Am J Transl Res 2026;18(3):1846-1862
www.ajtr.org /ISSN:1943-8141/AJTR0168992

Original Article

A novel prognostic signature based on epithelial-mesenchymal transition-associated genes for the prognosis and immune status in breast cancer

Zizheng Wu, Jie Zheng, Shuai Men, Shuangrui Sui, Weitao Yan, Yinfeng Liu, Meng Han

Breast Disease Diagnosis and Treatment Center, The First Hospital of Qinhuangdao, Qinhuangdao 066000, Hebei, China

Received September 24, 2025; Accepted January 29, 2026; Epub March 15, 2026; Published March 30, 2026

Abstract: Objective: Breast cancer (BC) incidence continues to rise, and recurrence and metastasis remain major contributors to mortality. The epithelial-mesenchymal transition (EMT), associated with the acquisition of invasive functions by epithelial cells, also promotes resistance to anticancer therapies. Here, an EMT-based prognostic model was developed to enhance BC outcome prediction. Methods: Clinical and gene expression data from the TCGA were randomly assigned to discovery and validation cohorts. Univariate Cox regression and LASSO analyses were utilized to develop a prognostic signature. The Cell-Type Identification by Estimating Relative Subsets of RNA Transcripts and ESTIMATE algorithms were applied to evaluate the tumor microenvironment (TME). Enriched immune-associated pathways were found using GSEA. SDC1 was knocked down and overexpressed in MCF-7 and MDA-MB-231 cells, and its effects on cell proliferation, apoptosis, migration, invasion and EMT key markers were evaluated by CCK-8, flow cytometry, Transwell and Western blot. Results: Ten EMT-related genes (TP63, TFPI2, ALX4, F2RL2, LEF1, PDLIM4, NDRG2, HMGB3, SDC1, and KRT17) showed significant links with overall survival. The resulting signature was used to allocate individuals into high- and low-risk groups with distinct prognoses in both cohorts. M0 macrophages, activated natural killer cells, memory-activated $CD4^+$ T cells, and immunological scores were all lower in high-risk patients. GSEA revealed that the low-risk group demonstrated greater enrichment in immune-related pathways, including cell adhesion molecules, cytokine-cytokine receptor interactions, and T-cell receptor signaling. SDC1 expression was markedly raised in tumor tissues and correlated with several clinical and pathological features. Knockdown of SDC1 in vitro inhibited the proliferation, migration and invasion of BC cells, induced apoptosis and reversed EMT process. Overexpression of SDC1 had the opposite cancer-promoting effect. Conclusion: This ten-gene EMT-based signature reliably predicts BC prognosis and offers valuable insight into the tumor immune microenvironment.

Keywords: Breast cancer, epithelial-mesenchymal transition, prognosis, tumor microenvironment, immune, prognostic signature

Introduction

Breast cancer (BC) is the most prevalent female cancer worldwide [1]. Moreover, its incidence continues to rise by roughly 0.5% annually, a trend partly linked to declining fertility rates and increasing body weight [1]. Despite considerable progress in diagnostic methods, therapeutic options, and overall survival improvements, BC still represents a major cause of cancer-related mortality in women [2]. Tumor size (T), lymph-node involvement (N), and distant metastasis (M) are all included in AJCC TNM staging, which is used extensively in clinical settings to evaluate prognosis in BC patients [3]. However, this staging approach does not capture the profound molecular heterogeneity of BC, resulting in significant survival differences even among patients with similar stages. Thus, identifying novel biomarkers and therapeutic targets remains essential for enhancing prognostic precision and informing personalized treatment plans.

Recurrence and metastasis are the leading contributors to BC mortality [4]. Multiple biological processes drive metastatic spread, including tumor invasion and migration, angiogenesis,

https://doi.org/10.62347/JQQX7009

immune evasion, and the epithelial-mesenchymal transition (EMT) [5-7]. Among these mechanisms, the EMT is considered an initiator of the metastatic cascade [8, 9]. The EMT involves the acquisition of mesenchymal features by epithelial cells, after which they lose both polarity and adhesiveness, and become more invasive and migratory [9, 10]. Cells undergoing EMT also demonstrate various forms of phenotypic plasticity, including stem-like properties, metastatic dormancy, and resistance to therapy [11]. Evidence suggests that BC cells can exploit EMT-related pathways to generate cancer stem-like cells, thereby promoting metastasis and contributing to treatment resistance [12-14]. As a result, EMT is crucial in determining the prognosis of BC patients.

In this work, a predictive signature, based on 10 EMT-associated genes, was developed for BC. We evaluated variations in immune cell subsets and signaling between the high- and low-risk cohorts and confirmed their predictive significance. This EMT-based model offers fresh insights for customized prognostic evaluation and may inform tailored treatment approaches for BC.

Methods

Data and preprocessing

The UCSC Xena platform was used to obtain the $\log_2$(FPKM+1) gene expression profiles and associated clinical data for breast invasive carcinoma (BRCA) from TCGA. The following criteria were applied to determine which samples were included: (a) a confirmed diagnosis of BRCA; (b) availability of both clinical and gene expression information; and (c) comprehensive clinical data, including survival status and duration.

Ensembl gene identifiers were expressed as gene symbols using the GENCODE annotation file (hg38, gencode v22 annotation gene probemap) to ensure accurate transcript mapping [15]. For further analysis, expression values expressed as FPKM were converted to transcripts per kilobase million (TPM) [16]. All TCGA data are publicly accessible, and this study adhered to TCGA data usage and publication guidelines. The EMT Gene Database provided EMT-related genes.

Differential expression of EMT-associated genes between tumor and normal tissues was assessed using "limma", with the criteria of $P < 0.05$ and $|\log_2 \text{fold change}| > 1$. To display differentially expressed genes (DEGs), volcano plots were generated using "ggrepel" and "ggplot2" in R.

Establishment and validation of EMT-associated prognostic signature

A total of 891 TCGA samples were randomly separated into discovery (n = 624) and validation (n = 267) cohorts in a ratio of 7:3. R was used to compile and compare the baseline information of members of the two cohorts (**Table 1**). Differences between groups were examined with chi-square tests, with statistical significance defined as $P < 0.05$.

In the discovery cohort, univariate Cox proportional hazards regression was applied to determine EMT-related mRNAs linked to overall survival (OS). Genes were deemed survival-related if their *p*-value was less than 0.01. LASSO regression was then utilized for further refinement [17], implemented with the "glmnet" package, where the minimum cross-validation error determined the optimal λ value.

Risk scores (RSs) for individual patients were computed as:

$$RS = coef(mRNA_1) \times expression(mRNA_1) + coef(mRNA_2) \times expression(mRNA_2) + \ldots + coef(mRNAn) \times expression(mRNAn).$$

Where each coefficient corresponds to the weight assigned by the LASSO model, and each expression value represents the gene's transcript level. Using the median RS, members of the discovery cohort were allocated to high-risk (HR) and low-risk (LR) groups. The same formula and cutoff were then utilized to stratify patients in the validation cohort. For comparison of OS between the HR and LR groups, survival was assessed using Kaplan-Meier curves with the "survival" and "survminer" packages. The "survivalROC" program was used to create ROC curves for evaluating the gene signature's prognostic ability in both cohorts.

RS distributions, OS scatter plots, and heatmaps displaying the expression of the 10 EMT-associated genes were constructed to illustrate differences between the HR and LR subgroups. In addition, the "survival" package was utilized for univariate and multivariate Cox regression

Table 1. Clinical characteristics of patients in discovery cohort and validation cohort (n = 891)

Variables	Discovery cohort n (%)	Validation cohort n (%)	*p*
Age, years			0.334
≤ 60	341 (54.647)	156 (58.427)	
> 60	283 (45.353)	111 (41.573)	
T			0.159
T1	169 (27.083)	74 (27.715)	
T2	366 (58.654)	148 (55.431)	
T3	69 (11.058)	41 (15.356)	
T4	20 (3.205)	4 (1.498)	
N			0.772
N0	299 (47.917)	132 (49.438)	
N1	210 (33.654)	89 (33.333)	
N2	64 (10.256)	30 (11.236)	
N3	51 (8.173)	16 (5.993)	
M			0.137
M0	611 (97.917)	266 (99.626)	
M1	13 (2.083)	1 (0.374)	
Stage			0.290
I	114 (18.269)	49 (18.352)	
II	361 (57.853)	154 (57.678)	
III	136 (21.795)	63 (23.596)	
IV	13 (2.083)	1 (0.374)	
Subtype			0.071
Non-triple negative	528 (84.615)	212 (79.401)	
Triple negative	96 (15.385)	55 (20.599)	

Abbreviations: T: tumor size; N: nodal status; M: metastases.

to evaluate the risk score's independence and relationship to other clinical variables.

Immunocyte infiltration analysis

The compositions of 22 immune cell subtypes in the discovery cohort were measured using CIBERSORT, with 1,000 permutations set [18]. For further analysis, only samples with $p < 0.05$ were kept. Wilcoxon tests were applied to assess immune cell composition differences between the two risk groups.

ESTIMATE was utilized to determine stromal scores, immunological scores, and tumor purity [19], with *p*-values < 0.05 deemed significant.

GSEA

Gene Set Enrichment Analysis (GSEA) [20] was conducted to examine pathways associated with the prognostic signature in the two groups. Unlike methods that focus solely on individual DEGs, GSEA evaluates predefined gene sets, enabling the detection of broader biological processes, such as cancer-related pathways, metabolic networks, transcriptional programs, and stress-response mechanisms. This approach also improves reproducibility and facilitates clearer interpretation of molecular profiling results. The criterion for statistical significance was an FDR *q*-value < 0.25 and a *p*-value < 0.05.

Sampling

Between November 2021 and March 2023, 100 BC tissues and the corresponding normal tissue samples were collected from the First Hospital of Qinhuangdao, China. The study was approved by the Ethics Committee of the First Hospital of Qinhuangdao, and all participants provided written informed consent.

qRT-PCR

Total RNA was collected from tissues using TRIzol (Thermo Fisher Scientific, USA), and its

concentration was assessed spectrophotometrically (NanoDrop 2000, Thermo Fisher Scientific). Taq Pro Universal SYBR qPCR Master Mix (Vazyme, Nanjing, China) was utilized for qRT-PCR as directed. The primer sequences were: *SDC1:* forward 5'-CCACCATGAGACCTCAACCC-3', reverse 5'-GCCACTACAGCCGTATTCTCC-3'; *β-actin:* forward 5'-CATGTACGTTGCTATCCAGGC-3', reverse 5'-CTCCTTAATGTCACGCACGAT-3'. Using β-actin for normalization, the $2^{-\Delta\Delta CT}$ technique was applied for calculating relative expression of SDC1.

Cell grouping

Human normal breast epithelial cells MCF10A and human BC cell lines (MCF-7, BT-549, MDA-MB-468 and MDA-MB-231) were purchased from SUNNCELL (SNL-225, SNL-060, SNL-271, SNL-061, SNL-073, Wuhan, China). All cells were cultured in DMEM complete medium (containing 10% fetal bovine serum (FBS), 1% penicillin-streptomycin solution and 90% DMEM medium). The culture conditions were 37°C, 5% CO_2 in a sterile cell incubator. When the cell confluence reached 70-80%, the passage operation was carried out, and then the cells in logarithmic growth phase were taken for subsequent experiments.

SDC1 overexpression plasmid (SDC1) and its no-load control (Vector), as well as SDC1 knockdown plasmid (sh-SDC1) and its negative control (sh-NC) were ordered by Jinkairui Bioengineering Co., Ltd. (Wuhan, China). MCF-7 and MDA-MB-231 cells were inoculated into the well plate at a density of 5×10^5 cells/well one day in advance. The cell density was about 80% and the cells were evenly distributed and in good condition. Before transfection, the cells were starved. According to the instructions of transfection reagent, the transfection reagent Lipofectamine™ 3000 (L3000001, Invitrogen, Waltham, MA, USA) was diluted with serum-free medium, and then mixed with sh-SDC1, sh-NC, SDC1 and Vector plasmids. After standing at room temperature for 5 min, the cells were incubated for 20 min to form a liposome-nucleic acid complex. The old medium was removed and replaced with fresh conventional medium. The cells that need to be transfected were cultured with the above working solution for 48 h for subsequent detection. At the same time, the cells were collected to extract total protein, and the efficiency of SDC1 overexpression and knockdown was detected by Western blot.

The transfected MCF-7 and MDA-MB-231 cells were seeded into 96-well plates at a density of 1×10^3 cells/well, and the cells were routinely cultured in the cell incubator for 24 h. Subsequently, 10 μL of CCK-8 solution (SNK-010, SUNNCELL) was added to each well and incubated in darkness for 2 h. The optical density of each group of cells at 450 nm was measured by a microplate reader (Cytation 5, Agilent, Santa Clara, CA, USA) to evaluate cell viability.

Western blot

Tumor tissues and cells were lysed with 900 μL RIPA (P0013B, Beyotime, Shanghai, China) and 10 μL PMSF (ST506, Beyotime), and incubated on ice for 30 min. After centrifugation (4°C, 12000 r/min, 15 min), the supernatant was collected for the total protein. A small amount of supernatant was used to determine the pro tein concentration in the sample using the BCA protein assay kit (P0010, Beyotime). Each well was loaded with 30 μg protein, and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed. The protein was transferred to polyvinylidene fluoride (PVDF) membrane, blocked with 5% skim milk for 2 h, and then incubated with primary antibody and secondary antibody in turn. The color was developed by electrochemical luminescence (ECL) (P0018M, Beyotime), and photographed in an automatic chemiluminescence imager (Chemi Doc XRS, Bio-Rad, Hercules, CA, USA). The image was analyzed by Image J, and the ratio of the gray value of the target protein to the internal reference protein (GAPDH) was semi-quantitatively analyzed. The antibody information used in the experiment was as follows: SDC1 (DF6367, 1:2000, Affinity, Jiangsu, China), Ki-67 (ab92742, 1:5000, Abcam), PCNA (ab152112, 1:3000, Abcam), caspase 3 (ab32351, 1:5000, Abcam), Cleaved-caspase 3 (ab32042, 1:500, Abcam), Bcl-2 (ab117115, 1:1000, Abcam), Bax (ab182734, 1:1000, Abcam), E-cadherin (ab40772, 1:10000, Abcam), N-cadherin (ab76011, 1:20000, Abcam), Vimentin (ab16700, 1:1000, Abcam), Slug (ab27568, 1:1000, Abcam), Twist 2 (ab66031, 1:1000, Abcam), ZEB1 (ab203829, 1:500, Abcam) and GAPDH (ab128915, 1:50000, Abcam).

Flow cytometry experiments

MCF-7 and MDA-MB-231 cells were digested with trypsin without EDTA and washed twice

with PBS. Then, 500 μL of binding buffer was added and gently blown to prepare a single cell suspension. 5 μL of PI and 5 μL of Annexin V-APC (E-CK-A217, Elabscience, Wuhan, China) were added and gently mixed, and incubated in dark at room temperature for 15 min. Cell apoptosis was analyzed by flow cytometry (Attune NxT, Thermo Fisher).

Transwell experiments

The Matrigel gel was diluted at a ratio of 1:9 and added to the Transwell chamber and incubated in the incubator for 4 h. The density of MCF-7 and MDA-MB-231 cells was adjusted to 1×10^4 cells/mL, and 200 μL cell suspension was inoculated into serum-free medium and added into Transwell chamber. 500 μL complete medium containing 20% FBS was added to the lower chamber. After 24 h of culture in the cell incubator, the Transwell chamber was carefully removed, the upper and lower chamber fluids were discarded, and PBS was washed twice to remove cells that did not cross the membrane. Subsequently, the chamber was fixed with 4% paraformaldehyde and the cells were stained with 0.1% crystal violet for 20 min. After washing away the crystal violet dye solution, six different fields of view were randomly selected under the microscope (CKX53, Olympus, Tokyo, Japan) for photo recording, and the number of invasive cells was analyzed by Image J software. For the cell migration experiment, the cell suspension was inoculated in the Transwell chamber without Matrigel gel coating, and the operation steps were exactly the same as the invasion experiment.

Immunofluorescence

MCF-7 and MDA-MB-231 cells were inoculated into 24-well plates with sterile coverslips for cell climbing, and cell transfection was performed after cell adherence. Cells were fixed with 4% paraformaldehyde for 20 min at room temperature, washed with PBS for 3 times, permeabilized with PBS containing 0.5% Triton X-100 for 15 min, and then blocked with blocking solution (P0260, Beyotime) for 1 h. Cells were incubated overnight with primary antibodies against E-cadherin (ab40772, 1:500, Abcam) and Vimentin (ab16700, 1:500, Abcam) at 4°C. After washing with PBS for 3 times, fluorescent antibody goat anti-rabbit IgG H&L (Alexa Fluor® 488) (ab150077, 1:1000, Abcam) was added and incubated at room temperature for 2 h. Finally, DAPI staining solution was added dropwise and incubated at room temperature in dark for 10 min. The cells were observed under a fluorescence microscope (BX53, Olympus) and the images were collected. The protein fluorescence intensity was analyzed by Image J software.

Statistical analysis

R (version 4.1.2) was utilized for all analyses. A two-tailed p-value < 0.05 was deemed statistically significant unless otherwise indicated. The OS rates in the two groups were compared using Kaplan-Meier curves. Furthermore, factors independently predictive of BC outcomes were investigated using univariate and multivariate Cox regression. For cell experiments, data statistical analysis was performed using GraphPad Prism 9.0 software. All data underwent normality and homogeneity of variance tests. One-way ANOVA analysis and Tukey post hoc tests were applied for comparisons among multiple groups, and the data of each group were expressed as mean ± standard deviation. A p value < 0.05 was considered statistically significant.

Results

Screening of EMT-associated genes and prognostic signature construction

The TCGA database was used to obtain transcriptome profiles from 891 BC samples and 113 normal breast tissues, together with comprehensive clinical data. EMT-associated DEGs were identified using “limma” in R. A volcano plot was used to display the 124 up-regulated and 186 down-regulated EMT-associated genes identified in this investigation (**Figure 1A**).

R software was then used to randomly divide the 891 BC patients into discovery (n = 624) and validation (n = 267) cohorts. The relationships between EMT-associated DEGs and prognosis were assessed using univariate regression, identifying 10 genes linked to OS (P < 0.01) (**Figure 1B**). These candidate genes were further refined using LASSO regression analysis (**Figure 1C**, **1D**), ultimately resulting in a ten-gene risk signature: NDRG2, ALX4, HMGB3, TP63, SDC1, LEF1, PDLIM4, TFPI2, KRT17, and F2RL2.

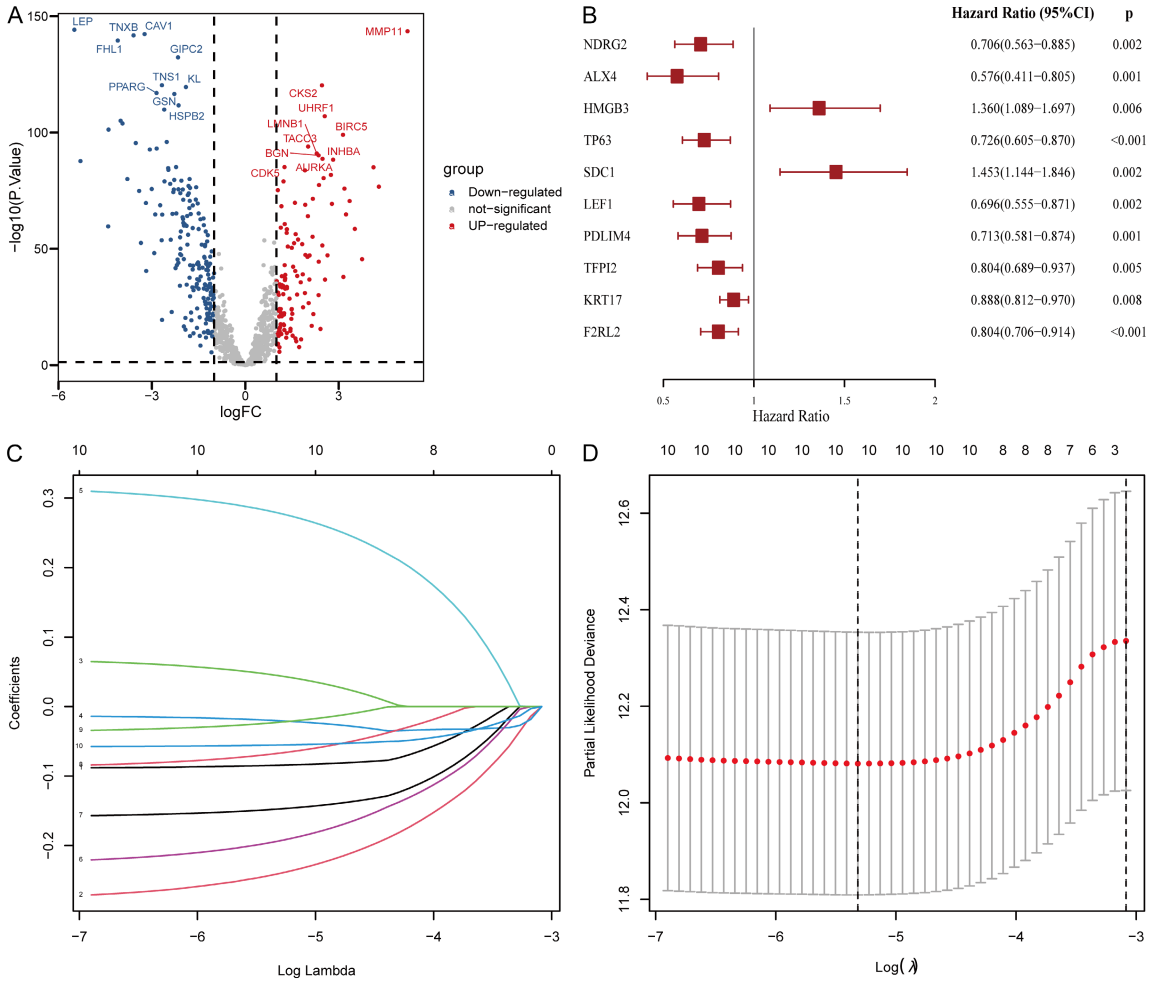

Figure 1. The process of EMT-associated genes selection. A: Volcano plot of differentially expressed EMT-associated mRNAs between breast cancer and normal tissues. The blue are down-regulated genes, and the red are up-regulated genes. B: Forest map of ten candidate EMT-associated genes selected by univariate Cox regression analysis associated with BC overall survival in the discovery cohort. TFPI2, ALX4, TP63, KRT17, NDRG2, PDLIM4, LEF1, F2RL2 are protective factors, and SDC1 and HMGB3 are risk factors. C and D: The performance of least absolute shrinkage and selection operator (LASSO) analysis.

Furthermore, the remaining genes functioned as protective factors; however, SDC1 and HMGB3 were identified as risk genes, with hazard ratios greater than 1. RS values for patients in the two cohorts were computed using the levels of these genes and the corresponding LASSO coefficients. RS = NDRG2 level × (-0.0844622322463379) + ALX4 level × (-0.0844622322463379) + HMGB3 level × 0.0445984754563584 + TP63 level × (-0.0203039658242269) + SDC1 level × (0.278688392836505) + LEF1 level × (-0.193642130476875) + PDLIM4 level × (-0.147473623029924) + TFPI2 level × (-0.0673056776543878) + KRT17 level × (-0.0227280886401868) + F2RL2 level × (-0.0553600089080876).

Patients were allocated to HR and LR groups using the median RS. The HR group showed markedly lower OS ($P < 0.0001$), as determined by Kaplan-Meier analysis (**Figure 2A**). With AUC values of 0.747 at three years and 0.745 at five years, the ROC analysis confirmed the good predictive accuracy of the model for BC survival (**Figure 2B**).

Plots of RS distributions and OS scatter plots illustrated the relationship between RS and survival, revealing that higher scores were associated with worse outcomes (**Figure 2C**, **2D**). A heatmap comparing gene expression patterns between the HR and LR subgroups indicated that HMGB3 and SDC1 were significantly

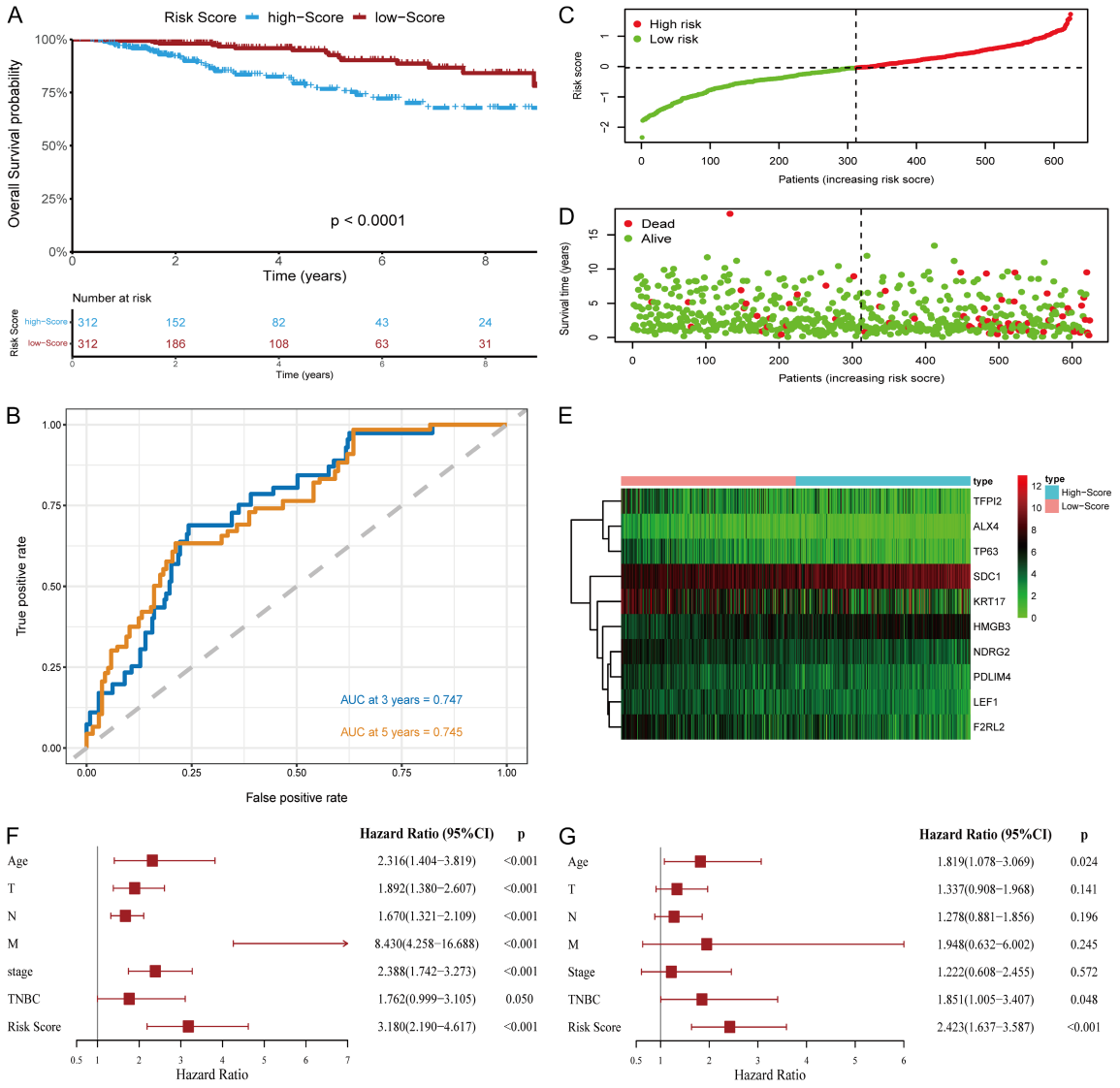

Figure 2. Evaluation of the prognostic signature in the discovery cohort. A: Kaplan Meier curve shows that the prognosis of patients between high and low risk groups is different. B: Time-dependent receiver operating characteristic (ROC) curve of the risk score. C: The risk score distribution. D: The survival status distribution. The green are alive patients, and the red are dead patients. E: Heat map of the ten EMT-associated risk genes expression differences between high and low risk groups. F and G: Univariate and multivariate analyses of the signature and clinical factors.

up-regulated in the HR group, whereas TFPI2, ALX4, TP63, KRT17, NDRG2, PDLIM4, LEF1, and F2RL2 showed higher expression in the LR group (**Figure 2E**).

The results of the univariate analysis indicated that the prognostic signature was strongly associated with OS (HR = 3.180, 95% CI = 2.190-4.617, *P* < 0.001) (**Figure 2F**). After adjusting for additional factors, including age, T, N, M, and overall stage, and triple-negative breast cancer (TNBC), the signature remained independently predictive of prognosis (HR = 2.423, 95% CI = 1.637-3.587, *P* < 0.001) (**Figure 2G**).

The validation cohort was treated similarly to the treatment cohort. HR patients demonstrated significantly lower survival outcomes (*P* = 0.0073), as indicated by Kaplan-Meier analysis (**Figure 3A**). ROC analysis once again indicated high predictive accuracy, with respective AUC values of 0.791 and 0.656 at 3 and 5 years (**Figure 3B**). Risk score and survival plots consistently showed that higher RS correlated with poorer prognosis (**Figure 3C**, **3D**). The heatmap

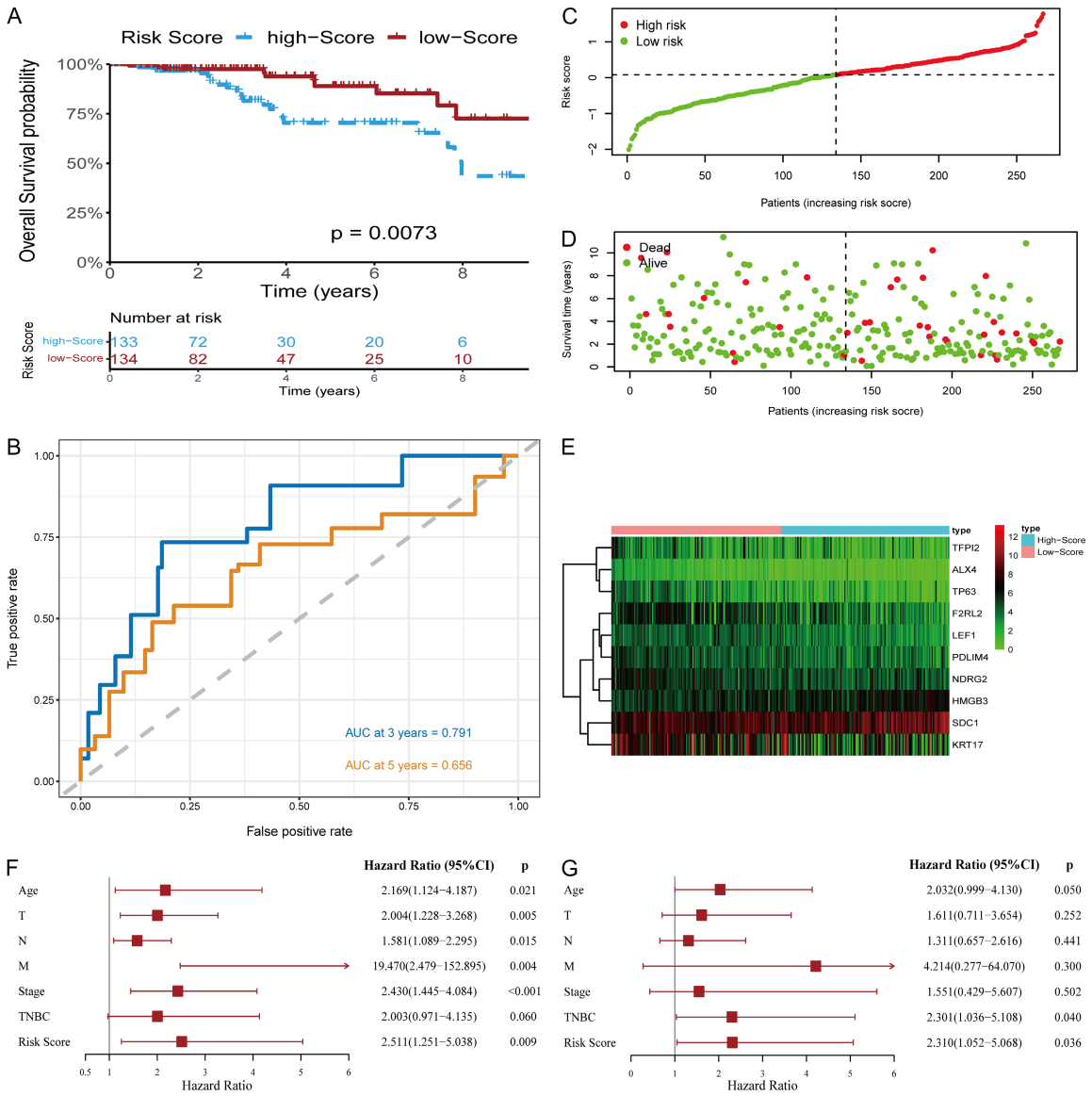

Figure 3. Evaluation of the prognostic signature in the validation cohort. A: Kaplan Meier curve shows that the prognosis of patients between high and low risk groups is different. B: Time-dependent receiver operating characteristic (ROC) curve of the risk score. C: The risk score distribution. D: The survival status distribution. E: Heat map of the ten EMT-associated risk genes expression differences between high and low risk groups. F and G: Univariate and multivariate analyses of the signature and clinical factors.

of gene expression in the validation cohort reflected the same pattern observed in the discovery cohort: HMGB3 and SDC1 were up-regulated in HR patients. LR patients also had higher levels of TFPI2, ALX4, TP63, KRT17, NDRG2, PDLIM4, LEF1, and F2RL2 (**Figure 3E**).

The ten-gene signature was substantially associated with overall survival, as indicated by univariate Cox analysis (HR = 2.511, 95% CI = 1.251-5.038, *P* = 0.009) (**Figure 3F**). After adjustments for age, tumor stage, nodal status, metastasis, clinical stage, and TNBC status, the risk signature remained independently predictive of survival in the validation cohort (HR = 2.310, 95% CI = 1.052-5.068, *P* = 0.036) (**Figure 3G**).

The prognostic signature and TME

The discovery cohort's HR and LR groups' immune cell proportions were compared using the Wilcoxon test. We found that the HR subgroup had significantly higher levels of $CD8^+$ T

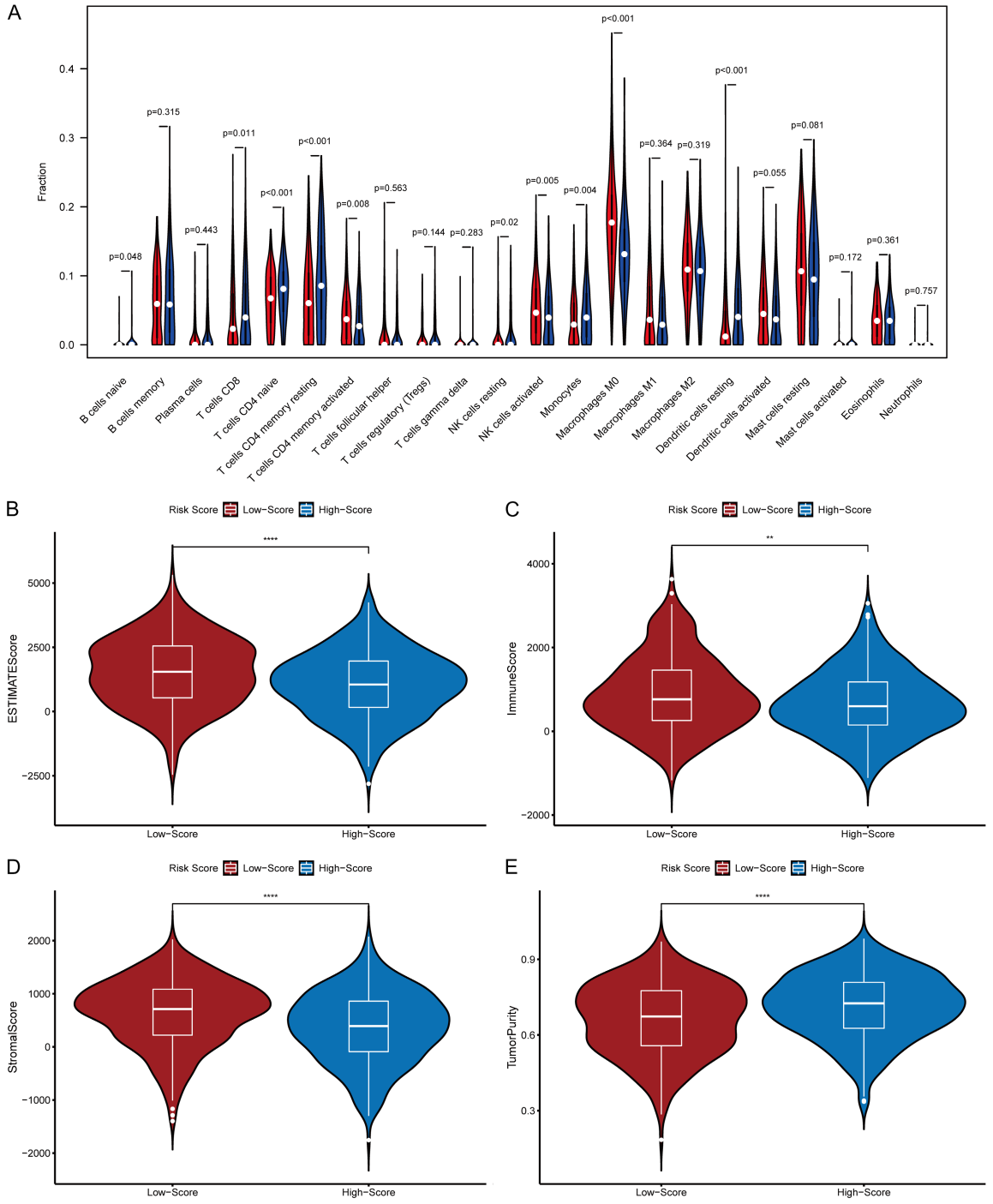

Figure 4. Association between prognostic signature and tumor microenvironment. A: Relationship between the prognostic signature and immune cell infiltration. B: ESTIMATE score between high and low risk groups. C: Immunity score between high and low risk groups. D: Stroma score between high and low risk groups. E: Tumor purity between high and low risk groups.

cells ($P = 0.011$), naïve CD4$^+$ T cells ($P < 0.001$), resting memory CD4$^+$ T cells ($P < 0.001$), monocytes ($P = 0.004$), and resting dendritic cells ($P < 0.001$). In comparison, the same group showed a significant decrease in activated memory CD4$^+$ T cells ($P = 0.008$), active NK cells ($P = 0.005$), and M0 macrophages ($P < 0.001$) (**Figure 4A**).

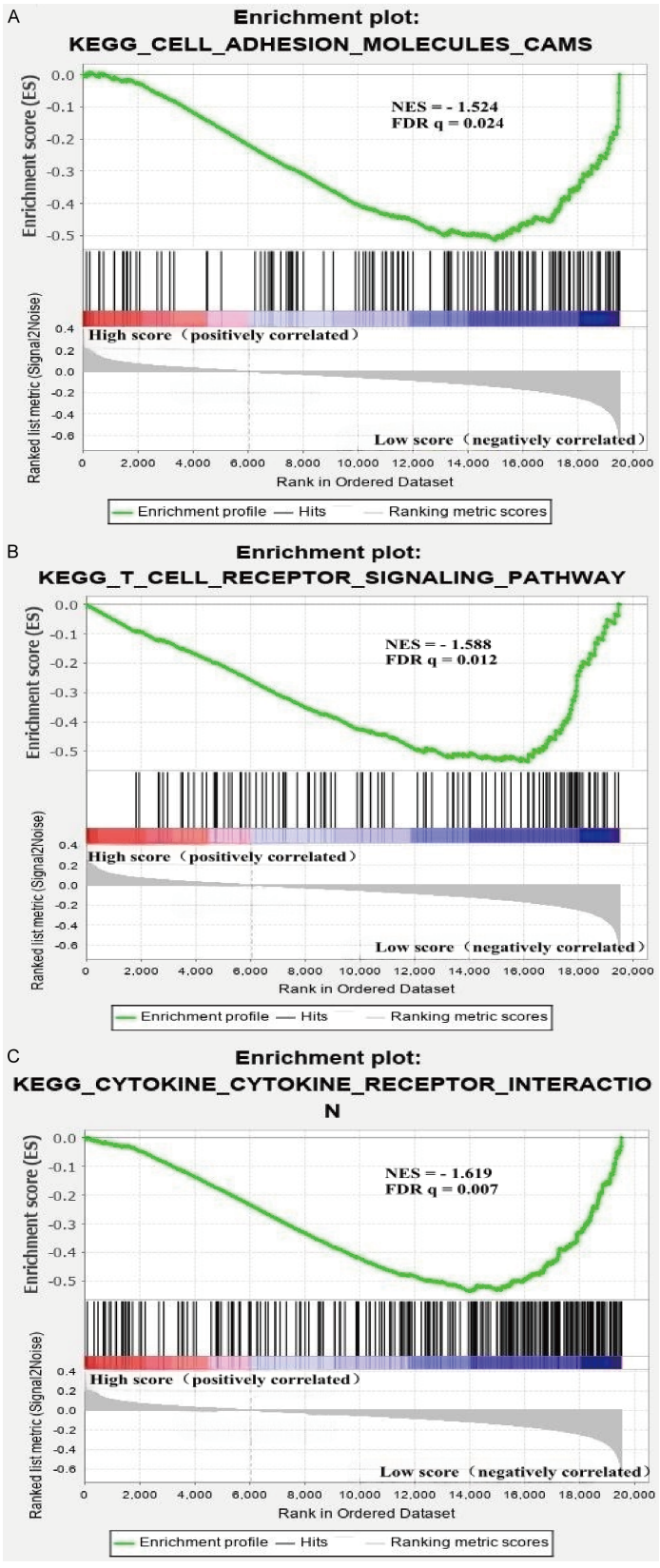

Figure 5. Gene set enrichment analysis (GSEA) for functional annotation of the prognostic signature. A: Cell adhesion molecules (CAMs). B: T cell receptor network. C: Cytokine-cytokine receptor association.

Elevated RS was negatively linked with ESTIMATE ($P < 0.0001$), immune ($P < 0.01$), and stromal scores ($P < 0.0001$), but positively associated with enhanced tumor purity ($P < 0.0001$), as determined by analysis using the ESTIMATE algorithm (**Figure 4B-E**). These results imply a strong correlation between the TME and the prognostic signature in individuals with BC.

GSEA for prognostic signature functional annotation

According to our study, the LR group demonstrated a greater enrichment of immune-related biochemical pathways relative to the HR group. GSEA further identified three immunity-associated KEGG pathways, CAMs, cytokine-cytokine receptor interactions, and the T-cell receptor signaling network, as significantly enriched (**Figure 5**).

External experimental validation of prognostic signature

The results of the GEPIA database showed that the expression level of SDC1 was significantly up-regulated in BC tissues (**Figure 6A**), suggesting that SDC1 might play an important role in the occurrence and development of BC. To verify the results of bioinformatics analysis, we collected 100 samples of BC tissues and adjacent normal tissues. The data showed that the mRNA and protein levels of SDC1 in BC tissues were significantly higher than those in adjacent normal tissues (**Figure 6B**, **6C**). Patients were allocated to two groups based on the median SDC1 level. Those in the high-expression group showed an increased risk of developing tumors larger than 2 cm, lymph node metastases, and TNM stage III relative to those with low expression ($P < 0.05$) (**Table 2**). Subsequently, the protein expression of SDC1 was detected at the cellular level. Compared with normal breast epithelial cells (MCF10A),

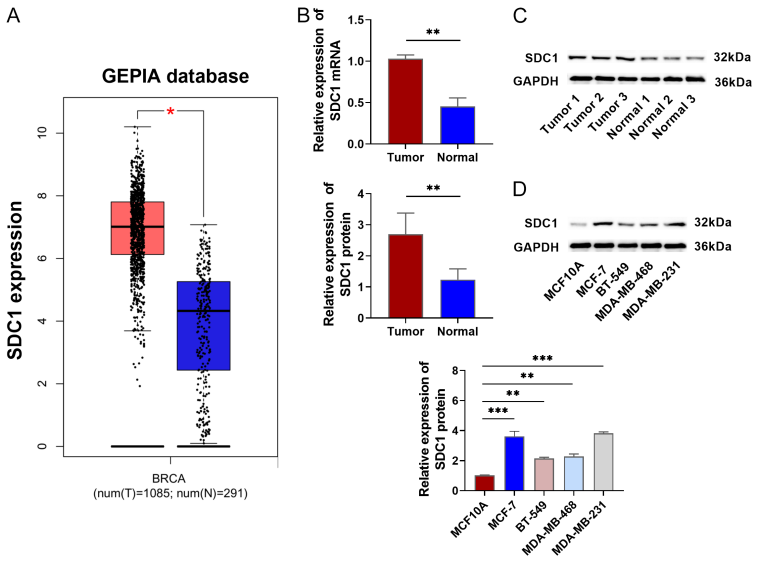

Figure 6. Up-regulation of SDC1 in BC. A: The expression level of SDC1 in BC tissues in the GEPIA database. B: qRT-PCR was used to detect the expression level of SDC1 in BC tissues and adjacent normal tissues. C: Western blot was used to detect the expression of SDC1 protein in BC tissues and adjacent normal tissues. D: Western blot was used to detect the protein expression of SDC1 in human normal breast epithelial cells (MCF10A) and BC cell lines (MCF-7, BT-549, MDA-MB-468 and MDA-MB-231). $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

SDC1 protein was highly expressed in four BC cell lines (MCF-7, BT-549, MDA-MB-468 and MDA-MB-231), and the expression level was the highest in MCF-7 and MDA-MB-231 cells (**Figure 6D**), so these two cells were selected for subsequent experiments. In summary, SDC1 might be involved in the progression of BC as an oncogene.

Knockdown of SDC1 expression inhibited BC cell proliferation and promoted apoptosis

We explored the effect of SDC1 on the malignant phenotype of BC cells. SDC1 was knocked down and overexpressed in MCF-7 and MDA-MB-231 cells, respectively, and the corresponding negative controls were used as controls. The level of SDC1 protein in sh-SDC1 group was significantly decreased, and the level of SDC1 protein in SDC1 group was significantly increased (**Figure 7A**), which confirmed the effectiveness of transfection. The viability of BC cells decreased significantly after knockdown of SDC1, but increased significantly after overexpression of SDC1 (**Figure 7B**, **7C**). The results of flow cytometry showed that the apoptosis rate increased significantly after knockdown of SDC1, and decreased significantly after overexpression of SDC1 (**Figure 7D**). This indicated that knockdown significantly inhibited the proliferation of BC cells and promoted apoptosis. After knocking down SDC1, the protein levels of Ki-67 and PCNA, the key markers of cell proliferation, were significantly reduced, the expression of pro-apoptotic proteins Bax and Cleaved-caspase 3 was increased, and the expression of anti-apoptotic protein Bcl-2 was decreased (**Figure 7E-I**). In contrast, overexpression of SDC1 had the opposite effect. These results indicated that knockdown of SDC1 effectively inhibited BC cell proliferation and induced apoptosis.

Knockdown of SDC1 expression inhibited BC cell migration, invasion and EMT process

Subsequently, Transwell assay results showed that after knocking down SDC1, the migration and invasion abilities of MCF-7 and MDA-MB-231 cells were significantly weakened, and the number of migration and invasion cells passing through the basement membrane was significantly reduced. In contrast, overexpression of SDC1 promoted cell migration and invasion (**Figure 8A**, **8B**). Western blot results showed that knockdown of SDC1 significantly increased the expression of epithelial marker E-cadherin, decreased the expression of mesenchymal markers N-cadherin, Vimentin and core transcription factors Slug, Twist 2, ZEB1. Overexpression of SDC1 showed a completely opposite trend, inducing EMT phenotype (**Figure 8C-E**). Immunofluorescence staining showed that in the sh-SDC1 group, the fluorescence intensity of E-cadherin was significantly enhanced, while the fluorescence intensity of Vimentin was significantly decreased. Overexpression of SDC1 resulted in a decrease in E-cadherin signal and an increase in Vimentin signal (**Figure 8F**, **8G**), suggesting a more invasive mesenchymal transition. In summary, knockdown of SDC1 effectively reversed the EMT process of BC cells and significantly inhibited their migration and invasion ability.

Table 2. Relationship between expression level of SDC1 and clinical features in breast cancer

Variables	*n*	Low expression of SDC1 (*n* = 50)	High expression of SDC1 (*n* = 50)	χ^2	*p* value
Age (yr)				0.043	0.836
≤ 60	63	32	31		
> 60	37	18	19		
Tumor size				8.319	0.004
≤ 2 cm	38	26	12		
> 2 cm	62	24	38		
Lymph node metastasis				17.361	< 0.001
No	64	42	22		
Yes	36	8	28		
Stage				8.392	0.004
I-II	78	45	33		
III	22	5	17		
Subtype				1.333	0.248
Triple negative	25	15	10		
Non-triple negative	75	35	40		

Discussion

Luminal A, Luminal B, HER2-positive, and TNBC are the four main subtypes of BC based on both histological and molecular features. The complex etiology and substantial heterogeneity of BC make accurate prognostic assessment particularly challenging. Furthermore, recurrence and metastasis continue to be significant causes of death for BC patients [21], and mounting data suggest that abnormal EMT activation is a critical factor in BC metastasis [22, 23]. To enhance the prediction of clinical outcomes in BC, we established a unique EMT-related prognostic signature in this study.

In this study, we identified ten EMT-related genes that showed strong associations with BC prognosis. Among them, SDC1 and HMGB3 acted as risk genes, with elevated levels in the HR group, whereas TFPI2, ALX4, TP63, KRT17, NDRG2, PDLIM4, LEF1, and F2RL2 functioned as protective genes, with raised levels in the LR group. The development of BC has been extensively linked to SDC1, a type I transmembrane proteoglycan that controls cell adhesion and migration. Qiao et al. reported that increased SDC1 expression correlates with poorer disease-free and overall survival, and is associated with more aggressive tumor phenotypes characterized by ER negativity and HER2 positivity [24].

Pham et al. further demonstrated that SDC1 modulates Wnt signaling and may influence BC cell migration [25]. Our research supported these conclusions by demonstrating that elevated SDC1 levels were linked with increased lymph node metastasis, larger tumor size, and more advanced TNM stages.

HMGB3, also known as HMG2a, is highly expressed in several tumor types and contributes to the growth of malignant cells as well as poor clinical outcomes [26]. HMGB3 knockdown has been demonstrated to inhibit tumor cell growth and increase chemotherapy sensitivity, and prior research indicates that it may function as an independent prognostic biomarker in BRCA+ BCs [27].

NDRG2, a member of the NDRG/α-β hydrolase superfamily, is recognized as a tumor suppressor involved in regulating proliferation and metastasis across multiple malignancies [28]. ALX4, a paired-like homeodomain transcription factor expressed primarily in mesenchymal tissues, has been shown to inhibit Wnt/β-catenin signaling in BC. A good prognostic predictor, elevated ALX4 expression is linked to decreased tumor development and metastasis [29, 30]. It has been revealed that TFPI2, a member of the serine protease inhibitor family, suppresses TWIST1-mediated integrin α5 expression, which lowers BC cell proliferation and metastatic potential [31, 32]. TP63 encodes

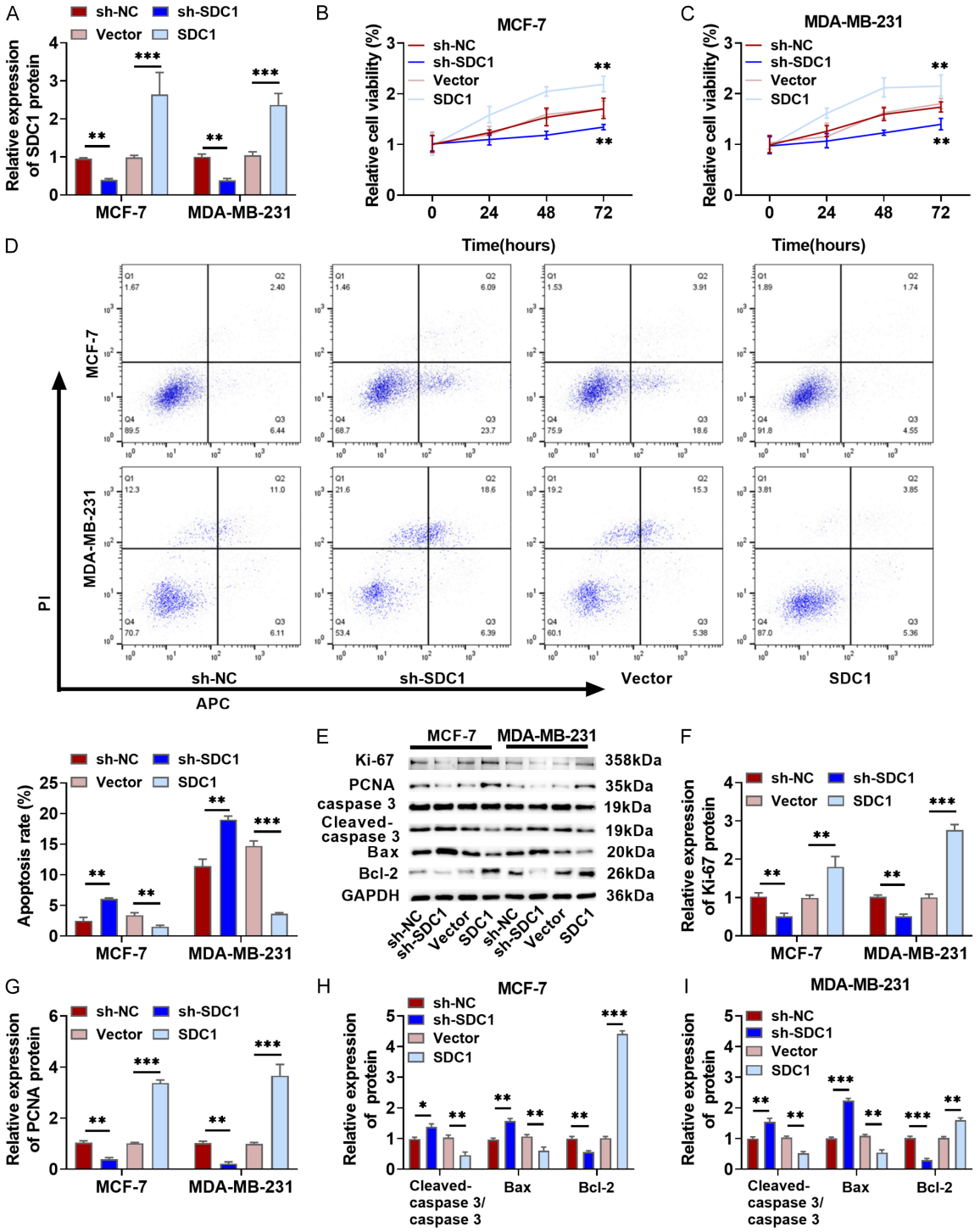

Figure 7. Knockdown of SDC1 expression inhibited BC cell proliferation and promoted apoptosis. A: sh-SDC1, sh-NC, SDC1 and vector were transfected into MCF-7 and MDA-MB-231 cells, and the transfection efficiency of SDC1 in cells was detected by Western blot. B, C: CCK-8 was used to detect the proliferation of MCF-7 and MDA-MB-231 cells. D: Flow cytometry was used to detect the apoptosis of MCF-7 and MDA-MB-231 cells. E-I: The protein levels of proliferation (Ki-67 and PCNA) and apoptosis (cleaved-caspase 3, caspase 3, Bcl-2 and Bax) were detected by Western blot. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

two major isoforms, TAp63 and ΔNp63. TAp63 has been linked to androgen receptor signaling, absence of BRCA mutations, PTEN expression, and improved patient survival [33]. In compari-

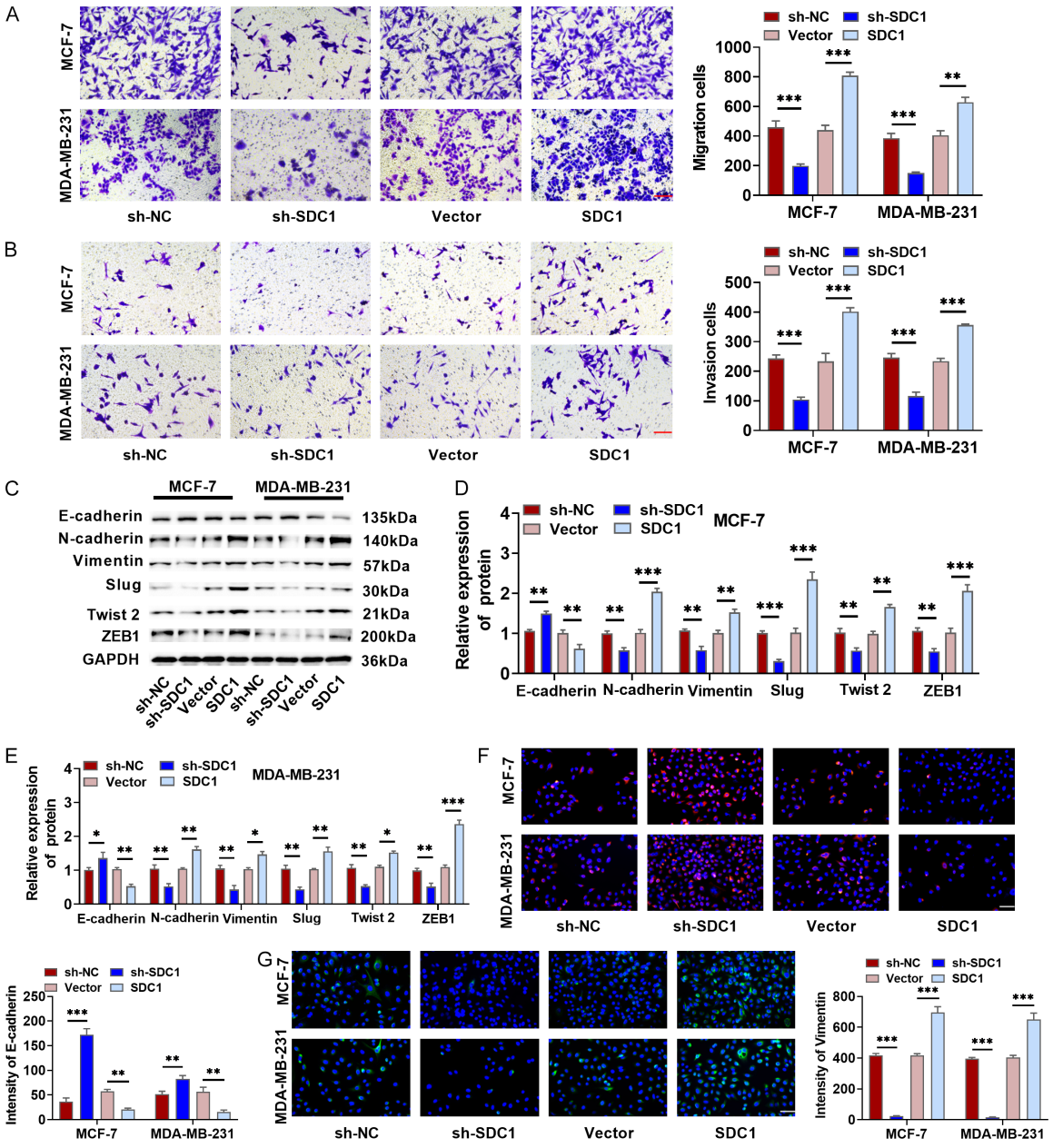

Figure 8. Knockdown of SDC1 expression inhibited BC cell migration, invasion and EMT process. A, B: Transwell assay was used to detect the migration and invasion ability of MCF-7 and MDA-MB-231 cells (×20, 100 μm). C-E: Western blot was used to detect the expression of EMT marker proteins (E-cadherin, N-cadherin, Vimentin, Slug, Twist 2 and ZEB1). F, G: Immunofluorescence was used to detect the expression of E-cadherin and Vimentin in MCF-7 and MDA-MB-231 cells (×40, 50 μm). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

son to normal breast tissue, BC tissues generally express KRT17, a member of the keratin family, at lower levels. Higher KRT17 expression has been associated with better outcomes, particularly in patients with HER2-positive tumors [34]. According to earlier research, PDLIM4 suppression increases BC cell proliferation and decreases apoptosis, suggesting that it may act as a tumor suppressor [35].

Although the roles of LEF1 and F2RL2 in BC have not been extensively characterized, accumulating evidence suggests that all ten of these EMT-associated genes contribute meaningfully to tumor development and progression.

Targeting this gene set may offer new opportunities to modulate EMT in BC. However, the exact mechanisms through which these genes affect prognosis need further exploration.

We successfully divided BC patients into HR and LR categories. Relative to those in the LR group, HR patients in the discovery cohort had markedly worse OS. Multivariate regression showed that the signature was independently predictive of BC outcomes. These findings were validated in the independent cohort, supporting the robustness of the model.

Overall, our results indicate that this prognostic signature can reliably predict overall survival in BC patients. When used in conjunction with conventional clinical indicators, it may enhance risk stratification, allowing for better differentiation between HR and LR individuals and reducing the likelihood of both overtreatment and undertreatment.

Our study revealed significantly higher concentrations of $CD8^+$ T cells ($P = 0.011$), naïve $CD4^+$ T cells ($P < 0.001$), resting memory $CD4^+$ T cells ($P < 0.001$), monocytes ($P = 0.004$), and resting dendritic cells ($P < 0.001$) in the high-risk group. In comparison, memory-activated $CD4^+$ T cells ($P = 0.008$), activated NK cells ($P = 0.005$), and M0 macrophages ($P < 0.001$) were considerably reduced in these patients.

Natural killer (NK) cells are cytotoxic effector lymphocytes that can eliminate cancer cells without prior sensitization [36]. Unlike $CD8^+$ T cells, NK cells do not need antigen presentation; instead, they are regulated through a complex network of activating and inhibitory receptors [37]. Previous research has reported an inverse relationship between NK cell activity and BC stage [38]. In TNBC, NK cells within the tumor microenvironment have been shown to suppress tumor cell invasion [39, 40].

The proportion of tumor cells typically decreases as the number of immune and stromal cells increases; therefore, higher immune and stromal cell infiltration generally reflects a lower tumor cell burden [41]. Higher risk scores were negatively linked with stromal, immunological, and ESTIMATE scores and positively related to increased purity of the tumor. This indicates that patients with HR scores had a greater proportion of tumor cells and fewer immune and stromal cells within their TMEs.

Furthermore, GSEA indicated greater enrichment in immune-related pathways in the LR group. Together, these findings indicate that members of the LR group show a more active immune microenvironment and higher antitumor immune responses compared to those in the HR group, which may contribute to their improved clinical outcomes. The interplay between these EMT-associated genes and the immune landscape in BC warrants further investigation.

We also found a significant increase in SDC1 expression in BC tissues. Subsequently, experiments confirmed that knockdown of SDC1 in BC cells could effectively inhibit cell proliferation, migration and invasion, promote cell apoptosis and reverse the EMT process; overexpression of SDC1 produced completely opposite cancer-promoting effects.

This study yielded promising findings; however, several limitations should be acknowledged. First, the data were retrieved from publicly available online databases, making the analysis a retrospective study. Second, we did not include sufficient experimental verification to support our bioinformatic results. Thus, more research is required to confirm the functions of the 10 EMT-related genes experimentally, test the prognostic signature's predictive value in BC, and investigate other immune-related pathways.

Conclusions

To summarize, the constructed prognostic signature effectively predicts clinical outcomes in BC, providing insight into the underlying immune landscape. High-risk patients revealed distinct alterations in immune-related pathways and tumor-infiltrating immune cell profiles compared to low-risk patients. The EMT-associated genes included in the model also hold promise as potential therapeutic targets for BC. However, these results offer fresh approaches to better BC management; further well-planned experimental research is required to confirm and support our findings.

Acknowledgements

The authors thank TCGA for sharing the BC data. This work was supported by Medical Science Research Project of Hebei (Grant No. 20231893).

All subjects gave their informed consent for inclusion before they participated in the study.

Disclosure of conflict of interest

None.

Address correspondence to: Meng Han, Breast Disease Diagnosis and Treatment Center, The First Hospital of Qinhuangdao, No. 258 Wenhua Road, Haigang District, Qinhuangdao 066000, Hebei, China. ORCID: 0000-0003-1752-7442; E-mail: menghan68527@163.com

References

[1] Siegel RL, Miller KD, Fuchs HE and Jemal A. Cancer statistics, 2021. CA Cancer J Clin 2021; 71: 7-33.

[2] Xu Y, Gong M, Wang Y, Yang Y, Liu S and Zeng Q. Global trends and forecasts of breast cancer incidence and deaths. Sci Data 2023; 10: 334.

[3] Giuliano AE, Edge SB and Hortobagyi GN. Eighth edition of the ajcc cancer staging manual: breast cancer. Ann Surg Oncol 2018; 25: 1783-1785.

[4] Liang Y, Zhang H, Song X and Yang Q. Metastatic heterogeneity of breast cancer: molecular mechanism and potential therapeutic targets. Semin Cancer Biol 2020; 60: 14-27.

[5] Novikov NM, Zolotaryova SY, Gautreau AM and Denisov EV. Mutational drivers of cancer cell migration and invasion. Br J Cancer 2021; 124: 102-114.

[6] Viallard C and Larrivée B. Tumor angiogenesis and vascular normalization: alternative therapeutic targets. Angiogenesis 2017; 20: 409-426.

[7] Dongre A and Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. Nat Rev Mol Cell Biol 2019; 20: 69-84.

[8] Mittal V. Epithelial mesenchymal transition in tumor metastasis. Annu Rev Pathol 2018; 13: 395-412.

[9] Ang HL, Mohan CD, Shanmugam MK, Leong HC, Makvandi P, Rangappa KS, Bishayee A, Kumar AP and Sethi G. Mechanism of epithelial-mesenchymal transition in cancer and its regulation by natural compounds. Med Res Rev 2023; 43: 1141-1200.

[10] Chen T, You Y, Jiang H and Wang ZZ. Epithelial-mesenchymal transition (EMT): a biological process in the development, stem cell differentiation, and tumorigenesis. J Cell Physiol 2017; 232: 3261-3272.

[11] Pastushenko I and Blanpain C. Emt transition states during tumor progression and metastasis. Trends Cell Biol 2019; 29: 212-226.

[12] Lu W and Kang Y. Epithelial-mesenchymal plasticity in cancer progression and metastasis. Dev Cell 2019; 49: 361-374.

[13] Debaugnies M, Rodríguez-Acebes S, Blondeau J, Parent MA, Zocco M, Song Y, de Maertelaer V, Moers V, Latil M, Dubois C, Coulonval K, Impens F, Van Haver D, Dufour S, Uemura A, Sotiropoulou PA, Méndez J and Blanpain C. RHOJ controls EMT-associated resistance to chemotherapy. Nature 2023; 616: 168-175.

[14] Lambert AW and Weinberg RA. Linking emt programmes to normal and neoplastic epithelial stem cells. Nat Rev Cancer 2021; 21: 325-338.

[15] Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigó R and Hubbard TJ. GENCODE: the reference human genome annotation for the encode project. Genome Res 2012; 22: 1760-1774.

[16] Li B, Ruotti V, Stewart RM, Thomson JA and Dewey CN. RNA-seq gene expression estimation with read mapping uncertainty. Bioinformatics 2010; 26: 493-500.

[17] Kidd AC, Mcgettrick M, Tsim S, Halligan DL, Bylesjo M and Blyth KG. Survival prediction in mesothelioma using a scalable lasso regression model: instructions for use and initial performance using clinical predictors. BMJ Open Respir Res 2018; 5: e240.

[18] Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M and Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015; 12: 453-457.

[19] Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB and Verhaak RG. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013; 4: 2612.

[20] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102: 15545-15550.

[21] Li Y, Yu M, Yang M and Yang J. The association of systemic immune-inflammation index with incident breast cancer and all-cause mortality: evidence from a large population-based study. Front Immunol 2025; 16: 1528690.

[22] Park M, Kim D, Ko S, Kim A, Mo K and Yoon H. Breast cancer metastasis: mechanisms and therapeutic implications. Int J Mol Sci 2022; 23: 6806.
[23] Hong D, Fritz AJ, Zaidi SK, van Wijnen AJ, Nickerson JA, Imbalzano AN, Lian JB, Stein JL and Stein GS. Epithelial-to-mesenchymal transition and cancer stem cells contribute to breast cancer heterogeneity. J Cell Physiol 2018; 233: 9136-9144.
[24] Qiao W, Liu H, Guo W, Li P and Deng M. Prognostic and clinical significance of syndecan-1 expression in breast cancer: a systematic review and meta-analysis. Eur J Surg Oncol 2019; 45: 1132-1137.
[25] Pham SH, Pratt K, Okolicsanyi RK, Oikari LE, Yu C, Peall IW, Arif KT, Chalmers TA, Gyimesi M, Griffiths LR and Haupt LM. Syndecan-1 and -4 influence wnt signaling and cell migration in human breast cancers. Biochimie 2022; 198: 60-75.
[26] Wen B, Wei YT and Zhao K. The role of high mobility group protein b3 (HMGB3) in tumor proliferation and drug resistance. Mol Cell Biochem 2021; 476: 1729-1739.
[27] Zhou X, Zhang Q, Liang G, Liang X and Luo B. Overexpression of HMGB3 and its prognostic value in breast cancer. Front Oncol 2022; 12: 1048921.
[28] Lee KW, Lim S and Kim KD. The function of n-myc downstream-regulated gene 2 (NDRG2) as a negative regulator in tumor cell metastasis. Int J Mol Sci 2022; 23: 9365.
[29] Chang H, Mohabir N, Done S and Hamel PA. Loss of alx4 expression in epithelial cells and adjacent stromal cells in breast cancer. J Clin Pathol 2009; 62: 908-914.
[30] Yang J, Han F, Liu W, Chen H, Hao X, Jiang X, Yin L, Huang Y, Cao J, Zhang H and Liu J. Alx4, an epigenetically down regulated tumor suppressor, inhibits breast cancer progression by interfering wnt/β-catenin pathway. J Exp Clin Cancer Res 2017; 36: 170.
[31] Chand HS, Schmidt AE, Bajaj SP and Kisiel W. Structure-function analysis of the reactive site in the first kunitz-type domain of human tissue factor pathway inhibitor-2. J Biol Chem 2004; 279: 17500-17507.
[32] Zhao D, Qiao J, He H, Song J, Zhao S and Yu J. TFPI2 suppresses breast cancer progression through inhibiting TWIST-integrin α5 pathway. Mol Med 2020; 26: 27.
[33] Coates PJ, Nenutil R, Holcakova J, Nekulova M, Podhorec J, Svoboda M and Vojtesek B. P63 isoforms in triple-negative breast cancer: δnp63 associates with the basal phenotype whereas tap63 associates with androgen receptor, lack of brca mutation, pten and improved survival. Virchows Arch 2018; 472: 351-359.
[34] Tang S, Liu W, Yong L, Liu D, Lin X, Huang Y, Wang H and Cai F. Reduced expression of KRT17 predicts poor prognosis in her2(high) breast cancer. Biomolecules 2022; 12: 1183.
[35] Xiao B, Li M, Cui M, Yin C and Zhang B. A large-scale screening and functional sorting of tumour microenvironment prognostic genes for breast cancer patients. Front Endocrinol (Lausanne) 2023; 14: 1131525.
[36] Valipour B, Velaei K, Abedelahi A, Karimipour M, Darabi M and Charoudeh HN. NK cells: an attractive candidate for cancer therapy. J Cell Physiol 2019; 234: 19352-19365.
[37] Wu SY, Fu T, Jiang YZ and Shao ZM. Natural killer cells in cancer biology and therapy. Mol Cancer 2020; 19: 120.
[38] Razeghian E, Kameh MC, Shafiee S, Khalafi F, Jafari F, Asghari M, Kazemi K, Ilkhani S, Shariatzadeh S and Haj-Mirzaian A. The role of the natural killer (NK) cell modulation in breast cancer incidence and progress. Mol Biol Rep 2022; 49: 10935-10948.
[39] Jin H, Choi H, Kim ES, Lee HH, Cho H and Moon A. Natural killer cells inhibit breast cancer cell invasion through downregulation of urokinase-type plasminogen activator. Oncol Rep 2021; 45: 299-308.
[40] Li F and Liu S. Focusing on NK cells and adcc: a promising immunotherapy approach in targeted therapy for HER2-positive breast cancer. Front Immunol 2022; 13: 1083462.
[41] Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH and de Reyniès A. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol 2016; 17: 218.