity of patients to drugs such as ABT.263, Thapsigargin, NSC.87877, and Lenalidomide (**Figure 15**).

Discussion

To our knowledge, TNBC accounts for 8-15% of all breast cancer cases worldwide. It is not responsive to hormonal or anti-HER2 therapies

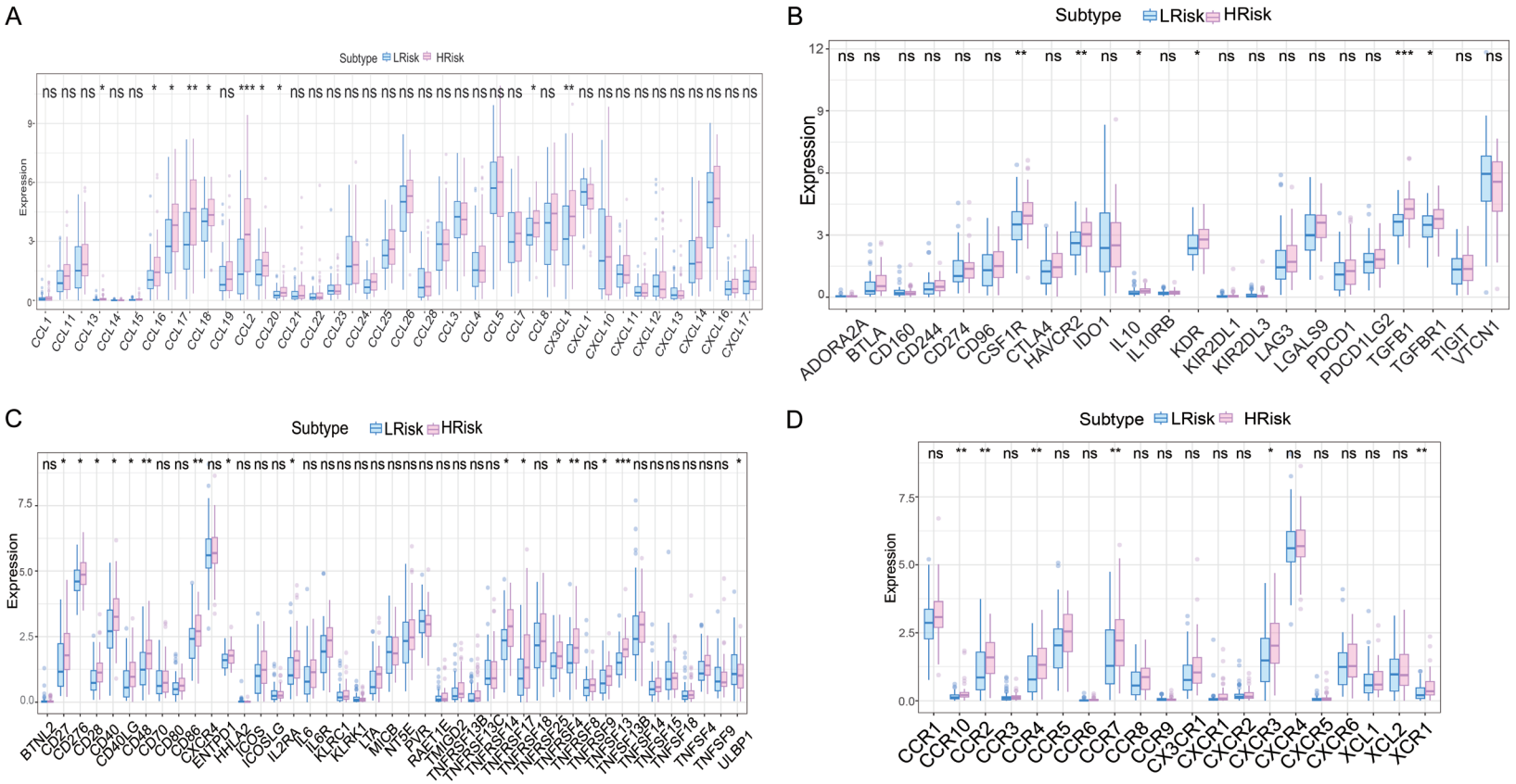

Figure 6. Relationship between prognostic models and immune regulatory genes. The differences in gene expression of immune-related chemokines, immunosuppressants, immunostimulatory factors, and immunoreceptors were described in high- and low-risk groups.

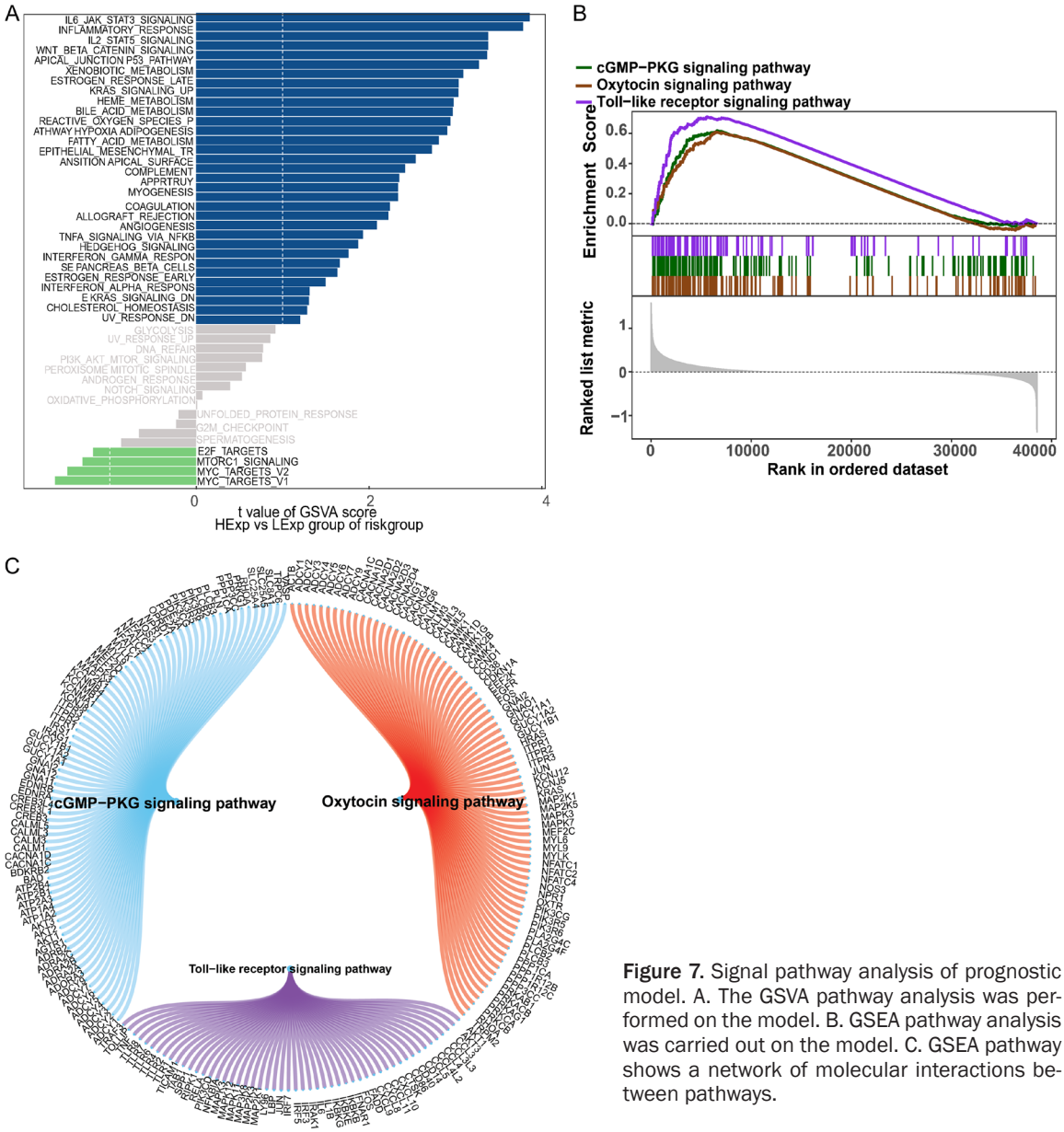

Figure 7. Signal pathway analysis of prognostic model. A. The GSVA pathway analysis was performed on the model. B. GSEA pathway analysis was carried out on the model. C. GSEA pathway shows a network of molecular interactions between pathways.

due to the absence of progesterone and estrogen receptors, and low HER2 expression levels. TNBC is an aggressive subtype with a poor prognosis compared to other breast cancer subtypes. Although it shows some sensitivity to chemotherapy, the efficacy is limited [13, 14]. Glycosylation of TNBC is a significant factor contributing to the poor prognosis of patients [15-17]. We hope that our TNBC glycosylation prognostic model will offer a new perspective on the prognosis of patients with triple-negative breast cancer and contribute novel insights to clinical research and treatment through the prediction of key genes.

We developed a TNBC glycosylation-related patient prognosis model using processed patient data from the TCGA database and validated it with an external database. The results confirm that the model effectively predicts patient disease progression with favorable prognostic accuracy. The model includes an immune assay, examining the immune cell ratio and the expression of immune-related regulatory genes. The analysis revealed significant differences in the populations of T cells follicular helper and activated dendritic cells between different risk score groups, with a higher proportion of activated cells in the low-risk group

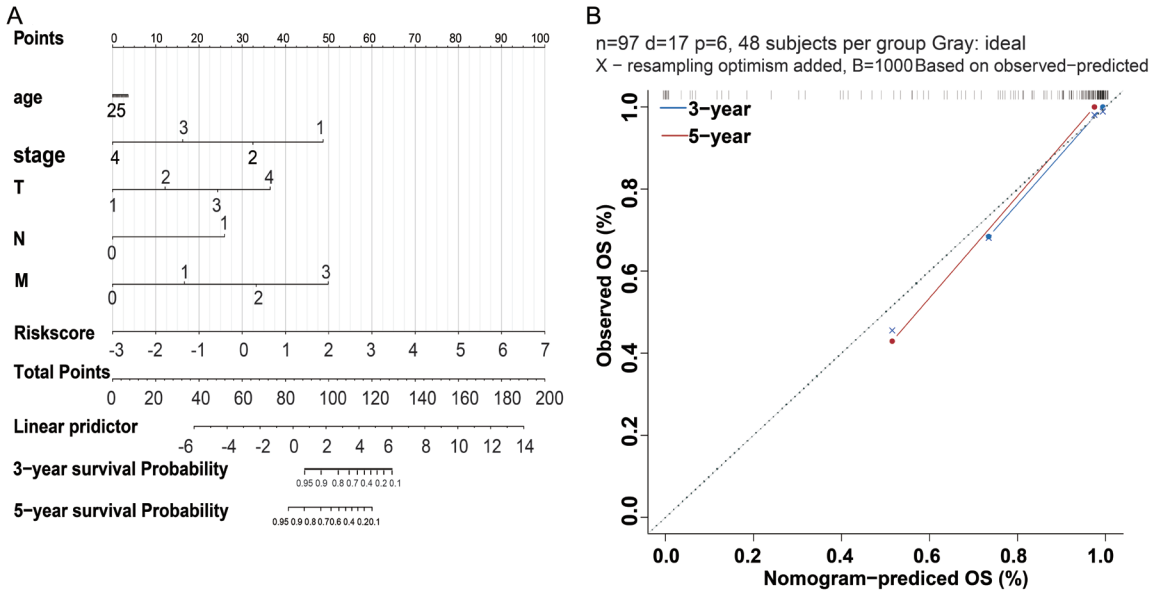

Figure 8. The clinical relevance of the prognostic model was analyzed. A. The risk scores for each stage of cancer. B. The OS of model patients was predicted by Norman diagram.

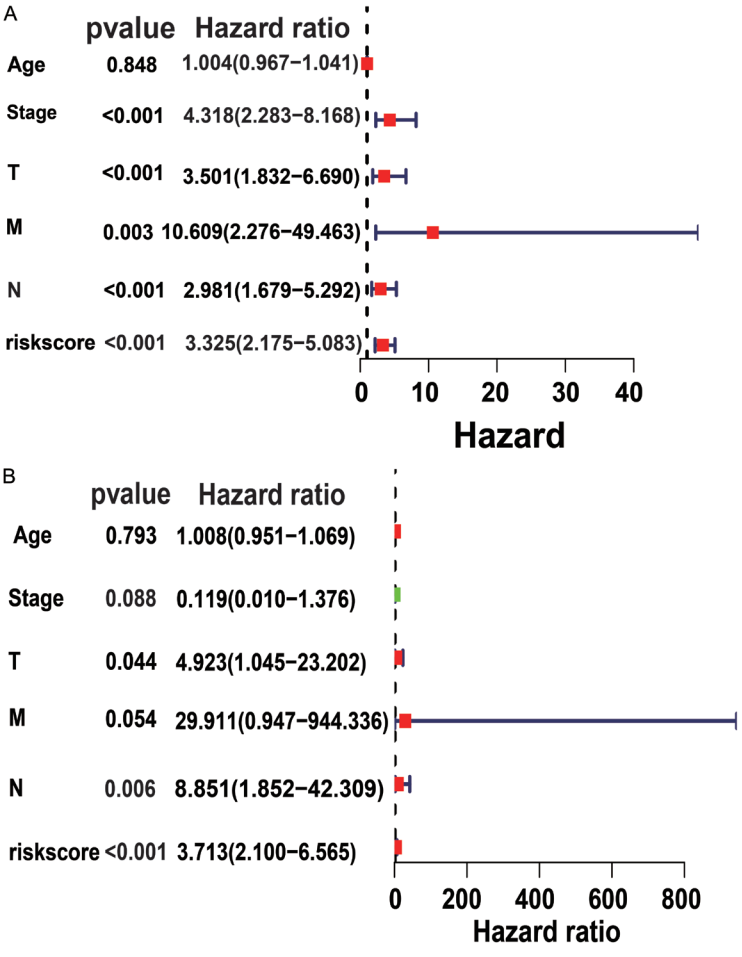

Figure 9. Risk score as an independent prognostic factor. A, B. Cox univariate and multivariate analyses were performed on the risk scores.

compared to the high-risk group. Both T follicular helper and activated dendritic cells are crucial for the protective role in the human immune environment and are central to initiating, regulating, and maintaining the immune response [18]. Further analysis of immunoregulatory genes showed differences in the expression of immune-related chemokines, immunosuppressants, immune-stimulating factors, and immune receptors across different risk-score groups. Notably, high expression of chemokines such as CCL18, CCL17, and CCL19 in the high-risk groups may be associated with a poorer prognosis for patients, and have been reported to play a role in tumor therapy [19].

The signaling pathways enriched through GSVA in the two groups with differing risk levels primarily involve IL2 (Interleukin-2) STAT5 SIGNALING, HEME METABOLISM, and IL6 (Interleukin-6) JAK/STAT3 SIGNALING. IL6 activates the JAK/STAT signaling

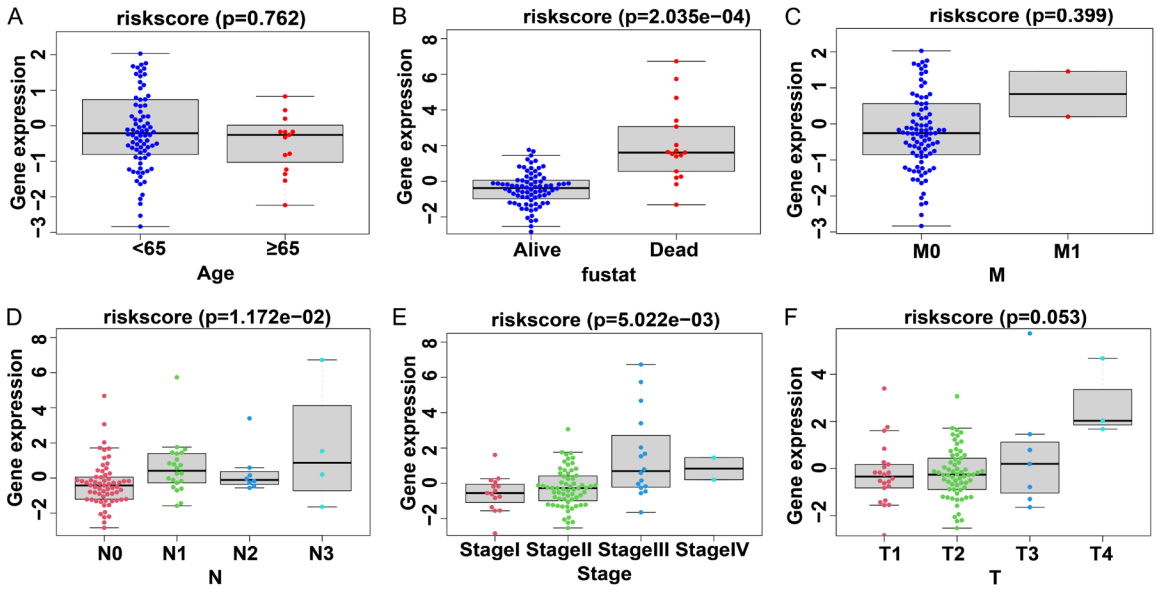

Figure 10. Clinical indicators of the differences between high- and low-risk groups.

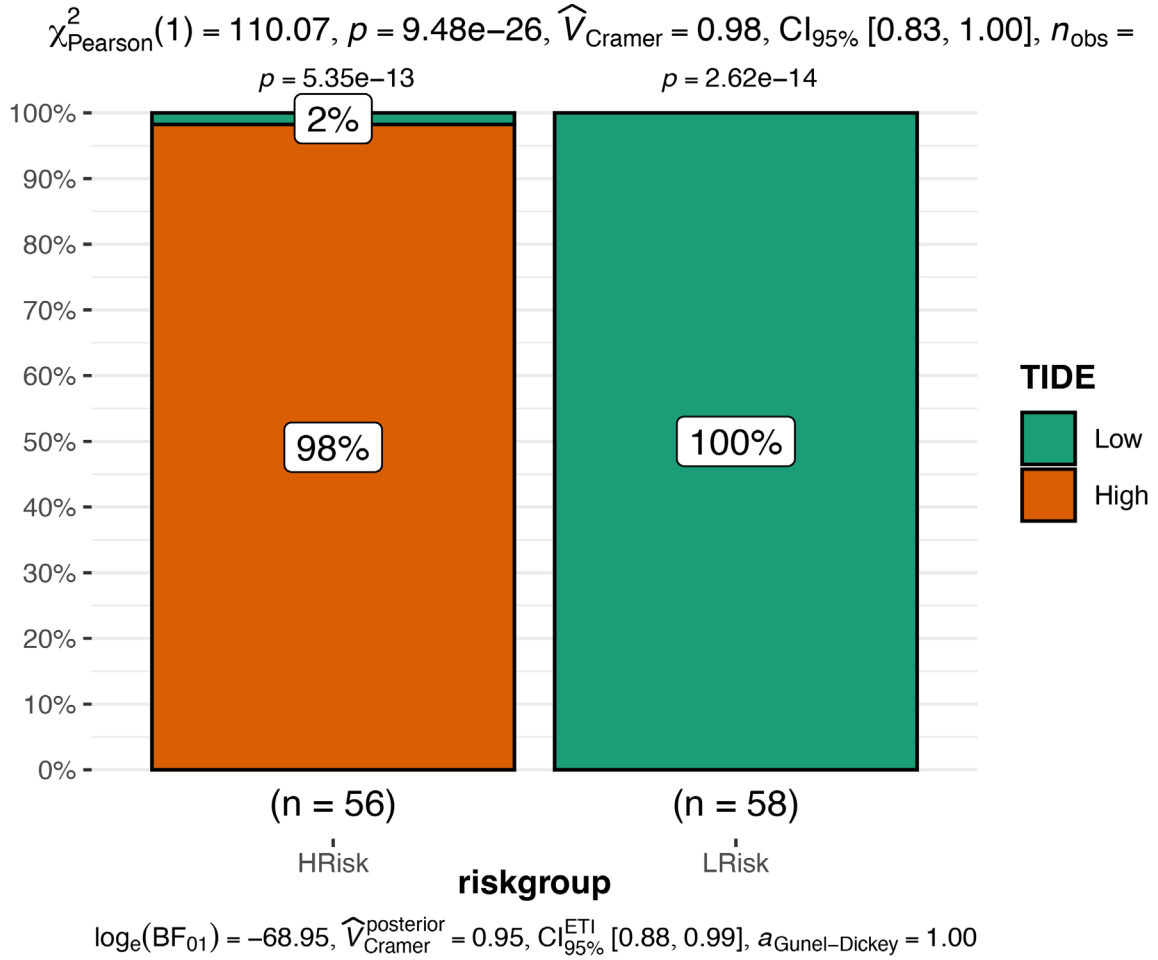

Figure 11. Difference of immunological efficacy in high- and low-risk groups.

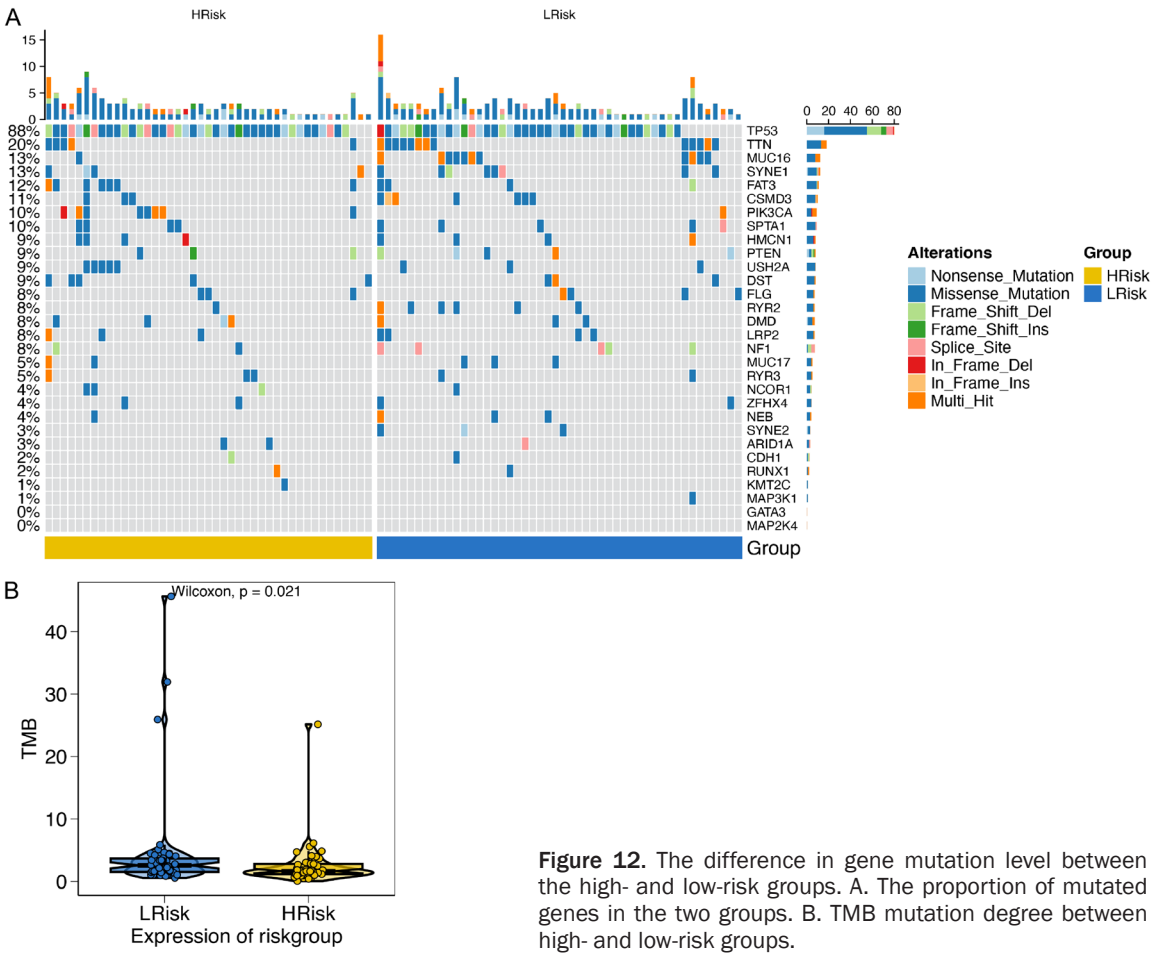

Figure 12. The difference in gene mutation level between the high- and low-risk groups. A. The proportion of mutated genes in the two groups. B. TMB mutation degree between high- and low-risk groups.

pathway, regulating cell proliferation, survival, and metastasis [20]. Hyperactivation of the IL6/JAK/STAT3 signaling pathway may be associated with tumor progression, invasion, and metastasis in TNBC. Moreover, IL2 within the IL2/STAT5 signaling pathway acts as a cytokine that regulates T cell development and function by stimulating the STAT5 signaling pathway [21]. Abnormal stimulation of the IL2/STAT5 signaling pathway in TNBC may be linked to immune escape and anti-apoptotic mechanisms in tumor cells [22]. Additionally, the HEME METABOLISM pathway in mammary cancer might influence tumor cell proliferation, invasion, drug resistance, and abnormal heme metabolism [23], which may be related to processes such as remodeling of tumor metabolism, oxidative stress, and immune escape [24]. The pathways implicated in GSEA analysis include the cGMP-PKG signaling pathway and the Toll-like receptor signaling pathway, which are associated with tumor immune response and inflammatory processes [25-27].

Through machine learning methods combined with Kaplan-Meier survival analysis, significant differences were identified in two key genes, DPM2 and PINK1. DPM2, located in the cytoplasm, encodes a hydrophobic protein with two predicted transmembrane domains and a putative endoplasmic reticulum (ER) localization signal near the C-terminus. This protein binds to DPM1 in vivo and is required for ER localization and stable expression. Its absence is one of the causes of congenital disorders of glycosylation in children; it is also one of the genes involved in glucose and lipid metabolism [28, 29]. PINK-1 is localized in the cytoplasm and is believed to be a mitochondrial membrane protein that maintains normal mitochondrial integrity, protecting cells against mitochondrial dysfunction [30]. Research indicates that PINK1 expression is associated with mitochondrial autophagy, regulating mitochondrial autophagy via the PARKIN pathway, playing a critical role in Parkinson's disease, neuroinflammatory disease, and activating E3 ubiquitin ligase in

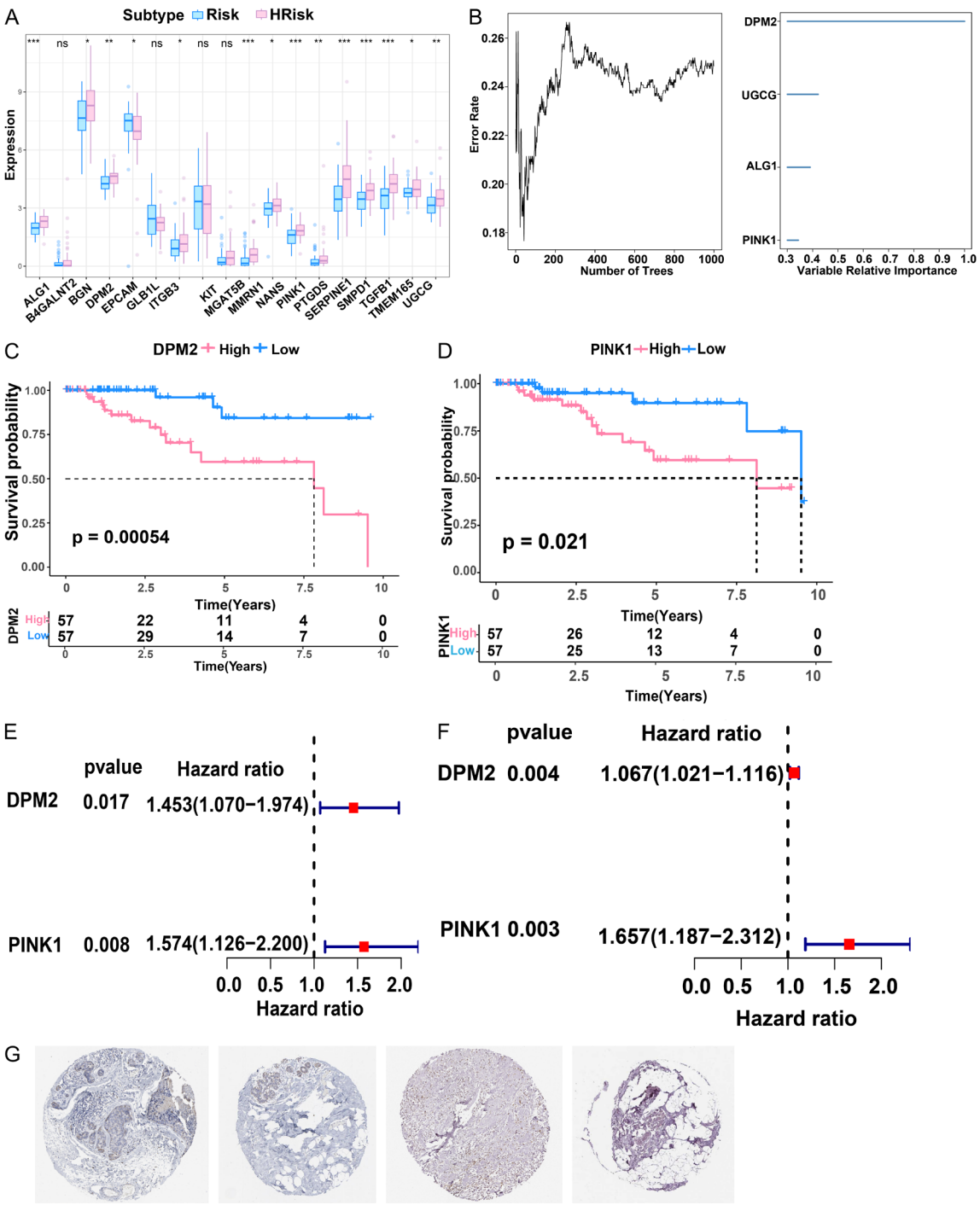

Figure 13. The prognostic model for screening key genes. A. The image shows the prognosis of gene expression in the model. B. The "random Forest-SRC" software package was used to screen and sequence the feature genes. C, D. The survival curves of the selected key genes were analyzed. E, F. Univariate and multivariate analysis of key genes. G. The Human Protein Atlas database's profiles of the key genes' protein expression.

Parkinson's disease [31]. In breast cancer, it is implicated in promoting tumor cell proliferation and migration [32]. These findings underscore the significance of our research, suggesting that the predicted key genes in TNBC glycosylation from our model may play a crucial role in tumor physiology, warranting further investigation in subsequent studies.

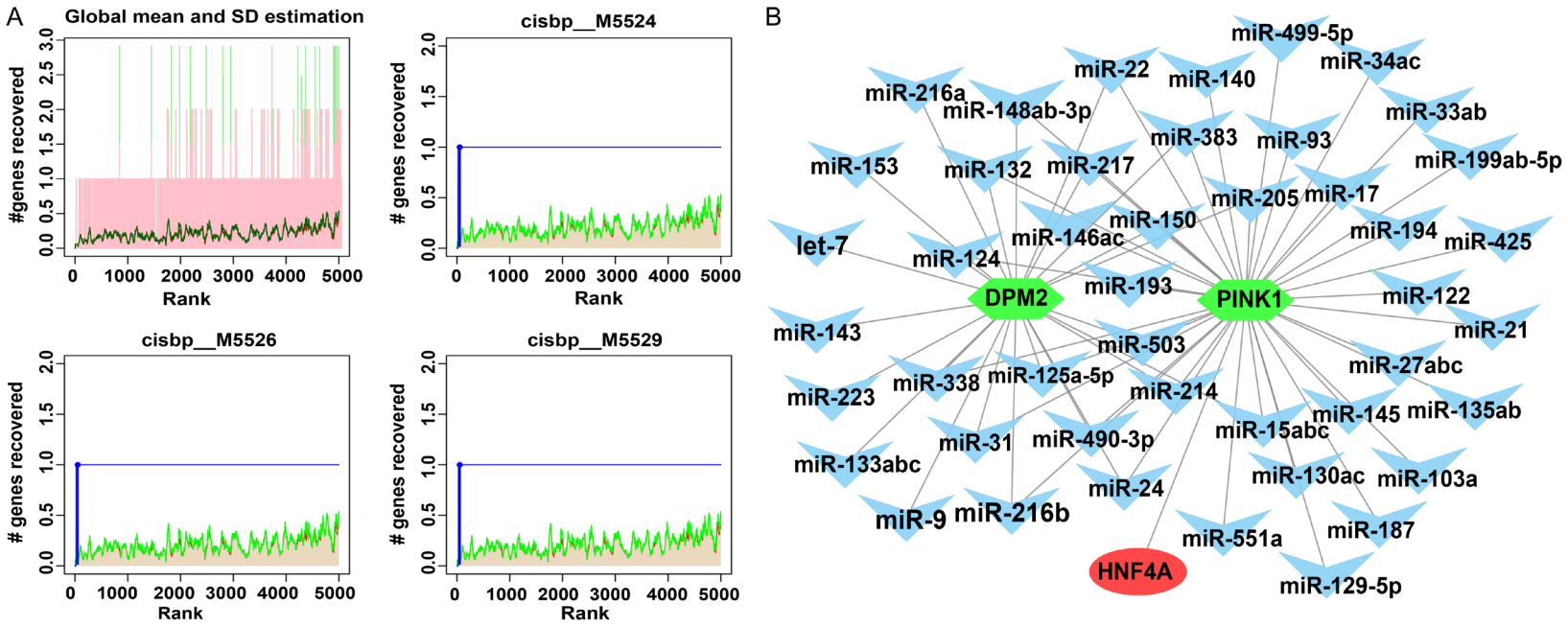

Figure 14. Transcription factors were screened and miRNA network was constructed. A. Transcription factors were predicted by the R package "RcisTarget". B. Visual gene networks of microRNAs built with cytoscape software.

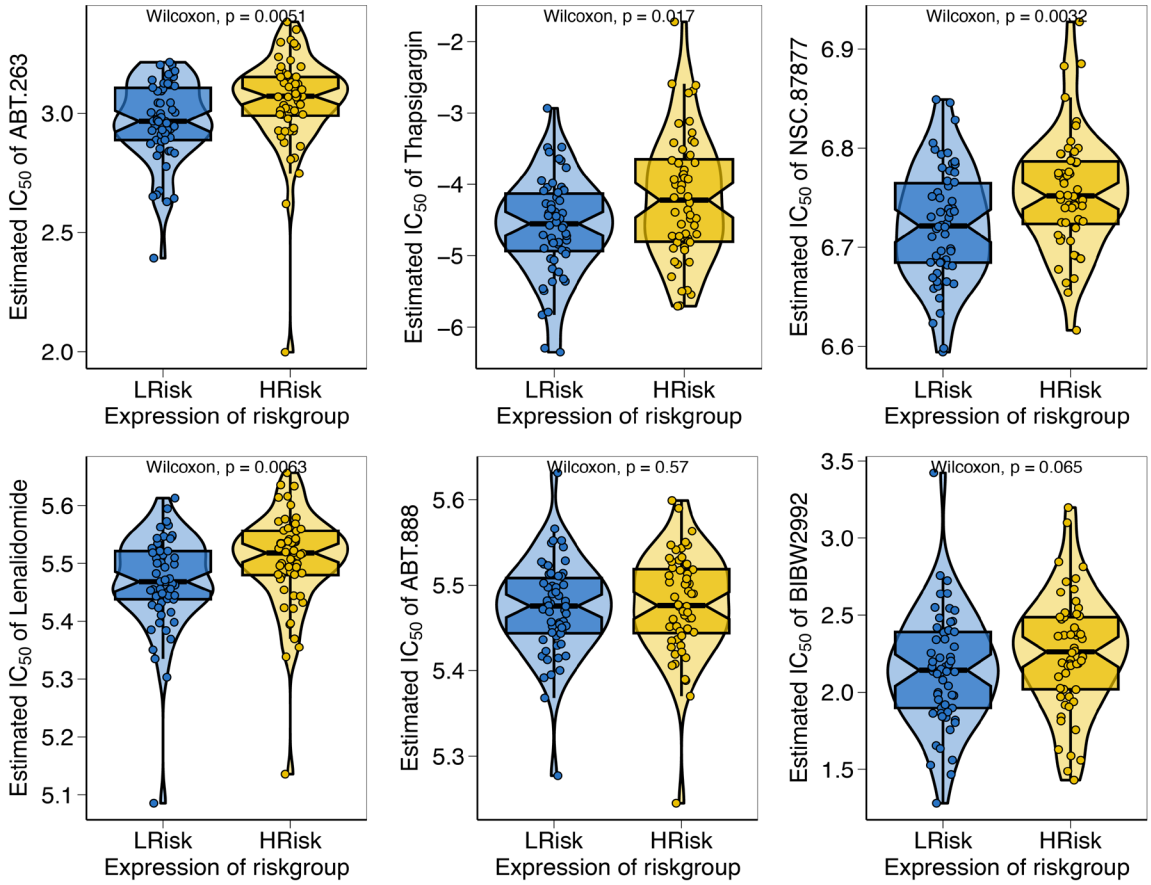

Figure 15. The sensitivity of the high- and low-risk groups to chemotherapy drugs in the model.

Show 5 entries Search:

	logo	geneSet	motif	NES	AUC	TF_highConf	nEnrGenes	enrichedGenes
	All	All	All	All	All	All	All	All
1		key_gene	cisbp__M5524	9.37	0.491	HNF4A (directAnnotation).	1	PINK1
2		key_gene	cisbp__M5526	9.35	0.49	HNF4A (directAnnotation).	1	PINK1
3		key_gene	cisbp__M5529	9.32	0.488	HNF4A (directAnnotation).	1	PINK1
4		key_gene	cisbp__M3415	8.91	0.467	HNF4A (directAnnotation).	1	PINK1
5		key_gene	cisbp__M1461	8.87	0.465		1	PINK1

Showing 1 to 5 of 10 entries Previous 1 2 Next

Figure 16. All enriched motifs of the key genes and their corresponding transcription factors.

Based on drug susceptibility data from the GDSC database, we conducted an intergroup analysis of sensitivity to cancer chemotherapy drugs according to the risk prediction score. Significant sensitivity differences were observed for drugs such as NSC.87877 [33], ABT-263 [34]. Lenalidomide, known for its efficacy in tumor treatment, exhibits multiple effects

including anti-tumor, immunomodulation, and anti-angiogenesis. It has shown positive outcomes in treating leukemia and lymphoma [35, 36]. Its combination with cisplatin analogs has demonstrated improved efficacy in TNBC treatment [37, 38]. Our findings suggest that the combination of triple-negative breast cancer with aerobic glycolysis negatively affects patient prognosis, immune cell infiltration, tumor microenvironment changes, and anti-tumor therapy efficacy. Additionally, we identified key genes influencing triple-negative breast cancer glycosylation, offering potential therapeutic opportunities for patients. However, our study has limitations: First, further clinical validation is needed due to the scarcity of cases involving TNBC combined with aerobic glycolysis. Second, drugs identified through GDSC database sensitivity data were not experimentally tested in this study. Third, the key genes obtained by constructing the prognostic model need to be verified through clinical trials to determine the reliability of the risk score model. Thus, our results require confirmation by subsequent clinical studies.

Conclusion

Our prognostic model was validated to predict outcomes for patients with glycosylated triple-negative breast cancer. We analyzed the correlation between prognostic factors and risk scores, including aspects such as immune cell infiltration, the tumor microenvironment, and the efficacy of antitumor therapy. In addition to providing useful indicators, a few key genes affecting the prognosis of glycosylation in TNBC were predicted. Thus, our findings will help to enhance therapeutic options for patients with glycosylated triple-negative breast cancer.

Acknowledgements

For their contributions to the current work, we are grateful to all of the authors cited. This study was supported by the National Natural Science Foundation of China (No. 82060520) and the Natural Science Foundation of Xinjiang Uygur Autonomous Region (2022D01C309).

Disclosure of conflict of interest

None.

Address correspondence to: Gang Sun, Department of Breast and Thyroid Surgery, The Affiliated Cancer Hospital of Xinjiang Medical University, Urumqi 830011, Xinjiang, China. E-mail: sungang@xjmu.edu.cn

References

[1] Garrido-Castro AC, Lin NU and Polyak K. Insights into molecular classifications of triple-negative breast cancer: improving patient selection for treatment. Cancer Discov 2019; 9: 176-198.

[2] Yin L, Duan JJ, Bian XW and Yu SC. Triple-negative breast cancer molecular subtyping and treatment progress. Breast Cancer Res 2020; 22: 61.

[3] Very N, Lefebvre T and El Yazidi-Belkoura I. Drug resistance related to aberrant glycosylation in colorectal cancer. Oncotarget 2017; 9: 1380-1402.

[4] Christiansen MN, Chik J, Lee L, Anugraham M, Abrahams JL and Packer NH. Cell surface protein glycosylation in cancer. Proteomics 2014; 14: 525-546.

[5] de Freitas Junior JC and Morgado-Díaz JA. The role of N-glycans in colorectal cancer progression: potential biomarkers and therapeutic applications. Oncotarget 2016; 7: 19395-19413.

[6] Deepak KGK, Vempati R, Nagaraju GP, Dasari VR, S N, Rao DN and Malla RR. Tumor microenvironment: challenges and opportunities in targeting metastasis of triple negative breast cancer. Pharmacol Res 2020; 153: 104683.

[7] Liubomirski Y, Lerrer S, Meshel T, Rubinstein-Achiasaf L, Morein D, Wiemann S, Körner C and Ben-Baruch A. Tumor-stroma-inflammation networks promote pro-metastatic chemokines and aggressiveness characteristics in triple-negative breast cancer. Front Immunol 2019; 10: 757.

[8] Chang R, Song L, Xu Y, Wu Y, Dai C, Wang X, Sun X, Hou Y, Li W, Zhan X and Zhan L. Loss of Wwox drives metastasis in triple-negative breast cancer by JAK2/STAT3 axis. Nat Commun 2018; 9: 3486.

[9] Lv Y, Wang X, Li X, Xu G, Bai Y, Wu J, Piao Y, Shi Y, Xiang R and Wang L. Nucleotide de novo synthesis increases breast cancer stemness and metastasis via cGMP-PKG-MAPK signaling pathway. PLoS Biol 2020; 18: e3000872.

[10] Fitzgerald KA and Kagan JC. Toll-like receptors and the control of immunity. Cell 2020; 180: 1044-1066.

[11] Han X, Chen J, Wang J, Xu J and Liu Y. TTN mutations predict a poor prognosis in patients with thyroid cancer. Biosci Rep 2022; 42: BSR20221168.

[12] Lin G, Liu X, Cong C and Xu L. Prognostic significance of long noncoding RNA TTN-AS1 in various malignancies. Cancer Rep (Hoboken) 2023; 6: e1876.
[13] Ferrari P, Scatena C, Ghilli M, Bargagna I, Lorenzini G and Nicolini A. Molecular mechanisms, biomarkers and emerging therapies for chemotherapy resistant TNBC. Int J Mol Sci 2022; 23: 1665.
[14] Li Y, Zhang H, Merkher Y, Chen L, Liu N, Leonov S and Chen Y. Recent advances in therapeutic strategies for triple-negative breast cancer. J Hematol Oncol 2022; 15: 121.
[15] Huang Y, Zhang HL, Li ZL, Du T, Chen YH, Wang Y, Ni HH, Zhang KM, Mai J, Hu BX, Huang JH, Zhou LH, Yang D, Peng XD, Feng GK, Tang J, Zhu XF and Deng R. FUT8-mediated aberrant N-glycosylation of B7H3 suppresses the immune response in triple-negative breast cancer. Nat Commun 2021; 12: 2672.
[16] Zhou R, Yazdanifar M, Roy LD, Whilding LM, Gavrill A, Maher J and Mukherjee P. CAR T cells targeting the tumor MUC1 glycoprotein reduce triple-negative breast cancer growth (vol 10, 1149, 2019). Front Immunol 2020; 11: 628776.
[17] Scott DA and Drake RR. Glycosylation and its implications in breast cancer. Expert Rev Proteomics 2019; 16: 665-680.
[18] Yudintceva N, Bobkov D, Sulatsky M, Mikhailova N, Oganesyan E, Vinogradova T, Muraviov A, Remezova A, Bogdanova E, Garapach I, Maslak O, Esmedlyaeva D, Dyakova M, Yablonskiy P, Ziganshin R, Kovalchuk S, Blum N, Sonawane SH, Sonawane A, Behl A, Shailja Singh and Shevtsov M. Mesenchymal stem cells-derived extracellular vesicles for therapeutics of renal tuberculosis. Sci Rep 2024; 14: 4495.
[19] Korbecki J, Kojder K, Siminska D, Bohatyrewicz R, Gutowska I, Chlubek D and Baranowska-Bosiacka I. CC chemokines in a tumor: a review of pro-cancer and anti-cancer properties of the ligands of receptors CCR1, CCR2, CCR3, and CCR4. Int J Mol Sci 2020; 21: 8412.
[20] Chen M and Wang S. Preclinical development and clinical studies of targeted JAK/STAT combined anti-PD-1/PD-L1 therapy. Int Immunopharmacol 2024; 130: 111717.
[21] Lutz V, Hellmund VM, Picard FSR, Raifer H, Ruckenbrod T, Klein M, Bopp T, Savai R, Duewell P, Keber CU, Weigert A, Chung HR, Buchholz M, Menke A, Gress TM, Huber M and Bauer C. IL18 receptor signaling regulates tumor-reactive CD8+ T-cell exhaustion via activation of the IL2/STAT5/mTOR pathway in a pancreatic cancer model. Cancer Immunol Res 2023; 11: 421-434.
[22] Marra P, Mathew S, Grigoriadis A, Wu Y, Kyle-Cezar F, Watkins J, Rashid M, De Rinaldis E, Hessey S, Gazinska P, Hayday A and Tutt A. IL-15RA drives antagonistic mechanisms of cancer development and immune control in lymphocyte-enriched triple-negative breast cancers. Cancer Res 2014; 74: 4908-4921.
[23] Kaur P, Nagar S, Bhagwat M, Uddin M, Zhu Y, Vancurova I and Vancura A. Activated heme synthesis regulates glycolysis and oxidative metabolism in breast and ovarian cancer cells. PLoS One 2021; 16: e0260400.
[24] Hardison RL, Heimlich DR, Harrison A, Beatty WL, Rains S, Moseley MA, Thompson JW, Justice SS and Mason KM. Transient nutrient deprivation promotes macropinocytosis-dependent intracellular bacterial community development (vol 3, e00286-18, 2018). mSphere 2018; 3: e00562-18.
[25] Kaur A, Baldwin J, Brar D, Salunke DB and Petrovsky N. Toll-like receptor (TLR) agonists as a driving force behind next-generation vaccine adjuvants and cancer therapeutics. Curr Opin Chem Biol 2022; 70: 102172.
[26] Piazza GA, Ward A, Chen X, Maxuitenko Y, Coley A, Aboelella NS, Buchsbaum DJ, Boyd MR, Keeton AB and Zhou G. PDE5 and PDE10 inhibition activates cGMP/PKG signaling to block Wnt/beta-catenin transcription, cancer cell growth, and tumor immunity. Drug Discov Today 2020; 25: 1521-1527.
[27] Shetab Boushehri MA and Lamprecht A. TLR4-based immunotherapeutics in cancer: a review of the achievements and shortcomings. Mol Pharm 2018; 15: 4777-4800.
[28] Barone R, Aiello C, Race V, Morava E, Foulquier F, Riemersma M, Passarelli C, Concolino D, Carella M, Santorelli F, Vleugels W, Mercuri E, Garozzo D, Sturiale L, Messina S, Jaeken J, Fiumara A, Wevers RA, Bertini E, Matthijs G and Lefeber DJ. DPM2-CDG: a muscular dystrophy-dystroglycanopathy syndrome with severe epilepsy. Ann Neurol 2012; 72: 550-558.
[29] Nagy S, Lau T, Alavi S, Karimiani EG, Vallian J, Ng BG, Noroozi Asl S, Akhondian J, Bahreini A, Yaghini O, Uapinyoying P, Bonnemann C, Freeze HH, Dissanayake VHW, Sirisena ND, Schmidts M, Houlden H, Moreno-De-Luca A and Maroofian R. A recurrent homozygous missense DPM3 variant leads to muscle and brain disease. Clin Genet 2022; 102: 530-536.
[30] Wang N, Zhu P, Huang R, Wang C, Sun L, Lan B, He Y, Zhao H and Gao Y. PINK1: the guard of mitochondria. Life Sci 2020; 259: 118247.
[31] Kane LA, Lazarou M, Fogel AI, Li Y, Yamano K, Sarraf SA, Banerjee S and Youle RJ. PINK1 phosphorylates ubiquitin to activate Parkin E3

ubiquitin ligase activity. J Cell Biol 2014; 205: 143-153.
[32] Li Q, Chu Y, Li S, Yu L, Deng H, Liao C, Liao X, Yang C, Qi M, Cheng J, Chen G and Huang L. The oncoprotein MUC1 facilitates breast cancer progression by promoting Pink1-dependent mitophagy via ATAD3A destabilization. Cell Death Dis 2022; 13: 899.
[33] Song M, Park JE, Park SG, Lee DH, Choi HK, Park BC, Ryu SE, Kim JH and Cho S. NSC-87877, inhibitor of SHP-1/2 PTPs, inhibits dual-specificity phosphatase 26 (DUSP26). Biochem Biophys Res Commun 2009; 381: 491-495.
[34] Ritter V, Krautter F, Klein D, Jendrossek V and Rudner J. Bcl-2/Bcl-xL inhibitor ABT-263 overcomes hypoxia-driven radioresistence and improves radiotherapy. Cell Death Dis 2021; 12: 694.
[35] Leonard JP, Trneny M, Izutsu K, Fowler NH, Hong X, Zhu J, Zhang H, Offner F, Scheliga A, Nowakowski GS, Pinto A, Re F, Fogliatto LM, Scheinberg P, Flinn IW, Moreira C, Cabecadas J, Liu D, Kalambakas S, Fustier P, Wu C and Gribben JG; AUGMENT Trial Investigators. AUGMENT: a phase III study of lenalidomide plus rituximab versus placebo plus rituximab in relapsed or refractory indolent lymphoma. J Clin Oncol 2019; 37: 1188-1199.
[36] Lonial S, Jacobus S, Fonseca R, Weiss M, Kumar S, Orlowski RZ, Kaufman JL, Yacoub AM, Buadi FK, O'Brien T, Matous JV, Anderson DM, Emmons RV, Mahindra A, Wagner LI, Dhodapkar MV and Rajkumar SV. Randomized trial of lenalidomide versus observation in smoldering multiple myeloma. J Clin Oncol 2020; 38: 1126-1137.
[37] Brosseau C, Colston K, Dalgleish AG and Galustian C. The immunomodulatory drug lenalidomide restores a vitamin D sensitive phenotype to the vitamin D resistant breast cancer cell line MDA-MB-231 through inhibition of BCL-2: potential for breast cancer therapeutics. Apoptosis 2012; 17: 164-173.
[38] Hao Ing Y, Md Salleh MS, Yahya MM, Ankathil R and Abdul Aziz AA. Association of ABCG2 polymorphisms on triple negative breast cancer (TNBC) susceptibility risk. Asian Pac J Cancer Prev 2023; 24: 3891-3897.