Original Article

Construction of a prognostic model and analysis of related mechanisms in breast cancer based on multiple datasets

Xiaofeng Wan[1], Jianmin Zhan[1], Shuke Ye[1], Chuanrong Chen[2], Runsheng Li[1], Ming Shen[1]

[1]National Health Commission (NHC) Key Lab of Reproduction Regulation (Shanghai Institute for Biomedical and Pharmaceutical Technologies), Shanghai, China; [2]Department of Oncology, Yijishan Hospital of Wannan Medical College, Wuhu, China

Contributions: (I) Conception and design: X Wan, M Shen, C Chen; (II) Administrative support: M Shen, R Li; (III) Provision of study materials or patients: C Chen; (IV) Collection and assembly of data: X Wan, J Zhan, S Ye; (V) Data analysis and interpretation: X Wan, M Shen; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Ming Shen, PhD. National Health Commission (NHC) Key Lab of Reproduction Regulation (Shanghai Institute for Biomedical and Pharmaceutical Technologies), 2140 Xietu Rd., Shanghai 200032, China. Email: shenming711@163.com.

Background: Breast cancer (BC) is a common tumor among women and is a heterogeneous disease with many subtypes. Each subtype shows different clinical presentations, disease trajectories and prognoses, and different responses to neoadjuvant therapy; thus, a new and universal prognostic biomarker for BC patients is urgently needed. Our goal is to identify a novel prognostic molecular biomarker that can accurately predict the outcome of all BC subtypes and guide their clinical management.

Methods: Utilizing data from The Cancer Genome Atlas (TCGA), we analyzed differential gene expression and patient clinical data. Weighted gene coexpression network analysis (WGCNA), Cox univariate regression and least absolute shrinkage and selection operator (LASSO) analysis were used to construct a prognostic model; the differential expression of the core genes in this model was validated via real-time quantitative polymerase chain reaction (RT-qPCR), and the reliability of the predictive model was validated in both an internal cohort and a BC patient dataset from the Gene Expression Omnibus (GEO) database. Further studies, such as gene set variation analysis (GSVA) and gene set enrichment analysis (GSEA), were performed to investigate the enrichment of signaling pathways. The CIBERSORT algorithm was used to estimate immune infiltration and tumor mutation burden (TMB), and drug sensitivity analysis was performed to evaluate the treatment response.

Results: A total of 1,643 differentially expressed genes were identified. After WGCNA and Cox regression combined with LASSO analysis, 15 genes were identified by screening and used to establish a prognostic gene signature. Further analysis revealed that the epithelial-mesenchymal transition (EMT) pathway gene signature was enriched in these genes. Each patient was assigned a risk score, and according to the median risk score, patients were classified into a high-risk group or a low-risk group. The prognosis of the low-risk group was better than that of the high-risk group ($P<0.01$), and analyses of two independent GEO validation cohorts yielded similar results. Furthermore, a nomogram was constructed and found to perform well in predicting prognosis. GSVA revealed that the EMT pathway, transforming growth factor β (TGF-β) signaling pathway and PI3K-Akt signaling pathway genes were enriched in the high-risk group, and the Wnt-β-catenin signaling pathway, DNA repair pathway and P53 pathway gene sets were enriched in the low-risk group. GSEA revealed genes related to TGF-β signaling and the PI3K-Akt signaling pathways were enriched in the high-risk group. CIBERSORT demonstrated that the low-risk group had greater infiltration of antitumor immune cells. The TMB and drug sensitivity results suggested that immunotherapy and chemotherapy are likely to be more effective in the low-risk group.

Conclusions: We established a new EMT pathway-related prognostic gene signature that can be used to effectively predict BC prognosis and treatment response.

Keywords: Breast cancer (BC); prognostic model; epithelial-mesenchymal transition pathway (EMT pathway)

Submitted May 23, 2024. Accepted for publication Dec 26, 2024. Published online Feb 24, 2025.
doi: 10.21037/tcr-24-838
View this article at: https://dx.doi.org/10.21037/tcr-24-838

Introduction

Breast cancer (BC) is a common form of cancer among women, with a relatively small percentage of cases occurring in men. Currently, there are more than 3.8 million BC survivors in the United States (1). As of 2023, its incidence rate is still increasing at an annual rate of approximately 0.6% (2). The pathogenesis of BC is complex and involves estrogen, genetic factors, family history, aging, the late menopause, reproductive factors, lifestyle, etc. (3-5).

BC is a heterogeneous disease. BC can be divided into several subtypes according to the expression of molecular markers, namely, estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2); the triple-negative BC (TNBC) subtype is negative for all three receptors. Compared with other BC subtypes, TNBC has a poorer prognosis, easy recurrence and a low overall survival rate. TNBC can be further divided into different subtypes (6-10). The treatment of BC is mainly based on staging, combining molecular and clinical characteristics; the treatment methods include surgical intervention, chemotherapy, radiotherapy, endocrine therapy, targeted therapy, antibody-drug conjugates and immune checkpoint inhibitors (11,12). Although advancements in early detection and treatment strategies have significantly improved clinical outcomes, major challenges remain. Specifically, studies are needed to address heterogeneity in terms of tumor features and neoadjuvant therapy response, improve the pathological complete response (pCR) rate, reduce the recurrence risk of high-risk early BC, and facilitate personalized treatment. Therefore, identification of a new universal biomarker for prognosis and survival prediction is urgently needed in order to guide the clinical management of BC. The use of such a biomarker may prevent overtreatment, minimize the toxicity and economic burden of treatment, and improve the curative effect, especially in cases with a good prognosis.

Here, we constructed a BC prognosis prediction model consisting of 15 genes based on the data of BC patients in The Cancer Genome Atlas (TCGA) database via various analysis methods and verified the stability of the model with BC patient data from the Gene Expression Omnibus (GEO) database. Our results indicate that the model performs well in predicting BC prognosis and has good reference significance for the selection of chemotherapy drugs and immunotherapy. A nomogram was also constructed and confirmed to have good predictive ability in the clinic. This manuscript is written following the TRIPOD reporting checklist (available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-838/rc).

Highlight box

Key findings

- Breast cancer (BC) is a heterogeneous disease, and patients respond differently to conventional treatments, with some experiencing overtreatment and others undertreatment. This study has developed a gene signature composed of 15 genes that can effectively predict the prognosis of BC patients.

What is known and what is new?

- The heterogeneity of BC is manifested in various aspects, bringing troubles to clinical management.
- We have established a prognostic molecular signature for BC that comprises 15 genes and is applicable to all subtypes of BC patients.

What is the implication, and what should change now?

- We have established a 15-gene signature that can predict the prognosis of all BC subtypes and guide their clinical management.

Methods

Data download

We downloaded processed raw RNA expression data of patients with BC from TCGA database (https://portal.gdc.cancer.gov/); the dataset included a normal group (n=113) and a tumor group (n=1,118). Additionally, we downloaded series matrix data files from the National Center for Biotechnology Information (NCBI) GEO public database (http://www.ncbi.nlm.nih.gov/geo/info/datasets.html); the obtained datasets (GSE20685 and GSE42568) were obtained via the GPL570 platform and include expression profile data for 327 patients and 104 patients, respectively. Clinical data and survival data were also available for these patients and were used to validate the prognostic model. The TCGA database is the largest cancer gene information database and includes gene expression, microRNA (miRNA)

expression, long noncoding RNA (lncRNA) expression, copy number variation, DNA methylation, single nucleotide polymorphism (SNP) data, etc. The GEO database is maintained by NCBI in the U.S.

Differential expression analysis

The Limma package in R was used to analyze differential gene expression to identify genes with significant differences between groups. By applying Limma, we identified differentially expressed genes between control and tumor samples with the criteria of P value <0.05 and |log fold change (FC)| >1. Furthermore, we visualized the results through differential gene volcano plots and heatmaps.

Differential gene function analysis

The Metascape database (https://www.Metascape.org) was used for functional annotation of the differentially expressed genes to investigate their functional relationships thoroughly. Gene ontology pathway analysis was performed on specific genes, and statistically significant results were indicated by a minimum overlap of ≥3 and a P value of ≤0.01.

Weighted gene coexpression network analysis (WGCNA)

A gene coexpression network was created via the WGCNA-R package, which involves identifying gene modules that exhibit coexpression patterns and investigating the connections between gene networks and key genes. Among all genes, the top 5,000 genes with the highest variances were selected, and a soft threshold of 4 was applied to establish the network. The weighted adjacency matrix was transformed into a topological overlap matrix (TOM) to evaluate network connectivity, and hierarchical clustering was used to construct a tree-like clustering structure of the TOM. Each branch of the tree signifies distinct gene modules, distinguished by unique colors. Genes are grouped into modules on the basis of their expression profiles, ensuring that genes with similar patterns are clustered together. This process allows the visualization of multiple modules that represent various gene expression patterns.

Model construction and prognosis

Genes at intersections were chosen and subjected to least absolute shrinkage and selection operator (LASSO) Cox regression analysis to construct prognostic models. The expression levels of each gene, weighted by the estimated regression coefficients derived from the LASSO analysis, were used to generate a formula for calculating the risk score for each patient: $risk\ score = \sum_{i=1}^{15}(coefficient_i \times expression\ of\ core\ gene_i)$, where i corresponds to each individual core gene. Patients were subsequently categorized into low-risk and high-risk group according to their risk scores, and the median was used as the threshold. Survival disparities between the two cohorts were assessed through Kaplan-Meier analysis and compared via log-rank tests. LASSO regression and stratified analysis were performed to evaluate the ability of the risk score to predict prognosis. Receiver operating characteristic (ROC) curves were employed to assess the precision of the model predictions.

Immune cell infiltration analysis between the high- and low-score groups

The CIBERSORT technique is widely utilized for assessing the various types of immune cells present in microenvironments. By leveraging support vector regression, deconvolution analysis can be conducted on matrices reflecting the expression of different immune cell subtypes. With this approach, a set of 547 biomarkers is used to distinguish 22 human immune cell phenotypes, including T cells, B cells, plasma cells, and various myeloid cell subsets. In this study, to evaluate the relationship between the prognosis-related gene score and immune cell infiltration, the CIBERSORT algorithm was used to estimate the relative proportions of 22 distinct types of infiltrating immune cells and perform correlation analysis between gene expression and immune cell composition.

Drug sensitivity analysis between the high- and low-score groups

We used the Genomics of Drug Sensitivity in Cancer (GDSC) database (http://www.cancerrxgene.org/), along with the R software package "pRRopetic", to predict the sensitivity of individual tumor samples to chemotherapy. Half-maximal inhibitory concentration (IC_{50}) values for specific chemotherapy drugs were estimated using regression analysis. To assess the accuracy of the regression and predictions, we performed 10 cross-validation tests on the GDSC training set, with default parameter settings such as "combo" to eliminate batch effects and to average

duplicate gene expression values.

Gene set variation analysis (GSVA)

By utilizing the GSVA algorithm, the biological functional changes in different samples on the basis of their gene expression profiles were evaluated by scoring gene sets obtained from the Molecular Signatures database (v7.0 version). GSVA is a nonparametric and unsupervised approach used to assess the enrichment of transcriptome gene sets through the conversion of individual gene changes into pathway-level alterations.

Gene set enrichment analysis (GSEA)

Patients were categorized into high- and low-risk groups according to the model risk score; then, the differentially expressed genes between groups were subjected to GSEA to reveal differentially enriched signaling pathways. Reference gene sets corresponding to pathways were obtained from the Molecular Signatures database. Subsequently, differential expression analysis of pathways between subtypes was performed, and significantly enriched gene sets (adjusted P value <0.05) were ranked on the basis of consistency scores. GSEA is a valuable tool for further analysis of the biological features and processes related to tumor classification.

Tumor mutation burden (TMB) analysis

TMB is determined by counting the total number of identified mutations in the coding sequences of somatic genes per million bases. In this research, TMB was calculated by dividing the number of nonsynonymous mutation sites by the total length of the protein coding region; the frequency of variations and the number of variations per exon length for each tumor sample were also calculated.

Nomogram model construction

A nomogram was generated via regression analysis to visualize the relationships between variables in the prediction model by combining the risk score and clinical features. In this process, a multifactor regression model was applied to assign scores to different factors according to their impact on the outcome variable, and the total score was calculated by summing the scores for each factor to predict the outcome.

Clinical sample acquisition and real-time quantitative polymerase chain reaction (RT-qPCR) analysis of core genes identified by LASSO analysis

We recruited three patients with TNBC from Yijishan Hospital of Wannan Medical College. Carcinoma tissues and adjacent tissues obtained by operation were each approximately 5–10 mm^3 in size and were placed into an EP tube filled with 1 mL of TRIzol, marked, and stored at –80 ℃ for subsequent total RNA extraction. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Medical Ethics board of Wannan Medical College (No. 2022-63) and informed consent was obtained from all individual participants.

Core genes were identified through LASSO analysis, and their expression levels were assessed via RT-qPCR. Total RNA was isolated from carcinoma tissues and adjacent tissues (as normal tissues) using TRIzol (CAT# 15596026, Invitrogen, CA, USA). Ophthalmic scissors were used to cut the tissues as much as possible; the tissues were homogenized, and extracted total RNA according to the extraction kit instructions. Complementary DNA (cDNA) (genome-free) was subsequently generated with the HiScript III 1st Strand cDNA Synthesis Kit (CAT# R312, Vazyme, Nanjing, China). Then, qPCR was carried out using Genious 2X SYBR Green Fast qPCR Mix (CAT# RK21205, ABclonal, Wuhan, China) on a LightCycler 480 II system (Roche, Basel, Switzerland). The relative expression levels were determined via the $2^{-\Delta\Delta Ct}$ method, with β-*actin* used as the internal reference. The primer sequences for nine genes (including β-*actin*) are detailed in *Table 1*, and the primer pairs for the other five genes were purchased from Beyotime Biotechnology Co. Ltd. (Shanghai, China) (*IL1R1*: QH03649S; *EZR*: QH06253S; *FZD7*: QH06513S; *TBC1D4*: QH07325S; *CAB39L*: QH09169S). For the remaining two genes, *HOXC13* and *PDLIM4*, there were no suitable primers; therefore, we did not compare the expression of these two genes in clinical samples.

Statistical analysis

The Kaplan-Meier method was used to generate survival curves, which were compared via the log-rank test. Multivariate analysis was performed via the Cox proportional hazards model. All the statistical analyses

Table 1 Primer sequences of several key genes used for qPCR

Target	Sequences of the primers (from 5'-3')
β-actin	Forward: ATTGGCAATGAGCGGTTCC
	Reverse: TGTGTTGGCGTACAGGTCTT
SIAH2	Forward: AGTGACTGTCAGCAGATTCCTT
	Reverse: TGTCTATTAGCCAGCCATCCAA
SPINT1	Forward: GCATTGTGGTGGTGGTAGC
	Reverse: CGGTAGTGGAGACAGTGGAG
SGCB	Forward: GCCGTGATTCGCATTGGA
	Reverse: TCGCCTTCCTCCTACTGTG
DAB2	Forward: GATGCCTTCACTGCCTTAGAC
	Reverse: GGAATGCCAACCTTGCTGTT
ANO6	Forward: GGCTCTCGGTGTTCATTGTATT
	Reverse: GTGGCTGTCTGTGGAGTCA
LIMCH1	Forward: CCTCATCAGAACCACAGCATT
	Reverse: ACAGCACCAACTCCACCTT
NEDD9	Forward: GACTGGCGGTGTTACGGATA
	Reverse: GCTTCATCTTGTTGTGGAGGAT
KLF2	Forward: GCAAGACCTACACCAAGAGTTC
	Reverse: GCACAGATGGCACTGGAATG

qPCR, quantitative polymerase chain reaction.

were conducted in R (version 4.3.0). We used the Wilcoxon test to compare two groups of data, with the threshold for statistical significance set at a two-sided P<0.05.

Results

Differential gene expression analysis results

The BC dataset was obtained from the TCGA public database; it includes data for 1,231 samples (113 normal breast tissue samples and 1,118 tumor samples from BC patients). Differential gene expression analysis using the limma package revealed a total of 1,643 differentially expressed genes, including 577 upregulated genes and 1,066 downregulated genes that met the screening criteria of P value <0.05 and |logFC| >1. Pathway analysis revealed enrichment of these genes in processes such as positive regulation of localization, the mitotic cell cycle, and cell-cell junctions (*Figure 1A-1C*).

WGCNA

In this study, WGCNA was performed to explore the regulatory network linked to BC. A soft threshold value of 4 (*Figure 2A*) was applied, leading to the identification of nine gene modules in BC: the black (192 genes), blue (985 genes), brown (697 genes), green (337 genes), grey (70 genes), pink (139 genes), red (213 genes), turquoise (1,751 genes), and yellow (616 genes) models. Notably, the yellow module demonstrated the strongest correlation with BC traits (cor =–0.78, P=5e–256) (*Figure 2B,2C*).

Subsequent analysis identified 325 overlapping differentially expressed genes between the yellow module and the differentially expressed gene set. These overlapping genes were used to construct a protein-protein interaction network via the STRING database (http://cn.string-db.org), and the network was visualized using Cytoscape (*Figure 2D,2E*).

Identification and validation of prognostic genes via regression analyses

The crucial genes within the shared gene set were further assessed by gathering clinical data from BC patients and employing Cox single-factor regression and LASSO regression feature selection algorithms for gene screening. A total of 23 prognosis-related genes were identified through Cox univariate regression analysis (see *Table 2*). LASSO regression was used to further filter these genes; ultimately, 15 prognosis-related genes were screened to construct a prognostic model (*Figure 3A-3C*). Further analysis revealed that the gene set was enriched in the EMT pathway. The TCGA patients were randomly divided into training and testing groups at a 4:1 ratio. The testing group was used as the internal validation cohort. LASSO regression analysis was used to ascertain the optimal risk score value for each sample, and the specific formula for calculating the risk score was as follows: risk score = *SIAH2* × (–0.4370976485195207) + *TBC1D4* × (–0.1987575763868614) + *FZD7* × (–0.134770148136664) + *NEDD9* × (–0.129988148893634) + *KLF2* × (–0.0754466294283182) + *PDLIM4* × (–0.049939298826895) + *IL1R1* × 0.00202129794153 + *CAB39L* × 0.103292486570372 + *LIMCH1* × 0.10601072066044 + *ANO6* × 0.138292169859593 + *EZR* × 0.160658836737492 + *DAB2* × 0.17809279540275 + *HOXC13* × 0.18825493493539 + *SPINT1* × 0.19848516664941 + *SGCB* × 0.198188469790138.

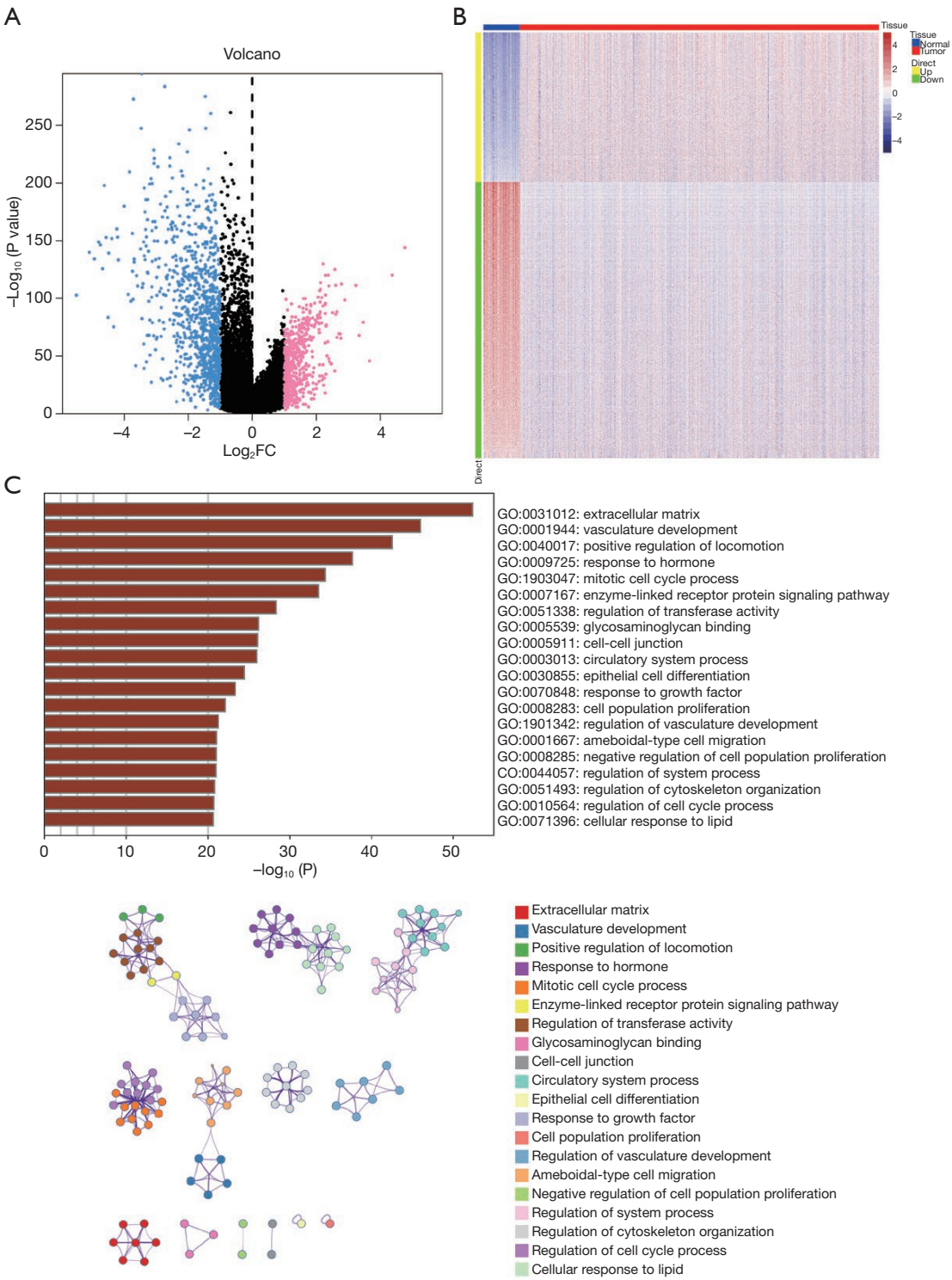

Figure 1 Differential gene expression and functional enrichment analysis in BC. (A) Volcano plot displaying significantly upregulated (pink), non-differentially expressed (black) and downregulated (blue) genes, between normal and tumor samples from BC patients. (B) Heatmap illustrating gene expression patterns across normal and tumor samples. (C) Bar chart demonstrating enriched GO terms, accompanied by a network plot. BC, breast cancer; FC, fold change; GO, Gene Ontology.

Table 2 The genes most closely related to prognosis in BC

Gene	HR (95% CI)	Z	P value
SPINT1	1.323652184 (1.143810274–1.531770733)	3.763368731	0.00016764
LIMCH1	1.222802008 (1.075254071–1.390596689)	3.06589217	0.002170216
EZR	1.221791831 (1.055100208–1.414818486)	2.676640834	0.007436431
FOLR2	1.141929246 (1.035846845–1.258875682)	2.667950269	0.007631555
SIAH2	0.744507166 (0.595656944–0.930553946)	–2.592396485	0.009530986
ERRFI1	0.746452027 (0.597515166–0.932512948)	–2.575313526	0.010014926
ANO6	1.224662609 (1.046785279–1.432766143)	2.53098583	0.011374244
NEDD9	0.767311504 (0.618855496–0.951380328)	–2.414279004	0.015766384
DAB2	1.187023704 (1.030457523–1.367378317)	2.37570234	0.017515586
GSN	0.798269122 (0.660348034–0.964996575)	–2.328138789	0.019904735
HOXC13	1.175839686 (1.020450777–1.354890407)	2.239905941	0.02509703
SORBS1	0.787909343 (0.639261128–0.971122919)	–2.234662927	0.025439486
TBC1D4	0.8234515 (0.693909952–0.977176319)	–2.224344325	0.026125287
KLF2	0.780113076 (0.62560716–0.972777248)	–2.20505384	0.027450317
CAB39L	1.17189094 (1.017549031–1.34964344)	2.201407925	0.027707158
CD248	1.142515367 (1.013834637–1.28752887)	2.185329937	0.028864664
PDLIM4	0.785595444 (0.629501248–0.980395516)	–2.135147943	0.032748928
SGCB	1.155730529 (1.007674704–1.325539928)	2.069271871	0.038520582
PLTP	1.159669405 (1.007611571–1.334674163)	2.065703434	0.038856491
RNASE1	1.181311248 (1.008364121–1.383920981)	2.063095005	0.039103603
IL1R1	1.120425206 (1.003902408–1.250472787)	2.029479095	0.042409516
NDRG2	0.800996662 (0.644168702–0.996005628)	–1.995965096	0.045937722
FZD7	0.802485674 (0.64569026–0.997356313)	–1.983830288	0.047274762

BC, breast cancer; CI, confidence interval; HR, hazard ratio.

Patients were divided into high-risk and low-risk groups on the basis of their risk scores. Kaplan-Meier curve analysis revealed a significantly lower overall survival rate for the high-risk group than for the low-risk group in both the training and testing sets (refer to *Figure 3D*,*3E*). Additionally, the ROC curve outcomes for both the training cohort and testing cohort confirmed the robust validation performance of the model [the area under the curve (AUC) values were all ≥0.7] (*Figure 3F*,*3G*).

External validation of the prognostic model using GEO datasets

We used two independent BC datasets from the GEO database (GSE20685 and GSE42568) for external validation to further confirm the predictive sensitivity and specificity of the model. GSE20685 was designated as Cohort 1/GEO1 and includes information on 327 patients; GSE42568 was designated at Cohort 2/GEO2 and includes data on 104 patients. Each patient was assigned a score according to the risk score formula, and patients in each cohort were subsequently divided into a high-risk group and a low-risk group according to the median score. Kaplan-Meier analysis was used to compare survival between the high-risk group and the low-risk group. The results revealed that the high-risk group had significantly shorter overall survival than the low-risk group did in the two independent GEO cohorts (*Figure 4A*,*4B*). ROC curve analysis of the GEO cohorts

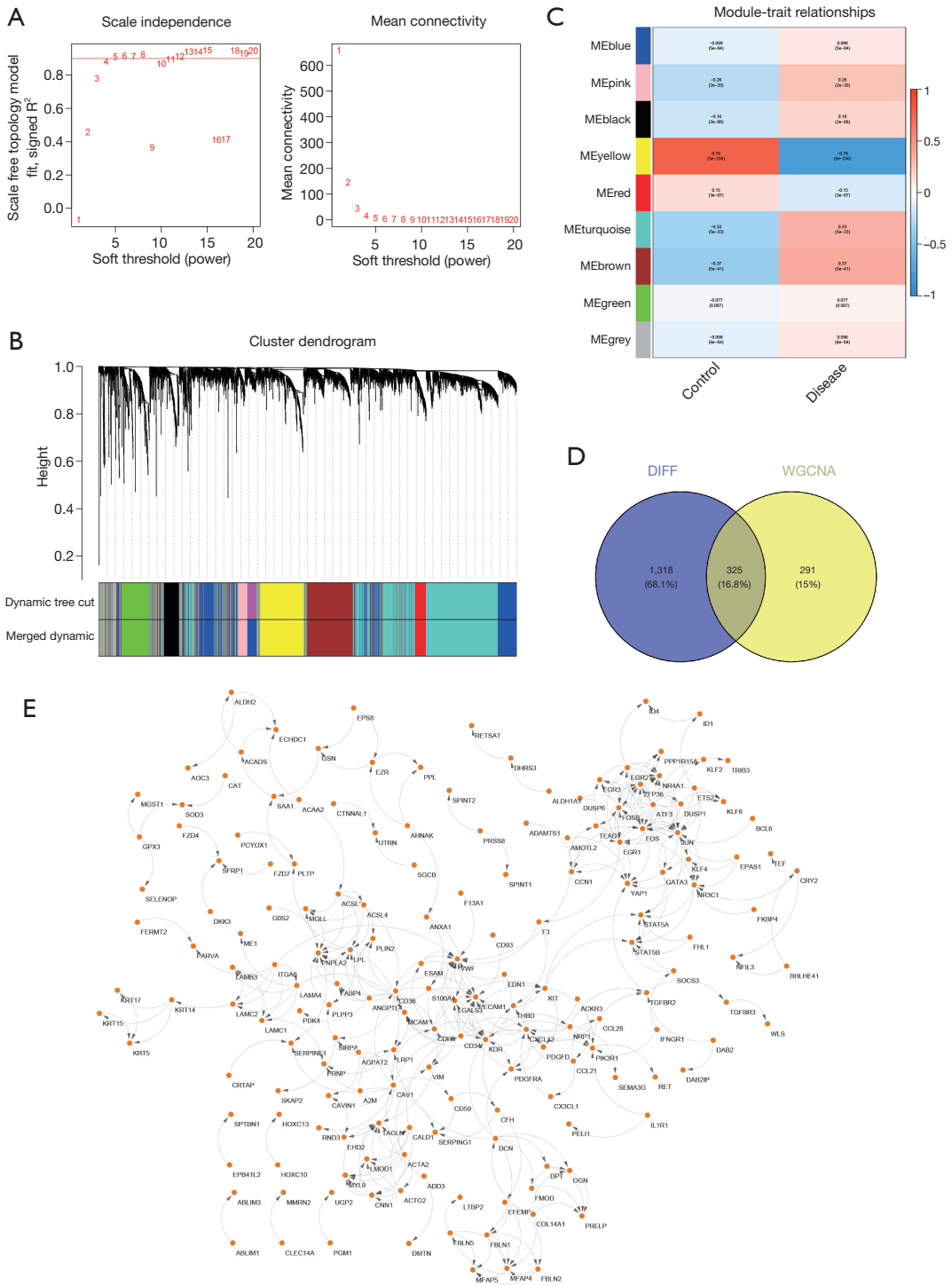

Figure 2 WGCNA for BC. (A) Scale independence and mean connectivity for selecting the appropriate soft-thresholding power in WGCNA. (B) Cluster dendrogram displaying hierarchical clustering of genes on the basis of their expression profiles. (C) Module-trait relationships showing the correlations between gene modules and clinical traits. (D) Venn diagram indicating the overlap between genes identified through differential expression analysis and those included in the WGCNA modules. (E) Gene network visualization illustrating interactions and connectivity between genes potentially involved in BC pathogenesis. BC, breast cancer; DIFF, differential expression analysis; WGCNA, weighted gene coexpression network analysis.

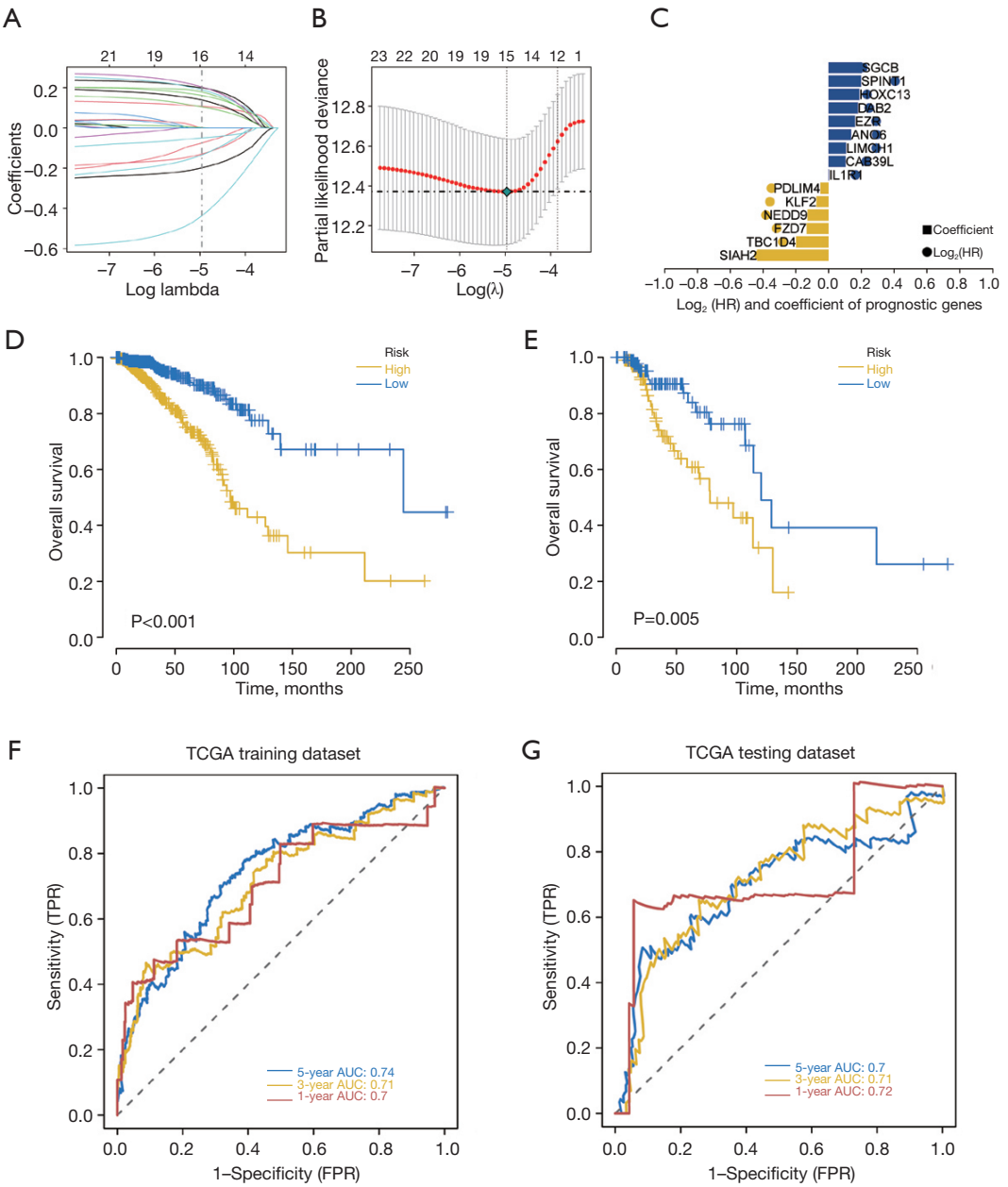

Figure 3 LASSO regression analysis and prognostic model validation for BC. (A) LASSO coefficient profiles showing variables whose coefficients decrease toward zero as lambda increases. (B) Model tuning parameter selection showing how the model was tuned for LASSO regression using cross-validation. (C) Bar plot of prognostic genes displaying coefficients of prognostic genes determined via LASSO regression, colored on the basis of the coefficients and hazard ratios. (D,E) Kaplan-Meier survival curves for the high-risk and low-risk groups in the training cohort (D) and testing cohort (E). (F,G) Receiver operating characteristic curves for validation of the predictive accuracy of the prognostic model at 1-year, 3-year, and 5-year intervals in the training cohort (F) and testing cohort (G). AUC, area under the curve; BC, breast cancer; FPR, false positive rate; HR, hazard ratio; LASSO, least absolute shrinkage and selection operator; TCGA, The Cancer Genome Atlas; TPR, true positive rate.

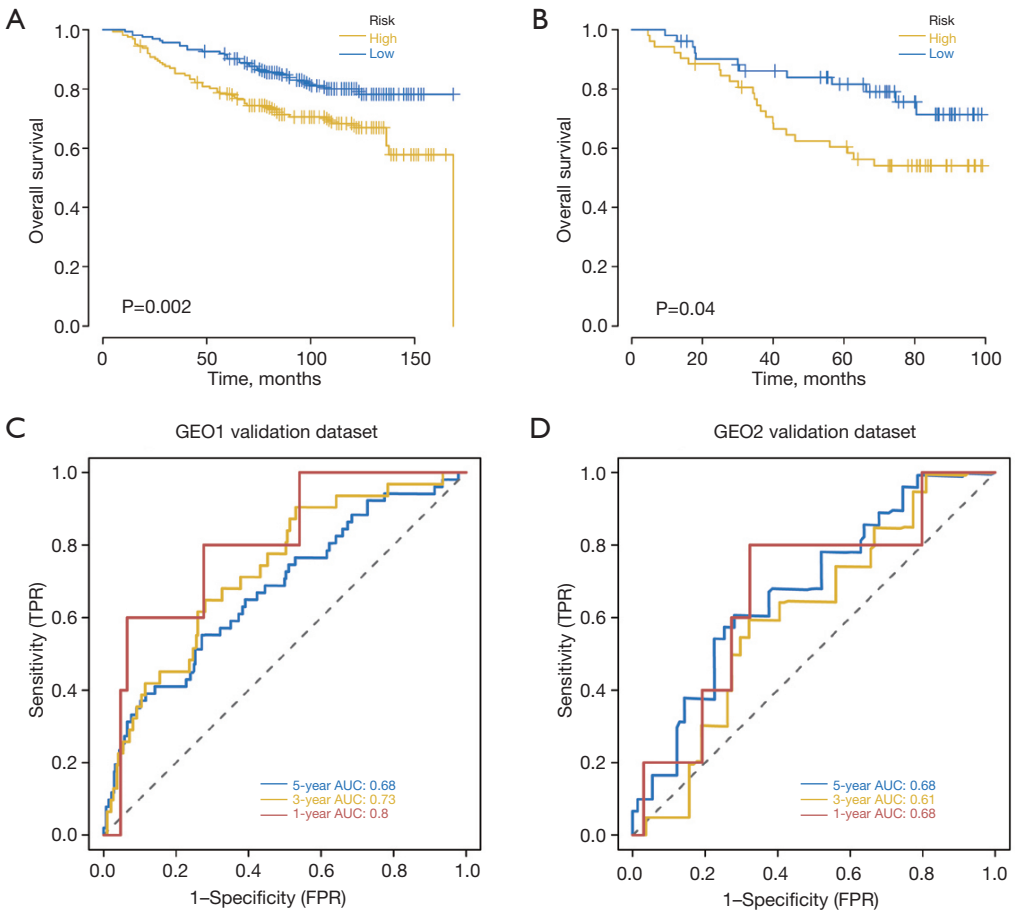

Figure 4 External validation of the BC prognostic model using GEO datasets. (A,B) Kaplan-Meier survival curves for two GEO cohorts. (C,D) ROC curves for two GEO cohorts. The curves were used to evaluate the predictive accuracy of the model at 1-year, 3-year, and 5-year intervals. AUC, area under the curve; BC, breast cancer; FPR, false positive rate; GEO, Gene Expression Omnibus; ROC, receiver operating characteristic; TPR, true positive rate.

confirmed the model's accuracy, demonstrating strong predictive performance for prognosis prediction; notably, the predictive effect of this model was better for the GEO1 dataset (*Figure 4C,4D*). In the GEO1 cohort, the predicted AUC values for 1-, 3- and 5-year survival were 0.8, 0.73 and 0.68, respectively, whereas in the GEO2 cohort, they were 0.68, 0.61 and 0.68, respectively. The AUC values indicate the model's performance at different time points.

Impact of the risk score on the tumor microenvironment (TME) and immune cell dynamics

The TME is a complex system consisting of tumor-associated fibroblasts, immune cells, the extracellular matrix, growth factors, inflammatory factors, and cancer cells. It plays a vital role in tumor diagnosis, survival rates, and treatment response. This study involved analysis of the influence of the risk score on immune cell infiltration in tumors and analysis of the molecular mechanisms involved in BC progression. *Figure 5A* shows the relative proportions of 22 immune cell subtypes in patients in the high-risk (HRisk) group and low-risk (LRisk) group. We also compared the expression levels of markers of 22 kinds of immune cells in the two groups. The results revealed that in the high-risk group, the estimated proportions of naive B cells, resting dendritic cells, monocytes, activated natural killer (NK) cells, plasma cells, CD8 T cells, follicular helper T cells and regulatory T cells (Tregs) were significantly decreased, whereas the proportions of M0 macrophages, M2 macrophages, resting mast cells and neutrophils were

significantly increased (*Figure 5B*). The correlation between the risk score and the proportions of immune cells was further analyzed. The analysis revealed a significant positive correlation between the risk score and the proportions of M2 macrophages, neutrophils, M0 macrophages, and resting mast cells; however, there was a significant negative correlation between the risk score and the proportions of CD8 T cells, follicular helper T cells, naive B cells, plasma cells, Tregs, monocytes, and activated NK cells (*Figure 5C*).

Analysis of genetic mutations and their correlations with the risk score

In our research focusing on BC, we analyzed processed SNP-related data to identify the top 30 genes in terms of mutation frequency. By comparing gene mutation types and frequencies between the HRisk and LRisk patient cohorts, we created a mutation landscape map via the R package ComplexHeatmap (*Figure 5D*). Although all samples have at least one mutation (HRisk: 536/536, LRisk: 537/537), The findings highlighted notable differences in the frequencies of gene mutations, such as *PIK3CA*, *TTN*, *CDH1*, *GATA3*, *KMT2C*, and *MAP3K1* mutations between the high-risk and low-risk groups.

Many studies have shown that TMB is strongly correlated with the response to immunotherapy. Thus, we explored the correlation between the risk score and TMB. Our analysis revealed a significant difference in TMB between the high-risk and low-risk groups (*Figure 5E*).

These findings provide insights into the intricate relationships among immune cell infiltration, immunotherapy response and the risk score in patients with BC and potential treatment strategies tailored to individual patient profiles.

Correlation between the risk score and chemotherapy sensitivity

Our research focused on predicting the response to chemotherapy in BC patients by utilizing drug sensitivity information from the GDSC database and the R package "pRRopetic". We investigated the correlation between the risk score and sensitivity to 6 chemotherapeutic drugs to gain insights into their influence on treatment effectiveness. We found that a low-risk score was associated with lower IC_{50} values for 5 chemotherapeutics (BAY.61.3606, BIRB.0796, AG.014699, CEP.701 and ABT.888), and high-risk patients were more sensitive to CCT007093 than low-risk patients were (*Figure 6*). Therefore, the risk score is useful for guiding chemotherapy decision-making in patients.

Functional enrichment analyses of the two risk groups

GSVA and GSEA were performed to study the specific signaling pathways involved in the prognostic model and explore the potential molecular mechanisms by which risk scores affect tumor progression and prognosis. GSVA revealed that the high-risk group exhibited high expression of genes related to the EMT pathway, transforming growth factor β (TGF-β) signaling pathway and PI3K-Akt signaling pathway, whereas the low-risk group exhibited high expression of genes related to the Wnt-β-catenin signaling pathway, DNA repair pathway and P53 signaling pathway (*Figure 7A*). GSEA revealed that the high-risk group exhibited high expression of genes related to the PI3K-Akt signaling pathway, the TGF-β signaling pathway, and the thyroid hormone signaling pathway (*Figure 7B*), and the network of molecular interactions among the three pathways is shown in *Figure 7C*.

Validation of the nomogram for predicting BC outcomes

In our BC study, patients were categorized into high-risk and low-risk groups according to the median risk score. Regression analysis was then conducted, and the results are presented in a bar chart. The logistic regression analysis demonstrated the significant impact of risk score values on the nomogram prediction model scoring process across all samples (*Figure 8A*).

Furthermore, we carried out survival prediction analysis for patients with BC at three and five years and generated ROC curves; both the calibration curves and the ROC curves indicated a high consistency between the actual and predicted OS rates (*Figure 8B,8C*). We also performed decision curve analysis (DCA) to assess the clinical validity of the nomogram's predictive performance (*Figure 8D*). We found that assessment based on the nomogram led to more net benefit than assessment based on clinical characteristics alone because the risk score contributed the most net benefit among the independent prognostic factors. The above results indicate that the risk score can contribute to

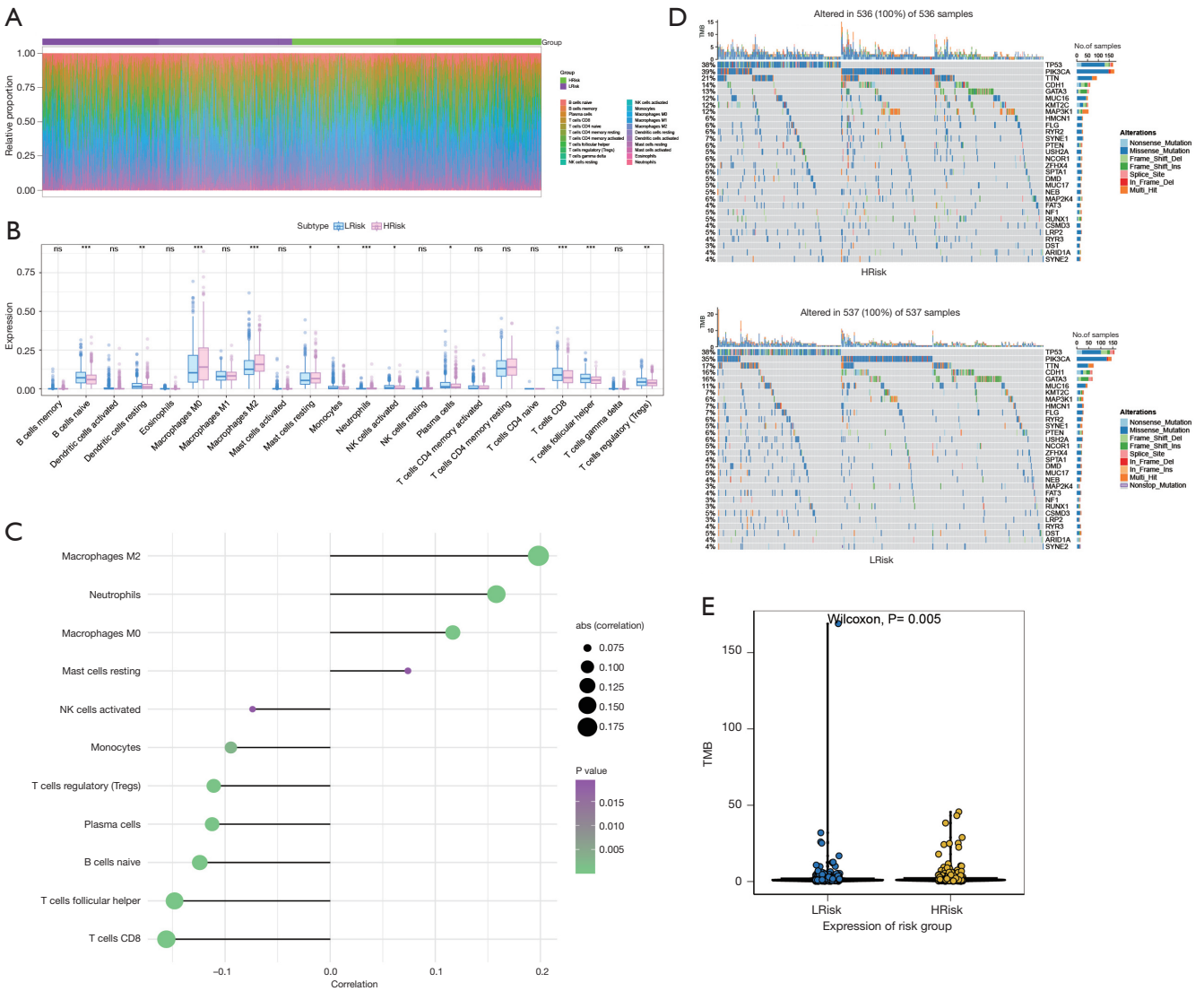

Figure 5 Analysis of immune cell infiltration and gene mutation in BC patients grouped according to the risk score. (A) Immune cell composition. Bar plot showing the relative proportions of different immune cell types in the LRisk and HRisk BC patients. (B) Boxplots of immune cell expression. Boxplots comparing the expression levels of immune cell markers between the LRisk and HRisk BC groups. Significant differences in immune cell content are annotated as follows: ns, P≥0.05; *, P<0.05; **, P<0.01; ***, P<0.001. (C) Plot of the correlations between immune-infiltrating cells and the risk score. The plot illustrates the correlations between the risk score and the proportions of various immune cell types. The size of each dot corresponds to the absolute value of the correlation, with larger dots indicating a stronger correlation. The color of the dots represents the statistical significance (P value), with a color gradient from purple (higher P values) to green (lower P values). (D) Mutation landscape profile. The left panel shows the frequency of specific gene mutations in this group of samples, whereas the right panel shows the names of the mutant genes. Different colors represent various types of mutations, such as nonsense, missense, and frameshift deletions and insertions. The top panel shows the number of mutations per million bases in each sample. (E) Bee swarm plot for TMB. A comparison of TMB between the LRisk and HRisk groups. Each dot represents the TMB of an individual sample, with significant differences in mutation loads observed between the groups (Wilcoxon test P value <0.05). abs, absolute; BC, breast cancer; HRisk, high risk; NK, natural killer; TMB, tumor mutation burden.

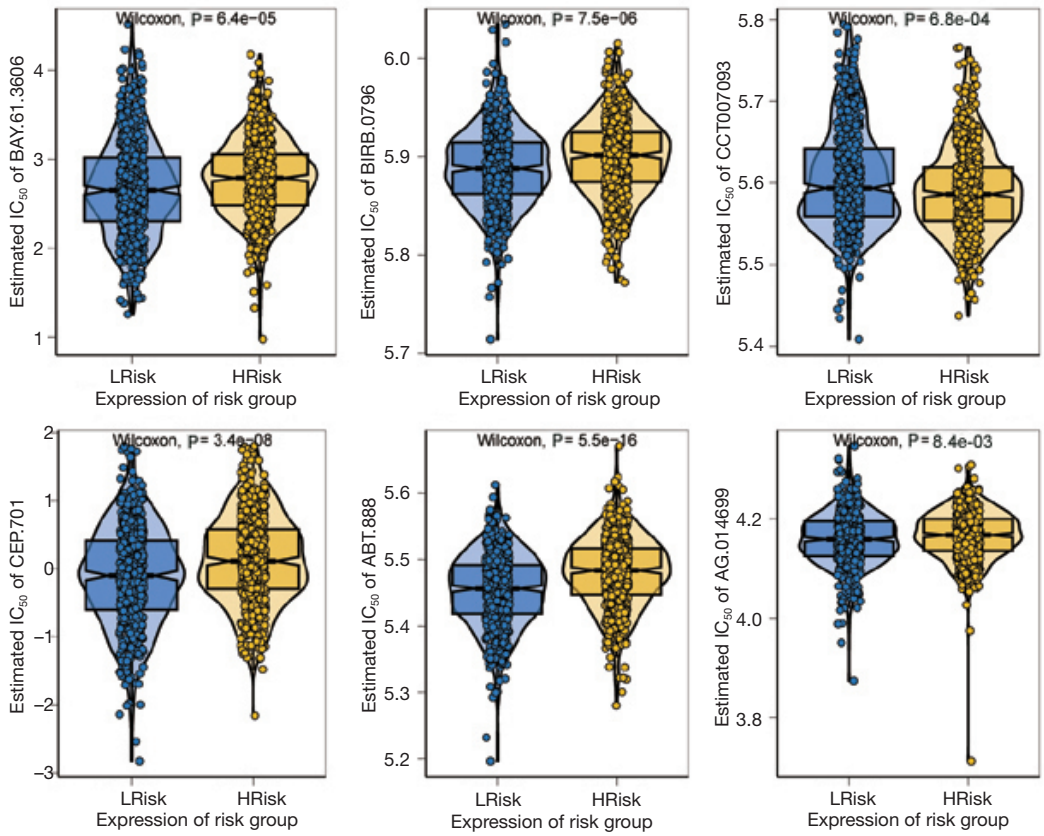

Figure 6 Sensitivity analyses of 6 chemotherapeutic drugs in two risk groups. Comparison of the IC50 values of chemotherapeutic drugs in the LRisk and HRisk BC groups. BC, breast cancer; HRisk, high risk; IC_{50}, half-maximal inhibitory concentration; LRisk, low risk.

better clinical management.

qPCR validation of core genes in the LASSO model

We analyzed key differentially expressed gene via the LASSO model. Clinical BC samples were collected for qPCR. Thirteen prognosis-related genes were verified via qPCR, and the expression of these genes was consistent with that in the TCGA database. Three prognosis-related genes (including *SPINT1*, *EZR*, and *SIAH2*) were significantly increased in clinical BRCA samples compared with normal samples; however, the expression of the remaining 10 prognosis-related genes (including *SGCB*, *DAB2*, *ANO6*, *LIMCH1*, *CAB39L*, *IL1R1*, *NEDD9*, *FZD7*, *KLF2* and *TBC1D4*) decreased dramatically in BRCA samples (*Figure 9A,9B*).

These findings highlight the dysregulation of specific genes in BC and provide insight into potential prognostic markers for this disease. qPCR validation further confirmed the relevance of these genes in BC progression and prognosis.

Discussion

BC represents a physical, psychological, social, economic, and spiritual burden for patients and their families. Cancer itself and cancer treatments can affect the psychological and mental health of patients and their families (13). Accurate diagnosis and proper monitoring of cancer patients are important for successful cancer treatment. Various biomarkers have improved the accuracy of the prediction of therapeutic effects or prognosis. The prediction of the prognosis of BC is also very important for the prevention and early treatment of recurrence.

The TME includes immune cells, fibroblasts, adipocytes, bone marrow-derived inflammatory cells, the extracellular matrix, vascular and lymphatic networks, various cytokines, inflammatory factors, and vesicles secreted by these cells (14,15). The various signaling factors and components in the TME are mostly beneficial for the growth and metastasis of cancers, leading to deterioration (16). Lymphocyte

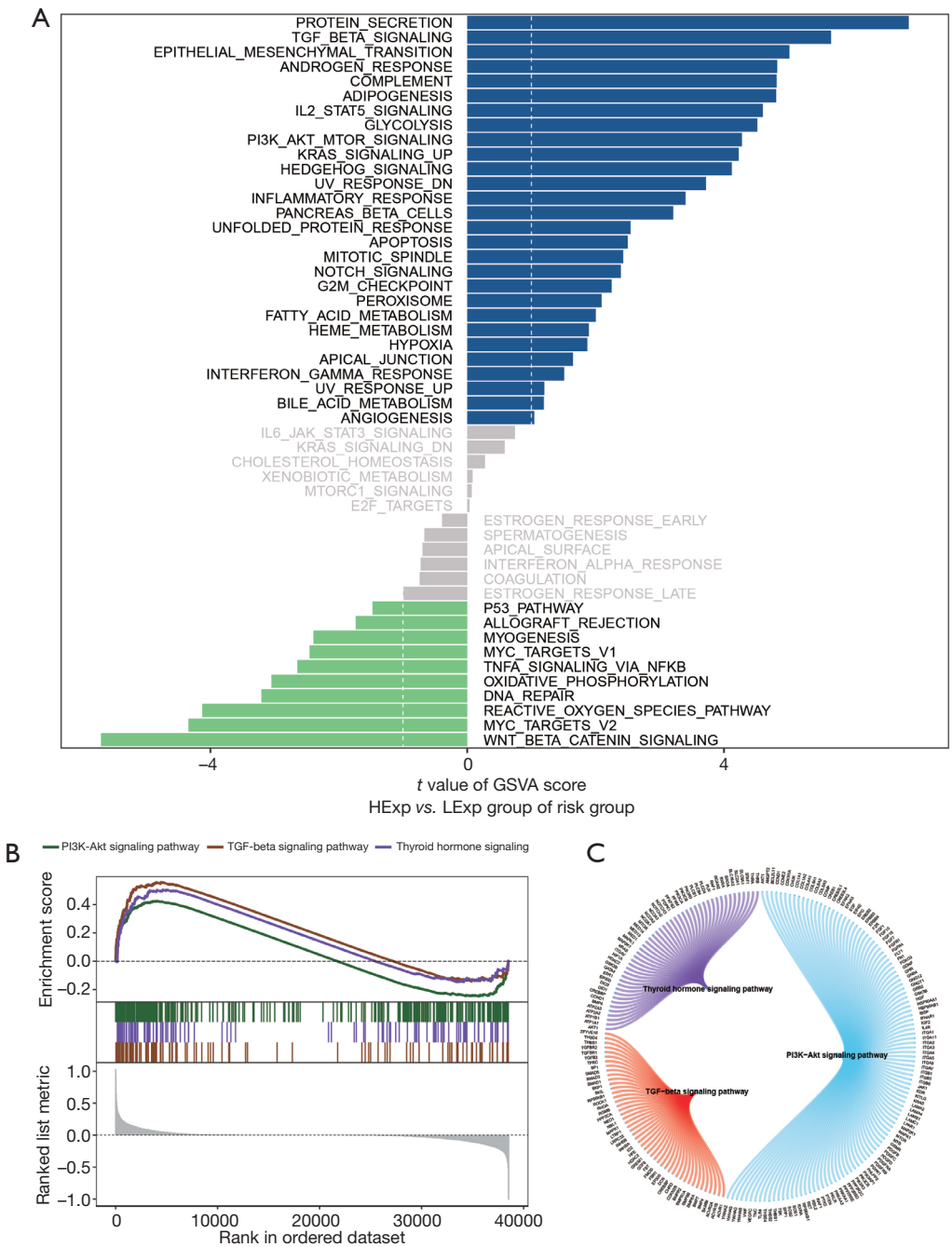

Figure 7 Pathway analysis and gene set enrichment based on risk groups in BC. (A) Pathway enrichment analysis. A bar plot illustrating the enrichment scores of various signaling pathways: significantly upregulated (blue), significantly downregulated (green), and not significant (gray). The length of the bars indicates the *t* value of the GSVA score, reflecting the degree of pathway enrichment or depletion. (B) GSEA. Plots showing the enrichment scores for the PI3K-Akt signaling pathway, TGF-beta signaling pathway, and thyroid hormone signaling pathway. The top plot displays the enrichment score across the ranked gene list, whereas the bottom plot highlights the ranked gene positions contributing to the enrichment signal. (C) Molecular interaction network. A chord diagram depicting the interactions between the three major signaling pathways analyzed. Each pathway is represented by a segment, with chords connecting segments to indicate shared genes and their interactions. BC, breast cancer; GSEA, gene set enrichment analysis; GSVA, gene set variation analysis; HExp, high expression; LExp, low expression; TGF-beta, transforming growth factor beta.

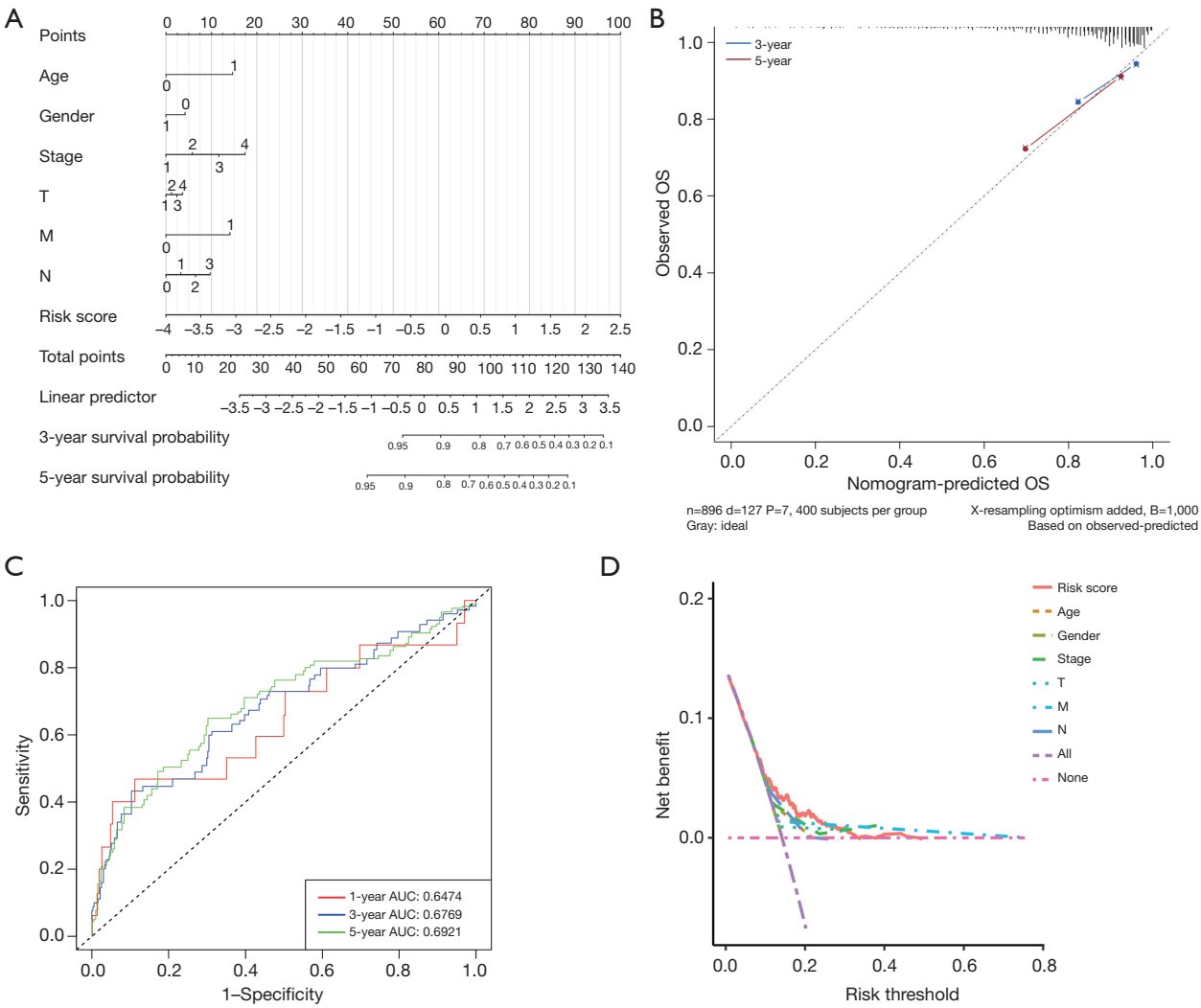

Figure 8 Nomogram and predictive analysis of survival in BC patients. (A) Nomogram. A predictive tool for estimating 3-year and 5-year survival probabilities in BC patients was developed. The nomogram incorporates multiple prognostic factors, such as age, sex, stage, tumor size (T), metastasis (M), node involvement (N), and the risk score (age: 0 means ≤65 years old, 1 means >65 years old; gender: 0 means female, and 1 means male). (B) Calibration curves. Comparison of predicted versus observed outcomes for 3-year and 5-year survival using the nomogram. (C) ROC curves. Evaluation of the performance of the nomogram for 1-, 3-, and 5-year survival prediction. The AUC is provided for each time point. (D) Decision curve analysis curves. Assessment of the clinical net benefit of utilizing the nomogram across various risk thresholds for decision-making. AUC, area under the curve; BC, breast cancer; OS, overall survival; ROC, receiver operating characteristic.

infiltration is a characteristic of malignant tumors (17,18). Low immunogenicity and immunosuppression in the malignant TME are important reasons for tumor resistance (19-22). The results of the prognosis analysis in this study verified this conclusion.

The postoperative treatment for hormone-dependent BC is generally chemotherapy/targeted drugs followed by hormone therapy (23,24). For TNBC, chemotherapy is still the standard postoperative treatment (25,26). Drug resistance often occurs during chemotherapy. Resistance is related to the molecular and biological properties of tumors. Studies have shown that resistance is related to signaling pathways, such as the PI3K-Akt and TGF-β pathways (27,28).

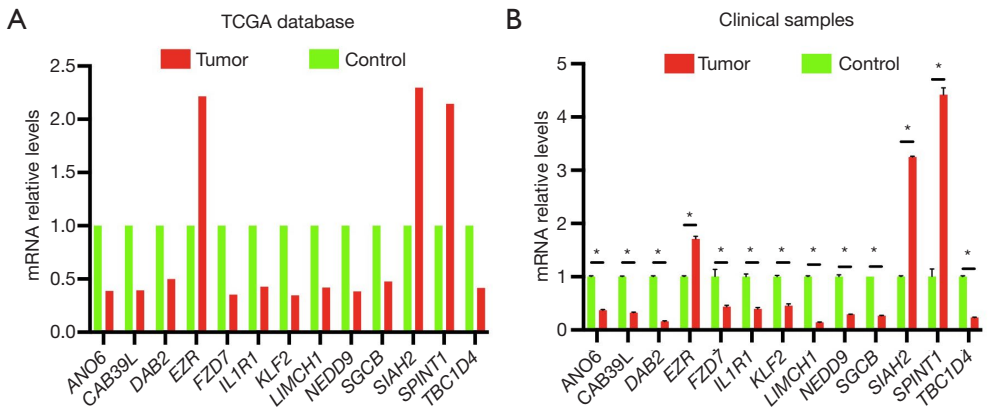

Figure 9 Characteristics of 13 prognostic genes in BC. (A) Differences in the expression levels of 13 prognosis-related genes between normal and BC tissues in the TCGA database. (B) The expression of 13 prognostic genes in clinical patient samples was verified by qPCR. * means P<0.01; *FZD7* P=0.002; the rest P<0.001. BC, breast cancer; qPCR, quantitative polymerase chain reaction; TCGA, The Cancer Genome Atlas.

Recently, there has been a surge of interest in cancer research surrounding the exploration of biological markers and functional genes. Polygenic risk scores (PRSs) could prove beneficial for guiding screening and decision making (29). Numerous signaling pathways, including those examined in this particular study, have been identified as having implications for the prognosis of patients with BC. TGF-β is a versatile cytokine that plays a key role in a variety of biological processes. Over the past two decades, TGF-β has been extensively researched and recognized as a promising biomarker with applications in early detection, disease monitoring, treatment selection, and tracking of tumor progression, offering valuable insights for disease management (30). The PI3K-Akt-mTOR signaling pathway plays a crucial role in driving carcinogenesis in various types of cancer (31). Some components of this pathway were also included in our prognostic model. The thyroid hormone signaling pathway is also related to BC prognosis and has potential as a target for therapy (32,33).

In this study, we established a new prognostic gene signature that can predict the prognosis and therapeutic effect of BC patients well. The model exhibited very stable predictive value in four independent cohorts from multiple databases. The only deficiency is that the predictive effect in the GEO2 cohort (with an AUC value between 0.6 and 0.7) was not as good as that in the other three cohorts (with AUC values greater than 0.7). This may be because the GEO2 cohort is relatively small, with a total of only 104 patients; in terms of bias, this dataset had a greater effect on the model than the datasets for the other three independent cohorts, which included more than 300 patients, and the training cohort, which included more than 900 patients. Relatively speaking, the bias of individual datasets had little effect on the model.

The core genes of our prognostic model (including *PDLIM4*, *DAB2* and *SGCB*) were enriched mainly in the EMT pathway. EMT is a biological process by which epithelial cells transform into mesenchymal cells, which can enhance the invasion and metastasis of tumor cells. EMT is also an important cause of drug resistance and radiation resistance in tumor cells. Targeted therapy for EMT can reduce BC metastasis and increase the chemosensitivity of tumor cells (34-36). Furthermore, GSVA also revealed that the EMT pathway was enriched in the high-risk group; TGF-β signaling, hedgehog signaling and Notch signaling, which are considered to activate EMT in the progress of BC, were also enriched in the high-risk group (37). The prediction model involves 15 genes, of which four (*SIAH2*, *EZR*, *SPINT1* and *HOXC13*) were relatively highly expressed in tumor tissues; the other 11 genes are expressed at relatively low levels. The expression of 13 genes was validated by qPCR in clinical TNBC tissues and adjacent normal tissues, and the results were consistent with the TCGA data. *PDLIM4* and *HOXC13* expression were not verified because there were no suitable primers. The seven-in-absentia homolog (SIAH) proteins are E3 ubiquitin ligases. SIAH1 and SIAH2 are two homologs in humans that are involved in different pathways, including cancers, hypoxia, inflammation, oxidative stress, DNA damage stress, and regulation of Treg recruitment

(38-40). Of the four genes that are highly expressed in BC tissue, *SIAH2* is the only one whose expression is negatively associated with risk, and therefore, in a sense, it may be a favorable gene for prognosis. *EZR* encodes Ezrin, Radixin and Moesin (ERM) family members. Ezrin promotes many signal transduction events in tumorigenesis. Moreover, Ezrin is an oncogenic protein related to the metastasis of various types of cancer, including BC, so Ezrin can be used as a therapeutic target (41,42). SPINT1, also known as HAI-1, is a Kunitz serine protease inhibitor that can inhibit many proteases, such as hepatocyte growth factor (HGF). HGF plays a variety of roles in tumor metastasis. HAI-1 regulates the activity of HGF by inhibiting HGF activator (HGFA), matrix enzymes and hepsin (43). The expression of homeobox C13 (HOXC13) is significantly increased in tumor tissues. HOXC13 increases cell viability, proliferation, migration, invasion, EMT and glycolysis in BC by regulating DNMT3A (44). Prostate cancer patients with high HOXC13 expression have the worst prognosis. Immune cell infiltration analysis revealed that high expression of HOXC13 inhibited the infiltration of γδ T cells and plasma cells (45).

There are some limitations to this study. First, the patients' data were obtained from public databases, and some data, such as history of other conditions and treatment interventions, were missing, and these factors may affect the prognosis of BC. Second, the molecular mechanism involved in this model needs to be verified and clarified by further experiments. Third, *PDLIM4* and *HOXC13* expression could not be validated by qPCR in TNBC tissue.

Conclusions

In summary, the prognostic model that we established herein can accurately predict the prognosis of BC. Moreover, the prognostic model has the potential to predict immune microenvironment features, the immunotherapy response and the chemotherapy response, which may be helpful for the clinical management of BC.

Acknowledgments

None.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-838/rc

Data Sharing Statement: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-838/dss

Peer Review File: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-838/prf

Funding: This work was supported by the National Natural Science Foundation of China (No. 82172753 to M.S.; No. 82271640 to R.L.), the Natural Science Foundation of Shanghai (No. 21ZR1454800 to M.S.) and Outstanding Youth Fund Project of Natural Science in University in Anhui Province (No. 2024AH020015 to C.C.).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-838/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Medical Ethics board of Wannan Medical College (No. 2022-63) and informed consent was obtained from all individual participants.

References

1. Childers CP, Childers KK, Maggard-Gibbons M, et al. National Estimates of Genetic Testing in Women With a History of Breast or Ovarian Cancer. J Clin Oncol 2017;35:3800-6.
2. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024;74:12-49.

3. Thakur A, Rana N, Kumar R. Altered hormone expression induced genetic changes leads to breast cancer. Curr Opin Oncol 2024;36:115-22.
4. Obeagu EI, Obeagu GU. Breast cancer: A review of risk factors and diagnosis. Medicine (Baltimore) 2024;103:e36905.
5. Moar K, Pant A, Saini V, et al. Potential diagnostic and prognostic biomarkers for breast cancer: A compiled review. Pathol Res Pract 2023;251:154893.
6. Li Y, Zhang H, Merkher Y, et al. Recent advances in therapeutic strategies for triple-negative breast cancer. J Hematol Oncol 2022;15:121.
7. Lehmann BD, Pietenpol JA. Clinical implications of molecular heterogeneity in triple negative breast cancer. Breast 2015;24 Suppl 2:S36-40.
8. Burstein MD, Tsimelzon A, Poage GM, et al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. Clin Cancer Res 2015;21:1688-98.
9. Dvir K, Giordano S, Leone JP. Immunotherapy in Breast Cancer. Int J Mol Sci 2024;25:7517.
10. Onkar SS, Carleton NM, Lucas PC, et al. The Great Immune Escape: Understanding the Divergent Immune Response in Breast Cancer Subtypes. Cancer Discov 2023;13:23-40.
11. Trayes KP, Cokenakes SEH. Breast Cancer Treatment. Am Fam Physician 2021;104:171-8.
12. Corti C, Batra-Sharma H, Kelsten M, et al. Systemic Therapy in Breast Cancer. Am Soc Clin Oncol Educ Book 2024;44:e432442.
13. Sunilkumar MM, Finni CG, Lijimol AS, et al. Health-Related Suffering and Palliative Care in Breast Cancer. Curr Breast Cancer Rep 2021;13:241-6.
14. Zhao C, Wu M, Zeng N, et al. Cancer-associated adipocytes: emerging supporters in breast cancer. J Exp Clin Cancer Res 2020;39:156.
15. Chen F, Zhuang X, Lin L, et al. New horizons in tumor microenvironment biology: challenges and opportunities. BMC Med 2015;13:45.
16. Guo S, Deng CX. Effect of Stromal Cells in Tumor Microenvironment on Metastasis Initiation. Int J Biol Sci 2018;14:2083-93.
17. Watanabe MA, Oda JM, Amarante MK, et al. Regulatory T cells and breast cancer: implications for immunopathogenesis. Cancer Metastasis Rev 2010;29:569-79.
18. Berghoff AS, Preusser M. The inflammatory microenvironment in brain metastases: potential treatment target? Chin Clin Oncol 2015;4:21.
19. Wang Z, Wu X. Study and analysis of antitumor resistance mechanism of PD1/PD-L1 immune checkpoint blocker. Cancer Med 2020;9:8086-121.
20. Patel SA, Minn AJ. Combination Cancer Therapy with Immune Checkpoint Blockade: Mechanisms and Strategies. Immunity 2018;48:417-33.
21. Lei Q, Wang D, Sun K, et al. Resistance Mechanisms of Anti-PD1/PDL1 Therapy in Solid Tumors. Front Cell Dev Biol 2020;8:672.
22. Zhao S, Ren S, Jiang T, et al. Low-Dose Apatinib Optimizes Tumor Microenvironment and Potentiates Antitumor Effect of PD-1/PD-L1 Blockade in Lung Cancer. Cancer Immunol Res 2019;7:630-43.
23. Lee EY, Lee DW, Lee KH, et al. Recent Developments in the Therapeutic Landscape of Advanced or Metastatic Hormone Receptor-Positive Breast Cancer. Cancer Res Treat 2023;55:1065-76.
24. Andrahennadi S, Sami A, Manna M, et al. Current Landscape of Targeted Therapy in Hormone Receptor-Positive and HER2-Negative Breast Cancer. Curr Oncol 2021;28:1803-22.
25. Schmadeka R, Harmon BE, Singh M. Triple-negative breast carcinoma: current and emerging concepts. Am J Clin Pathol 2014;141:462-77.
26. Millis SZ, Gatalica Z, Winkler J, et al. Predictive Biomarker Profiling of > 6000 Breast Cancer Patients Shows Heterogeneity in TNBC, With Treatment Implications. Clin Breast Cancer 2015;15:473-481.e3.
27. McCubrey JA, Abrams SL, Fitzgerald TL, et al. Roles of signaling pathways in drug resistance, cancer initiating cells and cancer progression and metastasis. Adv Biol Regul 2015;57:75-101.
28. Babyshkina N, Dronova T, Erdyneeva D, et al. Role of TGF-β signaling in the mechanisms of tamoxifen resistance. Cytokine Growth Factor Rev 2021;62:62-9.
29. McCarthy AM, Manning AK, Hsu S, et al. Breast cancer polygenic risk scores are associated with short-term risk of poor prognosis breast cancer. Breast Cancer Res Treat 2022;196:389-98.
30. Shukla N, Naik A, Moryani K, et al. TGF-β at the crossroads of multiple prognosis in breast cancer, and beyond. Life Sci 2022;310:121011.
31. Ung MH, Wang GL, Varn FS, et al. Application of pharmacologically induced transcriptomic profiles to interrogate PI3K-Akt-mTOR pathway activity associated with cancer patient prognosis. Oncotarget 2016;7:84142-54.

32. Jerzak KJ, Cockburn J, Pond GR, et al. Thyroid hormone receptor α in breast cancer: prognostic and therapeutic implications. Breast Cancer Res Treat 2015;149:293-301.
33. Davidson CD, Gillis NE, Carr FE. Thyroid Hormone Receptor Beta as Tumor Suppressor: Untapped Potential in Treatment and Diagnostics in Solid Tumors. Cancers (Basel) 2021;13:4254.
34. Park M, Kim D, Ko S, et al. Breast Cancer Metastasis: Mechanisms and Therapeutic Implications. Int J Mol Sci 2022;23:6806.
35. Hashemi M, Arani HZ, Orouei S, et al. EMT mechanism in breast cancer metastasis and drug resistance: Revisiting molecular interactions and biological functions. Biomed Pharmacother 2022;155:113774.
36. Tanabe S, Quader S, Cabral H, et al. Interplay of EMT and CSC in Cancer and the Potential Therapeutic Strategies. Front Pharmacol 2020;11:904.
37. Kumar A, Golani A, Kumar LD. EMT in breast cancer metastasis: an interplay of microRNAs, signaling pathways and circulating tumor cells. Front Biosci (Landmark Ed) 2020;25:979-1010.
38. Sabour Takanlou L, Cecener G, Sabour Takanlou M, et al. Correlation between Ubiquitin E3 Ligases (SIAHs) and Heat Shock Protein 90 in Breast Cancer Patients. Iran J Public Health 2022;51:1836-46.
39. Siswanto FM, Jawi IM, Kartiko BH. The role of E3 ubiquitin ligase seven in absentia homolog in the innate immune system: An overview. Vet World 2018;11:1551-7.
40. Scortegagna M, Hockemeyer K, Dolgalev I, et al. Siah2 control of T-regulatory cells limits anti-tumor immunity. Nat Commun 2020;11:99.
41. Song Y, Ma X, Zhang M, et al. Ezrin Mediates Invasion and Metastasis in Tumorigenesis: A Review. Front Cell Dev Biol 2020;8:588801.
42. Xiao G, Cheng F, Yuan J, et al. Integrative multiomics analysis identifies a metastasis-related gene signature and the potential oncogenic role of EZR in breast cancer. Oncol Res 2022;30:35-51.
43. Parr C, Jiang WG. Hepatocyte growth factor activation inhibitors (HAI-1 and HAI-2) regulate HGF-induced invasion of human breast cancer cells. Int J Cancer 2006;119:1176-83.
44. Li H, Gao P, Chen H, et al. HOXC13 promotes cell proliferation, metastasis and glycolysis in breast cancer by regulating DNMT3A. Exp Ther Med 2023;26:439.
45. Li M, Bai G, Cen Y, et al. Silencing HOXC13 exerts anti-prostate cancer effects by inducing DNA damage and activating cGAS/STING/IRF3 pathway. J Transl Med 2023;21:884.

Cite this article as: Wan X, Zhan J, Ye S, Chen C, Li R, Shen M. Construction of a prognostic model and analysis of related mechanisms in breast cancer based on multiple datasets. Transl Cancer Res 2025;14(2):930-948. doi: 10.21037/tcr-24-838