Heliyon 10 (2024) e37048

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Epithelial cell-related prognostic risk model in breast cancer based on single-cell and bulk RNA sequencing

Man-zhi Xia [a], Hai-chao Yan [b,*]

[a] General Surgery, Shaoxing Maternity and Child Health Care Hospital, Shaoxing, 312000, Zhejiang, China
[b] Department of Thyroid Surgery, The Second Affiliated Hospital of Zhejiang University College of Medicine, Hangzhou, 310009, Zhejiang, China

ARTICLE INFO

Keywords:
Breast cancer
Epithelial cell
Single-cell sequencing
Bulk RNA sequencing
Prognosis

ABSTRACT

Objective: This study aims to construct an epithelial cell-related prognostic risk model for breast cancer (BRCA) and explore its significance.
Methods: GSE42568, GSE10780, GSE245601, and TCGA-BRCA datasets were sourced from public databases. Epithelial cell-related differentially expressed genes were identified using single-cell data analysis. Venn diagrams determined the intersecting genes between epithelial cell-related and BRCA-related genes. Batch Kaplan-Meier (K-M) survival analysis identified core intersecting genes for BRCA overall survival. Consensus clustering, enrichment, LASSO, and COX regression analyses were performed on the core intersecting genes, and then a prognostic risk model was constructed. The diagnostic and prognostic effectiveness of the risk model was subsequently evaluated and immune infiltration analysis was conducted. Finally, qRT-PCR was used to verify the expression of genes in the risk model.
Results: There were 374 intersecting genes between epithelial cell-related and BRCA-related genes, among which 51 core intersecting genes were associated with BRCA prognosis. Consensus clustering categorized TCGA-BRCA into C1 and C2, with shared regulation of the estrogen signaling pathway. Three genes (DIRC3, SLC6A2, TUBA3D) were independent predictors of BRCA prognosis, forming the basis for a risk model. Except for exhibiting satisfactory diagnostic efficacy, the risk score elevation correlated with poor prognosis, elevated matrix, immune, and ESTIMATE scores, and negative correlation with microsatellite instability. The *in vitro* results confirmed the differential expression levels of DIRC3, SLC6A2, and TUBA3D.
Conclusion: The prognostic risk model associated with epithelial cells demonstrates effective diagnostic performance in BRCA, serving as an independent prognostic factor for BRCA patients. Additionally, it exhibits a correlation with immune scores.

1. Introduction

Breast cancer (BRCA) is the most common malignancy among women and a major threat to human health [1,2]. In 2023, approximately 297,790 new cases of BRCA were diagnosed in the United States among women, with 43,170 reported deaths [3]. Research findings in China focusing on women indicate a significant increase in the incidence of BRCA, accompanied by a rising trend in mortality [4]. Early diagnosis and prognosis screening for BRCA are believed to reduce mortality rates and improve patient

* Corresponding author. Department of Thyroid Surgery, The Second Affiliated Hospital of Zhejiang University College of Medicine, No. 88, Jiefang Road, Shangcheng District, Hangzhou, 310009, Zhejiang, China.
E-mail address: yanhaichao2@zju.edu.cn (H.-c. Yan).

https://doi.org/10.1016/j.heliyon.2024.e37048
Received 27 February 2024; Received in revised form 26 August 2024; Accepted 27 August 2024
Available online 28 August 2024

outcomes [5].

BRCA is a heterogeneous disease, exhibiting heterogeneity among patients and within each tumor [6]. Generally, BRCA mutations can manifest as non-invasive, confined to epithelial cells, or invasive, with over 95 % of BRCA originating from epithelial cells [7,8]. Research indicates that BRCA displays histological heterogeneity, where epithelial cell plasticity can give rise to distinct cellular subpopulations, contributing to intra-tumor heterogeneity [8]. The histological and phenotypic differences among BRCA play a crucial role in diagnosis, treatment, and prognosis [9]. The emergence of BRCA results from selective dysregulation of specific signals within mammary epithelial cells [10]. Additionally, increasing evidence suggests that epithelial-mesenchymal transition (EMT) and EMT-related mechanisms are associated with various diseases, including the initiation and progression of tumors [11]. Therefore, this study investigated the role of human mammary epithelial cells in the progression of BRCA to provide more options for the diagnosis, treatment, and prognosis of BRCA patients.

Traditional RNA sequencing (RNA-seq) involves extracting RNA from large cell populations to ensure an adequate amount for subsequent analysis. However, it only provides the average gene expression levels across different cell populations and fails to capture the transcriptional heterogeneity [12,13]. In contrast, single-cell RNA-seq enables high-throughput and high-resolution transcriptomic analysis of individual cells, allowing for the exploration of heterogeneity in different molecular subtypes of tumors. Therefore, this study utilizes single-cell RNA-seq data to identify genes associated with epithelial cells.

Based on single-cell and bulk RNA-seq data, this study utilizes bioinformatics and machine-learning algorithms to construct an epithelial cell-related prognostic model for BRCA patients. Additionally, the potential diagnostic, prognostic, and immunological implications of this model in BRCA patients were further explored. In conclusion, our findings suggest that the model may hold significant value in the diagnosis, prognosis, and treatment of BRCA patients.

2. Methods

2.1. Date source

The TCGA-BRCA expression profile data and clinical-pathological information were downloaded from the UCSC Xena platform (https://xenabrowers.net/datapage/). Samples with incomplete expression information and missing clinical data were excluded, including 920 cancer samples and 113 para-cancerous tissue samples. Subsequently, gene expression data and relevant clinical information from the GSE42568 dataset (including 17 normal and 104 tumor samples) and GSE10780 (including 143 normal samples and 42 tumor samples), collected on the GPL570 platform, were downloaded from the Gene Expression Omnibus (GEO) database. Single-cell sequencing data for human breast tissue were obtained from the GSE245601 dataset, including three tumor samples not subjected to Tamoxifen treatment.

2.2. Acquisition of epithelial cell-related genes

Initially, the Seurat package was employed to analyze the read count matrices for each gene in every sample. Subsequently, cells meeting the following three criteria were excluded, including $<$500 expressed genes, $>$20 % mitochondrial gene UMIs, and $>$50 % ribosomal gene UMIs. Following this, the gene expression matrix underwent normalization using the “NormalizeData” function, principal component analysis was performed on the expression matrix using the “RunPCA” function, the k-nearest neighbor graph was confirmed through the “FindNeighbors” function, graph-based clustering was executed with the “FindClusters” function, and the visualization of cell clustering results was achieved using the “RunTSNE” function. The “FindAllMarkers” function was employed to identify marker genes for each cluster. Subsequently, cell cluster annotation was carried out using the “SingleR” package. Finally, the “FindAllMarkers” function was utilized to identify differentially expressed genes (DEGs) between epithelial cells and other cell populations.

2.3. Functional enrichment analysis

The “limma” package in R was utilized to identify DEGs between normal and tumor tissues in TCGA-BRCA. Subsequently, the Venn package was employed to obtain the intersecting genes between epithelial cell-related genes and TCGA-BRCA-related genes. Following this, the “GO plot” package was utilized for functional enrichment analysis of the intersecting genes, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The GO analysis encompassed biological processes (BPs), molecular functions (MFs), and cellular components (CCs).

2.4. Consensus clustering analysis

First, we conducted a batch KM survival analysis on the intersecting genes using the "survminer" and "survival" software packages to obtain the core intersecting genes related to BRCA prognosis. The analysis method was to write a piece of code for batch KM survival analysis through R software, which can realize multiple KM survival analyses with one click. The “ConsensusClusterPlus” package was utilized to identify unsupervised subtypes and clusters within the TCGA-BRCA dataset based on the core intersecting genes.

2.5. Prognostic modeling

LASSO regression analysis was applied to the core intersecting genes to prevent overfitting when utilizing the "glmnet" package, followed by undergoing univariate and multivariate COX regression analyses sequentially to identify genes with a P value < 0.05. A risk model for BRCA patients was constructed from the screened genes and the risk score of each sample was calculated as follows: risk score = expRNA1*coefRNA1 + expRNA2*coefRNA2 + … + expRNAi*coefRNAi

2.6. Diagnostic value of the risk model

The risk score in normal and tumor tissues was visualized using the "vioplot" package. Subsequently, Decision Curve Analysis (DCA) and Receiver Operating Characteristic (ROC) analyses were conducted using the "ggDCA" and "pROC" packages, respectively, to assess the diagnostic performance of the risk model for BRCA patients.

2.7. Prognostic value of the risk model

K-M survival analysis was conducted using the "survminer" package to demonstrate the predictive efficacy of the risk model for overall survival in BRCA patients. To further explore the relationship between the model and the prognosis of BRCA patients, subgroup analyses were performed based on age, N stage, M stage, T stage, and pathological stage. Following this, univariate and multivariate COX regression analyses were separately employed to investigate the association between the risk score, clinical features, and the prognosis of BRCA patients. Finally, nomograms and corresponding calibration curves were constructed using the multivariate COX regression analysis results.

2.8. Immune infiltration analysis

Previous studies have indicated a close association between epithelial cells and immune cells [14]. Therefore, an immune infiltration analysis was conducted. The ESTIMATE algorithm was employed to calculate the stromal score, immune score, and estimate score in tumor samples. Subsequently, the relationship between the risk score and assessable immune prognostic indicators such as tumor mutational burden (TMB) and microsatellite instability (MSI) was explored using Pearson coefficient tests.

2.9. RNA extraction and qRT-PCR

In this experiment, we first used the TRIzol reagent (Invitrogen, USA) to isolate total RNA from cells according to the manufacturer's protocol. Secondly, we used a reverse transcription kit (Takara, Japan) to reverse transcribe RNA into cDNA. GAPDH was selected as the internal parameter. The expression of each gene was normalized to that of the internal control and quantified by the $2^{-\Delta\Delta CT}$ method [15]. The primers used were listed as follows: DIRC3 forward, 5′-TCACGGCAGCAGTATTCA-3′ and reverse, 5′-TCATTTCCACTCGCACAA-3'; and SLC6A2 forward, 5′-ACCAAGGGTGGAATTTACG-3′ and reverse, 5′-AAGGCAGGACTGACGAACT-3'; TUBA3D forward, 5′-CTCCATCCTGACCACCCA-3′ and reverse, 5′-GGACACGATCTGCCCAAT-3'; GAPDH forward, 5′-AGAAGGCTGGGGCTCATTTG -3′ and reverse, 5′- AGGGGCCATCCACAGTCTTC -3'.

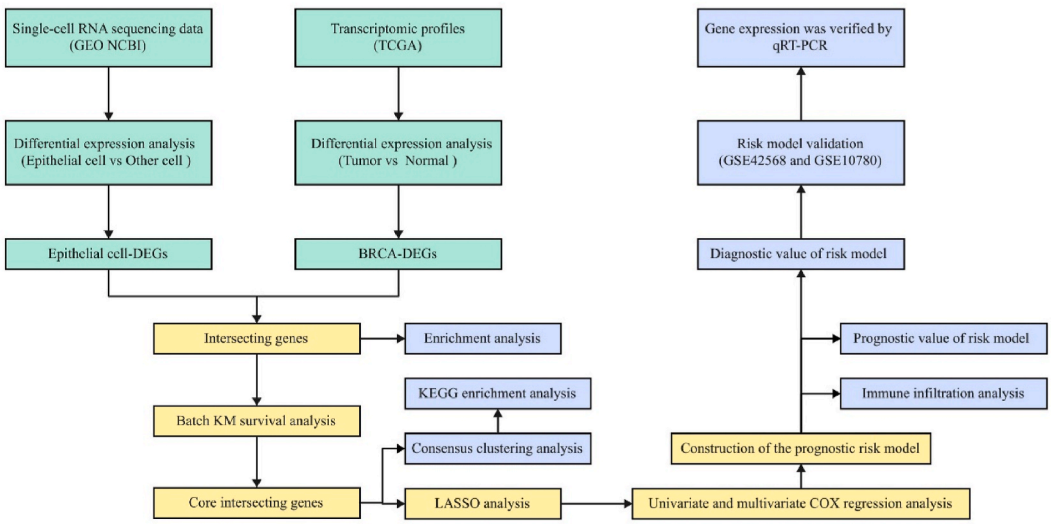

Fig. 1. The entire analytical process of the study.

2.10. Statistical analysis

Biostatistical analysis and mapping were performed using R language (version 3.14.3), SPSS 25.0 software, and GraphPad Prism 10.0. The Kolmogorov-Smirnov goodness-of-fit test was applied to handle continuous data. For data conforming to a normal distribution, the mean ± standard error was presented, and comparisons were conducted using the Student-t test or Analysis of Variance between groups. Non-normally distributed data were represented by quartiles (Q1-Q4), and differences between groups were analyzed through rank-sum tests. A significance level of $P < 0.05$ was considered statistically significant.

3. Results

3.1. Single-cell RNA-seq analysis

Fig. 1 displays the process of this research. To identify DEGs between epithelial cells and other cell types, 19 clusters were determined through unsupervised clustering of all cells (Fig. 2A). Subsequently, these clusters were annotated as B cells, endothelial cells, epithelial cells, mesenchymal cells, myeloid cells, and T cells based on the expression levels of typical marker genes (Fig. 2B). The proportions and distribution of different cells across three samples are shown in Fig. 2C and D. The results of the "FindAllMarkers" function indicated 1215 DEGs between epithelial cells and other cell populations.

3.2. Acquisition of intersecting genes

The differential expression analysis results revealed 4227 DEGs between BRCA and normal tissues, comprising 1413 upregulated genes and 2814 downregulated genes (Fig. 3A). The top 20 DEGs were visualized in the heatmap (Fig. 3B). Venn diagram identified 374 intersecting genes between epithelial cell-related genes and BRCA-associated genes (Fig. 3C).

3.3. Enrichment analysis of intersecting genes

GO and KEGG enrichment analyses were conducted to elucidate the functions and signaling pathways associated with the 374 intersecting genes. The GO functional annotation results revealed that the BPs, MCs, and CCs are relevant to the progression and differentiation of epithelial cells. The identified functions include extracellular region, extracellular region part, plasma membrane part, epithelium development, epithelial cell differentiation, epidermis development, cell-cell junction tight junction, bicellular tight junction apical, and junction complex (Fig. 4A). KEGG enrichment analysis demonstrated that the signaling pathways predominantly involved in the intersecting genes were also related to epithelial cells. These pathways encompass tight junction, PI3K-Akt signaling pathway, human papillomavirus infection, MAPK signaling pathway, estrogen signaling pathway, viral carcinogenesis, axon guidance, systemic lupus erythematosus, staphylococcus aureus infection, and caffeine metabolism (Fig. 4B).

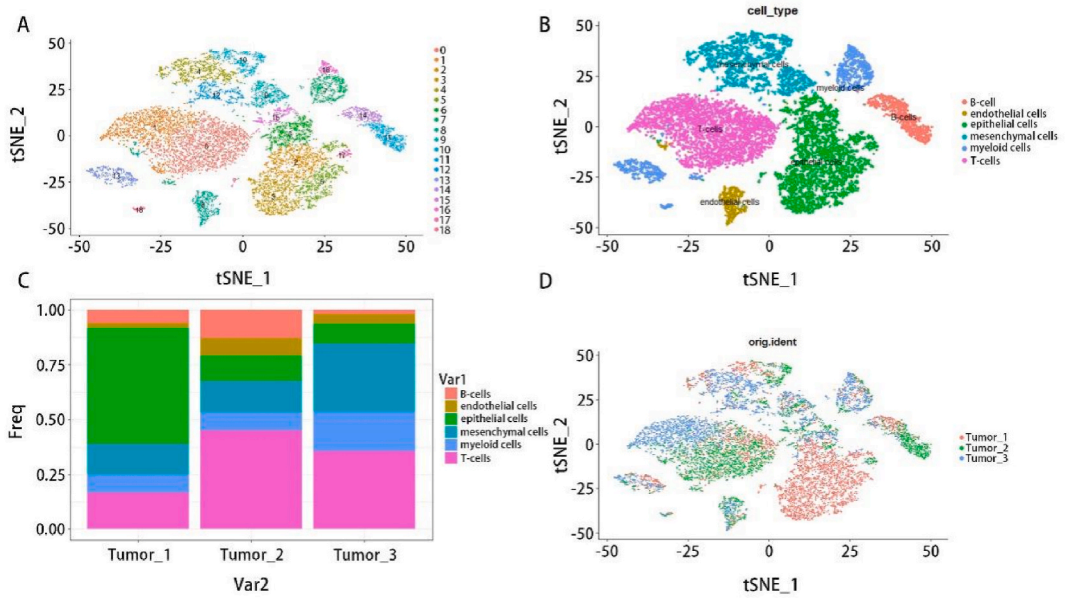

Fig. 2. Single-cell RNA sequencing analysis. (A) t-Distributed Stochastic Neighbor Embedding (tSNE) clustering plot of all cells. (B) Annotated tSNE plot post-annotation. (C) Proportional representation of major cell lines in each tumor sample. (D) Distribution plot of cells in each tumor sample.

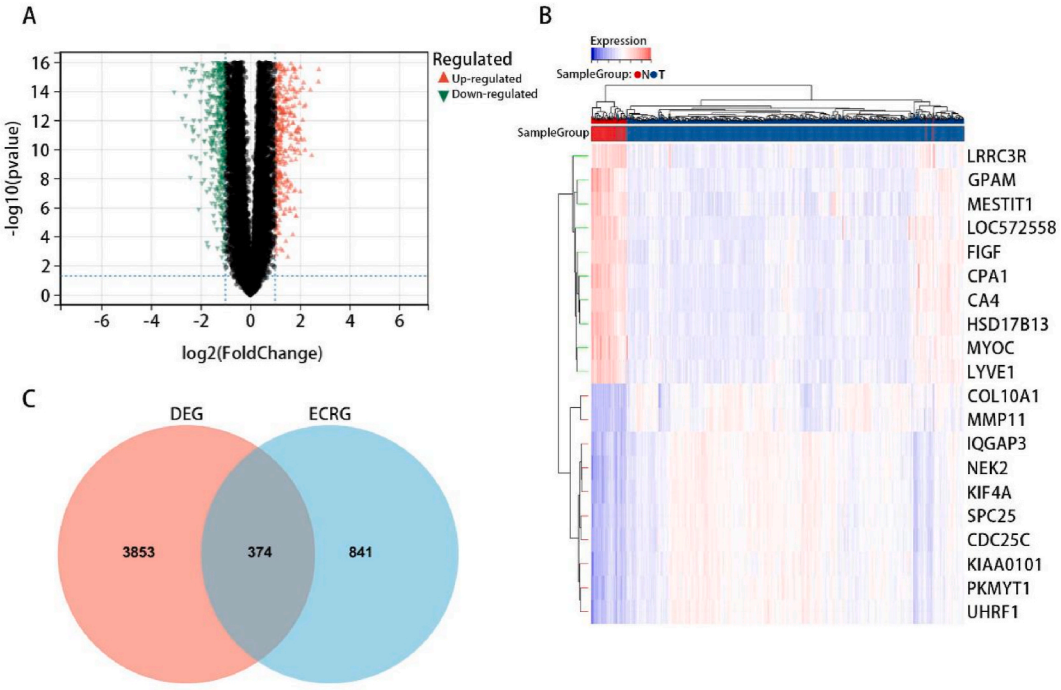

Fig. 3. Acquisition of intersecting genes between epithelial cell-related genes and BRCA-associated genes. (A) Volcano plots reveal the differentially expressed genes (DEGs) between BRCA tissue and normal tissue. (B) The top 20 DEGs were visualized in a heatmap. (C) Venn diagram identifies the intersecting genes between epithelial cell-related genes and BRCA-associated genes.

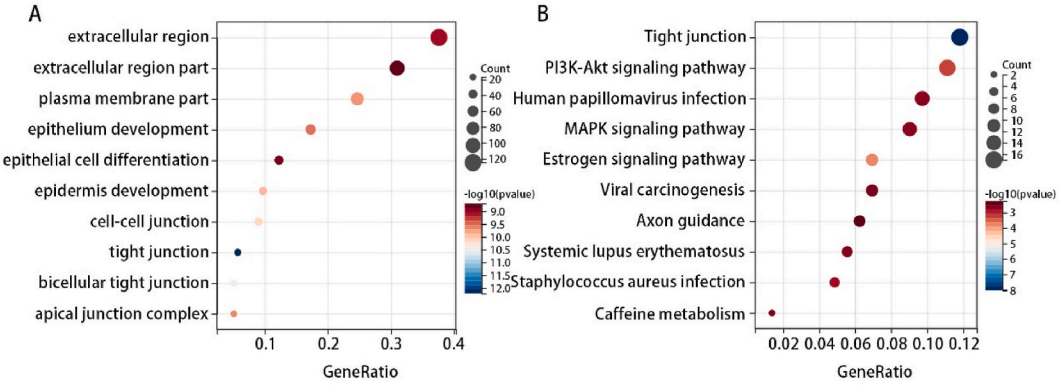

Fig. 4. Enrichment analysis of intersecting genes. (A) Bubble chart presenting the results of Gene Ontology (GO) functional annotation, including cellular components (CCs), molecular functions (MFs), and biological processes (BPs). (B) Bubble chart illustrating the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis results.

3.4. Identification of subgroups associated with core intersecting genes

First, batch K-M survival analysis on the intersecting genes yielded 51 core intersecting genes associated with the prognosis of BRCA patients ($P < 0.05$). Subsequently, based on the expression of the 51 core intersecting genes, unsupervised clustering analysis using the consensus clustering algorithm was conducted to explore potential clusters within the TCGA-BRCA dataset. The results indicated that k = 2 effectively divided the BRCA dataset into subgroups C1 and C2 (Fig. 5A and B). Furthermore, the results revealed 18 significantly downregulated and 19 upregulated genes in the C1 subgroup compared to the C2 subgroup, with no expression differences in the remaining 14 genes (Fig. 6A and B). KEGG enrichment analyses were performed for the downregulated and upregulated genes, showing that the estrogen signaling pathway was a common signaling pathway. Combined with the KEGG results of the intersecting genes, it is evident that this group of genes primarily participates in the progression of BRCA by modulating the estrogen signaling pathway.

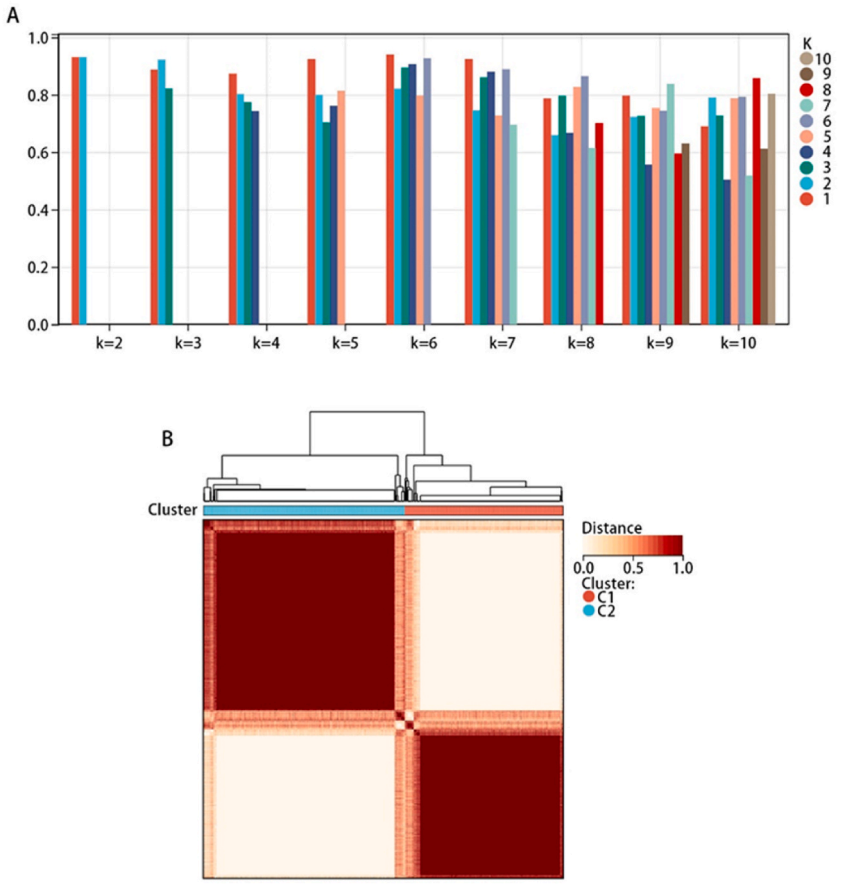

Fig. 5. Identification of subgroups associated with core intersecting genes. (A) The highest average consistency within the groups is achieved when k = 2. (B) Heatmap of the consensus matrix defining the two subgroups.

3.5. Prognostic model construction

To develop an epithelial cell-related prognostic model in BRCA, LASSO analysis was applied to 51 core intersecting genes, identifying 14 genes with non-zero regression coefficients (Fig. 7A and B). Subsequently, the multivariate COX regression analysis results indicated that three genes independently predicted the prognosis of BRCA patients. A risk model was then constructed with the risk score calculated using the following equation: risk score = 0.19 * DIRC3 + 0.235 * SLC6A2 - 0.107 * TUBA3D. As illustrated in Fig. 7C, low expression of DIRC3 and SLC16A2 and high expression of TUBA3D were observed in BRCA tissues ($P < 0.05$).

3.6. Prognostic model evaluation

Fig. 8A–C showed that the risk score was down-regulated in the BRCA group using the TCGA-BRCA, GSE42568, and GSE10780 datasets. ROC analysis showed that the model had a good diagnostic effect on BRCA patients, and the AUCs of the model in the TCGA-BRCA, GSE42568, and GSE10780 datasets were 0.84, 0.88, and 0.66, respectively (Fig. 8D–F). DCA results displayed that the risk model had good performance in predicting BRCA (Fig. 8G–I).

Moreover, the risk score was unrelated to age, M stage, and T stage. At the same time, it exhibited a significant association with histological type, N stage, and pathological stage ($P < 0.05$) (Fig. 9A–F).

3.7. Prognostic value of the risk model

Subsequently, we conducted a K-M survival analysis, showing that a high-risk score was unfavorable for overall survival (Fig. 10A). Subgroup analysis revealed that high-risk score in the M0 (HR = 2.17(1.51–3.11), $P < 0.001$), M1 (HR = 14.91(1.78–125.23), $P = 0.002$), N1+N2+N3 (HR = 3.19(2.05–4.97), $P < 0.001$), T1+T2 (HR = 2.21(1.48–3.3), $P < 0.001$), T3+T4 (HR = 3.07(1.55–6.11), $P < 0.001$), S1+S2 (HR = 1.58(1.01–2.47), $P < 0.04$), S3+S4 (HR = 4.35(2.38–7.93), $P < 0.001$), ≤55 (HR = 2.66(1.56–4.53), $P < 0.001$), and >55 (HR = 2.27(1.43–3.61), $P < 0.001$) subgroups was associated with adverse outcomes (Fig. 10B).

Furthermore, the univariate COX regression analysis results indicated that all the enrolled characteristics were significantly

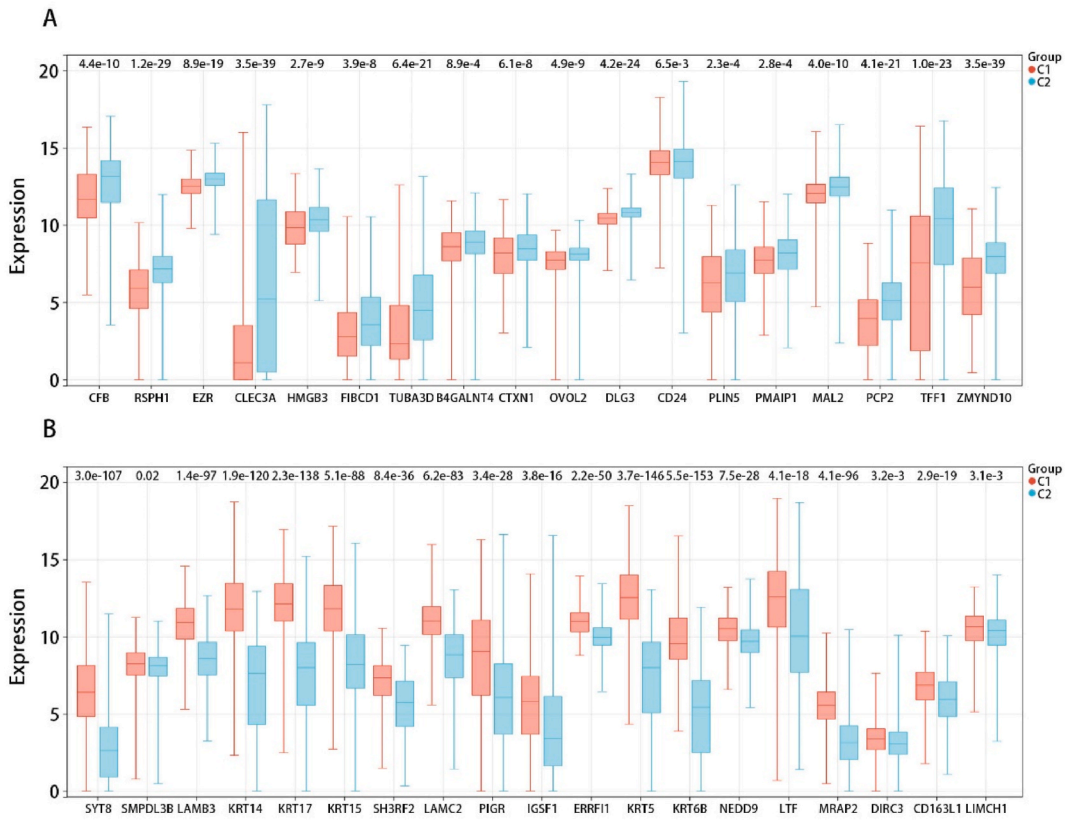

Fig. 6. Differential expression of core intersecting genes in two subgroups. (A) Genes that are downregulated among the core intersecting genes in the C1 subgroup. (B) Genes that are upregulated among the core intersecting genes in the C1 subgroup.

associated with BRCA prognosis ($P < 0.05$). The multivariate COX regression analysis revealed that the M stage, N stage, pathological stage, age, and risk score were independent prognostic factors for BRCA ($P < 0.05$) (Table 1).

Moreover, the risk score played a crucial role in predicting the overall survival rate of BRCA patients, with the highest contribution (Fig. 10C). Calibration curves demonstrated good alignment between predicted and observed values (Fig. 10D).

3.8. Immune infiltration analysis

An ESTIMATE analysis was performed to further explore the role of the risk model in immunotherapy for BRCA. The stromal, immune, and ESTIMATE scores exhibited significantly higher expression in the high-score group ($P < 0.05$, Fig. 11A). Additionally, the risk score showed a negative correlation with the MSI score ($P < 0.05$, Fig. 11B) and a slight correlation with the TMB score (Fig. 11C). This suggests that risk score was closely associated with immunotherapy in BRCA patients.

3.9. Validation of gene expression in risk model by qRT-PCR

To further verify the expression of genes in the risk model in BRCA cells and normal cells, we used qRT-PCR technology for validation. DIRC3 and SLC16A2 were lower expressed in BRCA cells; while TUBA3D was highly expressed in BRCA cells (all $P < 0.05$, Fig. 12A–C). The experimental results were consistent with the data analysis in this study.

4. Discussion

BRCA is a highly prevalent malignancy among women with a poor prognosis. Extensive research in recent years has indicated a close correlation between the regulation of epithelial cell plasticity and the progression of BRCA [8]. Despite continuous improvements and advancements in BRCA treatment methods and drugs, there has been limited improvement in the prognosis of BRCA patients. Therefore, this study establishes a prognostic risk model related to epithelial cells, exploring its role in the diagnosis, prognosis, and immunotherapy of BRCA. The findings aim to provide additional options for the treatment of BRCA patients.

Herein, we obtained 374 genes at the intersection of epithelial cell-related genes and DEGs in BRCA tissues and normal tissues, of which 51 core intersecting genes were related to BRCA prognosis. Enrichment analysis revealed the estrogen signaling pathway as the

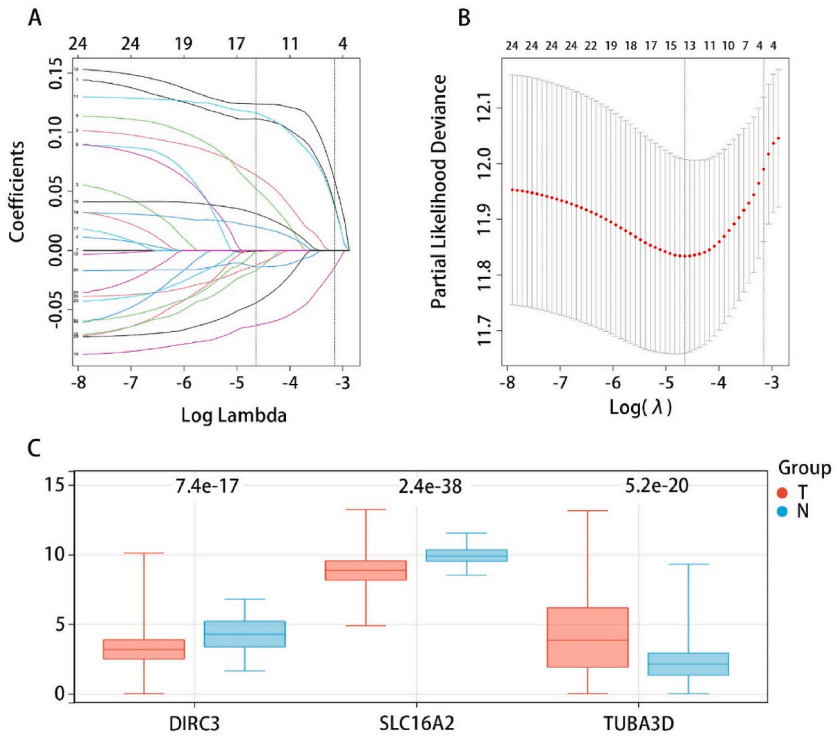

Fig. 7. Selection of prognosis-related genes to construct the risk model. (A–B) LASSO coefficient path plot for 24 risk factors, where 24 different colored lines represent the trajectories of the 24 independent variable coefficients. The ordinate refers to the value of the independent variable coefficient in the model; the lower horizontal axis is the logarithm of the penalty coefficient; the upper horizontal coordinate is the number of non-zero coefficients in the model. (B) LASSO regression analysis cross-validation curve. The vertical axis refers to the simulated deviation, the smaller the value, the better the fit; the lower horizontal axis refers to the logarithm of the penalty coefficient; the upper horizontal coordinate refers to the number of zero coefficients of the model. (C) Gene expression patterns associated with BRCA patient prognosis in both BRCA and normal tissues. T, tumor tissue; N, normal tissue.

potential regulatory pathway. Research indicated that 70 % of BRCA patients overexpress estrogen receptor alpha (ERα), and initially, most respond effectively to anti-estrogen therapy, inducing invasiveness in ERα-positive BRCA through promoting epithelial-mesenchymal transition [16]. Estrogen also induces proliferation in ERα-positive BRCA cells by directly activating cell cycle genes [17]. In conclusion, we infer that the epithelial cell-related genes obtained in this study may regulate the occurrence and progression of BRCA by modulating the estrogen signaling pathway in epithelial cells.

Previous studies have identified several biomarkers of BRCA and constructed related gene signatures for prognosis prediction. Li et al. detected prognostic biomarkers including ADRB1, SAV1, and TSPAN14 of BRCA by the regularized Cox proportional hazards model [18]. Lactylation-related gene signature (RAD51, CASP14, NEK10, PCP2, IDO1, CLSTN2, and IGHG1) was constructed to predict the prognosis of BRCA patients [19]. In addition, Sha et al. obtained four cuproptosis-related genes for prognostic signature construction and identified DLAT as an independent prognostic factor in BRCA [20]. Through various efficient bioinformatics tools, we established the risk model using three epithelial-related genes that independently predicted BRCA prognosis, including DIRC3, SLC16A2, and TUBA3D. Changes in DIRC3 expression levels may alter thyroid hormone (TH) production and lead to reduced epithelial differentiation, thereby indirectly promoting the development of thyroid cancer [21]. TH production is a complex and rigorous negative feedback regulation process, involving the hypothalamus, pituitary gland, and thyroid gland. First, the hypothalamus produces thyrotropin-releasing hormone (TRH) which acts on the TRH receptor of the pituitary gland and ultimately releases thyroid-stimulating hormone (TSH). In the thyroid, TSH binds to its receptor and produces TH. Triiodothyronine (T3) and tetraiodothyronine (T4) are also released into the pathway when the body needs them. T3 and T4 are formed by the iodination of tyrosine residues to monoiodotyrosine and diiodotyrosine. Interestingly, SLC16A2 encodes the plasma membrane transporter T3/T4, which enters the cell and binds to the TH receptor in the nucleus and mitochondria [22]. In cells, T3 can bind to TH receptor α and TH receptor β, and THRβ shows a high affinity for the DNA sequence of the TH reaction originals. The interaction of T4 with integrin αvβ3 promotes the production of new blood vessels during cancer development and wound healing [23]. At the mechanism level, T4 and integrin αvβ3 also promote cancer progression and block cell apoptosis by activating signaling pathways such as MAPK/ERK1 and PI3K [24]. In addition, T4 promotes BRCA through a discrete molecular mechanism derived from TH receptors on integrin αvβ3, dependent or independent of ER [25]. TUBA3D gene is a member of the α-tubulin protein family along with β-tubulin to constitute the primary structural components of microtubules [26]. They play a crucial role in cell movement, transport, mitosis, and cell structure [27]. Besides, TUBA3D is highly expressed in BRCA tissues and is associated with tumor aggregation, and it is down-regulated in

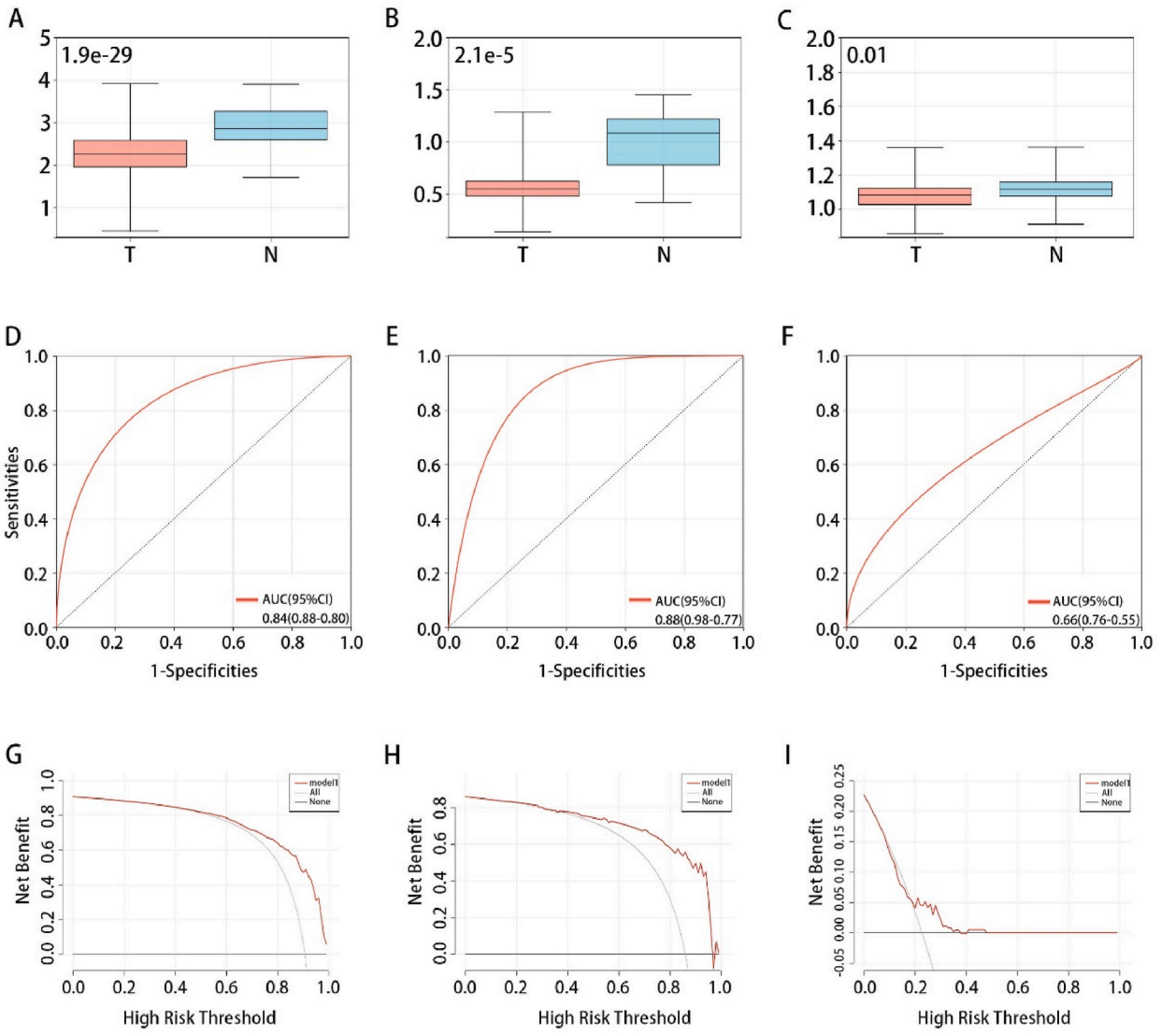

Fig. 8. Diagnostic performance of the risk model. (A–C) Expression of the risk score in BRCA and normal tissues in the TCGA-BRCA(A), GSE42568(B), and GSE10780(C) datasets. (D–F) Receiver operating characteristic curves for risk score differentiating BRCA from normal samples in the TCGA-BRCA (D), GSE42568 (E), and GSE10780 (F) datasets. (G–I) Decision curve analysis of the risk score in the TCGA-BRCA (G), GSE42568 (H), and GSE10780 (I) datasets. BRCA, breast cancer; T, tumor tissue; N, normal tissue; AUC, the area under the curve; 95 % CI, 95 % confidence interval.

paclitaxel-resistant cells [28]. In summary, DIRC3 and SLC16A2 may play a role in BRCA progression, either directly or indirectly, by regulating TH and thyroid receptors, while TUBA3D might be a therapeutic target for BRCA treatment.

Survival analysis demonstrated a higher risk score indicating an unfavorable prognosis. Immune analysis revealed elevated expression of stromal score, immune score, and ESTIMATE score in tissues with higher risk score. Additionally, the risk score showed a negative correlation with MSI. Studies suggest that the infiltration of immune cells and stromal cells plays a crucial role in the prognosis and treatment of cancer. MSI is a genomic biomarker for identifying patients who may benefit from immune checkpoint inhibitors [29,30]. Therefore, there is a close association between the scoring model and immune cells, providing new options for immunotherapy in BRCA.

For strengths, this study integrates efficient single-cell and bulk RNA sequencing approaches to construct and evaluate an epithelial-cell related prognostic model for BRCA. Secondly, the independent dataset and an *in vitro* experiment increased the generalization ability of the results. Moreover, we thoroughly explored the biological mechanisms underlying the identified genes' roles in BRCA development. However, the results should be validated in a large-scale cohort to test the robustness. More *in vitro* and *in vivo* experiments are required to explore its underlying mechanism in the future.

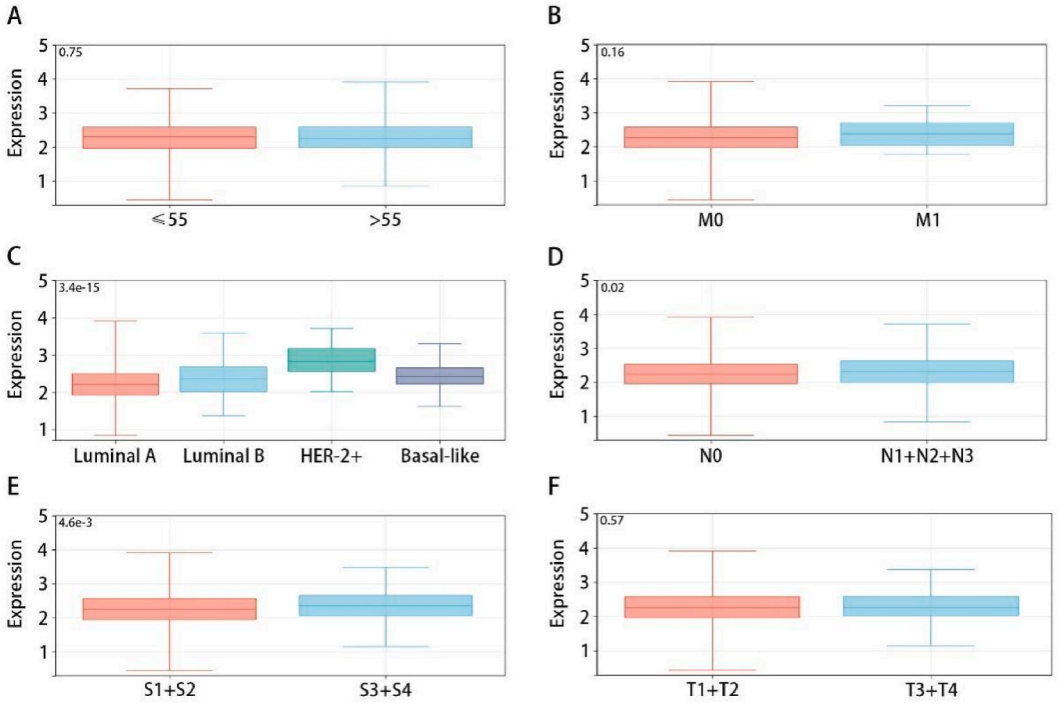

Fig. 9. Relationship between the risk score and clinical features. (A–F) Relationship between risk score and gender (A), M stage (B), histological type (C), N stage (D), pathological stage (E), and T stage (F).

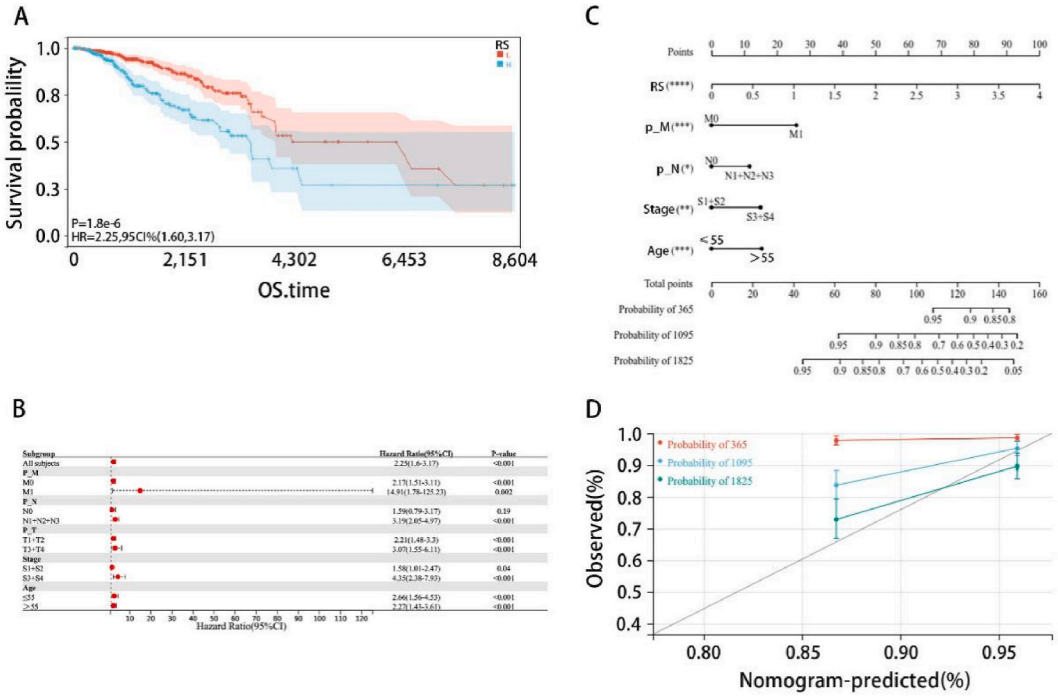

Fig. 10. Relationship between the risk score and prognosis in BRCA patients. (A) Kaplan-Meier curve analysis demonstrates an association between the high-risk score and adverse prognosis in BRCA patients. (B) The connection between risk score and prognosis in various clinical feature subgroups of BRCA patients. (C–D) Nomogram and corresponding calibration curve. OS, Overall Survival; RS, Risk Score; p_M, Pathologic M; p_N, Pathologic N; p_T, Pathologic T; Stage, Pathologic Stage.

Table 1
Relationship between risk score, clinical features, and prognosis in BRCA patients.

Characteristics	Univariate analysis		Multivariate analysis	
	HR (95%CI)	P-value	HR (95%CI)	P-value
P_M	0.231(0.134–0.400)	<0.001	0.329(0.182–0.593)	<0.001
P_N	0.453(0.314–0.653)	<0.001	0.621(0.399–0.964)	0.034
P_T	0.555(0.376–0.947)	0.003	0.969(0.593–1.584)	0.901
Stage	0.384(0.272–0.544)	<0.001	0.562(0.335–0.943)	0.029
Risk score	1.877(1.327–2.656)	<0.001	3.499(1.845–6.636)	<0.001
Age	1.584(1.119–2.242)	0.009	1.925(1.343–2.757)	<0.001

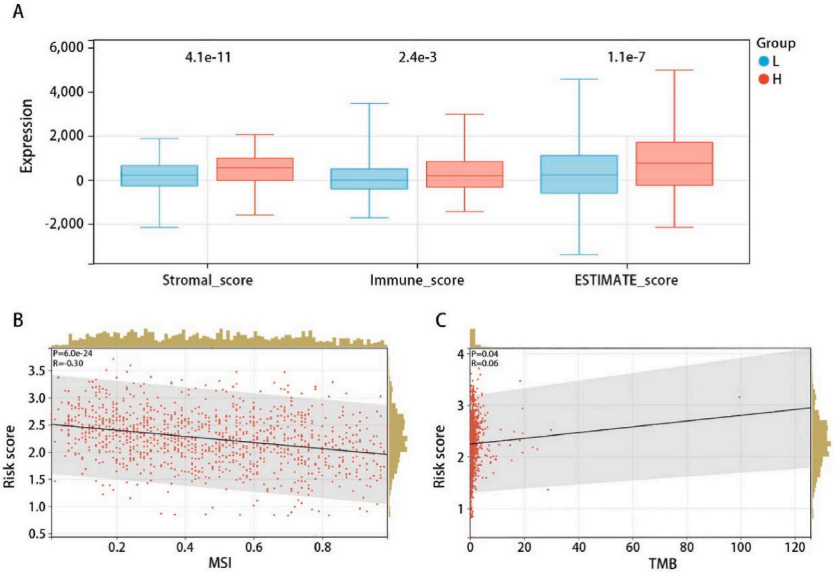

Fig. 11. Relationship between the risk score and immune cells. (A) Differential distributions of stromal score, immune score, and ESTIMATE score in two risk score groups. The association of the risk score with (B) MSI score and (C) TMB score. MSI, microsatellite instability; TMB, tumor mutation burden.

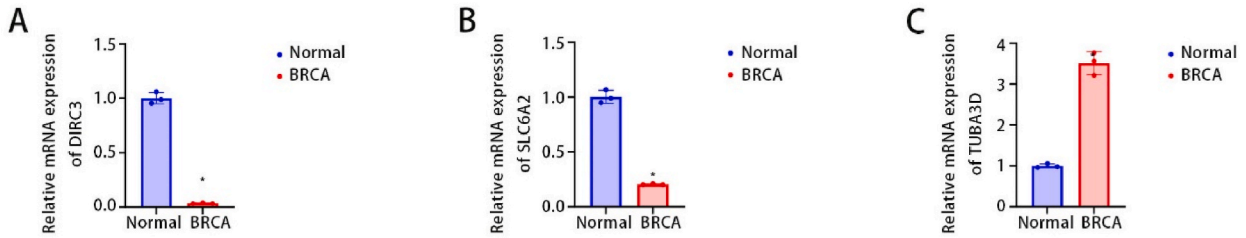

Fig. 12. qRT-PCR validation. (A–C) The qRT-PCR results of DIRC3, SLC16A2, and TUBA3D in normal cells and BRCA cells, in sequence. BRCA, breast cancer.

5. Conclusion

In summary, the epithelial cell-related prognostic model constructed in this study serves as a novel indicator for the diagnosis and prognosis of BRCA patients. DIRC3, SLC16A2, and TUBA3D within this model were differentially expressed in BRCA and normal groups validated *in vitro*. The three genes are closely associated with the role of epithelial cells in tumors. Furthermore, the model exhibits significant correlations with stromal score, immune score, ESTIMATE score, and MSI, providing more comprehensive information for BRCA.

Data availability

Data will be made available on request.

Ethics approval and consent to participate

The Ethics Committee of Shaoxing maternity and child health care hospital that this research is based on open-source data, so the need for ethics approval was waived.

Consent for publication

Not applicable.

Competing interests

The authors report no conflict of interest.

Funding

Not applicable.

CRediT authorship contribution statement

Man-zhi Xia: Writing – original draft, Methodology, Formal analysis, Data curation. **Hai-chao Yan:** Writing – review & editing, Writing – original draft, Supervision, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Not applicable.

References

[1] M. Zhang, et al., Bioinformatics analysis of prognostic significance of COL10A1 in breast cancer, Biosci. Rep. 40 (2) (2020).
[2] T. Yang, et al., Genetic testing enhances the precision diagnosis and treatment of breast cancer, Int. J. Mol. Sci. 24 (23) (2023).
[3] R.L. Siegel, et al., Cancer statistics, 2023, CA Cancer J Clin 73 (1) (2023) 17–48.
[4] O.M. Siggs, et al., Association of high polygenic risk with visual field worsening despite treatment in early primary open-angle glaucoma, JAMA Ophthalmol 141 (1) (2022) 73–77.
[5] J. Jiang, et al., Breast cancer screening should embrace precision medicine: evidence by reviewing economic evaluations in China, Adv. Ther. 40 (4) (2023) 1393–1417.
[6] S. Li, et al., Oncogenic transformation of normal breast epithelial cells co-cultured with cancer cells, Cell Cycle 17 (16) (2018) 2027–2040.
[7] P. Sarvari, et al., Advances of epigenetic biomarkers and epigenome editing for early diagnosis in breast cancer, Int. J. Mol. Sci. 23 (17) (2022).
[8] E. Tomaskovic-Crook, E.W. Thompson, J.P. Thiery, Epithelial to mesenchymal transition and breast cancer, Breast Cancer Res. 11 (6) (2009) 213.
[9] N. Harbeck, et al., Breast cancer, Nat Rev Dis Primers 5 (1) (2019) 66.
[10] C. de Castro Ferezin, et al., Identification of biological pathways and processes regulated by NEK5 in breast epithelial cells via an integrated proteomic approach, Cell Commun. Signal. 20 (1) (2022) 197.
[11] J.M. Reiman, K.L. Knutson, D.C. Radisky, Immune promotion of epithelial-mesenchymal transition and generation of breast cancer stem cells, Cancer Res. 70 (8) (2010) 3005–3008.
[12] S. Ding, X. Chen, K. Shen, Single-cell RNA sequencing in breast cancer: understanding tumor heterogeneity and paving roads to individualized therapy, Cancer Commun. 40 (8) (2020) 329–344.
[13] Y. Wang, et al., Changing technologies of RNA sequencing and their applications in clinical oncology, Front. Oncol. 10 (2020) 447.
[14] T. Risom, et al., Transition to invasive breast cancer is associated with progressive changes in the structure and composition of tumor stroma, Cell 185 (2) (2022) 299–310 e18.
[15] L. Dou, X. Zhang, Upregulation of CCT3 promotes cervical cancer progression through FN1, Mol. Med. Rep. 24 (6) (2021).
[16] J.Y. Ho, et al., Estrogen enhances the cell viability and motility of breast cancer cells through the ERalpha-DeltaNp63-integrin beta4 signaling pathway, PLoS One 11 (2) (2016) e0148301.
[17] B. Hanstein, et al., Insights into the molecular biology of the estrogen receptor define novel therapeutic targets for breast cancer, Eur. J. Endocrinol. 150 (3) (2004) 243–255.
[18] L. Li, Z.P. Liu, Detecting prognostic biomarkers of breast cancer by regularized Cox proportional hazards models, J. Transl. Med. 19 (1) (2021) 514.
[19] Y. Jiao, et al., Lactylation-related gene signature for prognostic prediction and immune infiltration analysis in breast cancer, Heliyon 10 (3) (2024) e24777.
[20] R. Sha, et al., Cuproptosis-related genes predict prognosis and trastuzumab therapeutic response in HER2-positive breast cancer, Sci. Rep. 14 (1) (2024) 2908.
[21] Z. Shen, et al., DIRC3 and near NABP1 genetic polymorphisms are associated laryngeal squamous cell carcinoma patient survival, Oncotarget 7 (48) (2016) 79596–79604.
[22] M.A. Prusinkiewicz, et al., Survival-associated metabolic genes in human papillomavirus-positive head and neck cancers, Cancers 12 (1) (2020).

[23] A. Al Husseini, et al., Thyroid hormone is highly permissive in angioproliferative pulmonary hypertension in rats, Eur. Respir. J. 41 (1) (2013) 104–114.
[24] C. Weingarten, et al., The interplay between epithelial-mesenchymal transition (EMT) and the thyroid hormones-alphavbeta3 Axis in ovarian cancer, Horm Cancer 9 (1) (2018) 22–32.
[25] A. Hercbergs, et al., Thyroid hormone in the clinic and breast cancer, Horm Cancer 9 (3) (2018) 139–143.
[26] Z. Marchocki, et al., Impact of neoadjuvant chemotherapy on somatic mutation status in high-grade serous ovarian carcinoma, J. Ovarian Res. 15 (1) (2022) 50.
[27] H.K. Berrieman, M.J. Lind, L. Cawkwell, Do beta-tubulin mutations have a role in resistance to chemotherapy? Lancet Oncol. 5 (3) (2004) 158–164.
[28] A.G. Murillo Carrasco, et al., Insights from a computational-based approach for analyzing autophagy genes across human cancers, Genes 14 (8) (2023).
[29] Y. Tong, et al., Comprehensive analyses of stromal-immune score-related competing endogenous RNA networks in colon adenocarcinoma, Dis. Markers 2022 (2022) 4235305.
[30] M. Palmeri, et al., Real-world application of tumor mutational burden-high (TMB-high) and microsatellite instability (MSI) confirms their utility as immunotherapy biomarkers, ESMO Open 7 (1) (2022) 100336.