# Identification of Molecular Subtypes and Prognostic Features of Breast Cancer Based on TGF-β Signaling-related Genes

Cancer Informatics
Volume 24: 1–14

DOI: 10.1177/11769351251316398

Jia Qu, Mei-Huan Wang, Yue-Hua Gao and Hua-Wei Zhang

Department of Ultrasound, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China.

## ABSTRACT

**OBJECTIVES:** The TGF-β signaling pathway is widely acknowledged for its role in various aspects of cancer progression, including cellular invasion, epithelial-mesenchymal transition, and immunosuppression. Immune checkpoint inhibitors (ICIs) and pharmacological agents that target TGF-β offer significant potential as therapeutic options for cancer. However, the specific role of TGF-β in prognostic assessment and treatment strategies for breast cancer (BC) remains unclear.

**METHODS:** The Cancer Genome Atlas (TCGA) database was utilized to develop a predictive model incorporating five TGF-β signaling-related genes (TSRGs). The GSE161529 dataset from the Gene Expression Omnibus was employed to conduct single-cell analyses aimed at further elucidating the characteristics of these TSRGs. Additionally, an unsupervised clustering algorithm was applied to categorize BC patients into two distinct groups based on the five TSRGs, with a focus on immune response and overall survival (OS). Further investigations were conducted to explore variations in pharmacotherapy and the tumor microenvironment across different patient cohorts and clusters.

**RESULTS:** The predictive model for BC identified five TSRGs: FUT8, IFNG, ID3, KLF10, and PARD6A. Single-cell analysis revealed that IFNG is predominantly expressed in CD8+ T cells. Consensus clustering effectively categorized BC patients into two distinct clusters, with cluster B demonstrating a longer OS and a more favorable prognosis. Immunological assessments indicated a higher presence of immune checkpoints and immune cells in cluster B, suggesting a greater likelihood of responsiveness to ICIs.

**CONCLUSION:** The findings of this study highlight the potential of the TGF-β signaling pathway for prognostic classification and the development of personalized treatment strategies for BC patients, thereby enhancing our understanding of its significance in BC prognosis.

**KEYWORDS:** TGF-β, breast cancer, immunotherapy, prognostic model, IFNG

**RECEIVED:** June 17, 2024. **ACCEPTED:** January 8, 2025.

**TYPE:** Original Research

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Natural Science Foundation of Shandong Province (grant number ZR2021QH347).

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Hua-Wei Zhang, Department of Ultrasound, Shandong Provincial Hospital affiliated to Shandong First Medical University, No. 324 Jing 5 Rd, Shandong Provincial Hospital, Jinan, Shandong 250021, China.
Email: zhanghuawei@sdfmu.edu.cn

## Introduction

Breast cancer (BC) represents the most prevalent malignancy among women, constituting approximately 30% of all cancer cases.[1] The prognosis for individuals diagnosed with BC remains unfavorable, primarily due to the disease's significant heterogeneity and propensity for early metastasis. Current therapeutic modalities, including surgical intervention, chemotherapy, radiotherapy, neoadjuvant therapy, targeted therapy, and immunotherapy, while effective in reducing mortality rates among BC patients, are insufficient in addressing the challenges posed by the high heterogeneity and early metastatic behavior of the disease.[2,3] Furthermore, existing classification systems, such as the TNM staging and PAM50 classification, fall short in accurately predicting patient outcomes.[4] Consequently, the identification and development of novel biomarkers to facilitate the reclassification of BC subtypes and to inform treatment strategies is essential for enhancing patient prognosis.

The tumor microenvironment (TME) comprises the extracellular matrix (ECM), stromal cells, and tumor-infiltrating immune cells (TIICs).[5] The type and functional mechanisms of TIICs are critical, as they not only affect the efficacy of anticancer immunotherapies such as immune checkpoint inhibitors (ICIs), but also serve as key determinants in predicting the clinical Research has established a robust correlation between the prognosis of triple-negative BC and the extent of T cell infiltration within the TME.[6-8] Additionally, there is increasing evidence that various signaling pathways modulate the composition and behavior of TIICs in the TME, with the transforming growth factor beta (TGF-β) signaling pathway emerging as particularly significant.[9,10]

The TGF-β signaling pathway exhibits a complex dual role in cancer progression.[11] In the early stages of cancer, TGF-β acts as a tumor suppressor by promoting cellular differentiation and inhibiting proliferation. Conversely, in advanced stages, TGF-β can function as a tumor promoter by facilitating epithelial-to-mesenchymal transition (EMT), thereby enhancing cancer initiation and enabling immune evasion.[12] TGF-β also promotes ferroptosis, a form of regulated cell death that inhibits cancer cell proliferation. Specifically, TGF-β1 activates Smad3, which in turn

suppresses the expression of the cystine/glutamate antiporter system Xc- (xCT), leading to decreased intracellular glutathione levels and increased lipid peroxidation, ultimately triggering cellular ferroptosis.[13] Moreover, excessive TGF-β activity within the TME can impair immune responses by inhibiting various immune cell types, including cytotoxic T cells and natural killer (NK) cells.[12,14] The interplay between TGF-β and the programmed cell death-1 receptor/programmed cell death ligand 1 (PD-1/PD-L1) pathways has been shown to be both independent and complementary, suggesting that TGF-β may serve as a predictive marker for the therapeutic efficacy of ICIs targeting PD-1/PD-L1.[15-17] Although TGF-β has been proposed as a potential therapeutic target for BC treatment in recent years,[18,19] further investigation into the intrinsic relationship between TGF-β and the TME is warranted to better assess the most effective treatment modalities for BC patients and to improve their overall prognosis.

This research established a predictive model for breast cancer that focuses on the TGF-β signaling pathway. A comprehensive analysis revealed a significant association between TGF-β and immune cells as well as cytokines within the tumor microenvironment (TME). The prognostic characteristics associated with TGF-β can be utilized to forecast and delineate prognostic outcomes and immune response status in breast cancer patients, as demonstrated by survival and drug susceptibility assessments. Additionally, cluster analysis based on five TGF-β signaling-related genes (TSRGs) enables the classification of breast cancer patients into two distinct molecular subtypes, which may enhance future clinical patient stratification and therapeutic strategies.

## Materials and Methods

### *Data sets collection*

The transcriptome expression matrix and clinical data from the TCGA-BRCA cohort, sourced from the Cancer Genome Atlas (TCGA) database, served as the training sets for this study. The TCGA-BRCA cohort comprised 1,113 BC samples alongside 113 normal breast tissue samples. For validation purposes, data sets GSE20685, GSE42568, and GSE58812 were obtained from the Gene Expression Omnibus (GEO) database. All three validation sets included overall survival (OS) data, with GSE42568 additionally providing recurrence-free survival (RFS) data and GSE58812 offering metastasis-free survival (MFS) data. Ultimately, a training set was established comprising 976 patients from the TCGA-BRCA cohort, while the validation sets included 327 patients from the GSE20685 cohort, 104 patients from the GSE42568 cohort, and 106 patients from the GSE58812 cohort. Furthermore, 225 TSRGs were extracted from a previously published article.[20]

### *Mutation analysis and immune cell infiltration analysis*

Somatic mutation data for the TCGA-BRCA cohort were retrieved from UCSC Xena, and the somatic mutation frequencies of TSRGs were visualized using the R package "maftools" in the form of a waterfall plot.[21] The single-sample Gene Set Enrichment Analysis (ssGSEA) method was employed to investigate variations in immune cell infiltration between normal breast tissue and BC tissue, with specific details regarding 28 immune cell types obtained from supplementary materials in a published article.[3] Additionally, differences in the TME between normal and BC groups were evaluated using the ESTIMATE algorithm to derive stromal and estimate scores.

### *Construction of TGF-β risk score model*

Initially, prognostic genes associated with OS were identified through univariate Cox regression analysis, applying a significance threshold of $P < .05$ within the TCGA-BRCA cohort. These OS-related genes were subsequently intersected with the 225 TSRGs, and the least absolute shrinkage and selection operator (LASSO) Cox regression analysis was utilized to minimize redundant genes and mitigate the risk of model overfitting.[22] Ultimately, multivariate Cox regression analysis was conducted to select the most effective prognostic model and to calculate coefficients for each candidate gene. The TGF-β risk score was determined as the product of the expression levels of each candidate gene and their corresponding coefficients. Patients were categorized into high- or low-risk groups based on their median TGF-β risk score. Additionally, boxplots were employed to illustrate the differences in candidate gene expression between the two groups, while Kaplan-Meier survival analysis was utilized to assess the impact of candidate gene expression on the prognosis of BC patients.

### *Validation of TGF-β risk score model*

The three previously mentioned datasets were utilized for validation purposes to further evaluate the accuracy of the TGF- risk score model. Patients were categorized into high- and low-risk groups based on the median risk scores within each validation set, with risk scores derived from the expression levels of selected candidate genes. Subsequently, the survival outcomes of patients in these risk categories were analyzed using Kaplan-Meier survival analysis. Given the constraints posed by variations in existing data, receiver operating characteristic (ROC) curves were employed in the GSE58812 validation set to assess the model's predictive accuracy for patient prognosis at 1, 3, and 5 years, while ROC curves from other training and validation sets were utilized to evaluate the model's prognostic accuracy at 3, 5, and 10 years.

### *Construction and validation of the TGF-β risk score related-clinical nomogram*

In the construction and validation of a TGF-β risk score-related clinical nomogram, independent prognostic factors were identified by examining the clinicopathological

characteristics and TGF-β risk scores of the TCGA-BRCA cohort through univariate and multivariate Cox regression analyses. The "rms" and "regplot" R packages were subsequently employed to develop the nomogram based on these findings.[23] To validate the predictive capability of the nomogram, calibration curves for 1, 3, and 5 years, ROC curves for 1, 2, and 5 years, and decision curve analysis (DCA) were conducted.

### *Gene set enrichment analysis*

Gene set enrichment analysis (GSEA) was performed using the R package "clusterProfiler," which included gene ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis for patients categorized into high- and low-risk groups. The results were visually represented through bubble plots and mountain plots, respectively.

### *Immune landscape estimation in BRCA TME*

Given the close association of emerging immunotherapies with immune checkpoints, the relationship between candidate genes and six immune checkpoints (PDCD1, CTLA4, IDO1, TIGIT, LAG3, and PVR) was assessed. Segmented violin plots illustrated the expression differences of these immune checkpoints across various risk groups and datasets. Additionally, to explore the role of TSRGs in the TME of BC patients, the Tumor Immune Estimation Resource (TIMER) database was utilized to analyze the correlation between the expression of these genes and the aforementioned immune checkpoints. Furthermore, transcriptome data from 976 BC patients were uploaded to the tumor immune dysfunction and exclusion (TIDE) website to evaluate responses to immune checkpoint blockade (ICB) in high- and low-risk patients.

### *Predicting the clinical sensitivity of BC patients to ICIs and chemotherapy drugs*

As the immunophenotype score (IPS, ranging from 0 to 10) from the Cancer Immunome Atlas database correlates with immunogenicity and can be used to predict sensitivity to ICIs treatment in patients with BC, IPS values from patients in the high- and low-risk groups were compared. Additionally, to forecast the responsiveness of BC patients to chemotherapeutic agents, drug susceptibility data were sourced from the Genomics of Drug Sensitivity in Cancer (GDSC) database, and the 50% inhibitory concentrations (IC50) of various drugs were predicted using the R package oncoPredict.[24]

### *Prediction of BRCA molecular subtypes by cluster analysis*

Clustering analysis was conducted utilizing the R package "ConsensusClusterPlus" to predict TGF-β related molecular subtypes in BC patients, based on candidate genes identified in previous studies.[25] Subsequently, principal component analysis (PCA) was employed to validate the accuracy of the classifications, while Kaplan-Meier curves were utilized to evaluate the prognostic outcomes of the subgroups. Additionally, relevant immunoassays were performed to investigate the differences in the TME among the identified subgroups.

### *Single-cell RNA-seq data acquisition and processing*

For the acquisition and processing of scRNA-seq data, the GSE161529 dataset from the GEO database was utilized to accurately assess the expression levels of five TSRGs at the cellular level. The "Seurat" software was employed for quality control, excluding cells with "nFeature" values below 200 and "percent.mt" values exceeding 20%. The data normalization was achieved using the "LogNormalize" method. Following normalization, PCA was applied for dimensionality reduction, and uniform manifold approximation and projection (UMAP) was utilized for unsupervised cell clustering. Finally, cell annotation was performed using the SingleR package.

### *Statistical analysis*

Statistical analyses and data visualizations were executed using R software (version 4.2.0). Both univariate and multivariate Cox regression analyses were conducted to identify independent prognostic factors, with a two-tailed $p$-value of less than .05 deemed statistically significant.

## Results

### *Identification of TSRGs in BC patients*

First, the proportion of 225 TSRGs with somatic mutations in the TCGA-BRCA cohort was assessed, and a waterfall plot was used to display the top 20 mutant genes, including TP53, CDH1, and others (Figure 1(a)). Furthermore, Figure 1(b) depicted the difference in TIICs in the TME between normal breast and BC tissue, with mast cells, eosinophils, natural killer T cells, and effector memory CD8 T cells being higher in normal tissues, while activated CD4 T cells and CD56 dim natural killer cells were higher in BC tissues. Lower stromal and ESTIMATE scores in BC tissues indicate a worse prognosis in BC patients (Figure 1(c)).

### *Construction of TGF-β risk score model for BC patients*

An optimal prognostic model was established through the integration of univariate COX regression analysis, LASSO regression analysis, and multivariate COX regression analysis (Figure 1(d) and (e)). This model comprises five TSRGs: FUT8, IFNG, ID3, KLF10, and PARD6A. Among these, ID3 and KLF10 are identified as adverse prognostic factors, indicated by a hazard ratio (HR) greater than 1, while FUT8, IFNG, and PARD6A are classified as protective prognostic

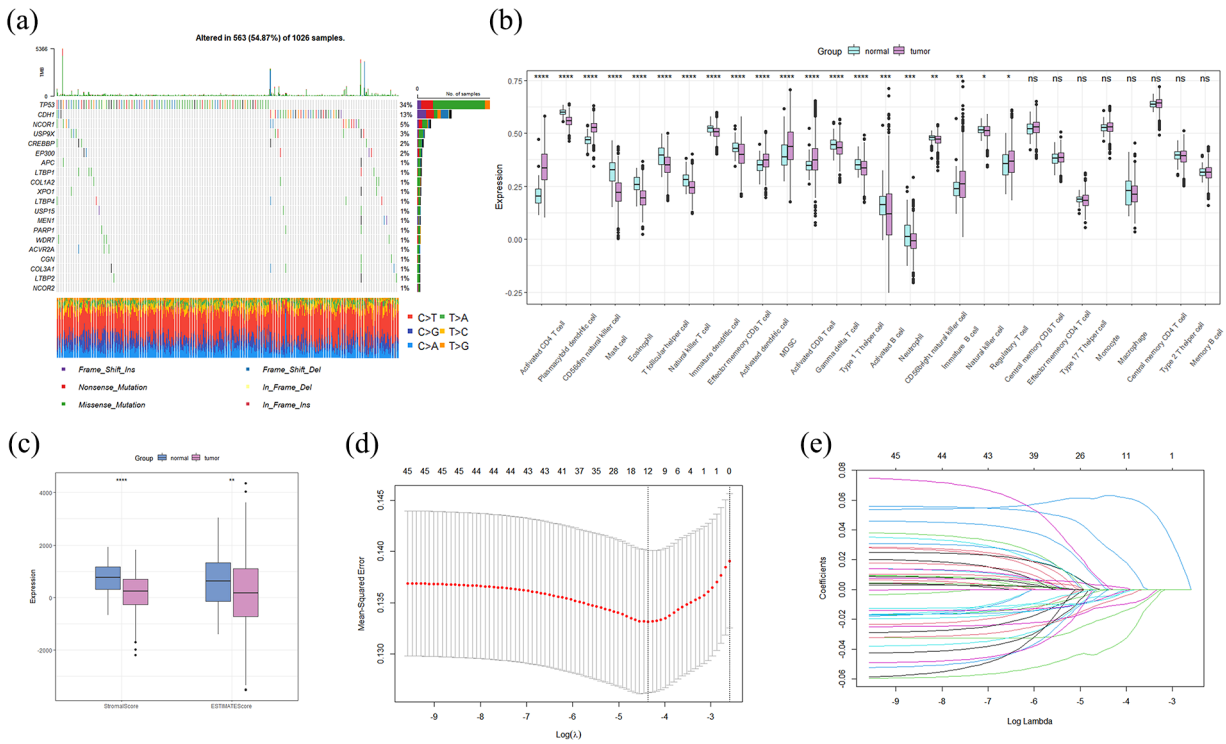


**Figure 1.** (a) Somatic mutation frequency of TGF-β in the TCGA-BRCA cohort, (b) Differences in the degree of invasion of 28 immune cells between normal and cancerous tissues in BC patients, (c) Variance of stromal score and ESTIMATE score in the different groups, (d) The LASSO Cox regression model was utilized to identify TSRGs, and (e) Select range of the optimal parameter (lambda) in the LASSO Cox regression model.

factors, as evidenced by an HR less than 1 (Figure 2(a)). The risk score is calculated using the formula: Risk score = −0.20464 * FUT8 - 0.65792 * IFNG + 0.21690 * ID3 + 0.37394 * KLF10 - 0.30315 * PARD6A. Based on the median risk score, subjects are categorized into high-risk and low-risk groups. Subsequent analysis revealed that ID3 and KLF10 exhibited higher expression levels in the high-risk group, whereas FUT8, IFNG, and PARD6A were more prominently expressed in the low-risk group (Figure 2(b)). The Kaplan-Meier survival curve analysis further assessed the independence of the five identified genes as predictive variables for BC. With the exception of FUT8 ($P = .13$), the other four genes were determined to be independent prognostic factors influencing BC outcomes ($P < .05$), as illustrated in Figure 2(c). Additionally, the differential expression of these five candidate genes in normal versus BC tissue is presented in Supplemental Figure S1 and S4.

### *Prognostic validation based on TGF-β risk score model*

The same prognostic analysis was performed in the training (TCGA-BRCA) and three validation sets (GSE20685, GSE42568, and GSE58812) to further investigate the accuracy of the TGF-β risk score model in BC prediction. Initially, patients within each dataset were categorized into high- or low-risk groups based on their median risk scores (Figures 3(b) and 4(b)). The results aligned with expectations, indicating an increase in mortality among BC patients corresponding to elevated hazard scores (Figures 3(c) and 4(c)). Consequently, individuals classified in the high-risk group exhibited poorer prognoses compared to those in the low-risk group, with statistical significance observed across various datasets (TCGA-BRCA: $P < .0001$; GSE58812-OS: $P = .003$; GSE58812-MFS: $P = .015$; GSE20685: $P = .0015$; GSE42568-OS: $P = .033$; GSE42568-RFS: $P = .033$) (Figures 3(a) and (a)). Heat maps illustrating the distribution of the five TSRG genes across the high- and low-risk groups are presented in Figures 3(d) and 4(d). Additionally, the predictive capability of the TGF-β risk score model was further assessed through ROC analysis, demonstrating satisfactory performance across all cohorts (Figures 3(e) and 4(e)).

### *Construction and validation of a nomogram for TGF-β risk score and clinical characteristics*

The relationship between the TGF-β risk score and various clinicopathological factors within the TCGA-BRCA dataset is illustrated in Supplemental Figure S2. Initial assessments utilizing both univariate and multivariate Cox regression analyses aimed to determine whether the variables T, N, M, age, stage, and TGF-β risk score functioned as independent prognostic indicators within the TCGA-BRCA cohort. The analysis revealed that age, M classification, and the TGF-β risk score were indeed independent prognostic factors (Figures 5(a) and (b)). Consequently, a clinicopathological nomogram for BC was developed incorporating these three variables to predict OS for BC patients at 1, 3, and 5 years (Figure 5(c)). The calibration curves for 1, 3, and 5 years demonstrated a strong concordance with the 45-degree

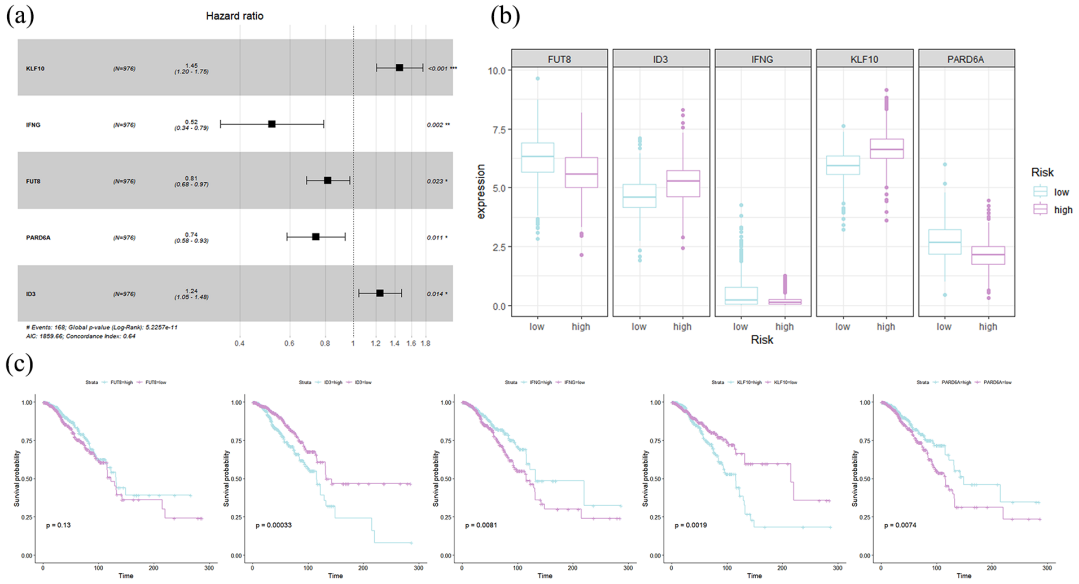


**Figure 2.** (a) The coefficient of the selected TSRGs, (b) Expression of 5 TSRGs in the high- and low-risk groups, and (c) The KM survival curves for 5 selected TSRGs based on expression level and OS.

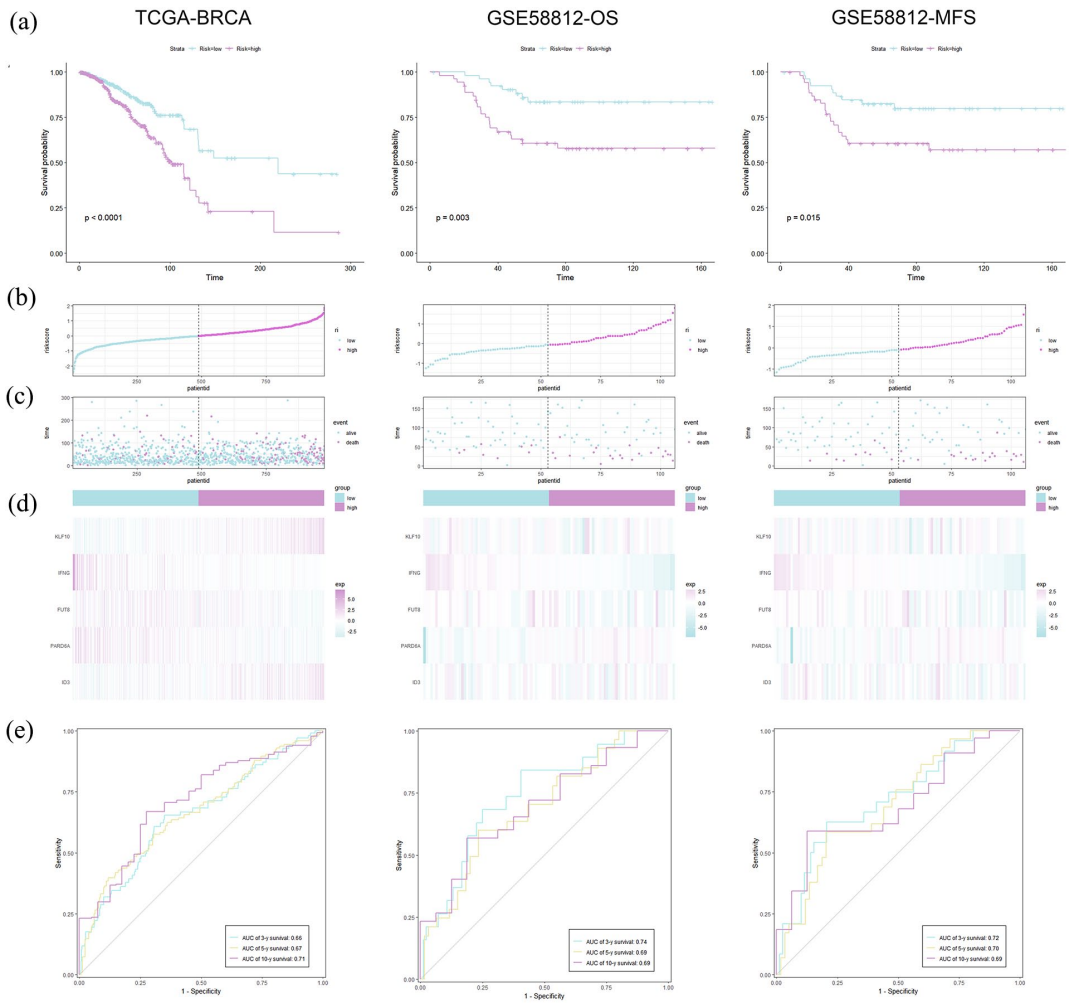


**Figure 3.** Assessment and verification of the efficiency of the TGF-β risk score: (a) The KM analysis of the overall survival in the TCGA-BRCA, GSE58812-OS and GSE58812-MFS cohorts, (b-d) Risk scores, survival status, and expression distributions of TGF-β risk score in different cohorts, and (e) The ROC curves of the TGF-β risk score in predicting 3, 5, and 10-year survival state of BC patients.

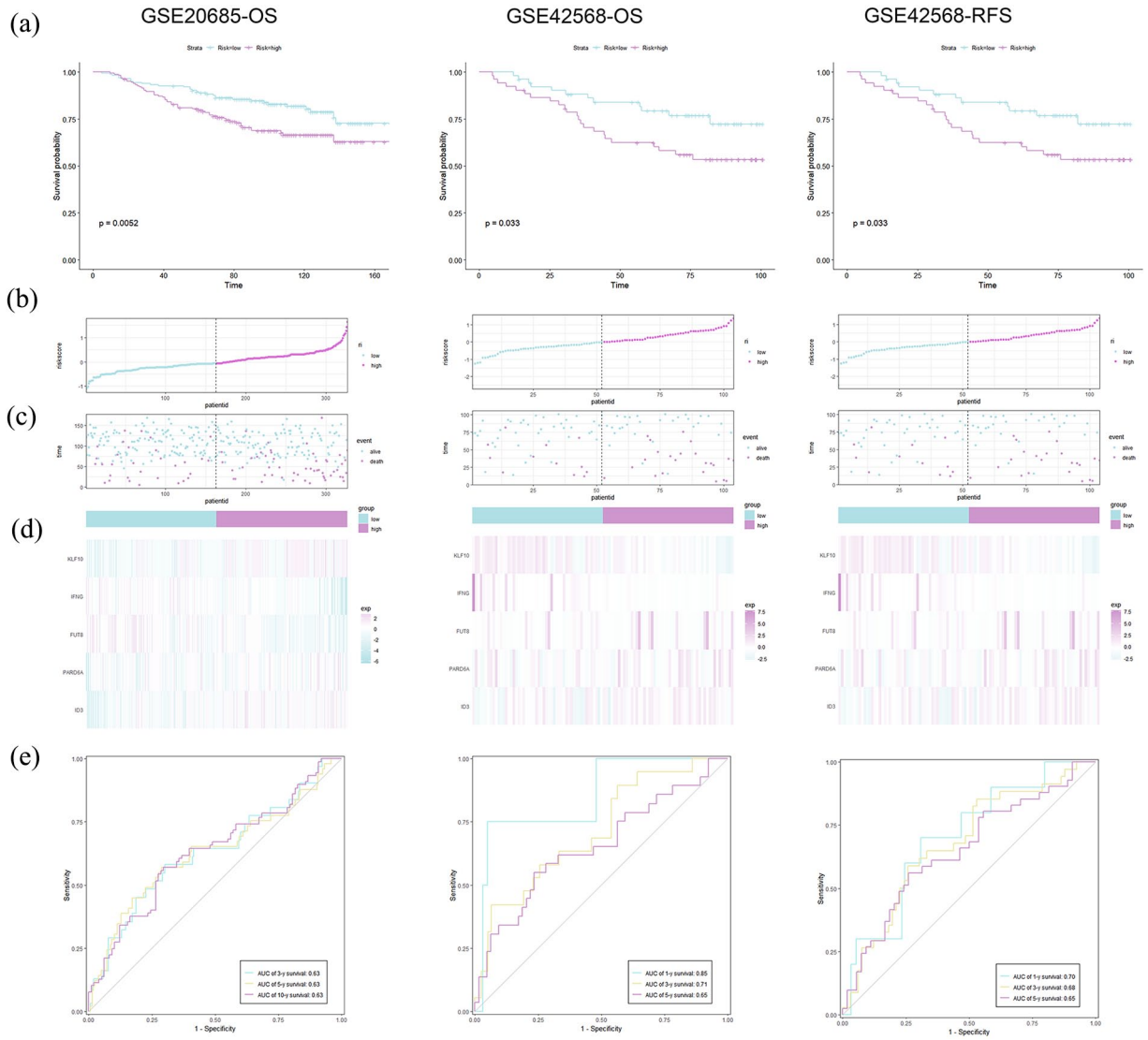


**Figure 4.** Additional validation set for the assessment and verification of the efficiency of the TGF-β risk score: (a) The KM analysis of the overall survival in the GSE20685-OS, GSE42568-OS and GSE42568-RFS cohorts, (b-d) Risk score, survival status, and expression distributions of TGF-β risk score in different cohorts, and (e) The ROC curves of the TGF-β risk score in predicting 3-, 5-, and 10-year OS in the GSE20685 cohort and 1,3,5 years survival state in the GSE42568 cohort.

reference line, thereby affirming the reliability of the constructed nomogram (Figure 5(d)). Additionally, the area under the curve (AUC) for the ROC analyses at 1, 2, and 5 years for the TCGA cohort were recorded at 0.78, 0.72, and 0.75, respectively (Figure 5(e)). These values surpassed the AUCs for the TGF-β risk score model alone, which were 0.66, 0.67, and 0.71 for the 3, 5, and 10-year intervals (Figure 3(e)). Furthermore, the DCA indicated that the nomogram provided a superior net benefit in prognostic prediction for BC patients compared to the use of clinicopathological factors in isolation (Figure 5(f)).

### *Gene set enrichment analysis based on TGF-β risk score model*

The results of the GSEA, conducted to elucidate the relationship between TGF-β and biological signaling pathways, are presented in Figure 6(a) and (b). Subsequent GO analysis revealed that the pathways exhibiting significant upregulation included the immunoglobulin complex, cilium assembly, and cilium organization. Conversely, the pathways that demonstrated significant downregulation encompassed the T cell receptor complex, plasma membrane signaling receptor complex, regulation of cytosolic calcium ion concentration, and cellular calcium ion homeostasis, among others. Furthermore, the KEGG analysis indicated that the high-risk group was notably enriched in pathways related to Herpes simplex virus 1 infection, chemokine signaling, neutrophil extracellular trap formation, neuroactive ligand-receptor interactions, cytokine-cytokine receptor interactions, and natural killer cell-mediated cytotoxicity.

### *Immune landscape analysis in BRCA TME*

The differences in TME between high-risk and low-risk groups were further explored. Firstly, except for PVR, the expression of immune checkpoints was higher in the

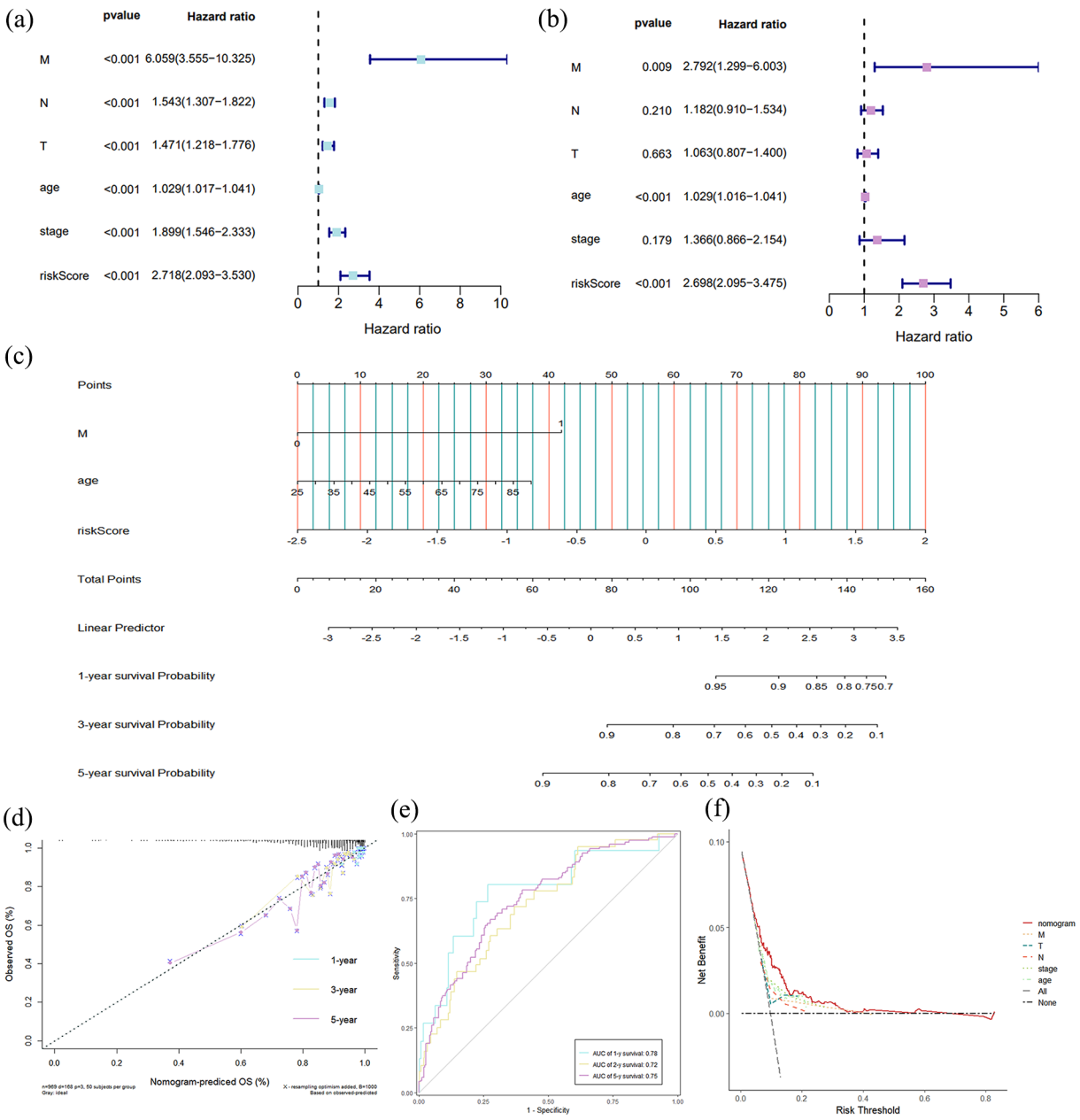


**Figure 5.** Establishment of a prognosis-related nomogram model based on TGF-β risk score: (a) Univariate Cox regression analysis of TGF-β risk score and clinicopathological characteristic, (b) Multivariate Cox regression analysis of TGF-β risk score and clinicopathological characteristic, (c) Development of the nomogram based on M, age and TGF-β risk score, (d) The Calibration plots for the nomogram, (e) The ROC curves and AUC values demonstrated favorable competence of the nomogram in predicting 1-, 2-, and 5-year OS of BC patients, and (f) The decision curve to evaluate the clinical decision effectiveness of the nomogram against other separate clinical parameters.

low-risk group (Figure 6(c)), which was also verified in the GEO dataset (Supplemental Figure S3). The chord plot showed that these six immune checkpoints are closely related to the patient's risk score (Figure 6(d)). Furthermore, Figure 6(e) depicts the link of five TSRGs to six immunological checkpoints, with IFNG showing a substantial association, indicating that it plays a key role in the immune cells of the patient's TME. Simultaneously, IFNG was found to have a substantial link with six immunoinfiltrating cells: CD4 T cells, B cells, CD8 T cells, neutrophils, macrophages, and dendritic cells (Supplemental Figure S4(a)). The differences in 28 types of immune cell infiltration in patients in the high and low risk groups are shown in the Supplemental Figure S4(b). In addition, patients in the high-risk group had higher TIDE scores, suggesting that they were more likely to have immune escape, which was associated with a poor prognosis (Figure 6(f)).

### *Drug sensitivity analysis in BC patients*

IPS values were used to predict how patients would respond to anti-CTLA-4 and anti-PD-1/PD-L1 treatments. The findings are consistent with our immune checkpoint analysis, which found that the low-risk group had higher IPS values,

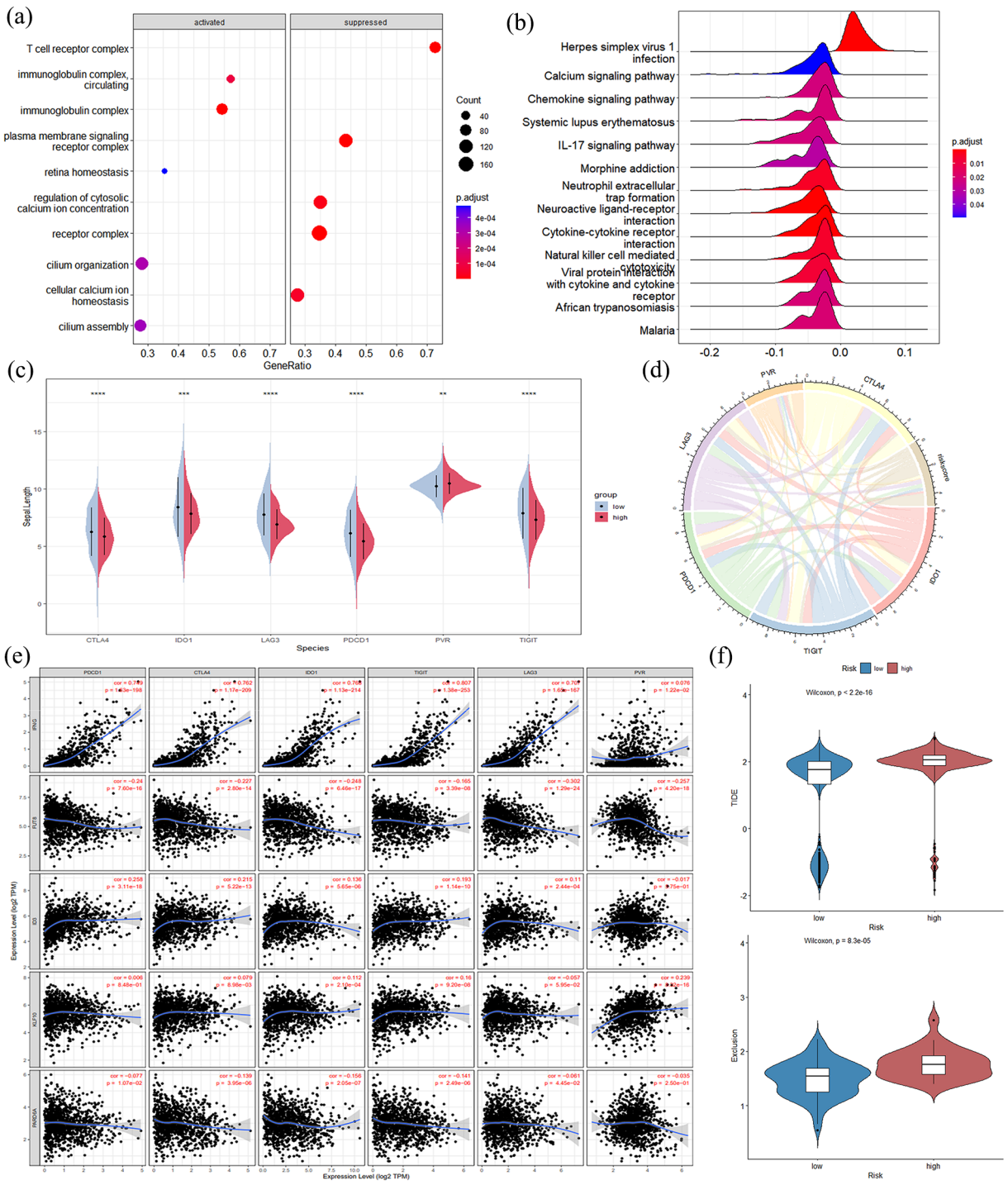


**Figure 6.** (a-b) Results of GO and KEGG in TCGA-BRCA cohort, (c) Expression level of PDCD1, CTLA4, IDO1, TIGIT, LAG3 and PVR in TCGA-BRCA, (d) Correlation chord plot between immune checkpoints and riskscore, (e) Correlation of five TSRGs with six immune cells, and (f) Differences in TIDE and Exclusion scores between high and low risk groups.

indicating that they may be more sensitive to ICIs treatment (Figure 7(a) and (b)). Furthermore, IC50 was used to estimate the response of BC patients to 198 chemotherapy drugs. The lower the IC50 value, the better the drug's effect. Chemotherapy drugs with a good treatment response in both high-risk and low-risk patients were chosen, and the results are shown in Figure 7(c) and (d). These medications include: Buparlisib, Dihydrorotenone, PD0325901, Rapamycin, Dactinomycin, AZD8055, CDK9.1, BMS.754807, Eg5, Epirubicin, GNE.317, Sabutoclax, BI.2536, and Rapamycin.

### *Description of TGF-β subtypes in BC*

The R package "ConsensusClusterPlus" was used to create consistent clustering of five genes for the TGF-β risk score model. When BC patients were divided into two clusters,

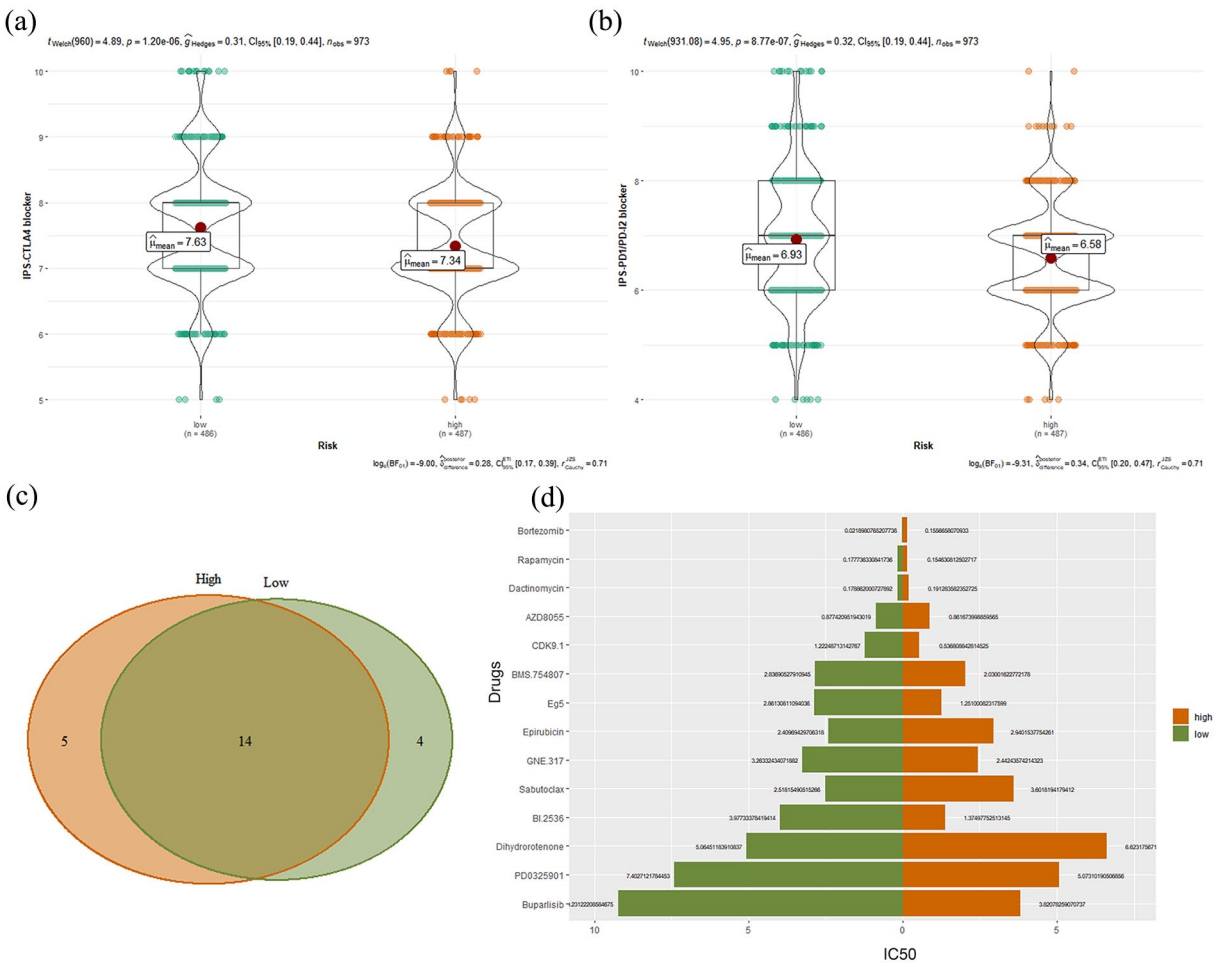


**Figure 7.** (a) Relative probability of response to CTLA-4 treatment in low-risk and high-risk groups, (b) Relative probability of response to PD-1/PD-L2 treatment in low-risk and high-risk groups, (c) A venn diagram of 14 drugs sensitive to BC patients, and (d) A bar plot of IC50 values for 14 drugs sensitive to BC patients.

there was good consistency and stability within subgroups (Figure 8(a)-(d)). The PCA results revealed that the two clusters of patients were clustered separately, confirming the reliability of the clustering results (Figure 8(e)). The survival analysis revealed that cluster A had a better prognosis than cluster B ($P = .0081$) (Figure 8(f)). A heat map also depicted the expression of five TSRGs between the two clusters (Figure 9(a)).

The immunoinfiltration landscape in TME in two subgroups was also examined. Cluster B had higher matrix, immunity, and ESTIMATE scores as determined by the ESTIMATE algorithm (Figure 9(b)). Furthermore, with the exception of CD56 dim natural killer cells, which were more abundant in cluster A, the majority of the remaining immune cells were expressed more strongly in cluster B (Figure 9(c)). In addition, the pathways that GSEA analysis showed differences between the two clusters were mainly antigen binding, immunoglobulin complex immunoglobulin complex,circulating (Figure 9(d)). Surprisingly, the expression of immune checkpoints was higher in Cluster B than in Cluster A, indicating that they may be more suitable for ICIs (Figure 9(e)).

### *Single-cell RNA-seq analysis*

Following quality control, 40,467 cells were kept for further investigation. UMAP is further subdivided into 14 cell clusters consisting of six different cells: HSCs, Erythroid cells, CD8+T cells, CD4+T cells, Eosinophils, Monocytes (Figure 10(a) and (b)). IFNG is mainly distributed in CD8+T cells, while KLF10, ID3, PARD6A are mainly distributed in HSCs and Erythroid cells (Figure 10(c)).

## Discussion

In recent years, the survival rates and prognoses of BC patients have significantly improved, largely due to the clinical implementation of ICIs. Nevertheless, it is important to note that not all patients exhibit favorable responses to ICIs, with a considerable proportion demonstrating either resistance or insensitivity to these therapies.[26] Research indicates that the TGF-β signaling pathway within the TME plays a critical role in immune evasion and resistance to ICIs in cancer. Consequently, the inhibition of TGF-β has been proposed as a strategy to enhance the efficacy of ICIs.[27,28] Furthermore, TGF-β is

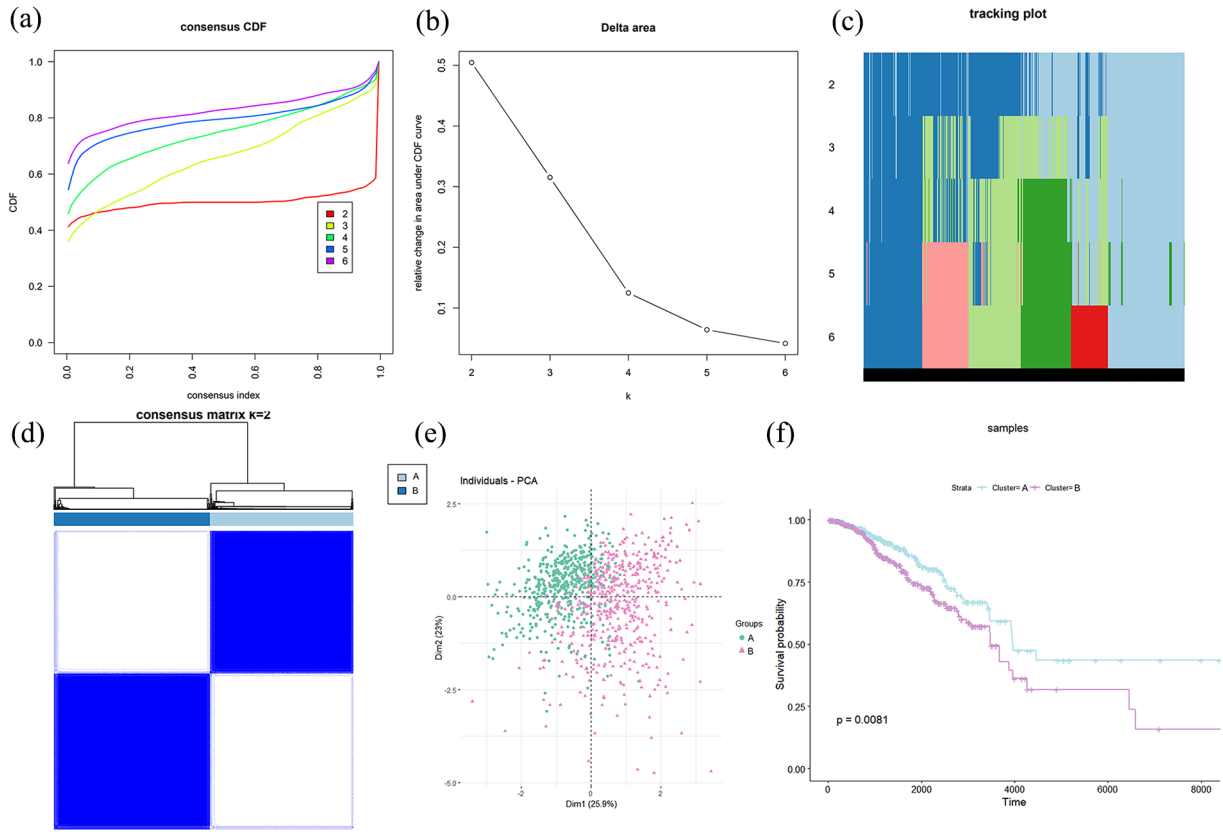


**Figure 8.** (a) The CDF value of consensus index, (b) Relative change in area under CDF curve for $k=2$–6, (c) The tracking plot for $k=2$ to $k=6$, (d) Consensus matrix for $k=2$, (e) Principal component analysis of the total RNA expression profile, and (f) KM curve of the survival difference between cluster A and cluster B.

implicated in the promotion of BC metastasis through the induction of epithelial-mesenchymal transition (EMT), which is another significant factor contributing to poor prognoses in patients.[29] In summary, the TGF-β signaling pathway is integral to the initiation, progression, and treatment of BC; however, its interactions with other elements within the TME and its implications for the prognosis of BC patients warrant further exploration.

In this study, genes associated with prognostic OS were identified through univariate Cox regression analysis utilizing the TCGA-BRCA dataset. These prognostic OS-related genes were subsequently cross-referenced with 225 TSRGs. Following this, LASSO regression analysis and multivariate Cox regression analysis were conducted on the intersecting genes, ultimately leading to the identification of an optimal prognostic model comprising five genes: FUT8, IFNG, ID3, KLF10, and PARD6A.

Fucosyltransferase 8 (FUT8) is frequently upregulated in various cancers due to its role in promoting core fucosylation, a process that is crucial for cancer cell metastasis and immune evasion.[30] Through core fucosylation, FUT8 not only facilitates the dissemination of BC cells but also enhances TGF-β signaling and EMT processes.[31] Numerous studies have demonstrated that the inhibition of FUT8 can decelerate the progression of BC and improve patient outcomes.[32,33] Similarly, PARD6A is associated with EMT during the invasion of BC cells.[34] The gene IFNG, which encodes interferon-gamma (IFN-γ), is vital for anti-tumor immunity, as it can be released by tumor-infiltrating immune cells (TIICs) to exert anti-tumor effects and can also induce the expression of PD-L1, thereby improving patient outcomes.[35,36] KLF10, a Krüppel-like zinc finger transcription factor, has been shown to play a significant role in inducing apoptosis via the Smad signaling pathway in response to TGF-β.[37] Additionally, ID3 has been implicated in promoting the proliferation and invasion of human MCF-7 breast cancer cells.[38] Notably, in our study, KLF10 was identified as a risk factor, while FUT8 and PARD6A were classified as protective factors. To further elucidate the underlying reasons for these findings, the expression levels of these five genes were examined in both normal and BC tissues. The results corroborate previous studies, indicating that FUT8 and PARD6A are more highly expressed in tumor tissues, whereas KLF10 exhibits greater expression in normal tissues. Therefore, it can be concluded that the observed results across different BC risk groups may be influenced by the molecular subtypes of BC, leading to variations in outcomes.

This study examined the potential association between TGF-β risk scores and the TME of BC patients. Initially, it was observed that patients with low TGF-β risk scores exhibited increased infiltration of gamma delta T cells (γδT cells)

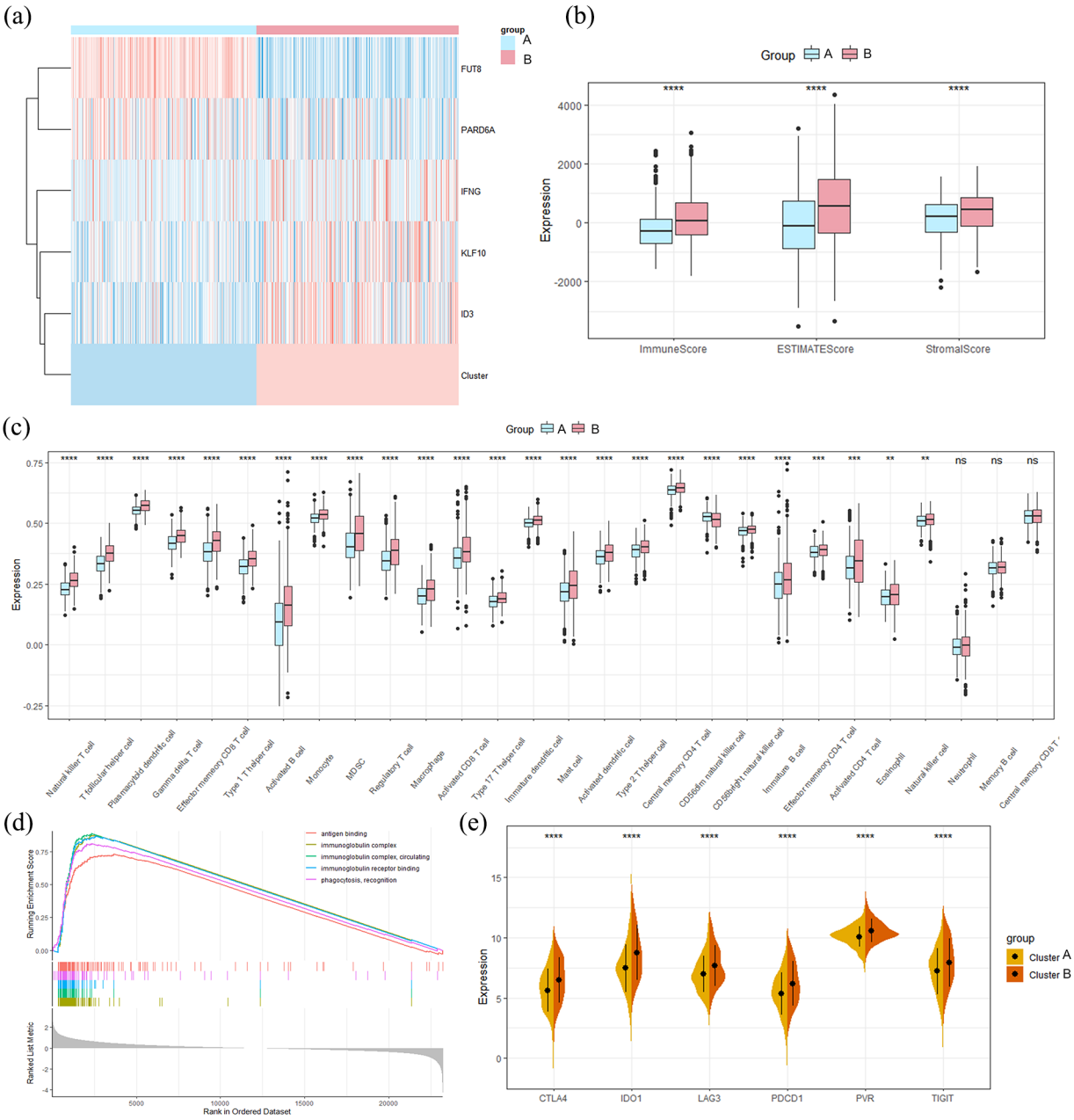


**Figure 9.** (a) Heatmap of the five genes between the two clusters, (b) Variance of immune score, stromal score, and ESTIMATE score in the two clusters, (c) Boxplots were used to depict the discrepancies in the infiltration extent of 28 immune cells between the two clusters, (d) GSEA analysis between Cluster A and Cluster B, and (e) Expression level of PDCD1, CTLA4, IDO1, TIGIT, LAG3 and PVR in the cluster A and cluster B groups.

and T follicular helper cells (Tfh cells) within the TME. This finding was consistent when comparing normal and tumor tissues. γδT cells play a pivotal role in both innate and adaptive immunity,[39] as they can recognize tumor cells independently of major histocompatibility complex (MHC) antigens and induce tumor cell death through potent anti-inflammatory and cytotoxic mechanisms.[40] Tfh cells enhance antitumor immunity in a CD8+-dependent manner and are significant effector cells in anti-PD-1/PD-L1 therapies.[41] Subsequently, the study assessed the correlation between TGF-β risk scores and six immune checkpoints within the TME, revealing a negative correlation across all datasets analyzed. Consequently, patients classified in the low TGF-β risk group were more likely to experience favorable outcomes from ICI treatments, a finding corroborated by comparisons using TIDE and IPS between the two groups. Furthermore, the prevailing understanding of IFNG suggests that it can augment the efficacy of anti-PD-1 therapy.[42] The study found a positive correlation between IFNG expression and several immune checkpoints, including It is important to note that IFNG, predominantly produced by CD8+ T cells, is integral to T cell immunity and cancer immunotherapy. The IFNγ produced by CD8+ T cells downregulates the expression of SLC3A2 and SLC7A11, leading to decreased cystine uptake by tumor cells, which in turn induces lipid peroxidation and ferroptosis.[43]

The analysis simulated the differential treatment responses to various drugs between high- and low-risk groups,

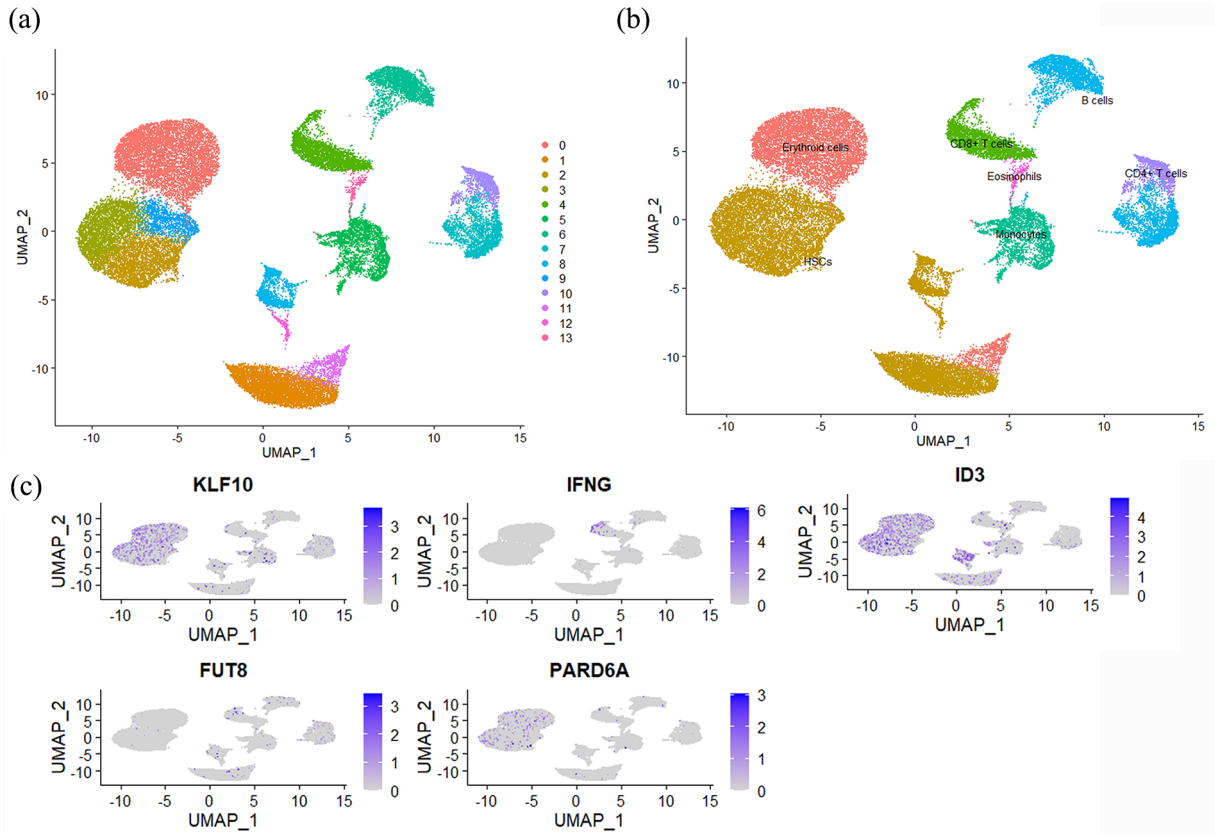


**Figure 10.** (a) The results of the reduced dimensionality clustering analysis are shown in the UMAP plot, (b) Annotate cells into 2 different types of cells, and (c) Expression distribution of FUT8, IFNG, ID3, KLF10, and PARD6A.

suggesting that low-risk patients exhibit greater sensitivity to ICIs. Additionally, it was predicted that targeted chemotherapy agents would be beneficial for both high- and low-risk groups to enhance survival in BC. Bortezomib, a proteasome inhibitor that impedes myeloma proliferation and exerts anabolic effects on bone, is a critical treatment for BC bone metastases.[44] Although the relationship between IFNG and bortezomib is unclear, bortezomib has been used in combination with other medicines in patients with amyloid light-chain (AL), and transcriptome analysis has revealed CD8 memory T cell activation and IFNG overexpression.[45] Everolimus, a derivative of rapamycin, inhibits BC cell proliferation and aggressiveness through the PI3K/AKT/mTOR signaling pathway.[46] Specifically, IFNG released by natural killer cells is inhibited by mTOR inhibitors.[47] To address tamoxifen resistance, the mTORC1/2 dual inhibitor (AZD8055) can downregulate HSPB8.[48] BMS.754807, by blocking the insulin-like growth factor 1 receptor (IGF1R), inhibits BC cell proliferation and enhances their sensitivity to the chemotherapeutic agent cisplatin.[49] Furthermore, sabutoclax effectively targets cancer stem cells in BC by inhibiting the IL-6/STAT3 signaling pathway.[50] Currently, there is insufficient evidence to support the efficacy of other drugs, such as CDK9.1, BI.2536, and Dihydrorotenone, in the treatment of BC, necessitating further investigation into their relationship with the disease.

The present study acknowledges several limitations. Firstly, this study was based on public databases, and the expression levels of five TSRG genes were limited by the study's objective experimental conditions and were not clinically verified, despite the fact that FUT8, IFNG, ID3, and KLF10 were all confirmed in other clinical breast cancer sample studies. Secondly, due to variations in the data within the existing database, the receiver operating characteristic (ROC) curve analysis in the GSE42568 validation set was limited to evaluating the model's predictive capabilities at 1, 3, and 5 years, rather than at 3, 5, and 10 years. Thirdly, further research is warranted to refine the classification and typing criteria for BC.

## Conclusion

In conclusion, this study presents a robust risk model for BC patients based on TSRGs. The model indicates that TGF-β risk scores can effectively characterize the TME in BC patients and align them with various sensitive therapeutic agents. Moreover, TGF-β has the potential to delineate molecular subtypes of BC, thereby paving the way for enhanced clinical patient classification and treatment strategies.

## Acknowledgements

The construction of the training set here are in whole based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga

## Author Contributions

JQ: Conceptualization, Formal analysis, Writing - original draft, Funding acquisition; MHW: Methodology, Software; YHG: Visualization; HWZ: Supervision, Writing - review & editing. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript.

## Data Availability

All data included in this study are available upon request by contact with the corresponding author.

## Ethics Approval and Consent to Participate

Not applicable.

## Consent for publication

Not applicable.

## ORCID iD

Hua-Wei Zhang https://orcid.org/0000-0002-0641-4680

## Supplemental Material

Supplemental material for this article is available online.

## REFERENCES


1. Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. Breast cancer. *Lancet*. 2021;397(10286):1750-1769. doi:10.1016/s0140-6736(20)32381-3
2. Redig AJ, McAllister SS. Breast cancer as a systemic disease: a view of metastasis. *J Int Med*. 2013;274(2):113-26. doi:10.1111/joim.12084
3. Charoentong P, Finotello F, Angelova M, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep*. 2017;18(1):248-262. doi:10.1016/j.celrep.2016.12.019
4. Mao YJ, Lim HJ, Ni M, Yan WH, Wong DW, Cheung JC. Breast tumour classification using ultrasound elastography with machine learning: a systematic scoping review. *Cancers*. 2022;14(2)doi:10.3390/cancers14020367
5. Roma-Rodrigues C, Mendes R, Baptista PV, Fernandes AR. Targeting tumor microenvironment for cancer therapy. *Int J Mol Sci*. 2019;20(4)doi:10.3390/ijms20040840
6. Byrne A, Savas P, Sant S, et al. Tissue-resident memory T cells in breast cancer control and immunotherapy responses. *Nat Rev Clin Oncol*. 2020;17(6):341-348. doi:10.1038/s41571-020-0333-y
7. Harao M, Forget MA, 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. doi:10.1158/2326-6066.Cir-16-0364
8. Baxevanis CN, Fortis SP, Perez SA. The balance between breast cancer and the immune system: challenges for prognosis and clinical benefit from immunotherapies. *Sem Cancer Biol*. 2021;72:76-89. doi:10.1016/j.semcancer.2019.12.018
9. Batlle E, Massagué J. Transforming growth factor-β Signaling in immunity and cancer. *Immunity*. 2019;50(4):924-940. doi:10.1016/j.immuni.2019.03.024
10. Li X, Xiang Y, Li F, Yin C, Li B, Ke X. WNT/β-Catenin signaling pathway regulating T cell-inflammation in the tumor microenvironment. *Front Immunol*. 2019;10:2293. doi:10.3389/fimmu.2019.02293
11. Massagué J. TGFbeta in cancer. *Cell*. 2008;134(2):215-30. doi:10.1016/j.cell.2008.07.001
12. Seoane J, Gomis RR. TGF-β family signaling in tumor suppression and cancer progression. *Cold Spring Harbor Perspect Biol*. 2017;9(12)doi:10.1101/cshperspect.a022277
13. Kim DH, Kim WD, Kim SK, Moon DH, Lee SJ. TGF-β1-mediated repression of SLC7A11 drives vulnerability to GPX4 inhibition in hepatocellular carcinoma cells. *Cell Death Dis*. 2020;11(5):406. doi:10.1038/s41419-020-2618-6
14. Yang L, Pang Y, Moses HL. TGF-beta and immune cells: an important regulatory axis in the tumor microenvironment and progression. *Trends Immunol*. 2010;31(6):220-7. doi:10.1016/j.it.2010.04.002
15. Chida K, Kawazoe A, Suzuki T, et al. Transcriptomic profiling of MSI-H/dMMR gastrointestinal tumors to identify determinants of responsiveness to anti-PD-1 therapy. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2022;28(10):2110-2117. doi:10.1158/1078-0432.Ccr-22-0041
16. Yi M, Li T, Niu M, Wu Y, Zhao Z, Wu K. TGF-β: A novel predictor and target for anti-PD-1/PD-L1 therapy. *Front Immunol*. 2022;13:1061394. doi:10.3389/fimmu.2022.1061394
17. Majidpoor J, Mortezaee K. The efficacy of PD-1/PD-L1 blockade in cold cancers and future perspectives. *Clin Immunol*. 2021;226:108707. doi:10.1016/j.clim.2021.108707
18. Panagi M, Voutouri C, Mpekris F, et al. TGF-β inhibition combined with cytotoxic nanomedicine normalizes triple negative breast cancer microenvironment towards anti-tumor immunity. *Theranostics*. 2020;10(4):1910-1922. doi:10.7150/thno.36936
19. Amerizadeh F, Bahrami A, Khazaei M, et al. Current status and future prospects of transforming growth factor-β as a potential prognostic and therapeutic target in the treatment of breast cancer. *Journal of cellular biochemistry*. 2019;120(5):6962-6971. doi:10.1002/jcb.27831
20. Liu Z, Qi T, Li X, et al. A Novel TGF-β risk score predicts the clinical outcomes and tumour microenvironment phenotypes in bladder cancer. *Front Immunol*. 2021;12:791924. doi:10.3389/fimmu.2021.791924
21. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Gen Res*. 2018;28(11):1747-1756. doi:10.1101/gr.239244.118
22. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Software*. 2010;33(1):1-22.
23. Zhang Z, Kattan MW. Drawing Nomograms with R: applications to categorical outcome and survival data. *Ann Translat Med*. 2017;5(10):211. doi:10.21037/atm.2017.04.01
24. Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief Bioinform*. 2021;22(6)doi:10.1093/bib/bbab260
25. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010;26(12):1572-3. doi:10.1093/bioinformatics/btq170
26. Yang Z, Sun JK, Lee MM, Chan MK. Restoration of p53 activity via intracellular protein delivery sensitizes triple negative breast cancer to anti-PD-1 immunotherapy. *J Immuno Cancer*. 2022;10(9):68. doi:10.1136/jitc-2022-005068
27. Bai X, Yi M, Jiao Y, Chu Q, Wu K. Blocking TGF-β signaling To enhance the efficacy of immune checkpoint inhibitor. *OncoTargets Therapy*. 2019;12:9527-9538. doi:10.2147/ott.S224013
28. Wu ZH, Li N, Gao ZZ, et al. Development of the novel bifunctional fusion protein BR102 that simultaneously targets PD-L1 and TGF-β for anticancer immunotherapy. *Cancers*. 2022;14(19):964. doi:10.3390/cancers14194964
29. Yu Y, Luo W, Yang ZJ, et al. miR-190 suppresses breast cancer metastasis by regulation of TGF-β-induced epithelial-mesenchymal transition. *Mol Cancer*. 2018;17(1):70. doi:10.1186/s12943-018-0818-9
30. Bastian K, Scott E, Elliott DJ, Munkley J. FUT8 Alpha-(1,6)-Fucosyltransferase in Cancer. *Int J Mol Sci*. 2021;22(1):455. doi:10.3390/ijms22010455
31. Tu CF, Wu MY, Lin YC, Kannagi R, Yang RB. FUT8 promotes breast cancer cell invasiveness by remodeling TGF-β receptor core fucosylation. *Breast Cancer Res BCR*. 2017;19(1):111. doi:10.1186/s13058-017-0904-8
32. Satoh M, Iida S, Shitara K. Non-fucosylated therapeutic antibodies as next-generation therapeutic antibodies. *Expert Opin Biol therapy*. 2006;6(11):1161-73. doi:10.1517/14712598.6.11.1161
33. Huang Y, Zhang HL, Li ZL, et al. FUT8-mediated aberrant N-glycosylation of B7H3 suppresses the immune response in triple-negative breast cancer. *Nat Commun*. 2021;12(1):2672. doi:10.1038/s41467-021-22618-x
34. Viloria-Petit AM, David L, Jia JY, et al. A role for the TGFbeta-Par6 polarity pathway in breast cancer progression. *Proceed Nat Acad Sci USA*. 2009;106(33):14028-14033. doi:10.1073/pnas.0906796106
35. Gao Y, Yang J, Cai Y, et al. IFN-γ-mediated inhibition of lung cancer correlates with PD-L1 expression and is regulated by PI3K-AKT signaling. *Int J Cancer*. 2018;143(4):931-943. doi:10.1002/ijc.31357
36. Shihab I, Khalil BA, Elemam NM, et al. Understanding the role of innate immune cells and identifying genes in breast cancer microenvironment. *Cancers*. 2020;12(8). 2226. doi:10.3390/cancers12082226
37. Subramaniam M, Hawse JR, Rajamannan NM, Ingle JN, Spelsberg TC. Functional role of KLF10 in multiple disease processes. *BioFactors*. 2010;36(1):8-18. doi:10.1002/biof.67
38. Chen YH, Wu ZQ, Zhao YL, Si YL, Guo MZ, Han WD. FHL2 inhibits the Id3-promoted proliferation and invasive growth of human MCF-7 breast cancer cells. *Chin Med J*. 2012;125(13):2329-33.
39. Song X, Wei C, Li X. Association between γδ T cells and clinicopathological features of breast cancer. *Int Immunopharmacol*. 2022;103:108457. doi:10.1016/j.intimp.2021.108457

40. Morrow ES, Roseweir A, Edwards J. The role of gamma delta T lymphocytes in breast cancer: a review. *Translat Res J Lab Clin Med*. 2019;203:88-96. doi:10.1016/j.trsl.2018.08.005
41. Niogret J, Berger H, Rebe C, et al. Follicular helper-T cells restore CD8(+)-dependent antitumor immunity and anti-PD-L1/PD-1 efficacy. *J Immunotherapy Cancer*. 2021;9(6)doi:10.1136/jitc-2020-002157
42. Yu M, Peng Z, Qin M, et al. Interferon-γ induces tumor resistance to anti-PD-1 immunotherapy by promoting YAP phase separation. *Mol Cell*. 2021;81(6):1216-1230.e9. doi:10.1016/j.molcel.2021.01.010
43. Wang W, Green M, Choi JE, et al. CD8(+) T cells regulate tumour ferroptosis during cancer immunotherapy. *Nature*. 2019;569(7755):270-274. doi:10.1038/s41586-019-1170-y
44. Suvannasankha A, Chirgwin JM. Role of bone-anabolic agents in the treatment of breast cancer bone metastases. *Breast Cancer Res BCR*. 2014;16(6):484. doi:10.1186/s13058-014-0484-9
45. Wang Y, Xu L, Zhao W, et al. T cell landscape and dynamics in immunoglobulin light chain amyloidosis before and after daratumumab-based therapy. *Clin Translat Med*. 2021;11(11):e582. doi:10.1002/ctm2.582
46. Du L, Li X, Zhen L, et al. Everolimus inhibits breast cancer cell growth through PI3K/AKT/mTOR signaling pathway. *Mol Med Reports*. 2018;17(5):7163-7169. doi:10.3892/mmr.2018.8769
47. Yao B, Yang Q, Yang Y, et al. Screening for active compounds targeting human natural killer cell activation identifying daphnetin as an enhancer for IFN-γ production and direct cytotoxicity. *Front Immunol*. 2021;12:680611. doi:10.3389/fimmu.2021.680611
48. Shi JJ, Chen SM, Guo CL, Li YX, Ding J, Meng LH. The mTOR inhibitor AZD8055 overcomes tamoxifen resistance in breast cancer cells by down-regulating HSPB8. *Acta Pharmacol Sin*. 2018;39(8):1338-1346. doi:10.1038/aps.2017.181
49. O'Flanagan CH, O'Shea S, Lyons A, et al. IGF-1R inhibition sensitizes breast cancer cells to ATM-related kinase (ATR) inhibitor and cisplatin. *Oncotarget*. 2016;7(35):56826-56841. doi:10.18632/oncotarget.10862
50. Hu Y, Yagüe E, Zhao J, et al. Sabutoclax, pan-active BCL-2 protein family antagonist, overcomes drug resistance and eliminates cancer stem cells in breast cancer. *Cancer Lett*. 2018;423:47-59. doi:10.1016/j.canlet.2018.02.036