Nomogram Personalizes and Visualizes the Overall Survival of Patients with Triple-Negative Breast Cancer Based on the Immune Genome

Peipei Wang, Yang Fu, Yueyun Chen, Qing Li, Ye Hong, Ting Liu, and Zhenyu Ding

Department of Biotherapy, Cancer Center, West China Hospital, West China Medical School, State Key Laboratory of Biotherapy, Sichuan University, Chengdu, China

Correspondence should be addressed to Zhenyu Ding; dingzhenyu@scu.edu.cn

Peipei Wang and Yang Fu contributed equally to this work.

Received 21 July 2020; Revised 2 October 2020; Accepted 12 November 2020; Published 24 November 2020

Academic Editor: Shijia Zhu

Copyright © 2020 Peipei Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Triple-negative breast cancer (TNBC) is usually poorly differentiated, highly invasive, susceptible to distant metastasis, and less responsive to endocrine and targeted therapy. However, immunotherapy is a promising treatment for TNBC patients recently. Methods. The prognostic value of immune-related genes (IRGs) was explored by using RNA sequencing and microarray data of 123 and 107 TNBC patients from TCGA and GEO databases, respectively. Results. In TCGA database, GO and KEGG pathway analysis of 119 differential IRGs indicated that they actively participate in the interaction of cytokines and receptors. A nomogram model constructed by the prognosis-related CCL25, IL29, TDGF3, GPR44, and GREM2 in the IRGs could personalize and visualize the 1-, 2-, 3-, 4-, and 5-year overall survival (OS) of TNBC patients. Moreover, TNBC patients could be defined as low-risk (risk score < 194) or high-risk (risk score ≥ 194) cohorts based on the risk score derived from the nomogram model. The results could be validated by the GSE58812 dataset. Furthermore, the risk score was an independent risk factor for TNBC patients (HR = 1.019, 95% CI 1.012-1.027, p < 0.001) and was positively related to stage (p = 0.017). Interestingly, the risk score could reflect the infiltration of B cells, CD4+ T cells, CD8+ T cells, dendritic cells, and neutrophils. Conclusion. These findings provided a reference for personalized OS prediction in TNBC patients and might be potential immune biomarkers for designing novel therapy.

1. Introduction

Breast cancer has high incidence in women. Different types of breast cancer have obvious differences in morphology, molecular pathological characteristics, clinical features, and responses to tumor treatment [1]. As a special subtype of breast cancer, triple-negative breast cancer (TNBC) lacks the expression of estrogen receptor (ER), progesterone receptor (PR), and epidermal growth factor receptor 2 (HER2), which severely limits the clinical usage of endocrine and targeted therapy.

TNBC is usually poorly differentiated, highly invasive, susceptible to distant metastasis, and less responsive to treatment than other hormone receptor-positive breast cancers, so it has a higher risk of early relapse [2]. However, due to its unstable genome and high mutation rate, TNBC is highly immunogenic [3]. At present, some progress has been made in immunotherapy for TNBC. Immunotherapy stimulates the immune response of TNBC patients through active immunity, such as cancer vaccines, or passive immunity, such as adoptive T cell therapy, tumor-specific antibodies, and immune checkpoint inhibitors [4]. A study has fused TNBC cells with peripheral blood monocyte-derived dendritic cells (DCs) to generate DC vaccines, which stimulate the proliferation of T lymphocytes and enhance the cytotoxic effect on breast cancer cells [5]. Emerging immune
checkpoint inhibitors have received increasing attention in numerous TNBC study. TNBC patients with higher PD-L1 expression levels and more tumor-infiltrating lymphocytes (TILs) have higher immunogenicity, which plays a crucial part in regulating the immune response [6]. FDA has approved PD-L1 blockade combined with chemotherapy for patients with PD-L1-positive TNBC [7].

Increasing evidence suggests that immune gene expression and TIL may be prognostic for TNBC. Rody et al. [8] analyzed the RNA sequencing data of 579 TNBC patients and found that the expression of immune cell metagenes is closely related to prognosis. A study found that TNBC patients with higher TILs have better overall survival (OS) [9]. Yeong et al. [10] found that high PD-1+ cell infiltration significantly improved disease-free survival in TNBC patients. Similarly, PD-1, IFNG, and IFN signaling genes are positively correlated with the improvement of clinical outcomes of TNBC patients. These studies suggest that immune genes and TILs play an important role in TNBC. However, there are currently no immune-related genes that individually and visually predict OS and TILs for patients with TNBC.

In this study, we analyzed the RNA-seq and microarray data in The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases to comprehensively evaluate immune-related genes (IRGs) expression levels and predict prognosis and immune cell infiltration in TNBC patients. The cell functions involved in IRGs were also explored. These explorations are especially important for the individual assessment of the prognosis of TNBC patients and the discovery of targeted immunization methods.

2. Materials and Methods

2.1. Acquisition ofDatasets and Patients’ Information. UCSC Xena (https://xenabrowser.net/datapages/) [11] was used to download the RNA-seq data of breast cancer in the form of log2 (norm_count + 1) from TCGA database (http://www.tcgag.org), including 123 newly diagnosed PR-, ER-, and Her2-negative breast cancers and 13 normal tissues adjacent to tumor. At the same time, the clinical information of patients was obtained, involving age, gender, tumor invasion depth (T), lymph node metastasis (N), distant metastasis (M), TNM stage, survival time, and status. The 123 TNBC patients in TCGA database served as a training cohort, and the clinical characteristics are listed in Table S1. In addition, a microarray dataset and clinical characteristics from 107 TNBC cases obtained from the GSE58812 dataset [12] in the GEO database were used as a validation cohort.

2.2. Differential Analysis of Genes. The limma package in R (version 3.6.1) was used to select differential genes between 123 cases of primary TNBC and 13 normal tissues adjacent to tumor. A total of 20,530 genes were included into the differential analysis, setting the adjusted p value < 0.05, log2 \( \text{fold change} > 1 \), and expression level > 0.2. And a total of 1076 genes with significant differences were finally determined. We focused on the role of IRGs in TNBC. The ImmPort database (https://www.immport.org/) [13] provides and updates 2498 IRGs for cancer research, all of which have been identified as being involved in the biological processes of immunity. Therefore, we downloaded the list of IRGs from the ImmPort database. Finally, 4.8% of IRGs were identified as those (a total of 119) differentially expressed genes and were included in subsequent analysis. Functional enrichment analysis of KEGG (Kyoto Encyclopedia of Genes and Genomes) and GO (Gene Ontology) was performed on 119 IRGs through the Database for Annotation, Visualization and Integrated Discovery (DAVID, https://david.ncifcrf.gov/) to explore potential molecular mechanisms.

2.3. Construction and Validation of Nomogram Model. The foreign and rms packages in R (version 3.6.1) were applied for establishing a nomogram model to analyze the role of IRGs in TNBC patients, which was described in detail in our previous study [14]. First, according to the nomogram, a point is given for each patient’s IRG expression level. Then, a total risk score is obtained by gathering the given points of all the IRGs of the patients, which can predict OS. The concordance index (C-index) of 1000-sample bootstrap and the receiver operating characteristic (ROC) curve were used to evaluate the prognostic prediction ability of the nomogram model, and the judgment criterion was Area Under the Curve (AUC) or C-index > 0.5. Model performance was assessed through both the internal and external calibration curve of 1000-sample bootstrap.

2.4. The Clinical Value of Risk Score. The risk score was divided into high or low based on the median value. Risk distribution, survival status, and IRG expression distribution of high- and low-risk TNBC were plotted by the heat map package in R (version 3.6.1). Survival and survminer packages were used to draw survival curves. We also explored the predictive value of risk score in TNBC for immune cell infiltration. The TIMER database (http://cistrome.org/TIMER/) [15] analyzed the TILs of 32 cancers in TCGA. Therefore, we obtained TIL abundances in TNBC patients from the TIMER and evaluated the correlation between the risk score and TILs.

2.5. Statistical Analysis. All statistical analysis was conducted in R (version 3.6.1, https://www.r-project.org/). Uni- and multivariate Cox regression analyses were used to screen prognostic variables. A log-rank test was used to compare the difference between survival curves. Two sets of quantitative data were compared by the Wilcoxon test. The correlation between the two sets of quantitative data was expressed by Spearman coefficient. A two-tailed p value < 0.05 was considered statistically significant.

3. Results

3.1. Identification of Differential IRGs. Differential analysis found that a total of 1076 genes were differentially expressed in TNBC, including 323 high expression and 753 low expression genes (Figure 1). Among them, a total of 119 differential IRGs, including 36 high expression and 83 low expression IRGs, were found. Their positions on the chromosome are shown in Figure 2. The KEGG showed that the enrichment
of differentially expressed IRGs was primarily in the “neuroactive ligand-receptor interaction” and “cytokine-cytokine receptor interaction” (Figure 3(a)). Biological processes, cellular components, and molecular functions are primarily enriched in “cell-cell signaling,” “extracellular region,” and “growth factor activity,” respectively (Figures 3(b)–3(d)). These findings suggested that cytokines and receptor pathways were most frequently implicated.

3.2. Identification of Prognosis-Related IRGs. Because prognostic molecular biomarkers are important for guiding treatment and disease monitoring, we focus on the impact of IRGs on the OS in TNBC. Prognostic analysis revealed that a total of 6 IRGs had significant impact on the OS of TNBC patients. Among them, high expression of C-C motif chemokine ligand 25 (CCL25), interleukin 29 (IL29), teratocarcinoma-derived growth factor 3 (TDGF3), and killer cell immunoglobulin like receptor, two Ig domains and long cytoplasmic tail 4 (KIR2DL4) predicted a favorable OS in TNBC patients (Figures 4(a), 4(d), 4(e), and 4(f)). Conversely, high expression of G protein-coupled receptor 44 (GPR44) and gremlin 2, DAN family BMP antagonist (GREM2) predicted a poor OS in TNBC patients (Figures 4(b) and 4(c)).

3.3. Establishment of Nomogram Model. CCL25, IL29, TDGF3, KIR2DL4, GPR44, and GREM2 obtained from the analysis of TCGA database were included into the establishment of a nomogram model. According to the expression levels of these IRGs, we got the total risk score of each individual, which could predict 1-, 2-, 3-, 4-, and 5-year OS (Figure S1). However, we could observe from Table S2 that the upregulated KIR2DL4 accounted for only 5 risk scores in the model. Compared with the risk scores of the other five IRGs, KIR2DL4 had a lower contribution to predicting the OS of TNBC patients, so we excluded KIR2DL4 in the establishment of the nomogram model. Finally, five prognosis-related IRGs including CCL25, IL29, TDGF3, GPR44, and GREM2 were included into the establishment of the nomogram model (Figure 5). To assess the predictive effect of the nomogram model on the OS of TNBC patients, we used the AUC of the ROC curve and the C-index of 1000-sample bootstrap for evaluation. When KIR2DL4 was
not removed, the AUC was 0.839 and C-index was 0.878 (Figure S2). When KIR2DL4 was excluded, AUC was 0.852 and C-index was 0.879 (Figure 6(a)). Further external validation of the nomogram model with the GSE58812 dataset in the GEO database showed that the AUC of the ROC curve was 0.619 and the C-index is 0.615 (Figure 6(b)). We also used the calibration curves to further validate the nomogram model. The calibration curves for the training group showed a good consistency between the predicted and actual 1-, 2-, 3-, 4-, and 5-year OS of the nomogram model (Figure 7(a)). At the same time, the calibration curves in the validation group were also well identified (Figure 7(b)). Through internal and external verification, it was proved that the nomogram model in this study could conduct relatively accurate prediction of the OS of TNBC patients.

3.4. Risk Stratification. In order to analyze the clinical value of the risk score, we compare the risk scores of different clinical information and found that compared with stage I/II
Figure 3: Continued.
TNBC patients with stage III/IV had higher risk scores ($p = 0.017$), and there was no significant difference between different ages, T, N, and M ($p > 0.05$) (Figure 8). After univariate and multivariate Cox regression analyses, we found that the risk score was an independent risk factor ($HR = 1.019$, 95% CI 1.012-1.027, $p < 0.001$, Table 1). The risk score was highly correlated with TNM stage, indicating that risk score could be used to construct risk stratification. TNBC patients can be defined as low-risk (risk score < 194) or high-risk (risk score ≥ 194) cohort based on the risk score (Figures 9(a)–9(d), left panel). As shown in Figure 9(d), left panel, the survival curve of low- and high-risk TNBC patients...
1.00

CCL25

High expression (n = 61)

Low expression (n = 62)

HR (High expression) = 0.161 (95% CI: 0.047–0.556)
Log rank test: p < 0.001

(a)

GPR44

High expression (n = 61)

Low expression (n = 62)

HR (High expression) = 3.958 (95% CI: 1.304–12.013)
Log rank test: p < 0.009

(b)

Figure 4: Continued.
OS

GREM2

High expression \((n = 62)\)

Low expression \((n = 61)\)

HR (High expression) = 3.107 (95% CI: 1.109–8.711)
Log rank test: \(p < 0.023\)

IL29

High expression \((n = 61)\)

Low expression \((n = 62)\)

HR (High expression) = 0.251 (95% CI: 0.083–0.758)
Log rank test: \(p < 0.008\)

Figure 4: Continued.
showed significant differences ($p < 0.001$). This finding could be validated by the external dataset GSE58812 ($p = 0.032$, Figures 9(a)–9(d), right panel).

3.5. **Correlation between Risk Score and TILs.** In order to investigate whether the risk score reflected by immune genes could accurately assess the state of TNBC immune microenvironment, we performed the correlation analysis between risk score and TILs (Figure 10) and found that risk score was negatively related to B cells ($R = -0.26, p = 0.005$), CD4+ T cells ($R = -0.21, p = 0.019$), CD8+ T cells ($R = -0.19, p = 0.034$), dendritic cells (DCs) ($R = -0.25, p = 0.005$), and neutrophils ($R = -0.27, p = 0.002$). However, the risk score had no significant correlation with macrophages ($p = 0.3$).
4. Discussion

The clonal proliferation and metastasis characteristics of cancer cells depend on genome changes. We focus our research on changes in the immune genome to reveal its relationship with the immune microenvironment. In this study, we extracted differential IRGs from two large TNBC cohorts in the GEO and TCGA databases and analyzed the underlying immune mechanism. Pathway and GO analysis found that TNBC patients primarily function through interactions between cytokines and receptors. Various studies have showed that cytokines and receptors are involved in the growth, invasion, and metastasis of breast cancer, and corresponding immune inhibitors against cytokines and receptors have been applied to the treatment of breast cancer [16, 17]. In addition, fibroblast growth factor receptors (FGFRs) are highly expressed in TNBC patients, and inhibitors against FGFRs have been tested in clinical trials [18]. Bioinformatics analysis provided clues that changes in the immune system of TNBC patients were initiated through cytokine and receptor pathways, which laid the foundation for in-depth immune-related research.

The nomogram model has been widely applied to systematically assess the outcome of cancer patients [14, 19]. At the same time, IRGs can provide personalized immune signature to assess the prognosis of lung cancer patients [20]. The prognosis-related CCL25, IL29, TDGF3, GPR44, and GREM2 in IRGs were used to construct a nomogram model to evaluate its clinical value in TNBC patients. The nomogram model we constructed can individualize and visualize 1-, 2-, 3-, 4-, and 5-year OS in TNBC patients. Evidence suggests that blocking the CCR9/CCL25 axis can promote tumor progression and distant metastasis [21, 22]. IL29 appears to inhibit tumor growth in a variety of cancers [23]. TDGF3, also known as TDGF1P3 or CRIPTO3, is expressed in cancer [24]. Our study found that high expression of CCL25, IL29, and TDGF3 predicted a good prognosis in TNBC patients.

Study has shown that GPR44 is involved in the release of cytokines from immune cells and the development of inflammation [25]. Silencing GREM2 can inhibit the JNK signaling pathway in gastric cancer, which inhibits tumor growth [26]. In this study, high expressions of GPR44 and GREM2 could predict adverse outcomes in TNBC patients.

Figure 5: Construction of nomogram model with prognostic IRGs for predicting 1-, 2-, 3-, 4-, and 5-year OS in TNBC patients. A higher risk score meant a lower survival probability.
The establishment of risk stratification can provide a reference for clinicians to more rationally manage cancer patients and individualized treatment plans [27]. We conducted risk stratification for TNBC patients based on the risk score and found that patients with a risk score $\geq 194$ belonged to a high-risk cohort, while patients with a risk score $< 194$ belonged to a low-risk cohort.

The risk score showed good clinical practicability. The risk score was an independent risk factor for TNBC patients and was positively related to TNM stage. In addition, the risk score was negatively correlated with B cells, CD4+ T cells, CD8+ T cells, dendritic cells, and neutrophils, which could reflect the level of TILs and provide a reference for clinicians to adjust the treatment plan.
Table 1: Uni- and multivariate Cox regression analyses of 123 TNBC patients in TCGA database.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Univariate Cox</th>
<th>Multivariate Cox</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>p value</td>
</tr>
<tr>
<td>Risk score</td>
<td>1.022 (1.014-1.031)</td>
<td>&lt;0.001***</td>
</tr>
<tr>
<td>Age (years)</td>
<td>1.004 (0.969-1.040)</td>
<td>0.842</td>
</tr>
<tr>
<td>Tumor invasion depth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1-2</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>T3-4</td>
<td>2.986 (0.976-9.140)</td>
<td>0.055</td>
</tr>
<tr>
<td>Distant metastasis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>54.325 (4.926-599.140)</td>
<td>0.001**</td>
</tr>
<tr>
<td>Lymph node metastasis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>N1-3</td>
<td>3.785 (1.437-9.967)</td>
<td>0.007**</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I/II</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>III/IV</td>
<td>5.441 (2.113-14.012)</td>
<td>&lt;0.001***</td>
</tr>
</tbody>
</table>

HR: hazard ratio; CI: confidence interval; *p < 0.05; **p < 0.01; ***p < 0.001.

Figure 8: The distribution of risk scores in TNBC patients among different age, tumor invasion depth (T), distant metastasis (M), lymph node metastasis (N), and TNM stage.
Figure 9: Continued.
Figure 9: Risk stratification of TNBC patients by risk score. (a) Risk score, (b) survival time, and (c) gene expression distribution of low- and high-risk groups in TCGA dataset (left panel) and GSE58812 (right panel). (d) Plotting survival curves according to low- and high-risk scores in TCGA dataset (left panel) and GSE58812 (right panel).

Figure 10: Spearman correlation between risk score and tumor-infiltrating lymphocytes (TILs).
Studies indicate that a favorable prognosis is observed in TNBC patients with high B cells and CD4+ T cells [28, 29]. Furthermore, it has been reported that activated CD8+ T cells have been shown to kill cancer cells through various mechanisms [30]. Not surprisingly, less infiltration of CD8+ T cells into tumors in TNBC patients is often related to worse outcomes [31]. In addition, a study suggests that neutrophil infiltration is a favorable prognostic factor for colon cancer [32], which is consistent with our findings. As we know that DCs take a vital part in the tumor microenvironment, infiltration of activated DCs into tumors can improve the antitumor effect of immune cells [33]. In this study, high-risk patients might have the lower infiltration levels of B cells, CD4+ T cells, CD8+ T cells, dendritic cells, and neutrophils, which was associated with a poor OS.

The limitation of this study was that no clinical samples and corresponding clinical information were used to validate the nomogram model and risk stratification constructed by CCL25, IL29, TDGF3, GPR44, and GREM2. Additionally, the reliability of the results was still challenged because we also lacked validation in vitro and in vivo.

5. Conclusion

A nomogram model constructed by CCL25, IL29, TDGF3, GPR44, and GREM2 could predict the 1-, 2-, 3-, 4-, and 5-year OS of TNBC patients and perform risk stratification. The risk score derived from the nomogram model could also predict the level of immune cell infiltration in tumors. These findings provided a reference for personalized prognosis prediction in TNBC patients and might be potential immune biomarkers for designing novel therapy.

Data Availability

The data used and analyzed during the current study are available from UCSC Xena platform (https://xenabrowser.net/datapages/) and GEO database (https://www.ncbi.nlm.nih.gov/geo/). The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that no competing interest exists.

Authors’ Contributions

ZYD contributed to the concept development and study design. PPW analyzed the data, performed bioinformatics training, and wrote the manuscript. YF analyzed the data and helped to write the manuscript. YYC, QL, and YH carried out data management and statistical analysis. TL helped with cohort identification and data management. All authors read and approved the final manuscript. Peipei Wang and Yang Fu contributed equally to this work.

Supplementary Materials

Supplementary 1. Figure S1: construction of a nomogram model with CCL25, IL29, TDGF3, KIR2DL4, GPR44, and GREM2 for predicting 1-, 2-, 3-, 4-, and 5-year OS in TNBC patients.

Supplementary 2. Figure S2: ROC curve verified the nomogram model constructed by CCL25, IL29, TDGF3, KIR2DL4, GPR44, and GREM2.

Supplementary 3. Table S1: the clinical information of TNBC patients.

Supplementary 4. Table S2: the risk scores for variables in nomograms.

References


