



A Novel Signature Integrated of Immunoglobulin, Glycosylation and Anti-Viral Genes to Predict Prognosis for Breast Cancer

Shengshan Xu^{1,2†}, Yuchen Liu^{2*†}, Hansu Ma², Shuo Fang³, Shoupeng Wei², Xiaoping Li⁴, Zhuming Lu¹, Youbin Zheng⁵, Tong Liu⁶, Xiaojian Zhu², Dongming Xu^{7*} and Yihang Pan^{2*}

OPEN ACCESS

Edited by:

Ruowen Zhang,
Jiahehongsheng (Shenzhen) Health
Industry Group, China

Reviewed by:

Udo Jeschke,
Ludwig Maximilian University of
Munich, Germany
Zhu Wang,
Shenzhen Longhua People's Hospital,
China

*Correspondence:

Yuchen Liu
liuych83@mail.sysu.edu.cn
Dongming Xu
xudm0807@163.com
Yihang Pan
panyih@mail.sysu.edu.cn

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Cancer Genetics and Oncogenomics,
a section of the journal
Frontiers in Genetics

Received: 13 December 2021

Accepted: 14 March 2022

Published: 01 April 2022

Citation:

Xu S, Liu Y, Ma H, Fang S, Wei S, Li X,
Lu Z, Zheng Y, Liu T, Zhu X, Xu D and
Pan Y (2022) A Novel Signature
Integrated of Immunoglobulin,
Glycosylation and Anti-Viral Genes to
Predict Prognosis for Breast Cancer.
Front. Genet. 13:834731.
doi: 10.3389/fgene.2022.834731

¹Department of Thoracic Surgery, Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen, China, ²Scientific Research Center, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China, ³Department of Oncology, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China, ⁴Department of Breast, Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen, China, ⁵Department of Radiology, Jiangmen Wuyi Hospital of Traditional Chinese Medicine, Jiangmen, China, ⁶Department of Thoracic Surgery, First Hospital of Jilin University, Changchun, China, ⁷Department of Neurosurgery, The Country Hospital of Qianguo, Songyuan, China

Background: Aberrant glycosylation is significantly related to the occurrence, progression, metastasis, and drug resistance of tumors. It is essential to identify glycosylation and related genes with prognostic value for breast cancer.

Objective: We aimed to construct and validate a prognostic model based on glycosylation and related genes, and further investigate its prognosis values in validation set and external independent cohorts.

Materials and Methods: The transcriptome and clinical data of breast cancer patients were downloaded from The Cancer Genome Atlas (TCGA, $n = 1072$), Molecular Taxonomy of Breast Cancer International Consortium (METABRIC, $n = 1451$), and GSE2741 ($n = 120$). Glycosylation-related genes were downloaded from the Genecards website. Differentially expressed glycosylation-related genes were identified by comparing the tumor tissues with the adjacent tissues. The TCGA data were randomly divided into training set and validation set in a 1:1 ratio for further analysis. The glycosylation risk-scoring prognosis model was constructed by univariate and multivariate Cox regression analysis, followed by confirmation in TCGA validation, METABRIC, and GEO datasets. Gene set enrichment analysis (GSEA) and Gene ontology analysis for identifying the affected pathways in the high- and low-risk groups were performed.

Results: We attained 1072 breast cancer samples from the TCGA database and 786 glycosylation genes from the Genecards website. A signature contains immunoglobulin, glycosylation and anti-viral related genes was constructed to separate BRCA patients into two risk groups. Low-risk patients had better overall survival than high-risk patients ($p < 0.001$). A nomogram was constructed with risk scores and clinical characteristics. The area under time-dependent ROC curve reached 0.764 at 1 year, 0.744 at 3 years, and 0.765 at 5 years in the training set. Subgroup analysis showed differences in OS between the high- and low-risk patients in different subgroups. Moreover, the risk score was

confirmed as an independent prognostic indicator of BRCA patients and was potentially correlated with immunotherapy response and drug sensitivity.

Conclusion: We identified a novel signature integrated of immunoglobulin (IGHA2), glycosylation-related (SLC35A2) and anti-viral gene (BST2) that was an independent prognostic indicator for BRCA patients. The risk-scoring model could be used for predicting prognosis and immunotherapy in BRCA, thus providing a powerful instrument for combating BRCA.

Keywords: breast cancer, gene signature, prognosis mode, immunotherapy, glycosylation

INTRODUCTION

Globally, breast cancer is the most common cancer among women, and is the leading cause of cancer deaths among women aged 20 to 59 (R. L. Siegel et al., 2021). The cumulative risk of breast cancer is about 5% in women, and the risk of death is 1.4% (E. Hadadi et al., 2020). In recent years, the incidence of breast cancer has continued to increase by about 0.5% annually (R. L. Siegel et al., 2021), which seriously affects women's health and quality of life. Breast cancer is a highly heterogeneous disease. The main treatments for breast cancer include systemic therapy (chemotherapy, endocrine therapy, targeted therapy, and immunotherapy) and local treatment (surgery and radiotherapy). Breast cancer molecular subtypes were essential indicators for treatment and prognosis. Currently, the majority of patients diagnosed with a specific breast cancer subtype receive the same treatment, even though it has been repeatedly proven that they should adopt differential strategies (S. A. Eccles et al., 2013; S. Fallahpour et al., 2017; W. J. Gradishar et al., 2020). Triple-negative breast cancer (TNBC) is a type of breast cancer that lacks expression of human epidermal growth factor receptor 2 (HER2), progesterone receptor, and estrogen receptor (G. Bianchini et al., 2016). Typically, the prognosis for women with TNBC following metastatic recurrence is much poorer than other subtypes (M. Smid et al., 2008). It is a heterogeneous disease representing about 15% of total breast cancer incidents, which is difficult to treat as lack of available targeted therapies. Chemotherapy remains to be the preferred systemic treatment for TNBC (D. P. Silver et al., 2010), and particularly in those carrying BRCA1 mutations (J. Collignon et al., 2016). This demonstrates the reliable predictive biomarkers are necessary for precise diagnosis and individualized treatment for breast cancer patients, and the precision medicine progress have been fueled by the continuous development of new sequencing and computational technologies.

Genetic and epigenetic alterations are considered the primary causes of cancer development, and the downstream phenotypic changes at the protein level are amongst the driving forces (A. Peixoto et al., 2019). Glycosylation is the most common and complicated post-translational modification for membrane-bound proteins. More than 50–70% of proteins in the circulation are glycosylated, which play important roles in various cellular activities, such as cell growth, differentiation, transformation, and adhesion (Ohtsubo and Marth, 2006). Aberrant glycosylation has been identified as a hallmark of cancer and intimately correlated with cancer occurrence, progression, metastasis, tumor recurrence, and drug resistance (Pinho and Reis, 2015; Rao et al., 2017; A.

Chakraborty et al., 2018; Cui et al., 2018). There was also a correlation between glycosylation and antitumor immunity. For example, Freire et al. demonstrated that Tn glycosylation of the MUC6 protein strongly affects the immunogenicity of its B and T cell, and might enable immune escape of tumor cells (T. Freire et al., 2011). Bone marrow stromal antigen 2 (BST2) was a type II transmembrane protein, also known as tethered protein, HM1.24 or CD317. BST2 homodimer promotes cancer cell adhesion and enhances cancer cell survival and growth by enhancing proteasomal degradation of pro-apoptotic proteins (Mahauad-Fernandez and Okeoma, 2017). Sayeed et al. indicated that aberrant BST2 overexpression promoted the disappearance of TGF β -mediated tumor-suppressive effects in breast cancer as a consequence and the ensuing loss of the differentiation program (A. Sayeed et al., 2013). Mahauad Fernandez and Okeoma suggested that BST-2 targets breast cancer cells that are resistant to anoikis via the GRB2/ERK/BIM/Cas3 pathway. Almost all breast tumors express BST2 to a certain level, and the high expression level of BST2 was related to progressive malignancies (W. D. Mahauad-Fernandez et al., 2014). Several studies have shown that BST2 regulates the occurrence of gastric cancer, oral squamous cell carcinoma, lung cancer and is involved in tumor metastasis and invasion (W. Wang et al., 2009a; K. H. Fang et al., 2014; W. Liu et al., 2018; W. Liu et al., 2020). Immunoglobulin heavy constant alpha 2 (IGHA2), located on chromosome 14, expressed in breast cancer cells and upregulated in advanced breast tumor tissues by comparison with early tumors, was involved in the early stage of the tumor microenvironment remodeling and has been identified as a marker of regional metastasis in lymph nodes. Suki Kang et al. reported that IGH2 might protect the cells against physiological stresses during the neoplastic process and promote tumor growth in the advanced stages of cancer (S. Kang et al., 2012). Exploring the role of glycosylation, immunoglobulin, and anti-viral in BRCA and related molecules will help us investigate strategies to combat for BRCA. Currently, there is still no study exploring the prognosis value of glycosylation, immunoglobulin, and anti-viral -related genes. Therefore, this study aims to establish a prognostic model for breast cancer based on glycosylation-related genes and evaluate it from multiple dimensions.

MATERIALS AND METHODS

Data Source and Processing

Breast cancer datasets were downloaded from the Molecular Taxonomy of Breast Cancer International Consortium

(METABRIC, <http://www.cbioportal.org>), The Cancer Genome Atlas (TCGA, <https://www.cancer.gov>), and the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo>). Patients who met the following selection criteria were included: 1) histologically diagnosed with malignant breast cancer; 2) available RNA expression data; and 3) available OS data. After screening, this study included 1,451 patients from METABRIC, 1,072 patients from TCGA, and 120 patients from GSE2741. Glycosylation-related genes were downloaded from the website of Genecards (<https://www.genecards.org>). A total of 786 glycosylation genes were analyzed by comparing the tumor tissues with the adjacent tissues to obtain differentially expressed glycosylation-related genes (DEGRGs). A *p* value less than 0.05 was considered statistically significant.

Construction and Validation of the Risk-Scoring Model

The TCGA BRCA data were randomly assigned into training set and validation set according to the ratio of 1:1. In order to determine the survival-related glycosylation genes, we performed univariate Cox regression, in which *p* < 0.05 was set as a cut-off criterion. Subsequently, multivariate Cox regression was performed to construct a prognostic risk-scoring model, in which the risk score for each patient was calculated according to the following formula:

$$RiskScore = [(0.9138 \times Expression\ value\ of\ SLC35A2) + [(-0.2483) \times Expression\ value\ of\ BST2] + [(-0.1002) \times Expression\ value\ of\ IGHA2]]$$

All BRCA patients were assigned to the high- and low-risk groups according to the median risk score in the training set. The difference in OS between these two groups was investigated by the log-rank test and Kaplan-Meier survival analysis. In addition, the distributions of survival status, OS, and risk score in the training set were also plotted.

Independent Prognostic Analysis

To evaluate the relationship between clinicopathological factors and risk scores on survival time, we used the “Survival” R package to perform univariate and multivariate Cox regression. The time-dependent receiver operator characteristic (ROC) curve was drawn, and the R package “timeROC” was applied to determine the prognostic performance of either clinicopathological factors or risk

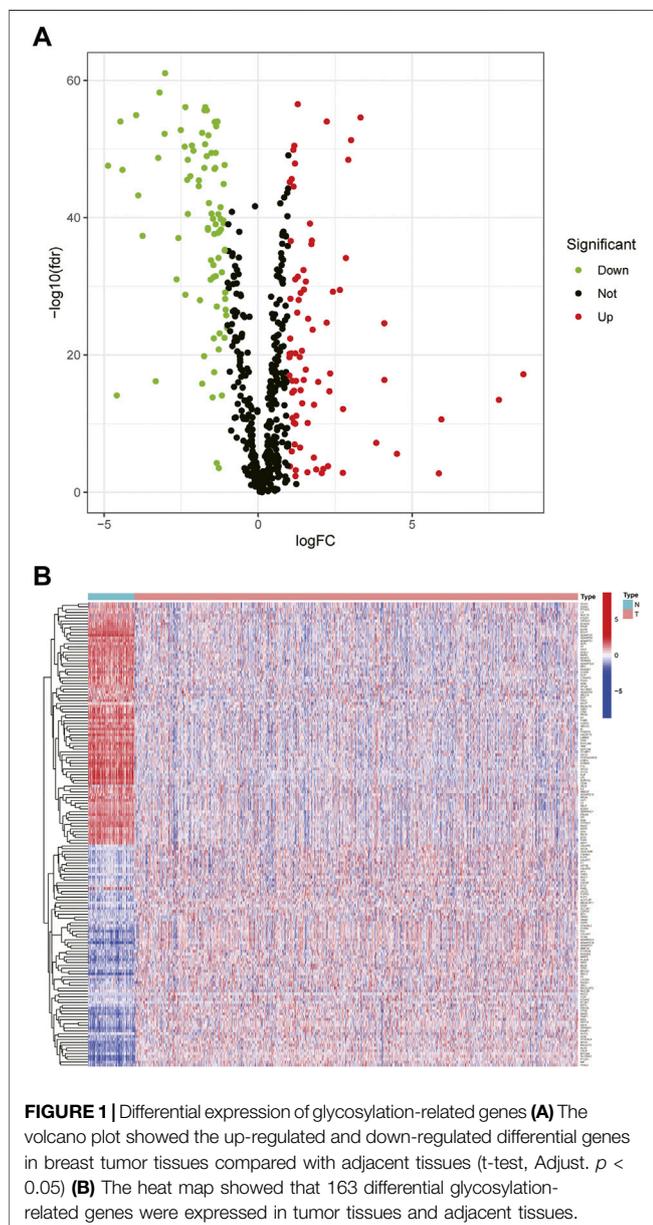
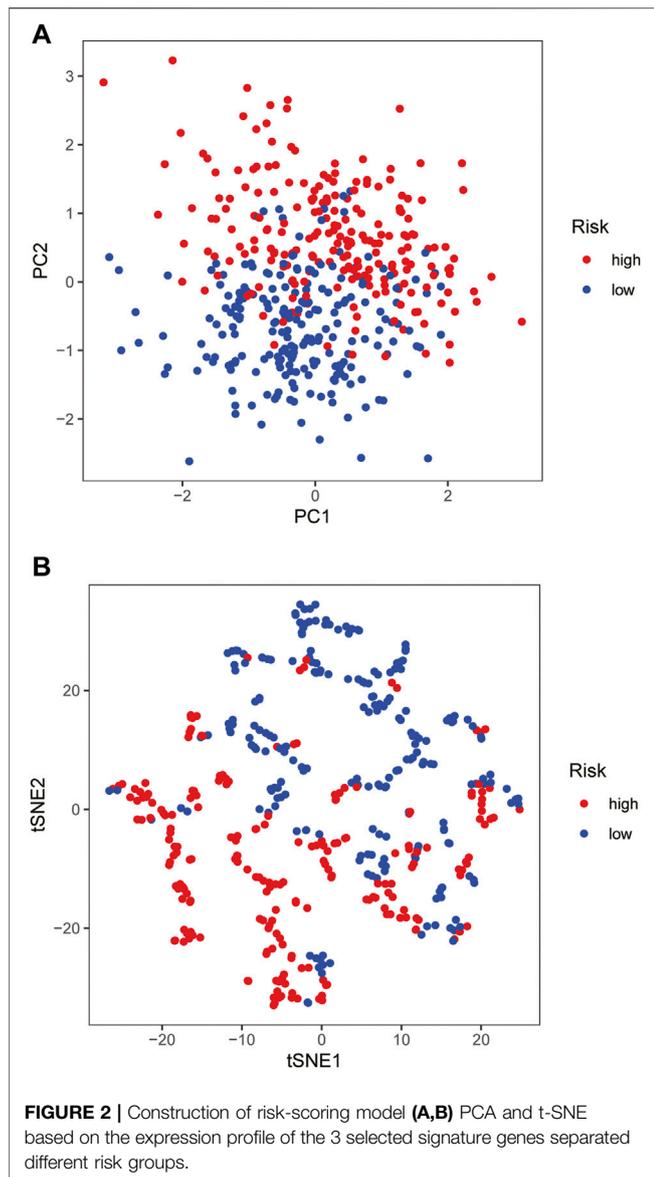


TABLE 1 | The association between risk score and patients’ clinical features in the training set.

Variables	Training Set		Validation Set		<i>p</i> -Value
	(n = 432)		(n = 429)		
	No.	%	No.	%	
Age					0.986
≤65	319	73.8	318	74.1	
>65	113	26.2	111	25.9	
Stage					0.870
I	72	16.7	79	18.4	
II	254	58.8	251	58.5	
III	97	25.5	92	21.5	
IV	9	2	7	1.6	
T stage					0.622
T1	111	25.7	115	26.8	
T2	253	58.6	258	60.1	
T3	51	11.8	45	10.5	
T4	17	3.9	11	2.6	
N stage					0.504
N0	218	50.5	201	46.9	
N1	141	32.6	147	34.2	
N2	51	11.8	50	11.7	
N3	22	5.1	31	7.2	
M stage					0.812
M0	423	97.9	422	98.4	
M1	9	2.1	7	1.6	



scores on survival time. $p < 0.05$ was considered statistically significant.

Subgroup Analysis

We further evaluated the model's predictive ability by stratifying patients into various subgroups. These variables include age (≤ 65 and > 65 years), tumor stage (I-II and III-IV), T stage (T1-2 and T3-4), N stage (N0 and N1-3), M stage (M0 and M1), estrogen receptor (ER) state (positive and negative), human epidermal growth factor receptor (HER-2) state (positive and negative), triple-negative breast cancer (yes and no). To further confirm the value of the risk score in TNBC and HER-2 positive subgroups, the TNBC and HER-2 positive samples from METABRIC database were used for re-verification. Through clinical survival analysis, the predictive ability of the risk-scoring model in various clinical subgroups was clarified. $p < 0.05$ was considered statistically significant.

Exploration of the Value of the Risk-Scoring Model in Clinical Utility

To further improve the practical value of the risk-scoring model, a nomogram was constructed by integrating age, tumor stage, T stage, N stage, M stage, and risk score to predict the OS of patients at 1, 3, and 5 years. In addition, the C-index was used to measure the accuracy of the nomogram, and the calibration curve was drawn to evaluate the calibration of the model. The ROC curves of various clinical characteristics were drawn, and the AUC was calculated to judge the performance of the prognosis model. In addition, decision curve analysis (DCA) was used to estimate the maximum clinical benefit by logistic regression analysis.

Functional Enrichment Analysis

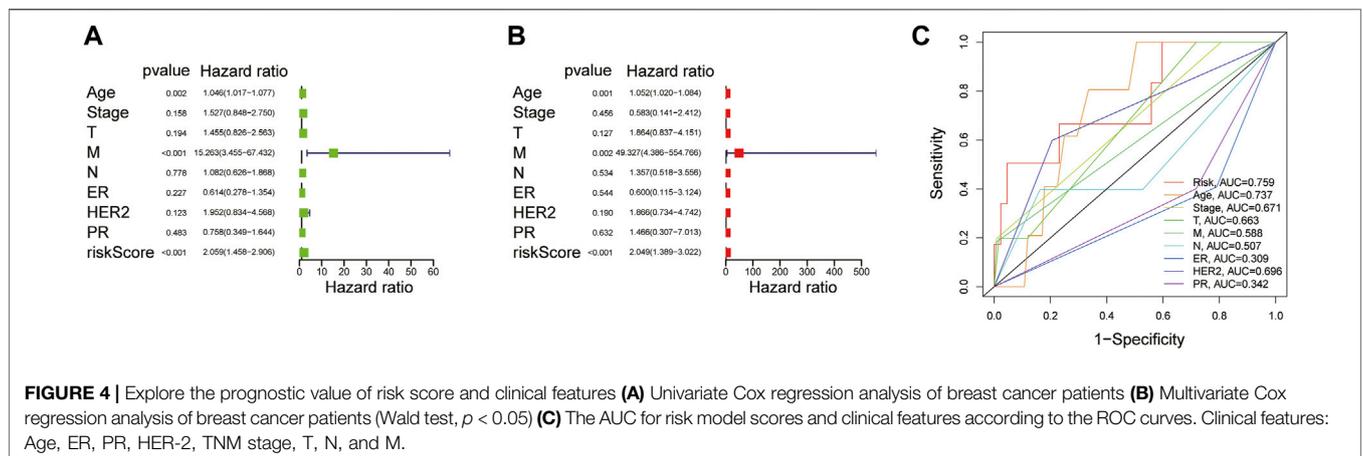
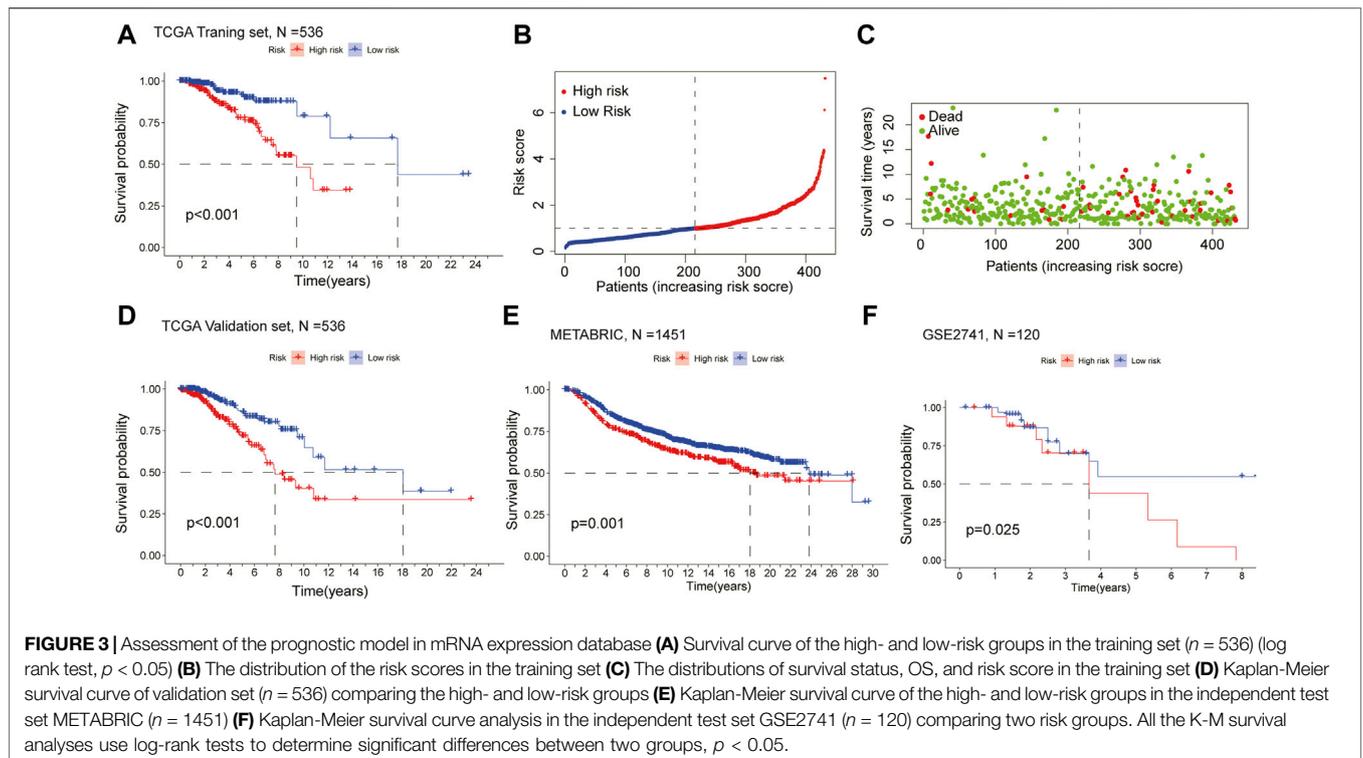
To explore the difference molecular pathways underlying survival prognosis between the high- and low-risk groups, we used the “clusterProfiler” R package to perform Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. Bubble Diagram visualized the most important pathways in KEGG and each GO category. Gene Enrichment Score Analysis (GSEA) was also applied to determine signaling pathways regulated in patients of the high- and low-risk groups. $p < 0.05$ and $FC > 1.2$ (or $FC < 0.83$) were set as the cut-off values.

Analysis of Tumor Immune Microenvironment

The tumor microenvironment score of each single BRCA patient was estimated using the ESTIMATE algorithm. CIBERSORT algorithm was utilized to evaluate the proportion of 22 human immune cell subsets in the high- and low-risk groups. The “GSVA” R package was applied to perform single-sample gene set enrichment analysis (ssGSEA) to quantify the GSVA scores of the 13 immune signatures. The difference between the expression levels of immune checkpoints in the high- and low-risk groups was further evaluated. p -value < 0.05 was considered statistically significant.

Tumor Mutation Burden Analysis

After obtaining the somatic mutation data in the TCGA BRCA dataset, the “maftools” R package was applied to analyze the tumor mutation burden (TMB) of the training set, TMB value was calculated, and a waterfall chart was drawn. Then it would be assessed whether the TMB scores were related to the risk scores and patient survival probability. We found the median value of TMB and divided TCGA BRCA dataset into high TMB group and low TMB group. Patients were stratified into the new groups by integrating TMB and risk scores, and Kaplan-Meier survival analysis was used. The “Survival” and the “survminer” R packages were used for joint survival analysis. The CBIOPORTAL database (<https://www.cbiportal.org/>) was used to analyze the mutations profiles in the high-and low-risk groups. Finally, the protein domains where the mutations were located were clarified. $p < 0.05$ was considered statistically significant.



Relationship Between Risk Scores and Immunotherapy

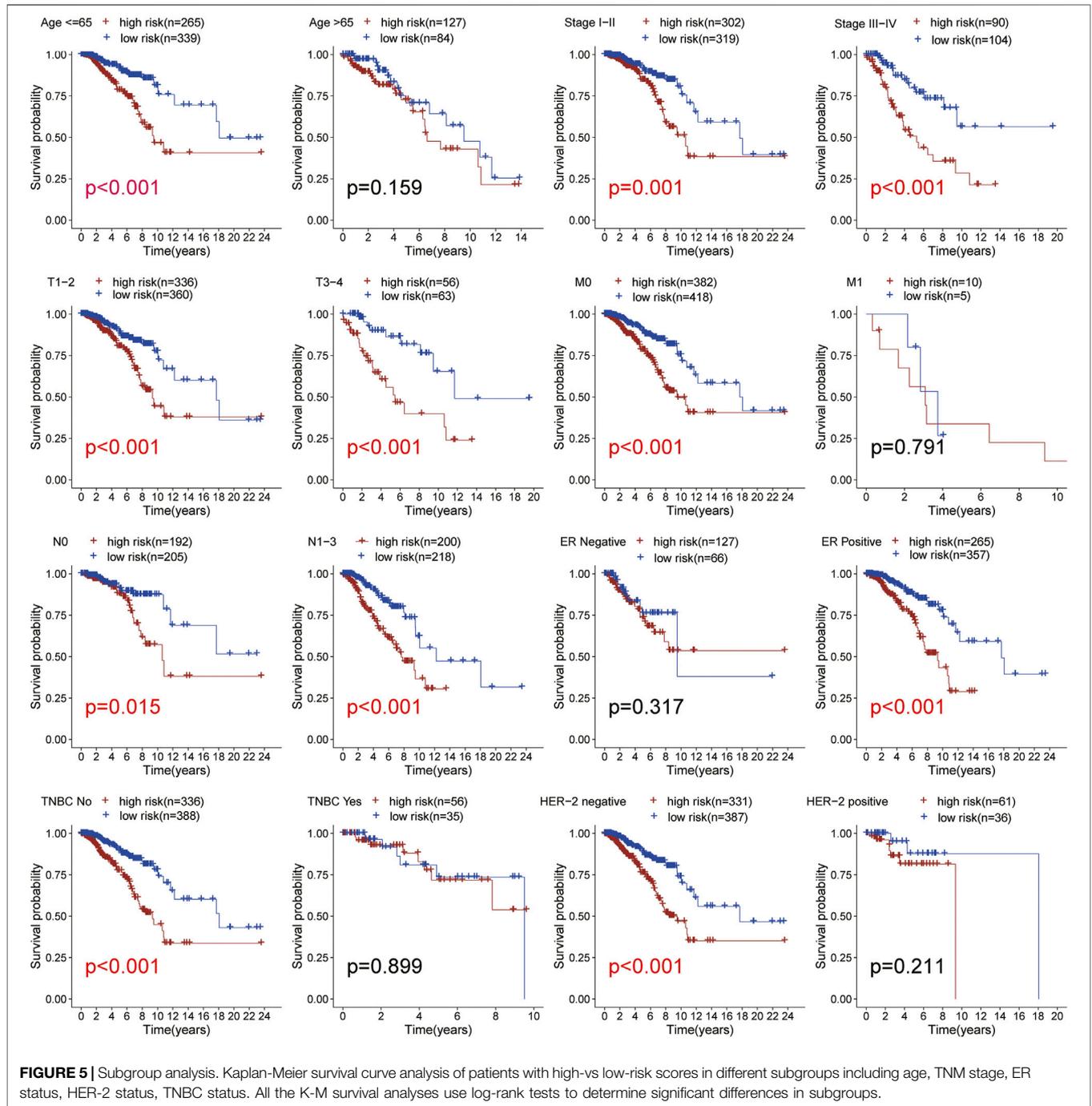
IMvigor210, a phase II trial of atezolizumab (MPDL3280A) in platinum-treated locally advanced or metastatic urothelial carcinoma. It was downloaded to assess the correlation between risk scores and immunotherapy response. The tumor immune dysfunction and exclusion (TIDE) score was calculated online (<http://tide.dfc.harvard.edu/>) to assess the immune checkpoint inhibitor response between the high- and low-risk groups (Z. Lin et al., 2021). We downloaded RNA-seq and compound activity: DTP NCI-60 through the CellMiner database (<https://discover.nci.nih.gov/cellminer>) and excluded FDA status as empty or clinical trial data to explore the relationship between the expression of genes and

drug sensitivity. The three-dimensional structures of drugs were obtained through the PubChem database (<https://pubchem.ncbi.nlm.nih.gov>). $p < 0.05$ was considered statistically significant.

RESULTS

Identification of Differential Expression of Glycosylation-Related Genes in Breast Cancer

We compiled the gene expression data of breast cancer from the TCGA database and finally got 1,072 tumor tissues and 99

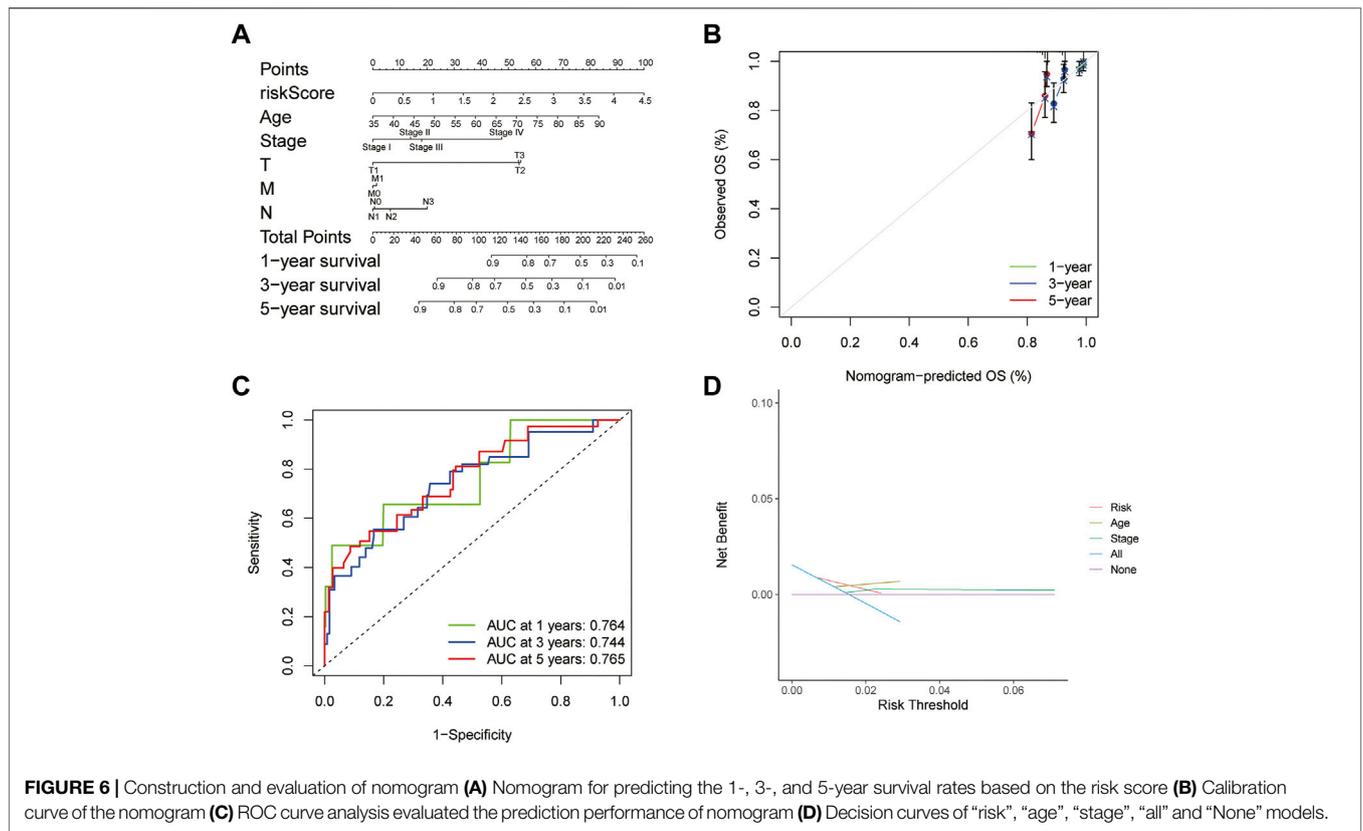


adjacent tissue samples. Around 786 glycosylation-related genes were obtained by the Genecards website with a correlation >2.0 as the cut-off value. Based on standard cut-off values for fold-change in gene expression ($|\log(FC)| > 1$) and false discovery rate (FDR <0.05), the breast cancer tissues had 163 DEGRGs, with 86 down-regulated genes and 77 up-regulated genes (Figures 1A,B).

Construction of a Risk-Scoring Model

The samples screened from the TCGA database were randomly assigned to the training set for the construction of the model and

the validation set for accuracy estimation according to the ratio of 1:1. The clinical features of all patients are shown in detail in Table 1. There were no statistically significant differences in clinical features between patients in the training set and validation set. Based on the training data set, the prognosis-related glycosylation genes were screened by univariate Cox regression ($p < 0.05$), including two low-risk genes [hazard ratio (HR) < 1] and one high-risk gene [risk ratio (HR) > 1]. Then, multivariate Cox analysis was performed to screen three genes related to the glycosylation with prognostic significance



($p < 0.05$), namely *BST2*, *IGHA2*, and *SLC35A2*. These three genes were used to construct a risk-scoring model. According to the risk score formula and the median risk score, the patients with breast cancer were divided into the high- and low-risk groups. The PCA and t-SNE showed that the high- and low-risk groups had different distribution directions, suggesting that the risk-scoring model could clearly divide BRCA patients into two groups (Figures 2A,B).

Evaluation and Validation of the Aberrant Glycosylation-Related Risk-Scoring Model

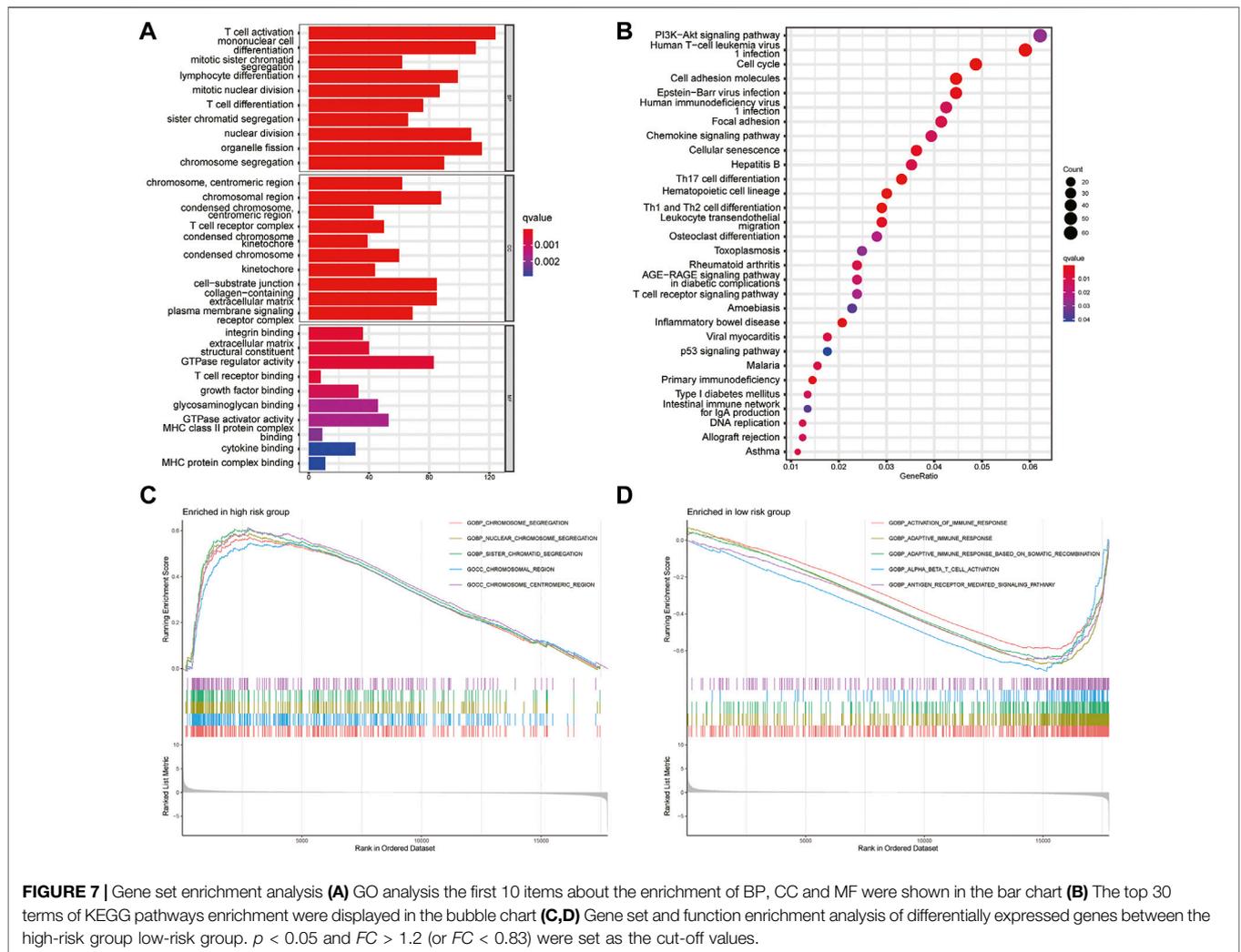
Kaplan-Meier survival analysis showed that the OS of the high-risk group was lower than that of the low-risk group in the training cohort ($p < 0.001$) (Figure 3A), demonstrating the excellent predictive value of the risk-scoring model in the training set. The risk curves showed the survival status and risk scores of each breast cancer sample, which was calculated and ranked based on the signature model (Figure 3B). The scatter plot represented the OS status of BRCA patients according to the risk score, suggesting that the higher the risk scores were, the higher the number of death was (Figure 3C). Similarly, with Kaplan-Meier survival analysis using the validation cohort and external independent test sets, the OS rate of the high-risk group was lower than that of the low-risk group ($p < 0.05$), confirming that the risk-scoring model had a robust prognostic value (Figures 3D-F).

Explore the Independent Prognostic Factors of Breast Cancer

To evaluate whether the established risk-scoring model is an independent prognostic factor for breast cancer, univariate and multivariate Cox regression was performed. The HR of the risk scores and 95% CI were 2.059 and 1.458–2.906 ($p < 0.001$), respectively, in univariate Cox regression analysis (Figure 4A), which were 2.049 and 1.389–3.022 ($p < 0.001$) in multivariate Cox regression analysis, respectively (Figure 4B). The result demonstrated that the risk score was a significant prognostic factor independent of multiple clinicopathological parameters such as the expression level of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor (HER-2) as well as the M stage, N stage, T stage, tumor stage. Additionally, compared with other clinicopathological factors, the AUC of the risk score for 1-year OS shown by ROC analysis reached 0.759, which was superior to other clinicopathological variables (Figure 4C). In summary, it can be concluded that the aberrant glycosylation-related risk scoring model is a significant independent prognostic factor for BRCA patients.

Subgroup Analysis

We used the TCGA BRCA samples to verify the relationship between the risk score and the prognosis of clinical features and emphasize the molecular heterogeneity of BRCA. After comparing a statistical difference in OS between the two risk



groups with Kaplan Meier survival curve, the results showed that the OS of the high-risk group was lower than that of the low-risk group among age ≤ 65 , ER positive, HER-2 negative, M0, N0, N1-N3, stage I-II, stage III-IV, T1-2, T3-4, and non-TNBC subgroups ($p < 0.05$; **Figure 5**). In addition, the low-risk group had higher OS in the TNBC and HER-2 positive subgroups from METABRIC (**Supplementary Figure S1A,S2**). The results suggested that the risk score is closely related to the clinical features of BRCA and can be used as an effective auxiliary tool to predict the BRCA prognosis.

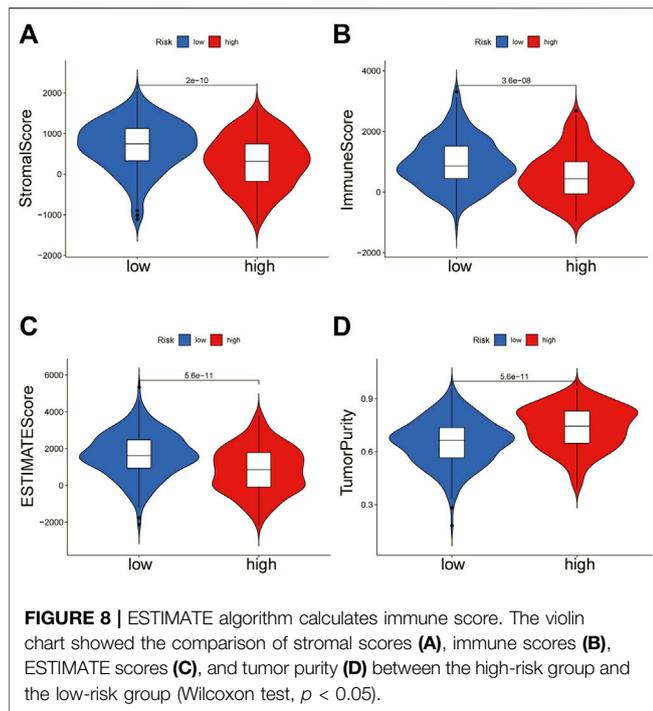
Clinical Evaluation Ability of the Risk-Scoring Model

A predictive nomogram based on the integration of risk score, pathological stage, and age integration was established in the TCGA cohort (**Figure 6A**). The C-index of the nomogram was 0.676 in the TCGA cohort, indicating that the nomogram had a good predictive performance. The calibration curve analysis verified that the nomogram was reliable and accurate, which demonstrated that the predictive probability of 1-, 3- and 5-year

OS was ideally consistent with actual observation (**Figure 6B**). With the analysis of the ROC curve, the AUC value was 0.764, 0.744, and 0.765 for 1-, 3-, and 5-year OS, respectively, indicating that the risk-scoring model had excellent predictive accuracy (**Figure 6C**). In addition, DCA analysis was performed to evaluate the predictive value of the nomogram in clinical decision-making (**Figure 6D**). The above all indicated that the risk-scoring model and nomogram had high reliability.

Gene Set and Function Enrichment Analysis

To analyze the pathways related to the risk-scoring model, GO and KEGG enrichment analysis were performed. GO analysis showed that the differential genes between the high- and low-risk groups were enriched in T cell activation, T cell receptor binding, integrin binding, and nuclear division (**Figure 7A**). KEGG enrichment showed the differential genes between two groups related to PI3K-Akt signaling pathways, cell adhesion molecules, and leukocyte differentiation (**Figure 7B**). GSEA was further used to investigate the key signaling pathways in different risk groups (**Figures 7C,D**). The results revealed that chromosome segregation was enriched in the high-risk group while T cell



activation, immune response, antigen receptor-mediated signaling pathway were enriched in the low-risk group. These findings explained the poor survival in the high-risk group and may help us gain insight into the implication of aberrant glycosylation-related signature.

Tumor Immune Microenvironment of BRCA

To further explain the difference in survival between the two groups, we explored the relationship between glycosylation and the tumor immune microenvironment. Stromal scores, immune scores, and ESTIMATE scores were evaluated by the ESTIMATE package, all of which were higher in the low-risk group ($p < 0.001$; Figures 8A–C). In contrast, patients in the high-risk group were associated with significantly higher tumor purity ($p < 0.001$) (Figure 8D). After the proportion of 22 immune cell types was calculated by CIBERSORT, there were significant differences in the infiltration scores of nine immune cells between the two groups (Figure 9A), including B cell naive, plasma cell, T cell CD8+, T cell CD4+ memory resting, T cell gamma delta, Macrophage M0, Macrophage M2, mast cell resting, and T cell follicular helper ($p < 0.05$). Subsequently, we used Spearman Correlation Analysis to explore the relationship between risk score and immune cell infiltration. The results showed that the low-risk group had a higher infiltration level of B cells naive, mast cells resting, plasma cells, T cells CD4 memory resting, T cells CD8, and T cells gamma delta ($p < 0.05$), while higher macrophages M0, macrophages M2 and T cells follicular helper were found in the high-risk group ($p < 0.05$) (Supplementary Figure S3). The ssGSEA was performed to quantify the enrichment scores of 13 immune cell-related functions between the two risk groups. The results showed that the scores of APC co-inhibition, CCR, Checkpoint, cytolytic activity, HLA, Inflammation-promoting, Para inflammation, T cells co-stimulation, Type I IFN response, Type

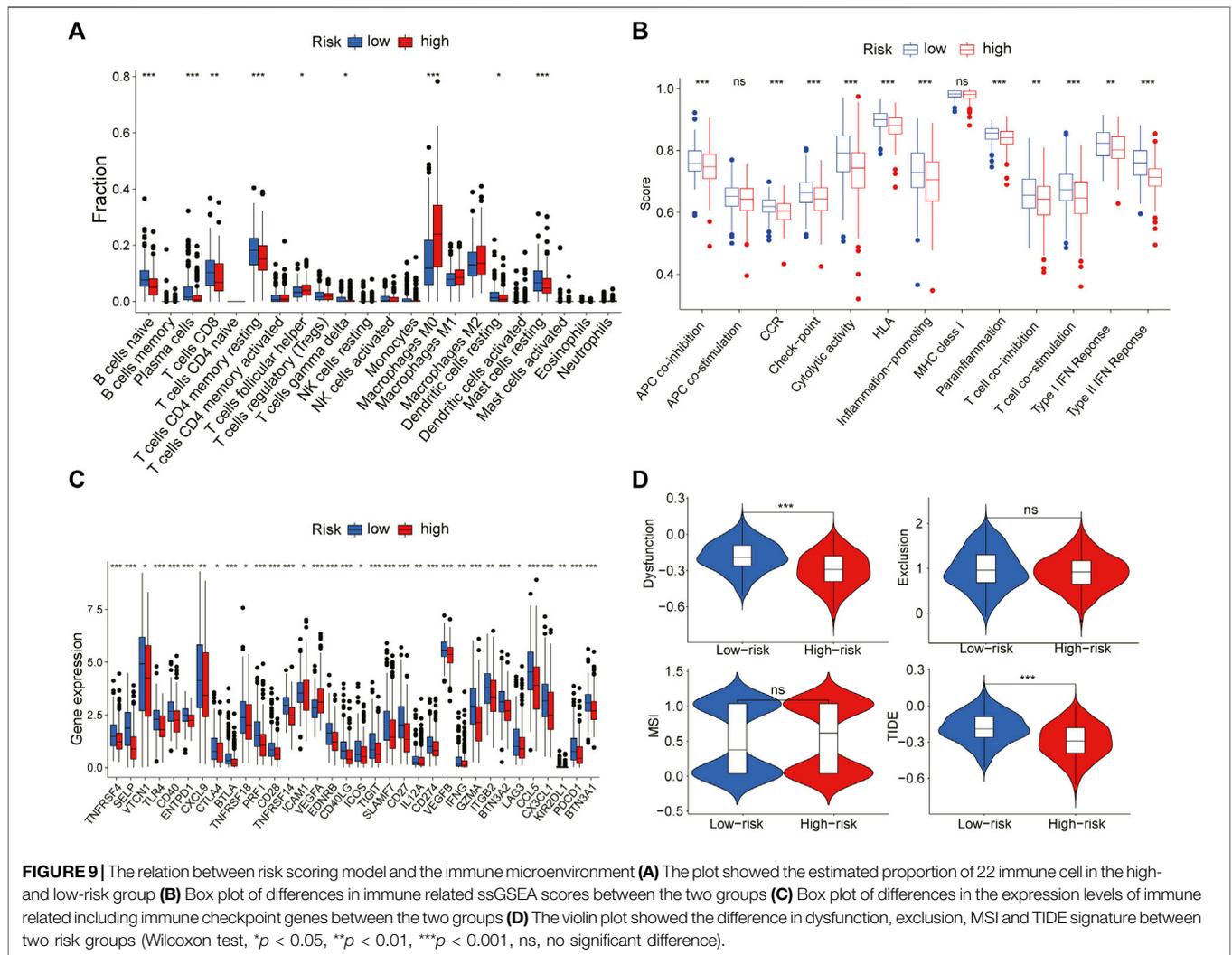
II IFN response of patients in the low-risk group were higher ($p < 0.01$, Figure 9B), indicating that the low-risk group had higher immune infiltration than the high-risk group did. We further explored the immune checkpoints, and the result showed that the distribution of immune checkpoint-related molecule expression was significantly different between the high- and low-risk groups. The difference analysis confirmed that the expression of 22 immune promoting checkpoints (TNFRSF4, SELP, TLR4, CD40, ENTPD1, CXCL9, TNFRSF18, PRF1, CD28, TNFRSF14, ICAM1, CD40LG, ICOS, CD27, IL12A, IFNG, GZMA, ITGB2, BTN3A2, CCL5, CX3CL1, BTN3A1) in the low-risk group was higher, indicating that the prognosis of the low-risk group was better than that of the high-risk group ($p < 0.05$) (Figure 9C), that may provide the potential targets of immunotherapy for BRCA patients. In conclusion, the low-risk group had higher Tumor infiltrating lymphocytes (TILs), of which the prognosis was better, compared with the high-risk group.

Genomic Mutation Analysis

Through the downloaded somatic mutation data, the mutation frequency of the high-risk group and the low-risk group was calculated, and the waterfall chart was drawn to confirm the difference in the distribution of somatic mutations. It found that 169 of 198 (85.35%) BRCA samples in the high-risk group and 165 of 191 (86.39%) BRCA samples in the low-risk group displayed genetic mutations, and missense mutation was the most common variant classification. Moreover, in the high-risk group, PIK3CA had high genetic alterations (25%), which was just junior to the genetic alterations of TP53 (43%). In the low-risk group, PIK3CA had the most genetic alterations (41%) (Figures 10A,B). We also found that the TMB of patients in the high-risk group was significantly higher than that in the low-risk group ($p < 0.001$, Figure 10C), indicating that BRCA patients in the high-risk groups may derive good outcomes from immune checkpoint inhibitor treatments. The combined survival analysis showed that the prognosis of low-risk and low TMB patients was significantly better than that of high-risk and high TMB patients ($p < 0.01$, Figure 10D). Mutations of the genes for constructing the risk-scoring model and the genes with higher mutation frequency in the high-risk group showed that mutations in *SLC35A2* and *BST2* were mainly related to gene amplification, while mutations in *IGHA2*, *TP53*, and *TTN* were mainly related to missense mutations. By the CBIOPORTAL database, the domain where the mutations were located was defined, including *SLC35A2* mutations were localized in the nucleotide sugar transporter domain, *IGHA2* mutations were enriched in the immunoglobulin domain, the majority of *TP53* mutations were centralized in the P53 DNA binding domain, and *TTN* mutations were localized in the immunoglobulin I-set domain, fibronectin type III domain, protein kinase domain, titin Z, and PPAK motif (Figure 10E).

Analysis of Immunotherapy Response and Drug Sensitivity

We explored the OS difference after immunotherapy between the high- and low-risk group in the phase II trial, in which



atezolizumab was treated with platinum-treated locally advanced or metastatic urothelial carcinoma (IMvigor210). The patients in the high-risk group had a better effect in receiving immunotherapy ($p < 0.05$, **Figure 11A**). Although there was no significant statistical difference, it was found that the prognosis of the high-risk group was better than that of the low-risk group after immunotherapy ($p = 0.095$, **Figure 11B**). The better response to immunotherapy in the high-risk group may be due to the relatively higher TMB that we had demonstrated before. A similar correlation between TMB and our risk scoring model was identified in TCGA BLCA (**Supplementary Figure S4**). Then, TIDE was further used to assess the potential immunotherapy effect in the high- and low-risk groups. The high-risk group had a lower TIDE score, which represented a lower possibility of immune escape, suggesting the BRCA patients in the high-risk group could benefit more from immune checkpoint inhibitor therapy (**Figure 9D**). Besides, the low-risk group got a higher T-cell dysfunction score (**Figure 9D**). Through further analyzing the drug sensitivity by comparing the

expression levels of model genes and drug response data from CELLMINER, we obtained the drugs with the most statistical significance (**Figure 12A**). The expression of *SLC35A2* was positively correlated with the sensitivity of two drugs (Vismodegib and Abiraterone) ($HR > 1$, $p < 0.05$), in turn, more sensitive in the high-risk group than in the low-risk group. The three-dimensional structures of the two drugs were obtained from the PubChem database, which provided potential guidance for chemotherapy in high-risk BRCA patients (**Figures 12B,C**).

DISCUSSION

Breast cancer is the most common cancer in women, accounting for nearly 25% of all cancer cases in women. It is also the leading cause of cancer deaths among elderly women (Hadadi et al., 2020; Siegel et al., 2021). The 5-year survival rate for patients with metastatic or stage IV breast cancer is 22% (www.cancer.org). The

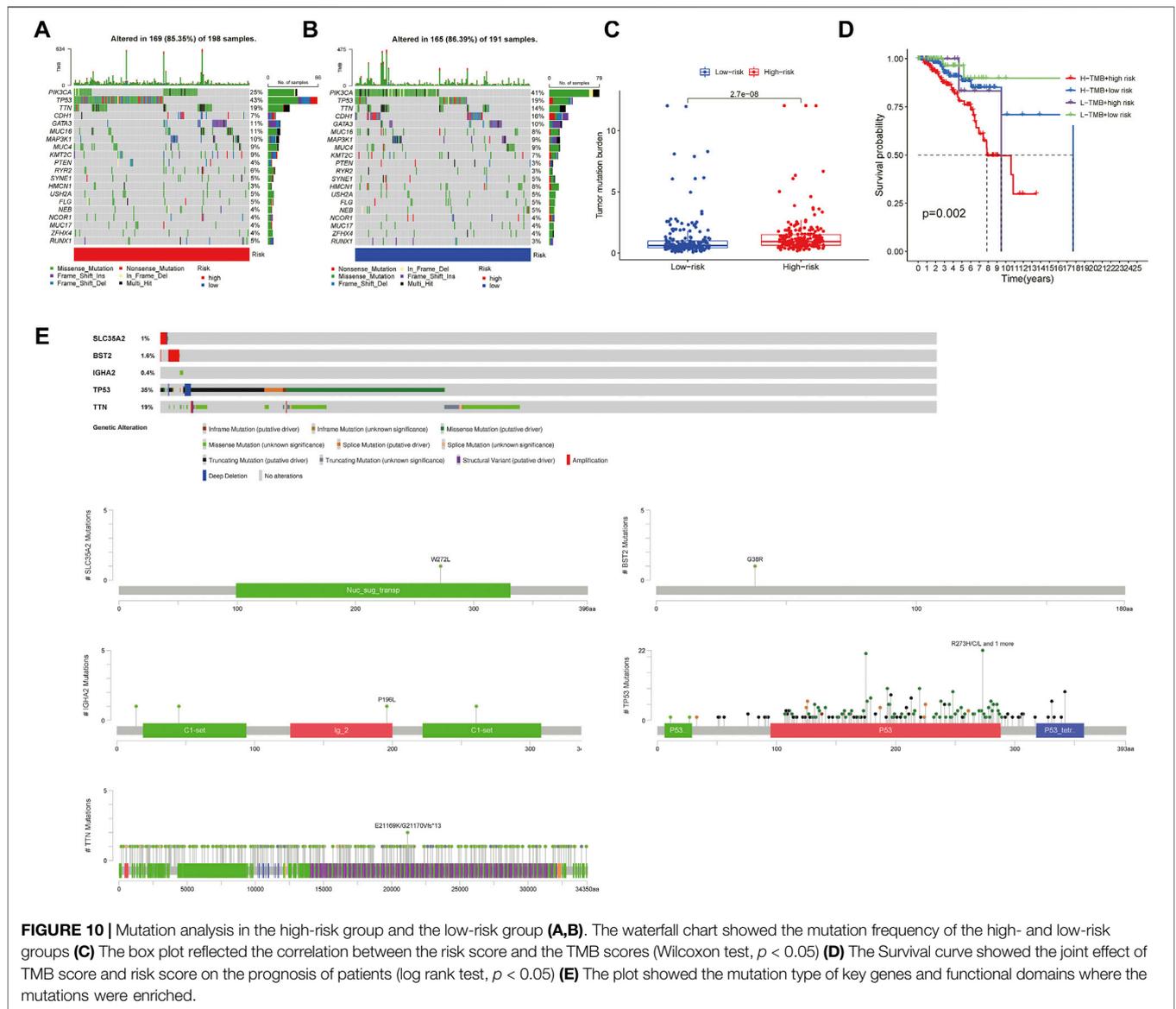
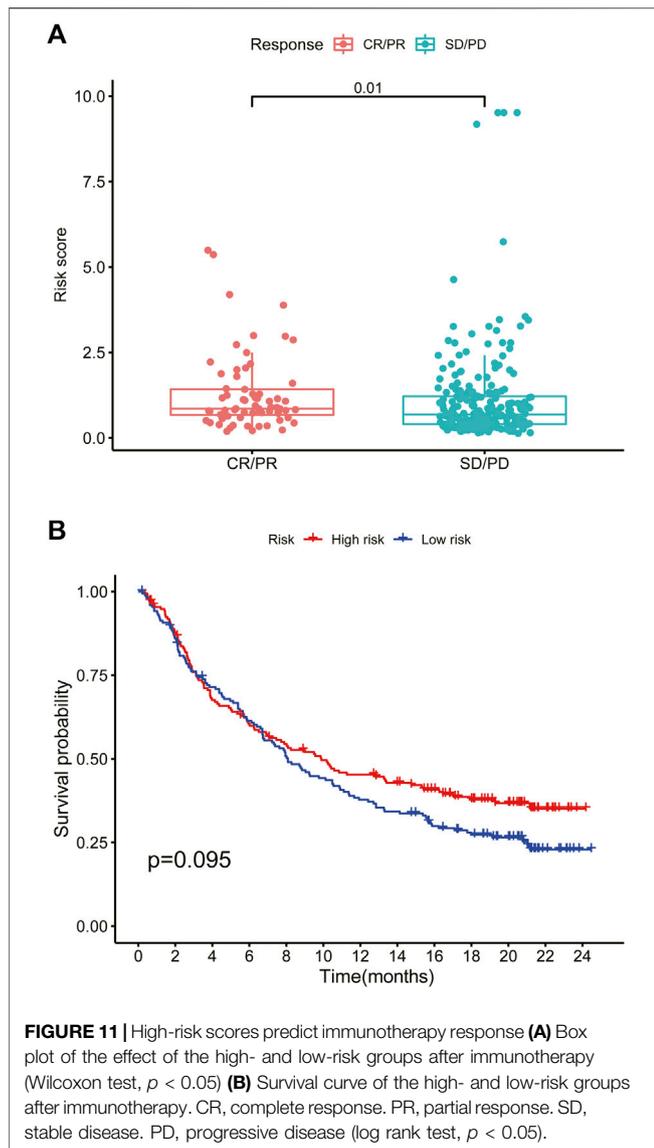


FIGURE 10 | Mutation analysis in the high-risk group and the low-risk group (A,B). The waterfall chart showed the mutation frequency of the high- and low-risk groups (C) The box plot reflected the correlation between the risk score and the TMB scores (Wilcoxon test, $p < 0.05$) (D) The Survival curve showed the joint effect of TMB score and risk score on the prognosis of patients (log rank test, $p < 0.05$) (E) The plot showed the mutation type of key genes and functional domains where the mutations were enriched.

disease has heterogeneity and possesses a diverse mutational landscape, suggesting differences in patients' response to treatments and lack of targeted treatment for patients in specific breast cancer subtypes, which indicates the need to improve the guidance for treatment strategies. In recent years, with the progress of high-throughput sequencing and data analysis, it has become a vital biomedical research tool, which can be used for prognosis prediction, recurrence monitoring, and clinical stratification (Z. Wang et al., 2009b; I. D. Kyrochristos et al., 2019). Therefore, it is urgent to apply this tool to prevent and treat breast cancer. Many shreds of evidence had shown that aberrant glycosylation had multiple effects on cancer's occurrence, progression, invasion, and metastasis (T. D. Rao et al., 2017; J. Cui et al., 2018). Potapenko et al. reported not only that there are significant differences in the expression

characteristics of glycosylation-related genes in breast cancer compared to normal breast tissue, but also that glycosylation-related genes show significant differences in expression between breast cancer subtypes and may be associated with patient prognosis and suggested that alterations in glycosylation pathways may occur at different time points in the carcinogenesis process (I. O. Potapenko et al., 2010; I. O. Potapenko et al., 2015). At present, there was still no prognostic model based on glycosylation-related genes for breast cancer. Given the critical impact of glycosylation on tumors, we had developed a prognostic model based on three glycosylation-related genes (*BST2*, *IGHA2*, *SLC35A2*).

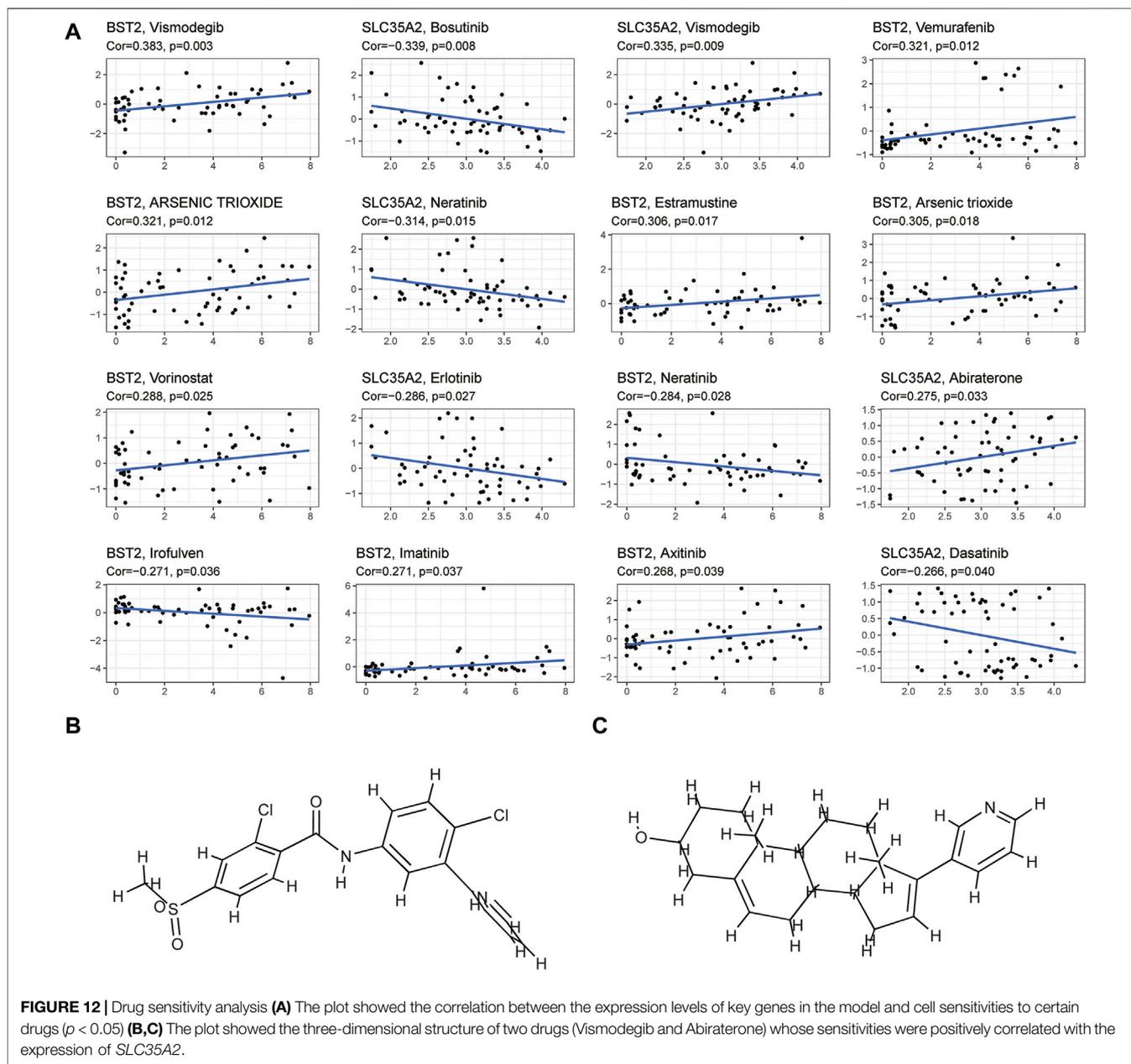
Evidence showed that *SLC35A2* belonged to the solute carrier family *SLC35* of human nucleoside sugar transporters, and encoded an X-linked transporter that



transports uridine diphosphate—galactose from the cytoplasm to the lumen of the Golgi apparatus and endoplasmic reticulum. Pathogenic variation was associated with congenital glycosylation disorder characterized by epileptic encephalopathy (D. Quelhas et al., 2021). However, rare research had shown its relationship with cancer. Therefore, exploring the mechanism of *SLC35A2* is in high demand. In our study, the established prognostic model showed that the OS rate of patients in the high-risk group was significantly lower than that in the low-risk group through Kaplan Meier survival analysis ($p < 0.001$). Then, we testified the reliability of the model by using the validation set and external independent test sets. By comparing the risk score with other clinicopathological factors with a ROC curve, it was found that the risk model had a higher prognostic value than other clinicopathological factors. We also constructed a nomogram based on the age, TMN stage, and risk score to improve the

accuracy of clinical decision-making. The risk score and nomogram had high reliability by calculating C-index and decision curve analysis. Subsequently, we divided the samples into multiple groups to demonstrate the application of the model in specific categories. The risk model had an excellent predictive value among age ≤ 65 , ER positive, HER-2 negative, M0, N0, N1-N3, Stage I-II, Stage III-IV, T1-2, T3-4, non-TNBC subgroups ($p < 0.05$). In M1 and HER-2 positive subgroups, although there was no significant difference between the high- and low-risk groups, which might be ascribed to the small number of samples and the short follow-up time of patients in the subgroups, the survival time of the low-risk group was higher than that of the high-risk group. There was no difference in the ER negative and TNBC subgroups, resulting from the small sample size. We successfully validated the predictive value of the model in HER-2 positive and TNBC subgroups by METABRIC. Regarding the immunobiology of BRCA, TILs as an important biomarker in predicting the efficacy and outcome of treatment were worth exploring in depth. In breast cancer patients, loss of the anti-HER-2 CD4⁺ Th1 immune response is independently correlated with disease recurrence (J. Datta et al., 2016). CD8⁺ TILs, Th1 CD4⁺ TILs can influence anti-tumor immune response in breast cancer (A. Basu et al., 2019). Several clinical trials have also shown an increased pathological complete response associated with a high density of TILs (Y. Issa-Nummer et al., 2013; C. Denkert et al., 2015). We explored the relationship between model grouping and scores of immune infiltrating cells. The results showed that the low-risk group had a higher infiltrative proportion of B cells naive, Mast cells resting, Plasma cells, T cells CD4 memory resting, T cells CD8, T cells gamma delta, which was in keeping with the previous investigations, suggesting a correlation between glycosylation and TILs. Therefore, impaired anti-tumor immune function may account for the poor prognosis in high-risk patients. Subsequently, we explored the differences in somatic mutations between the high- and the low-risk groups. The results showed that the TMB of the high-risk group was higher, indicating that patients in this group were more likely to receive the benefits of immune checkpoint inhibitor treatment (Steuer and Ramalingam, 2018). Moreover, the high-risk group had a lower TIDE score, which represented a lower possibility of immune escape, suggesting the BRCA patients in the high-risk group could benefit more from immune checkpoint inhibitor therapy (P. Jiang et al., 2018). The combined survival analysis showed that the prognosis of low-risk and low TMB patients was significantly better than that of high-risk and high TMB patients ($p < 0.01$). Then we analyzed chemotherapy based on the key genes. Two drugs (Vismodegib, Abiraterone) were expected to be validated and applied for the high-risk breast cancer patients.

This study also has some limitations. First, as it is a retrospective study derived from public data, it lacks some information such as recurrence time and treatment records. Second, clinical trials are urgently needed to confirm whether inducing glycosylation could improve the efficacy of immunotherapy in human BRCA patients.



Furthermore, the specific molecular mechanism of the gene in the risk-scoring model has not been fully explored.

CONCLUSION

In conclusion, a novel prognostic model integrating glycosylation-related genes was firstly constructed. Through verification in the validation set and external independent test sets, the risk-scoring model has been proved to be an independent prediction model for predicting the prognosis of patients, which was correlated with immunotherapy effect and drug sensitivity. Moreover, we established a prognostic nomogram to predict the

OS of patients with BRCA. The novel model might provide insights for predicting the prognosis of BRCA patients and suggestions to guide individual therapeutic strategies.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The data could be downloaded at (<https://www.cancer.gov>; <http://www.ncbi.nlm.nih.gov/geo>; <http://www.cbioportal.org>; <https://www.genecards.org>) and the code used during the current study are available from the corresponding author on reasonable request.

AUTHOR CONTRIBUTIONS

SX and YL performed the data analysis and interpreted the results. SX and YL prepared the draft. YP, SX, and YL designed the research and revised the draft. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key Research and Development Program (2018YFA0902801), Guangdong

Provincial Key Laboratory of Digestive Cancer Research (No. 2021B1212040006), Natural Science Foundation of Guangdong Province, China (Grant No. 2020A1515010028), and the Scientific Technology Project of Shenzhen City (JCYJ20210324124003008).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.834731/full#supplementary-material>

REFERENCES

- Basu, A., Ramamoorthi, G., Jia, Y., Faughn, J., Wiener, D., Awshah, S., et al. (2019). Immunotherapy in Breast Cancer: Current Status and Future Directions. *Adv. Cancer Res.* 143, 295–349. doi:10.1016/bs.acr.2019.03.006
- Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E., and Gianni, L. (2016). Triple-negative Breast Cancer: Challenges and Opportunities of a Heterogeneous Disease. *Nat. Rev. Clin. Oncol.* 13 (11), 674–690. doi:10.1038/nrclinonc.2016.66
- Chakraborty, A., Dorsett, K. A., Trummell, H. Q., Yang, E. S., Oliver, P. G., Bonner, J. A., et al. (2018). ST6Gal-I Sialyltransferase Promotes Chemoresistance in Pancreatic Ductal Adenocarcinoma by Abrogating Gemcitabine-Mediated DNA Damage. *J. Biol. Chem.* 293 (3), 984–994. doi:10.1074/jbc.M117.808584
- Cui, J., Huang, W., Wu, B., Jin, J., Jing, L., Shi, W.-P., et al. (2018). N-Glycosylation by N-acetylglucosaminyltransferase V Enhances the Interaction of CD147/basigin with Integrin $\beta 1$ and Promotes HCC Metastasis. *J. Pathol.* 245 (1), 41–52. doi:10.1002/path.5054
- Datta, J., Fracol, M., McMillan, M. T., Berk, E., Xu, S., Goodman, N., et al. (2016). Association of Depressed Anti-HER2 T-Helper Type 1 Response with Recurrence in Patients with Completely Treated HER2-Positive Breast Cancer. *JAMA Oncol.* 2 (2), 242–246. doi:10.1001/jamaoncol.2015.5482
- Denkert, C., von Minckwitz, G., Brase, J. C., Sinn, B. V., Gade, S., Kronenwett, R., et al. (2015). Tumor-infiltrating Lymphocytes and Response to Neoadjuvant Chemotherapy with or without Carboplatin in Human Epidermal Growth Factor Receptor 2-positive and Triple-Negative Primary Breast Cancers. *Jco* 33 (9), 983–991. doi:10.1200/jco.2014.58.1967
- Eccles, S. A., Aboagye, E. O., Ali, S., Anderson, A. S., Armes, J., Berdichevski, F., et al. (2013). Critical Research Gaps and Translational Priorities for the Successful Prevention and Treatment of Breast Cancer. *Breast Cancer Res.* 15 (5), R92. doi:10.1186/bcr3493
- Fallahpour, S., Navaneelan, T., De, P., and Borgo, A. (2017). Breast Cancer Survival by Molecular Subtype: a Population-Based Analysis of Cancer Registry Data. *CMAJ Open* 5 (3), E734–e739. doi:10.9778/cmajo.20170030
- Fang, K.-H., Kao, H.-K., Chi, L.-M., Liang, Y., Liu, S.-C., Hseuh, C., et al. (2014). Overexpression of BST2 Is Associated with Nodal Metastasis and Poorer Prognosis in Oral Cavity Cancer. *The Laryngoscope* 124 (9), E354–E360. doi:10.1002/lary.24700
- Freire, T., Lo-Man, R., Bay, S., and Leclerc, C. (2011). Tn Glycosylation of the MUC6 Protein Modulates its Immunogenicity and Promotes the Induction of Th17-Biased T Cell Responses. *J. Biol. Chem.* 286 (10), 7797–7811. doi:10.1074/jbc.M110.209742
- Gradishar, W. J., Anderson, B. O., Abraham, J., Aft, R., Agnese, D., Allison, K. H., et al. (2020). Breast Cancer, Version 3.2020, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Canc Netw.* 18 (4), 452–478. doi:10.6004/jnccn.2020.0016
- Hadadi, E., Taylor, W., Li, X.-M., Aslan, Y., Villote, M., Rivière, J., et al. (2020). Chronic Circadian Disruption Modulates Breast Cancer Stemness and Immune Microenvironment to Drive Metastasis in Mice. *Nat. Commun.* 11 (1), 3193. doi:10.1038/s41467-020-16890-6
- Issa-Nummer, Y., Darb-Esfahani, S., Loibl, S., Kunz, G., Nekljudova, V., Schrader, I., et al. (2013). Prospective Validation of Immunological Infiltrate for Prediction of Response to Neoadjuvant Chemotherapy in HER2-Negative Breast Cancer - A Substudy of the Neoadjuvant GeparQuinto Trial. *PLoS One* 8 (12), e79775. doi:10.1371/journal.pone.0079775
- Jerusalem, G., Collignon, J., Schroeder, H., and Lousberg, L. (2016). Triple-negative Breast Cancer: Treatment Challenges and Solutions. *Bctt* 8, 93–107. doi:10.2147/bctt.s69488
- Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T Cell Dysfunction and Exclusion Predict Cancer Immunotherapy Response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1
- Kang, S., Maeng, H., Kim, B. G., Qing, G. M., Choi, Y. P., Kim, H. Y., et al. (2012). *In Situ* Identification and Localization of IGHA2 in the Breast Tumor Microenvironment by Mass Spectrometry. *J. Proteome Res.* 11 (9), 4567–4574. doi:10.1021/pr3003672
- Kyrcostas, I. D., Ziogas, D. E., and Roukos, D. H. (2019). Dynamic Genome and Transcriptional Network-based Biomarkers and Drugs: Precision in Breast Cancer Therapy. *Med. Res. Rev.* 39 (3), 1205–1227. doi:10.1002/med.21549
- Lin, Z., Xu, Q., Miao, D., and Yu, F. (2021). An Inflammatory Response-Related Gene Signature Can Impact the Immune Status and Predict the Prognosis of Hepatocellular Carcinoma. *Front. Oncol.* 11, 644416. doi:10.3389/fonc.2021.644416
- Liu, W., Cao, Y., Guan, Y., and Zheng, C. (2018). BST2 Promotes Cell Proliferation, Migration and Induces NF-Kb Activation in Gastric Cancer. *Biotechnol. Lett.* 40 (7), 1015–1027. doi:10.1007/s10529-018-2562-z
- Liu, W., Li, Y., Feng, S., Guan, Y., and Cao, Y. (2020). MicroRNA-760 Inhibits Cell Viability and Migration through Down-Regulating BST2 in Gastric Cancer. *J. Biochem.* 168 (2), 159–170. doi:10.1093/jb/mvaa031
- Mahaud-Fernandez, W. D., DeMali, K. A., Olivier, A. K., and Okeoma, C. M. (2014). Bone Marrow Stromal Antigen 2 Expressed in Cancer Cells Promotes Mammary Tumor Growth and Metastasis. *Breast Cancer Res.* 16 (6), 493. doi:10.1186/s13058-014-0493-8
- Mahaud-Fernandez, W. D., and Okeoma, C. M. (2017). Cysteine-linked Dimerization of BST-2 Confers Anoikis Resistance to Breast Cancer Cells by Negating Proapoptotic Activities to Promote Tumor Cell Survival and Growth. *Cell Death Dis* 8 (3), e2687. doi:10.1038/cddis.2017.68
- Ohtsubo, K., and Marth, J. D. (2006). Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* 126 (5), 855–867. doi:10.1016/j.cell.2006.08.019
- Peixoto, A., Relvas-Santos, M., Azevedo, R., Santos, L. L., and Ferreira, J. A. (2019). Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks. *Front. Oncol.* 9, 380. doi:10.3389/fonc.2019.00380
- Pinho, S. S., and Reis, C. A. (2015). Glycosylation in Cancer: Mechanisms and Clinical Implications. *Nat. Rev. Cancer* 15 (9), 540–555. doi:10.1038/nrc3982
- Potapenko, I. O., Haakensen, V. D., Lüders, T., Helland, Å., Bukholm, I., Sørli, T., et al. (2010). Glycan Gene Expression Signatures in normal and Malignant Breast Tissue; Possible Role in Diagnosis and Progression. *Mol. Oncol.* 4 (2), 98–118. doi:10.1016/j.molonc.2009.12.001
- Potapenko, I. O., Lüders, T., Russnes, H. G., Helland, Å., Sørli, T., Kristensen, V. N., et al. (2015). Glycan-related Gene Expression Signatures in Breast Cancer Subtypes; Relation to Survival. *Mol. Oncol.* 9 (4), 861–876. doi:10.1016/j.molonc.2014.12.013
- Quelhas, D., Correia, J., Jaeken, J., Azevedo, L., Lopes-Marques, M., Bandeira, A., et al. (2021). SLC35A2-CDG: Novel Variant and Review. *Mol. Genet. Metab. Rep.* 26, 100717. doi:10.1016/j.ymgmr.2021.100717
- Rao, T. D., Fernández-Tejada, A., Axelrod, A., Rosales, N., Yan, X., Thapi, S., et al. (2017). Antibodies against Specific MUC16 Glycosylation Sites Inhibit Ovarian Cancer Growth. *ACS Chem. Biol.* 12 (8), 2085–2096. doi:10.1021/acscchembio.7b00305
- Sayeed, A., Luciani-Torres, G., Meng, Z., Bennington, J. L., Moore, D. H., and Dairkee, S. H. (2013). Aberrant Regulation of the BST2 (Tetherin) Promoter Enhances Cell Proliferation and Apoptosis Evasion in High Grade Breast Cancer Cells. *PLoS One* 8 (6), e67191. doi:10.1371/journal.pone.0067191

- Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer Statistics, 2021. *CA A. Cancer J. Clin.* 71 (1), 7–33. doi:10.3322/caac.21654
- Silver, D. P., Richardson, A. L., Eklund, A. C., Wang, Z. C., Szallasi, Z., Li, Q., et al. (2010). Efficacy of Neoadjuvant Cisplatin in Triple-Negative Breast Cancer. *Jco* 28 (7), 1145–1153. doi:10.1200/jco.2009.22.4725
- Smid, M., Wang, Y., Zhang, Y., Sieuwerts, A. M., Yu, J., Klijn, J. G. M., et al. (2008). Subtypes of Breast Cancer Show Preferential Site of Relapse. *Cancer Res.* 68 (9), 3108–3114. doi:10.1158/0008-5472.can-07-5644
- Steuer, C. E., and Ramalingam, S. S. (2018). Tumor Mutation Burden: Leading Immunotherapy to the Era of Precision Medicine? *Jco* 36 (7), 631–632. doi:10.1200/jco.2017.76.8770
- Wang, W., Nishioka, Y., Ozaki, S., Jalili, A., Abe, S., Kakiuchi, S., et al. (2009a). HM1.24 (CD317) Is a Novel Target against Lung Cancer for Immunotherapy Using anti-HM1.24 Antibody. *Cancer Immunol. Immunother.* 58 (6), 967–976. doi:10.1007/s00262-008-0612-4
- Wang, Z., Gerstein, M., and Snyder, M. (2009b). RNA-seq: a Revolutionary Tool for Transcriptomics. *Nat. Rev. Genet.* 10 (1), 57–63. doi:10.1038/nrg2484

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Xu, Liu, Ma, Fang, Wei, Li, Lu, Zheng, Liu, Zhu, Xu and Pan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.