

Autoimmunity



ISSN: (Print) (Online) Journal homepage: informahealthcare.com/journals/iaut20

Identification of prognostic biomarkers of breast cancer based on the immune-related gene module

Ruijuan Wang, Huanhong Zeng, Xueming Xiao, Junjie Zheng, Naizhuo Ke, Wenjun Xie, Qiang Lin & Hui Zhang

To cite this article: Ruijuan Wang, Huanhong Zeng, Xueming Xiao, Junjie Zheng, Naizhuo Ke, Wenjun Xie, Qiang Lin & Hui Zhang (2023) Identification of prognostic biomarkers of breast cancer based on the immune-related gene module, Autoimmunity, 56:1, 2244695, DOI: <u>10.1080/08916934.2023.2244695</u>

To link to this article: <u>https://doi.org/10.1080/08916934.2023.2244695</u>

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



0

Published online: 16 Aug 2023.

Submit your article to this journal 🕑





View related articles 🗹



Uiew Crossmark data 🗹

RESEARCH ARTICLE

Taylor & Francis Taylor & Francis Group

OPEN ACCESS OPEN ACCESS

Identification of prognostic biomarkers of breast cancer based on the immune-related gene module

Ruijuan Wang^a*, Huanhong Zeng^a*, Xueming Xiao^a, Junjie Zheng^a, Naizhuo Ke^a, Wenjun Xie^a, Qiang Lin^a and Hui Zhang^b

^aDepartment of Basic Surgery, Fujian Provincial Hospital, Shengli Clinical College of Fujian Medical University, Fuzhou, Fujian, China; ^bDepartment of Surgical Oncology, Fujian Provincial Hospital, Shengli Clinical College of Fujian Medical University, Fuzhou, Fujian, China

ABSTRACT

Breast cancer (BC) is highly malignant and its mortality rate remains high. The development of immunotherapy has gradually improved the prognosis and survival rate of patients. Therefore, identifying molecular markers concerned with BC immunity is of great importance for the treatment of this disease. The Cancer Genome Atlas-breast invasive carcinoma (TCGA-BRCA) was utilized as the training set while the BC expression dataset from the gene expression omnibus database was taken as the validation set here. Weighted gene co-expression network analysis combined with Pearson analysis and Tumor immune estimation resource (TIMER) was used to obtain immune cell-related hub gene module. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed on this module. Then, receiver operating characteristic curves combining Kaplan-Meier was used to evaluate the effectiveness of the model. Feature genes were screened and the independence of risk score was evaluated by univariate and multivariate Cox analyses. Differences in immune characteristics were analyzed via single-sample gene set enrichment analysis and CIBERSORT, and differences in gene mutation frequency were assessed via GenVisR analysis. Finally, the expression levels of prognostic feature genes in BC cells were validated by quantitative reverse transcription polymerase chain reaction (gRT-PCR). In this study, cell immune-related gene modules in TCGA-BRCA were successfully excavated, and a five-gene (TNFRSF14, NFKBIA, DLG3, IRF2, and CYP27A1) prognostic model was established. The prognostic model could effectively forecast the prognosis and survival rate of BC patients. The result showed that human leukocyte antigen-related proteins and macrophage M2 scores were remarkably highly expressed in the high-risk group, whereas CD8+ T cells, natural killer cells, M1, and other anti-tumor cells were lowly expressed. The model could be used as an independent prognostic factor to predict the prognosis of BC patients. The results of qRT-PCR validation were consistent with the results in the database, that is, except DLG3, the other four feature genes were lowly expressed in BC. The five-gene model established in this study can predict the prognostic and immune mode of BC patients effectively, which is anticipated to become a feasible molecular target for BC therapy.

ARTICLE HISTORY

Received 20 March 2023 Revised 3 July 2023 Accepted 31 July 2023

KEYWORDS

Breast cancer; weighted gene co-expression network analysis; immune microenvironment; HLA; gene mutation

1. Introduction

World widely, breast cancer (BC) is one of the most frequently diagnosed malignancies, which also ranks as the number one cause of cancer-related deaths in women. Based on the American Cancer Society, the incidence of BC in women increased by about 0.05% annually from 2014 to 2018, and it is estimated that 31% of BC new cases will be diagnosed in women in 2022 [1]. BC often shows intratumoral and intertumoral heterogeneity, which seriously interferes with clinicians' judgment of diagnosis, treatment, and prognosis [2]. Most BC patients have advanced disease or metastatic lesions when diagnosed, and for patients who have tumor within 2 cm without metastasis, their 5-year survival rates are more than 80%, whereas the 5-year survival rate of metastatic BC remains 26% [3,4]. Difficult early diagnosis and rapid tumor progression are not conducive to patient prognosis management. Consequently, it is very important to provide potent and accurate prognostic biomarkers for prognosis prediction and treatment response evaluation of BC patients. These prognostic tools will further guide accurate and personalized therapy.

CONTACT Hui Zhang Alara huizhang5466@163.com Department of Surgical Oncology, Fujian Provincial Hospital, Shengli Clinical College of Fujian Medical University, 134 East Street, Gulou District, Fuzhou, Fujian 350001, China *These authors contributed equally to this work.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Based on the anti-tumor immune response, cancer immunotherapy excludes tumor cells via stimulating the host immune system [5]. Anti-tumor responses are induced by T-cell-related immune responses via increasing immune checkpoint inhibitors, but only a few cancer patients benefit from them [6]. More and more evidence has confirmed that immune cell infiltration in the tumor immune microenvironment (TIME) is an essential factor in predicting BC patients' prognoses. Recent research has reported that T-cell infiltration level is dramatically concerned with the prognosis or treatment of BC patients. For example, studies have proved that highly infiltrated CD8⁺ T cells in tumor tissues are in connection with modified outcomes in triple-negative BC patients, providing a premise for the use of CD8⁺ T cells in autologous adoptive cell therapy [7]. However, the specific mechanism of TIME in BC is still ill-defined. Consequently, it is of huge significance to dive into mechanisms of tumor immune regulation and to explore new effective prognostic biomarkers for BC.

Here, the gene co-expression network of The Cancer Genome Atlas-breast invasive carcinoma (TCGA-BRCA) set was constructed via weighted gene co-expression network analysis (WGCNA). After applying Tumor Immune Estimation Resource (TIMER) to assess the relevance between feature modules and infiltration levels of six immune cells, the target module, as the module that was significantly related to the infiltration of six immune cells was retained. Feature genes were screened by univariate and multivariate Cox analyses. Then, a five-gene prognostic model was constructed and its effectiveness was verified. Single-sample gene set enrichment analysis (ssGSEA) and CIBERSORT were used to analyze the differences in immune characteristics between groups. GenVisR was used to evaluate the differences in gene mutation frequency between groups, which verified that the model could predict the differences in immune patterns and gene mutation frequency of samples. Finally, forest plots were constructed to verify the clinical significance of the model. Viewed in total, the five-gene prognostic model manufactured in this study can predict the prognosis and immune pattern of BC patients and is expected to be a potential prognostic biomarker for BC.

2. Materials and methods

2.1. Data download

The training set sample expression together with clinical data were derived from TCGA-BRCA dataset (https://portal. gdc.cancer.gov/), involving 1109 cases of tumor samples and 113 cases of para-carcinoma samples. The validation set sample data came from GSE42568 and GSE20685 datasets in the gene expression omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20685). The content data of six kinds of immune cells (B cells, CD4⁺ T cells, CD8⁺ T cells, neutrophils, macrophages, myeloid dendritic cells) in each sample were extracted from TIMER (https://cistrome.shinyapps.io/timer/).

2.2. Target module gene obtained by WGCNA and functional annotation

In the data preprocessing stage, samples with missing clinical information in TCGA-BRCA were deleted, genes with zero expression level in 80% of samples were excluded, and the top 5000 genes according to median absolute deviation (MAD) were retained as the test target genes. Pearson correlation analysis was conducted on the target genes to construct the similarity matrix, and the optimal soft threshold was selected to power the matrix and construct a weighted adjacency matrix. The selection criterion of the optimal soft threshold was the value when the scale-free fitting index reached 0.9. The adjacency matrix was transformed into topological overlap matrix (TOM). Hierarchical clustering was carried out based on the dissimilarity of TOM, and highly similar gene modules were merged (the minimum module contained 50 genes, dissimilarity threshold was 0.25). The correlation between each module and TIMER immune cell abundance and T, N, clinical stage was analyzed. The module that had the highest correlation with the level of immune infiltration was taken as the target module. Finally, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses on target module gene were performed by using R package clusteProfiler, with the screening criteria of p < 0.05 and q < 0.05.

2.3. Construction and verification of the prognostic model

TCGA-BRCA data were used as the training set, and R package survival (https://cran.r-project.org/web/packages/survival/ survival.pdf) was applied for univariate Cox regression analysis, with p < 0.05 as the standard. Genes notably associated with prognosis were retained for subsequent analyses. For the purpose of improving the interpretability and prediction accuracy of statistical models, Least absolute shrinkage and selection operator (LASSO) Cox regression analysis was accomplished on obtained genes concerned with prognosis by using the R package glmnet [8]. Besides, multivariate Cox regression analysis was executed by using R package survival to build the prognostic model. Samples were divided into high- and low-risk groups in accordance with a median risk score, R package FactoMineR [9] was used for dimension reduction of feature genes, and R package survival was used for Kaplan-Meier (K-M) analysis. Receiver operating characteristic (ROC) curves were plotted using R package timeROC [10], and the area under the curve (AUC) value was obtained. The above methods were repeated in the validation set.

2.4. TIME assessment

The ssGSEA analysis was implemented in high- and low-risk groups utilizing the R package GSVA [11]. R package estimation (https://bioinformatics.mdanderson.org/estimate/rpackage. html) was used to calculate the stromal score, immune score, ESTIMATE score, and tumor purity score in the two risk groups. R package CIBERSORT [12] was utilized to evaluate differences in immune cell infiltration levels between two risk groups. *T*-test was performed on human leukocyte antigen (HLA) protein expression level in the two groups to reveal differences between groups, and correlation analysis was done on the expression levels of immune checkpoint-related molecules (PDL1, PD1, CTLA4, and LAG3) in TCGA-BRCA samples and prognostic model risk score.

2.5. Analysis of gene mutation frequency in high- and low-risk groups

R package GenVisR [13] was utilized to draw a waterfall map of the top 10 high-frequency mutation genes in two risk groups, and differences in gene mutation frequency in different groups were analyzed.

2.6. GSEA analysis of high- and low-risk groups

TCGA-BRCA high- and low-risk group samples were analyzed by GSEA software [14]. The Wilcox test was used to compare risk scores between BC stages or subtypes.

2.7. Correlation between risk scores and clinical traits

TCGA clinical information (Stage, Subtype) was obtained through cBioPortalData package [15,16]. The Wilcoxon test was used to detect differences in risk scores among different stages and BC subtypes.

2.8. Nomogram construction and verification

Taking risk score as a prognostic factor, univariate and multivariate Cox regression analyses were fulfilled by R package survival combined with clinical characteristics (age, N stage, T stage, Stage) of TCGA-BRCA samples, to further evaluate the independence of multivariate model in predicting the prognosis of patients. R package rms [17] was used to forecast the feasibility of 1-, 3-, and 5-year overall survival (OS) by combining clinical traits (age, N stage, T stage, Stage) and risk score values of the genes. Corresponding calibration curves were drawn to evaluate the prediction effect.

2.9. Cell culture

Human normal mammary epithelial cells MCF 10A (CTCC-001-0045) and human BC cells MDA-MB-231(CTCC-001-0019) and MCF7 (CTCC-001-0042) were purchased from Meisen Cell Biotechnology Co. Ltd. All

the above cell lines were cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS) at $37 \,^{\circ}$ C in a 5% CO₂ incubator.

2.10. Quantitative reverse transcription polymerase chain reaction (qRT-PCR)

Total RNA was extracted from cells using Trizol reagent (Invitrogen, USA) and reversely transcribed into cDNA. cDNA was used as a template for PCR amplification using the One Step SYBR PrimeScript RT-PCR Kit (Takara, Japan). PCR was performed using a fluorescent quantitative PCR instrument. Gene expression levels were quantitatively calculated by the $2^{-\Delta\Delta Ct}$ method. GAPDH was the internal reference. Primer sequences are shown in Table 1.

2.11. Statistical analysis

All data analyses and visualization were accomplished by R Studio (version 3.4.3), and then the Wilcoxon rank sum test or Kruskal–Wallis test was adopted to compare differences. The Cox regression model was applied to execute univariate and multivariate analyses, and the log-rank test was utilized to evaluate the difference in survival time. Spearman's rank-order correlation test was used to analyze the correlation between risk score and immune cell infiltration. Here, p < 0.05 means statistical significance.

Each experiment for qRT-PCR was repeated at least three times. Data were analyzed using GraphPad Prism 8.0 software. A comparison between the two groups was conducted by Student's *t*-test. p < 0.05 was considered a significant difference.

3. Results

3.1. WGCNA is used to mine the immune-related target module

The prognosis of BC patients is closely related to the infiltration of immune cells. To identify the genes associated with immunity in BC patients, WGCNA was performed. In TCGA-BRCA dataset, the MAD top 5000 genes in 1046 samples obtained after data preprocessing were used to construct a co-expression network. Here, the soft threshold value was set to 6 (scale-free R2=0.91) to construct a scale-free network (Figure 1(A–D)). Subsequently, altogether 12 gene modules were acquired based on average hierarchical clustering and dynamic splicing method, modules and the number of genes included were as follows: black (198), blue (814), brown (626), green (296), green–yellow (57), gray (993), magenta

Table 1. Primer sequences for gRT-PCR.

Primer sequence (5'-3')			
TNFRSF14	CATCGTCATTGTTTGCTCCAC	CTCCTGTCTTTTCCGCTGG	
NFKBIA	GAGTCAGAGTTCACGGAGTTC	CATGTTCTTTCAGCCCCTTTG	
DLG3	TCGGTGAATGGAGTGAATCTG	TGTTCTCGTAAGTCATGTATCTTCG	
IRF2	GATCTTTCTCCTGAGTATGCGG	GCAAATGTCTGGCGGATTG	
CYP27A1	CATGGAGCTATGGAAGGAGC	TGAAAGCATCCGTATAGAGCG	
GAPDH	CAATGACCCCTTCATTGACC	GACAAGCTTCCCGTTCTCAG	



Figure 1. Identification of soft threshold power in WGCNA and screening of immune-related gene modules. (A) Scale-free index analysis of various soft threshold power β . (B) Average connectivity analysis of each soft threshold power β . (C) When $\beta = 6$, the connectivity distribution of the co-expression network. (D) Scale-free topological performance verification of co-expression network when $\beta = 6$. (E) Dynamic splicing clustering tree based on intergenic dissimilarity to measure 1-TOM clustering. Each color represents a gene module. (F) Correlation heat map between module principal components and BC immune traits. The numbers above the parentheses in each square indicate the correlation coefficient and the numbers in parentheses indicate the degree of significance

(119), pink (127), purple (111), red (215), turquoise (1004), yellow (440) (Figure 1(E)). Correlation analysis between modules and infiltration levels of six types of immune cells displayed that the green module was signally relevant to the cells (Figure 1(F)). Therefore, this study took the green module as follow-up research object and named it target module.

3.2. GO and KEGG analyses of target module gene

In order to explore the function of target module genes, GO and KEGG enrichment analyses were performed on the green module genes using the clusterProfiler package. GO enrichment analysis of target module gene portrayed that target module gene was related to positive regulation of cytokine production, defense response to the virus, type I interferon signaling pathway, and other cytokines, interferon, and virus defense biological functions in biological process regulation. While in terms of molecular function, target module gene was associated with the vesicle lumen, lysosomal lumen, secretory granule membrane, and other cellular components such as the vesicle membrane and lysosome cavity. In terms of cellular component regulation, target module gene was related to GTPase activity, immune receptor activity, cytokine receptor activity, chemokine activity, and other cytokines, chemokines, and immune receptor activity (Figure 2(A)). In KEGG analysis, target module gene was mainly enriched in cytokine–cytokine receptor interaction, chemokine signaling pathway, viral protein interaction with cytokine and cytokine receptor, complement and coagulation cascades, antigen processing and presentation, interaction between cytokines and cytokine receptors, chemokine signaling pathway, complement and coagulation cascade, antigen processing and other signaling pathways (Figure 2(B)).

3.3. Construction of the prognostic model

We expect to construct a model that can predict the prognosis of BC patients through these immune-related module genes, so as to accurately predict the prognosis of patients and guide clinical treatment. Univariate Cox analysis was carried out on target module gene in TCGA-BRCA, and 85 genes markedly associated with prognosis were obtained as the research objects. These genes were selected by LASSO regression analysis and 11 genes were identified (Figure 3(A,B)). Five



Figure 2. Enrichment analysis of target module gene.

(A) GO functional enrichment analysis. (B) KEGG analysis. The color of the dots indicates the degree of enrichment significance. The size of the dots indicates the number of enriched genes.



Figure 3. Construction of the risk assessment model based on the target module.

(A) LASSO coefficient spectrum of LASSO regression analysis. (B) Coefficient distribution map for logarithmic (A) sequence in LASSO model. (C) Forest plot of five genes obtained by multivariate Cox regression analysis. When hazard ratio (HR) >1, the characteristic gene is a risk factor. When HR <1, the trait gene is a protective factor. When HR = 1, the characteristic gene does not play a role in survival time.



Figure 4. Validation of the five-feature gene risk assessment model.

(A) Expression heat map of five-feature genes in high- and low-risk groups of TCGA-BRCA training set. (B) Distribution of five-feature gene prognostic model risk scores in TCGA-BRCA training set. (C) Distribution of survival status in TCGA-BRCA training set. Red dots indicate that the patient is dead at a fixed time, and green dots indicate that the patient is alive. (D) PCA reduced-dimension map of five-feature genes in TCGA-BRCA high- and low-risk groups. (E–G) ROC curves of a prognostic model in TCGA-BRCA training set, GSE42568 validation set. (H–J) K-M survival curves in high- and low-risk groups of TCGA-BRCA training set, GSE42568 validation set.

prognostic feature genes (TNFRSF14, NFKBIA, DLG3, IRF2, and CYP27A1) were screened by multivariate Cox stepwise regression method from these 11 genes to construct a prognostic risk assessment model. In the constructed model, TNFRSF14, NFKBIA, IRF2, and CYP27A1 were the protective factors of the model, while DLG3 was the risk factor. Risk score = $-0.1504 \times \text{TNFRSF14} - 0.21661 \times \text{NFKBIA} + 0.4818 \times \text{DLG3} - 0.6047 \times \text{IRF2} - 0.1576 \times \text{CYP27A1}$ (Figure 3(C)).

3.4. Validation of the prognostic model

Next, we validated the performance of the prognostic model in internal and external queues. TCGA-BRCA samples were put into high- and low-risk groups based on the median prognostic model risk score. Analysis of expression levels of feature genes in two groups showed that the expression of five feature genes TNFRSF14, NFKBIA, IRF2, and CYP27A1 in the low-risk group was higher than that in the high-risk group (Figure 4(A)). Distribution of risk score and survival status revealed that risk score and the number of deaths were positively correlated while the survival time decreased (Figure 4(B,C)). Principal component analysis (PCA) dimension reduction revealed that the five-feature gene risk assessment model could better distinguish the samples of two groups (Figure 4(D)). For the purpose of further assessing the predictive effect of the prognostic model, ROC curves were drawn to evaluate the 1-, 3-, and 5-year survival rates of patients in TCGA training set and two GEO validation sets. AUC values of 1-, 3-, and 5-year survival rates in

TCGA-BRCA training set (1 year: 0.68, 2 years: 0.67, 3 years: 0.67; Figure 4(E)), GSE42568 (1 year: 0.83, 0.74, and 0.7; Figure 4(F)) and GSE20685 validation sets (1 year: 0.83, 2 years: 0.74, 3 years: 0.71; Figure 4(G)). K-M survival curve displayed that in TCGA training set, GSE42568, and GSE20685 validation sets, patients in the low-risk group had significantly higher overall prognostic survival than those in the high-risk group (Figure 4(H–J)). The above results supported the rationality of the risk assessment model based on five feature genes in this study.

3.5. Differential analysis of immune mode and gene mutation mode between high- and low-risk groups

Based on the above analysis, we found that there was a significant difference in survival rate between the high- and low-risk groups divided by the model constructed by immune-related genes. Therefore, we further explored the differences in immunity and other aspects between the high- and low-risk groups. The ssGSEA results in two groups established a higher level of immune cell infiltration in the low-risk group (Figure 5(A)). Comparison of tumor purity between the two risk groups revealed higher tumor purity in the high-risk group (Figure 5(B)). *T*-test for HLA-related proteins was performed in two risk groups, whose results showed that expression levels of HLA-related proteins were markedly high in the low-risk group (Figure 5(C)). CIBERSORT algorithm was applied to score the immune infiltration abundance of 22 immune cells. *T*-test



Figure 5. Analysis of immune characteristics and gene mutation differences in TCGA-BRCA high- and low-risk groups. (A) Heat map of 29 immune features analyzed by ssGSEA. (B) Violin plot of tumor purity score for samples from high- and low-risk groups. ***p < 0.0001. (C) Box-plot of HLA-related protein expression in high- and low-risk groups. ***p < 0.0001. (D) CIBERSORT was used to analyze differences in 22 immune cell infiltration scores between high- and low-risk groups. **p < 0.0001. (E) Pearson correlation analysis of programmed cell death-Ligand 1(PD-L1) and risk score. (F) Pearson correlation analysis between programmed death-1(PD-1) and risk score. (G) Pearson correlation analysis between LAG3 and risk score. (H) Pearson correlation analysis between CTLA4 and risk score. (I) Waterfall map of 10 high-frequency mutation genes in the high-risk group. The ordinate represents the gene, the abscissa represents the patient, and the color of the thin bars in the coordinate represents the mutation type of the gene corresponding to the abscissa

showed that CD8⁺ T cells, natural killer (NK) cells, and M1 were dramatically lower in the high-risk group than in another group, and M2/M0 was prominently higher in the high-risk group (Figure 5(D)). These results indicated that there were differentiae in the types and levels of immune cell infiltration between the two groups. Further studies conducted Pearson correlation analysis between risk score and genes in connection with the immune checkpoint. Results showed that the risk score of the five-gene prognostic model was notably negatively correlated with the expression of immune checkpoint-related genes PDL1, PD1, LAG3, and CTLA4 (Figure 5(E-H)). In view of relevant reports suggesting that differences in immune mode involved gene mutation regulation, this study visualized the top 10 genes with high-mutation frequency in high- and low-risk groups. The results showed that the frequency of gene mutation was higher in the high-risk group, especially the frequency of TP53 mutation in the high-risk group was dramatically higher than that in another group (Figure 5(I,J)). These above results supported the feasibility of a prognostic model to forecast the immune pattern as well as gene mutation frequency of samples.

3.6. GSEA analysis

To explore the functional differences between the two risk groups, GSEA analysis was conducted on two risk groups of TCGA-BRCA samples. The results showed that the gene sets of the two groups were mainly different in immune-related pathways such as CHEMOKINE_SIGNALING_PATHWAY (ES = -0.57), B_CELL_RECEPTOR_SIGNALING_PATHWAY (ES = -0.60), and T_CELL_RECEPTOR_SIGNALING_PATHWAY (ES = -0.57) (Figure 6(A–C)). Among them, the low-risk group was mainly enriched in these three immune-related signaling pathways. And these pathways were significantly enriched as shown by false discovery rate (FDR) q-val.

3.7. Construction and verification of the nomogram

We further explored whether the prognostic model had the ability to independently predict the prognosis of BC patients.

Firstly, the analysis found that patients in stages II-IV corresponded to higher risk score (Figure 7(A)). More interestingly, there were significant differences in risk scores between BC patients with different BC subtypes (Figure 7(B)). In this study, univariate and multivariate Cox regression analyses were fulfilled on TCGA-BRCA training set in combination with clinical traits (age, T stage, N stage, Stage) and risk score. Univariate Cox analysis indicated that age, T stage, N stage, Stage, and risk score had significant effects on prognosis (Figure 7(C)). Multivariate Cox analysis clarified that age and stage were significant prognostic factors, and risk score could be employed as prognostic factors independent of clinical characteristics (Figure 7(D)). By combining the clinical characteristics of genes (age, T stage, N stage, Stage) and risk score values, generated nomogram can be used to forecast 1-, 3-, and 5-year OS rates of patients (Figure 7(E)), and corresponding calibration curve fitting was ideal (Figure 7(F-H)). Overall, these results confirmed that the five-feature gene prognostic model had good predictive performance.

3.8. Expression-level validation of prognostic feature genes

We further validated the expression levels of the model feature genes using qRT-PCR. The validation results found that consistent with the previous analysis of gene expression in the database, all the feature genes except DLG3 were lowly expressed in BC (Figure 8).

4. Discussion

In recent years, immunotherapy research aimed at mobilizing cell activity in TIME has achieved fruitful results. Although only part of patients with BC benefit from it, the importance of TIME in the development of BC and prognoses of patients has been extensively recognized [18,19]. Su et al. [20] reported that tumor-associated macrophages in BC can induce the differentiation of circulating naïve CD4⁺ T cells into Tregs and lead to an immunosuppressive response. A thorough exploration of the correlation between



Figure 6. Enrichment results of GSEA biological pathway in high- and low-risk groups. (A-C) GSEA results showed enriched signaling pathways. ES indicates enrichment scores and FDR q-val indicates significance.





(A) Correlation between risk score and clinical tumor stage. (B) Association between risk scores and breast cancer subtypes. (C) The forest plot was constructed in line with univariate Cox regression analysis. (D) The forest plot was constructed in line with multivariate Cox regression analysis. (E) Nomogram for predicting OS in TCGA-BRCA training set. (F–H) Calibration curves for predicting 1-, 3-, and 5-year OS rates of BC patients.

immune cells and prognosis and treatment prediction is essential for the precise management of cancer prognosis. Therefore, in this study, we combined WGCNA and TIME assessment to mine the gene modules related to the immune regulation of BC and established a five-gene prognostic model to predict the prognostic survival rate of BC patients. In addition, differences in immune cell infiltration, immune checkpoint genes, and gene mutation frequency between the two risk groups were analyzed.

Five genes utilized to construct the prognostic risk assessment model in this study were TNFRSF14, NFKBIA, DLG3, IRF2, and CYP27A1. Among them, only DLG3 was considered a prognostic risk factor, and the rest were prognostic protective factors of BC. Interestingly, we found that all the feature genes except DLG3 were lowly expressed in BC in our qRT-PCR experimental validation. Previous research has shown that these genes have an association with immune regulation, cancer progression, or prognosis. TNFRSF14 gene is associated with good prognosis in a variety of cancer patients, but its role in BC is unknown [21–23]. Functionally, TNFRSF14 is the link of NF- κ B, RELA, AP-1, AKT, and other signaling pathways, which plays a crucial part in the immune system, such as promoting T-cell co-stimulation, regulating dendritic cell homeostasis, activating autoimmunemediated inflammatory response and host defense against pathogens [24]. Specifically, the typical function of the gene in immune regulation is to promote the activation of NK cells, thereby promoting the anti-tumor activity of



Figure 8. Expression levels of prognostic feature genes in the high- and low-risk groups were verified by qRT-PCR. *p < 0.05 indicates a significant difference in gene expression between the tumor and normal groups.

tumor-specific CD8⁺ T cells [25]. Its immune regulation is reflected in multifarious cancers like lymphatic cancer, ovarian cancer, and bladder cancer [21-23]. NFKBIA is the encoding gene of nuclear factor κB inhibitor (ikba), which is the inhibitor of NF-kB, a typical inflammatory factor [26]. It is well known that the activation of NF-KB is closely related to inflammation-mediated tumor progression [27]. It has been shown that NFKBIA is related to the good prognosis of patients. BC From the perspective of mechanism, NFKBIA-encoded protein IkBa binds to NF-kB in the cytoplasm and inhibits BC metastasis by inhibiting macrophage M2 polarization [28,29]. DLG3 (Discs large homolog), also known as SAP102, is a member of the membraneassociated guanylate kinases (MAGUKs) family [30]. This gene is commonly seen in the study of neurological diseases, which plays a signal transduction role mainly by connecting N-methyl-D-aspartate (NMDA) receptors to the submembrane cell matrix related to excitatory synaptic density [31]. However, recently, related studies have gradually emerged that this gene can foretell the prognosis of cancer patients. For example, previous studies have proved that high expression of DLG3 is related to the decrease in prognosis and

survival rate of BC. As a promoter, this gene up-regulates the PI3K/AKT signaling pathway activated by RAC1 to promote the proliferation, invasion, and epithelial-mesenchymal transition of BC cells [32,33]. IRF1 encodes interferon regulatory factor, and IRF1 recruits and activates tumor-infiltrating dendritic cells in an NF-kB-dependent manner to exert anti-tumor activity [34]. The high expression of this factor may affect tumor progression through immunosuppression, which is consistent with the results of this study that this factor serves as a prognostic protective factor for BC. The encoded protein of CYP27A1 is a hydroxylase, which acts as a homeostasis regulator of estrogen receptor and inhibits the growth and metastasis of BC [35]. It has been reported that this factor is a biomarker with good prognosis in BC patients. This report revealed that the high expression of CYP27A1 in BC patients was concerned with a 4-fold (HRadj = 0.26, 95% CI = 0.07-0.93) and 8-fold (HRadj = 0.13, 95% CI = 0.03-0.60) reduction in the incidence of distant recurrence-free survival events. In vitro studies have shown that 27HC catalyzed by CYP27A1 can effectively inhibit the proliferation of BC cells [36]. Importantly, if verified, these results may have an impact on the decision making of adjuvant therapy in premenopausal patients, especially when considering downgrade treatment. This is accordant with the outcomes of this study, but the specific mechanism is unknown [36]. This study revealed that the association of five genes with the prognosis of patients was highly consistent with the gene function described in previous reports, supporting its reliability as a prognostic biomarker for BC.

To further explore the correlation between the prognosis of the five-gene model and immune regulation, based on ssGESA, this study disclosed that the difference in survival rate between the two groups may be driven by differences in immune cell infiltration (CD8+ T cells, NK cells, M1 macrophages, M2 macrophages). This study exhibited that CD8⁺ T cells in the tumor microenvironment (TME) of the high-risk group were observably lower than those of another group. The function of CD8⁺ T cells was to activate other cytotoxic T lymphocytes in the tumor immune cycle, such as CD4+ T cells, and mediate anti-tumor immune response [37]. In clinical studies, NK cells often act synergically with CD8+ T cells in the anti-tumor immune process, and both have similar cytotoxic mechanisms [38,39]. M1 macrophages are the pro-inflammatory phenotype of macrophages, which is correlated with the mediation of anti-tumor immune response. The typical function of this cell in the TIME of BC is to inhibit the NF-kB signal transduction to promote the pyrophosphorylation of BC cells, thereby promoting the apoptosis of cancer cells [40]. While M2 macrophages in BC mediate cell proliferation, migration, and tumor growth in mice by reducing expression of interferon regulatory factor IRF7 and up-regulating the oncogene miR-1587 [41]. This study revealed the down-regulation of multiple immune cell infiltration levels (CD8⁺ T cells, NK cells, M1 macrophages) and the up-regulation of M2 macrophages that promoted the malignant progression of tumors in the high-risk group, suggesting that TIME in this group was immunosuppressive. Perhaps this model could use the state of TIME to determine the prognosis of patients.

In conclusion, we screened BC immune-associated gene modules *via* WGCNA network and built a five-gene prognostic model based on the screening of prognostic biomarkers. Based on clinical survival analysis, immune cell infiltration level, and pathway enrichment analysis, it provides a reliable theoretical reference for the study of BC immune mechanism. However, this study is purely bioinformatic and has not been proved by subsequent molecular experiments. Therefore, further molecular and cell experiments are required to support the results here.

Author contributions

Conceptualization and writing – original draft: Ruijuan Wang and Huanhong Zeng; data curation: Junjie Zheng and Naizhuo Ke; formal analysis and visualization: Junjie Zheng; funding acquisition and supervision: Naizhuo Ke; investigation and project administration: Xueming Xiao; methodology and software: Qiang Lin; resources: Wenjun Xie; validation: Hui Zhang; writing – review and editing: Ruijuan Wang, Huanhong Zeng, and Hui Zhang.

Ethics approval and consent to participate

Not applicable.

Disclosure statement

The authors report no conflict of interest.

Funding

This study was supported in part by grants from the Young and Middle-Aged Talents Training Project of Fujian Provincial Health Commission (2021GGA001) and the Natural Science Foundation of Fujian Province of China (2021J01379).

Data availability statement

The datasets generated and analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

References

- [1] Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2022. CA Cancer J Clin. 2022;72(1):1–13.
- [2] Koren S, Bentires-Alj M. Breast tumor heterogeneity: source of fitness, hurdle for therapy. Mol Cell. 2015;60(4):537-546.
- [3] Peart O. Metastatic breast cancer. Radiol Technol. 2017;88(5):519M– 539M.
- [4] Liu N, Johnson KJ, Ma CX. Male breast cancer: an updated surveillance, epidemiology, and end results data analysis. Clin Breast Cancer. 2018;18(5):e997–e1002.
- [5] Zhang Y, Zhang Z. The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. Cell Mol Immunol. 2020;17(8):807–821.
- [6] Mirgayazova R, Khadiullina R, Mingaleeva R, et al. Novel isatin-based activator of p53 transcriptional functions in tumor cells. Mol Biol Res Commun. 2019;8(3):119–128.
- [7] Harao M, Forget M-A, Roszik J, et al. 4-1BB-enhanced expansion of CD8⁺ TIL from triple-negative breast cancer unveils mutation-specific CD8⁺ T cells. Cancer Immunol Res. 2017;5(6):439–445.
- [8] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1–22.
- [9] Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. J Stat Softw. 2008;25(1):1–18.
- [10] Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Stat Med. 2013;32(30):5381–5397.
- [11] Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.
- [12] Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453–457.
- [13] Skidmore ZL, Wagner AH, Lesurf R, et al. GenVisR: genomic visualizations in R. Bioinformatics. 2016;32(19):3012–3014.
- [14] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102(43):15545–15550.
- [15] Cerami E, Gao J, Dogrusoz U, et al. The cBio Cancer Genomics Portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401–404.
- [16] Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269)pl:pl1.
- [17] Huang C, Liu Z, Xiao L, et al. Clinical significance of serum CA125, CA19-9, CA72-4, and fibrinogen-to-lymphocyte ratio in gastric cancer with peritoneal dissemination. Front Oncol. 2019;9:1159.

- [18] Disis ML, Stanton SE. Immunotherapy in breast cancer: an introduction. Breast. 2018;37:196–199.
- [19] Cassetta L, Pollard JW. Repolarizing macrophages improves breast cancer therapy. Cell Res. 2017;27(8):963–964.
- [20] Su S, Liao J, Liu J, et al. Blocking the recruitment of naive CD4⁺ T cells reverses immunosuppression in breast cancer. Cell Res. 2017;27(4):461–482.
- [21] Carreras J, Lopez-Guillermo A, Kikuti YY, et al. High TNFRSF14 and low BTLA are associated with poor prognosis in follicular lymphoma and in diffuse large B-cell lymphoma transformation. J Clin Exp Hematop. 2019;59(1):1–16.
- [22] Choi E-K, Kim W-K, Sul O-J, et al. TNFRSF14 deficiency protects against ovariectomy-induced adipose tissue inflammation. J Endocrinol. 2014;220(1):25–33.
- [23] Zhu YD, Lu MY. Increased expression of TNFRSF14 indicates good prognosis and inhibits bladder cancer proliferation by promoting apoptosis. Mol Med Rep. 2018;18(3):3403–3410.
- [24] Steinberg MW, Cheung TC, Ware CF. The signaling networks of the herpesvirus entry mediator (TNFRSF14) in immune regulation. Immunol Rev. 2011;244(1):169–187.
- [25] Fan Z, Yu P, Wang Y, et al. NK-cell activation by LIGHT triggers tumor-specific CD8⁺ T-cell immunity to reject established tumors. Blood. 2006;107(4):1342–1351.
- [26] Hayden MS, Ghosh S. Signaling to NF-κB. Genes Dev. 2004;18(18): 2195–2224.
- [27] DiDonato JA, Mercurio F, Karin M. NF-κB and the link between inflammation and cancer. Immunol Rev. 2012;246(1):379–400.
- [28] Li Y, Zhao X, Liu Q, et al. Bioinformatics reveal macrophages marker genes signature in breast cancer to predict prognosis. Ann Med. 2021;53(1):1019–1031.
- [29] Liu B, Sun L, Liu Q, et al. A cytoplasmic NF-κB interacting long noncoding RNA blocks IκB phosphorylation and suppresses breast cancer metastasis. Cancer Cell. 2015;27(3):370–381.
- [30] Zheng CY, Seabold GK, Horak M, et al. MAGUKs, synaptic development, and synaptic plasticity. Neuroscientist. 2011;17(5):493–512.

- [31] Müller BM, Kistner U, Kindler S, et al. SAP102, a novel postsynaptic protein that interacts with NMDA receptor complexes in vivo. Neuron. 1996;17(2):255–265.
- [32] Liu J, Li P, Wang R, et al. High expression of DLG3 is associated with decreased survival from breast cancer. Clin Exp Pharmacol Physiol. 2019;46(10):937–943.
- [33] Liu B, Xu Y, Zhang L, et al. Hypermethylation of DLG3 promoter upregulates RAC1 and activates the PI3K/AKT signaling pathway to promote breast cancer progression. Evid Based Complement Alternat Med. 2021;2021:8428130.
- [34] Ghislat G, Cheema AS, Baudoin E, et al. NF-κB-dependent IRF1 activation programs cDC1 dendritic cells to drive antitumor immunity. Sci Immunol. 2021;6(61):eabg3570.
- [35] Kimbung S, Chang C-Y, Bendahl P-O, et al. Impact of 27-hydroxylase (CYP27A1) and 27-hydroxycholesterol in breast cancer. Endocr Relat Cancer. 2017;24(7):339–349.
- [36] Inasu M, Bendahl P-O, Fernö M, et al. High CYP27A1 expression is a biomarker of favorable prognosis in premenopausal patients with estrogen receptor positive primary breast cancer. NPJ Breast Cancer. 2021;7(1):127.
- [37] Farhood B, Najafi M, Mortezaee K. CD8⁺ cytotoxic T lymphocytes in cancer immunotherapy: a review. J Cell Physiol. 2019;234(6):8509–8521.
- [38] Xie G, Dong H, Liang Y, et al. CAR-NK cells: a promising cellular immunotherapy for cancer. EBioMedicine. 2020;59:102975.
- [39] André P, Denis C, Soulas C, et al. Anti-NKG2A mAb is a checkpoint inhibitor that promotes anti-tumor immunity by unleashing both T and NK cells. Cell. 2018;175(7):1731–1743.
- [40] Tan Y, Sun R, Liu L, et al. Tumor suppressor DRD2 facilitates M1 macrophages and restricts NF-κB signaling to trigger pyroptosis in breast cancer. Theranostics. 2021;11(11):5214– 5231.
- [41] Tu D, Dou J, Wang M, et al. M2 macrophages contribute to cell proliferation and migration of breast cancer. Cell Biol Int. 2021;45(4):831–838.