Immune Microenvironment-Related Genes Contribute to Clinical Prognosis in Patients with Triple-Negative Breast Cancer

Dandan Feng
Tianjin Medical University
Jie Gao
Capital Medical University
Tianshu Yang
Capital Medical University
Yanli Ren
Tianjin Medical University
Wei Liu
Tianjin Medical University
Gen Li
Capital Medical University
Jihao Wu
University of Science and Technology of China
Lin Zhou
University of Science and Technology of China
Yajuan Feng
University of Science and Technology of China
Xu Teng
Capital Medical University
Hefen Yu (yhf23@ccmu.edu.cn)
Capital Medical University https://orcid.org/0000-0002-1139-3976
Wei Huang
Capital Medical University

Primary research

Keywords: triple-negative breast cancer, tumor microenvironment, immune microenvironment, immune score, stromal score

Posted Date: June 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-630909/v1

Abstract

Background:

Triple-negative breast cancer (TNBC) is a high-risk breast cancer subtype, which accounts for 15% to 20% of all breast cancers and generally has a poor prognosis. TNBC patients have high recurrence post-surgery and high risk of metastasis to other organs. The tumor microenvironment (TME) plays important roles in the carcinogenesis, development, and metastasis of tumors.

Methods:

This study aimed to investigate the effects of immune and stromal cell-related genes on TNBC prognosis. The ESTIMATE algorithm was used to calculate the immune/stromal scores of TNBC samples from the GEO database. The samples were divided into high- and low- score groups and the differential expression genes were identified using the limma package within R. Functional enrichment and protein-protein interaction (PPI) network analyses revealed that these genes are primarily involved in immune responses.

Results:

Survival analysis in both GSE21653 and KM-plotter website showed that 36 genes were significantly related to disease-free survival (DFS) of TNBC. Another survival analysis by R package survival in GSE58812 indicated that 14 genes of 36 were greatly interrelated to DFS of TNBC. Moreover, the expression levels of some of these genes were verified through immunohistochemical staining and RT-qPCR. Finally, four genes importantly associated with TNBC prognosis were identified with Cox-LASSO analysis. Time-dependent receiver-operating characteristic (ROC) analysis displayed an area under the curve (AUC) of 0.95 for one-year survival rate, indicating the four genes performed very well for prognosis prediction.

Conclusions:

In conclusion, we identified four genes, including BIRC3, CD8A, GNLY and TRIM22, that are possibly associated with the TME and are potential prognostic and therapeutic markers of TNBC.

Introduction

Breast cancer is the cancer with the highest morbidity and mortality among women [1]. Globally, there were estimated 2.1 million newly diagnosed female breast cancer cases in 2018, accounting for nearly a quarter of female cancer cases [2]. TNBC was diagnosed via immunohistochemical methods through the lack of amplified expression of three receptors - estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) [3]. TNBC accounts for 15–20% of all breast cancers and has a more obvious pattern of metastasis than other subtypes of the disease, with poor prognosis in patients [4, 5]. Many efforts have been made to understand the molecular mechanism of TNBC. The Cancer Genome Atlas (TCGA) project contributes to a comprehensive understanding of breast cancer-

specific molecular heterogeneity and driver mutations, including TNBC [6]. So far, the etiology and molecular mechanism of TNBC have not been well explained. Therefore, it is particularly important to find the prognostic biomarkers in TNBC patients and explore the molecular mechanisms underlying the high morbidity and mortality of TNBC.

TME refers to the internal and external environment of tumor cells that are closely associated with the occurrence, growth, and metastasis of tumors. In the TME, tumor cells are able to adapt and proliferate with greatly reduced detection and eradication by host immune surveillance [7]. In addition to tumor cells, the TME mainly includes two non-tumor components immune cells and stromal cells, which are considered to be of great significance for tumor diagnosis and prognosis. Currently, the most promising way to activate therapeutic anti-tumor immunity is blocking immune checkpoints to reduce immunosuppression. Cancer immunotherapies targeting programmed cell death protein 1 (PD-1) and programmed cell death 1 ligand 1 (PD-L1) are changing traditional tumor treatment. Moreover, anti-PD-1 / PD-L1 antibodies have been proven to be of great clinical significance in more than 15 types of cancers, including melanoma, non-small cell lung cancer (NSCLC), and renal cell carcinoma (RCC) [8, 9]. The diverse immune microenvironment in TNBC greatly influences the risk of relapse, response to chemotherapy, and applications of immunotherapy. An immune checkpoint inhibitor is now ready for use in the study of neoadjuvant and adjuvant therapies for TNBC [10].

ESTIMATE algorithm is a tool that uses gene expression profile characteristics to predict the proportion of stromal and immune cells in tumors, and infer the purity of tumors in tissues [11]. Current ESTIMATE analysis has shown that stromal/immune cell infiltration is associated with improved prognosis in patients with various types of tumors, including glioblastoma and cutaneous melanoma [12, 13]. Nonetheless, there is currently no detailed analysis of TNBC immune/stromal scores.

In this study, in order to better understand the effects of immune and stromal cell-related genes on TNBC prognosis, we systematically analyzed gene expression profiles and excavated TME-related genes with poor prognosis to explore potential regulatory mechanisms.

Materials And Methods

Raw data

RNA-seq data and corresponding clinicopathological information of TNBC patients in GSE21653 and GSE58812 were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). Immune/stromal/estimate scores of the samples in GSE21653 were calculated using the ESTIMATE algorithm (https://r-forge.r-project.org) [11]. We selected 85 samples containing both DFS information and ESTIMATE scores for further analysis. The E-MTAB-365 dataset from KM-plotter website (http://kmplot.com/analysis/index.php?p=service&cancer=breast), containing 48 TNBC samples, was used to further confirm the survival analysis [14]. Survival curve parameters were provided in Table 1. The GSE58812 dataset was used for validation, Cox-Lasso regression analysis and ROC curve. The Human

Protein Atlas (http://www.proteinatlas.org) was used to validate the immunohistochemistry of genes with prognostic values.

Table 1
Survival curve parameters

Item	Parameter
Split patients by	Median
Survival	RFS
ER status	ER negative
PR status	PR negative
HER2 status	HER2 negative
Use following dataset for the analysis	E-MTAB-365

Survival Analysis

The survival curves used for GSE21653 and GSE58812 were plotted using the R package survival to analyze the relationship between patients' disease-free survival (DFS) rate (months) and high-low score group or high-low gene expression group, and compared using a Log Rank test whether there are differences in survival curves between the two groups of patients.

Differential Expression Analysis

Differential expression analysis was performed using the R package limma (Version: 3.42.2) [15]. The screening conditions for DEGs were | log2FC (fold change) | >1and FDR < 0.05.

Protein-protein Interaction Analysis (Ppi)

PPI analysis was conducted via STRING [16]. Then PPI network was reconstructed using Cytoscape software [17], which showed a network of more than 10 nodes. The node size was related to the node degree, with the thickness of the edges reflecting a score of the degree of interaction between the nodes, and the color of the nodes reflecting the differential expression degree. The MCODE tool of Cytoscape was used for PPI network analysis and showed the PPI modules of the DEGs.

Functional And Pathway Enrichment Analysis

STRING was used to perform functional enrichment analysis of DEGs, including GO, KEGG pathways, Reactome pathways, UniProt keywords, PFAM protein domains, INTERPRO protein domains and features,

and SMART protein domains. The GO analysis contains molecular functions (MF), biological processes (BP), and cellular components (CC).

Immunohistochemical Staining

Images were downloaded from The Human Protein Atlas (https://www.proteinatlas.org/). RNA expression of genes with prognostic values between normal tissues and TNBC were presented by GraphPad Prism (https://www.graphpad.com/).

Cell Culture

The cell lines used were obtained from the American Type Culture Collection. MCF-10A cells were cultured with the mammary epithelium growth factor medium (MEGM) kit supplemented with growth factors (Lonza). Hs 578T cells were maintained in Dulbecco's modified Eagle's medium (DMEM). Cells were maintained in a humidified incubator equilibrated with 5% CO_2 at 37°C. MDA-MB-231 cells were cultured in L-15 medium without CO_2. In addition to MEGM, all the other media were supplemented with 10% fetal bovine serum (FBS), 100 units/ml penicillin, and 100 mg/ml streptomycin (Gibco).

Rt-qpcr

Total RNA was extracted with Trizol reagent according to the manufacturer's instructions (Invitrogen). Potential DNA contamination was avoided with RNase-free DNase treatment (Promega). cDNA was prepared with MMLV Reverse Transcriptase (Roche). Relative quantitation of gene expression was performed using the ABI PRISM 7500 sequence detection system (Applied Biosystems) measuring real-time SYBR green fluorescence. The results were acquired using the comparative Ct method ($2^{-\triangle\triangle}$Ct) with GAPDH as an internal control. This experiment was performed at least thrice independently. The primer sequences used are listed in Supplementary Table 1.

Statistical analysis

We fit Ten-fold cross-validated Cox Survival Analysis and Least Absolute Shrinkage and Selection Operator Regression (Cox-LASSO regression) model, as implemented in R package glmnet and survival. The corresponding hazard ratio (HR), 95% confidence interval (CI), and p value were collected. The forest plot was drawn using the R package survminer.

Time-dependent receiver operating characteristic (ROC) curves were carried out to estimate the predictive accuracy for prognosis and describe the associated AUCs at years 1 ~ 10 based on the risk score by the R package pROC. From the curves, sensitivity, specificity, likelihood ratio, predictive value and their respective 95% confidence intervals can be seen. p-value < 0.05 were considered statistically significant.

Results

ESTIMATE and immune scores are closely related to the survival of TNBC

We downloaded clinical information of TNBC samples (GSE21653) from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). We analyzed the samples using ESTIMATE, immune and stromal scores. The distribution of scores of TNBC samples were shown in Fig. 1A.

Then we divided the TNBC samples into high- and low- score groups based on the median, and conducted survival analysis using the clinical follow-up information corresponding to the three scores for each group. As shown in Fig. 1B, the survival curve showed that the high- score group of the ESTIMATE score has a higher survival rate than the low- score group ($p = 0.0028$ in log-rank test), indicating the ESTIMATE score was significantly associated with disease-free survival (DFS) of TNBC samples from the GEO database ($p < 0.05$). Similar phenomena were observed in the high- and low- score groups of the immune score (Fig. 1C-1D).

Identification of differentially expressed genes based on the TNBC ESTIMATE score

Since the ESTIMATE score is a comprehensive evaluation of the immune and stromal score, we further explored the genes associated with TNBC prognosis based on the ESTIMATE score. Differentially expressed genes (DEGs) were first screened between the high- and low- ESTIMATE score groups. A total of 278 DEGs (276 upregulated and 2 downregulated) were identified ($|log2FC|>1$ and $p < 0.05$) (Supplementary Table 2). The DEGs in the high- and low- score groups were shown in Fig. 2A.

In order to study the function of these DEGs, we performed functional enrichment analysis on the 276 upregulated and 2 downregulated genes via the STRING database, including GO (molecular function, biological process and cellular component) and KEGG pathway analyses. The top ten enrichment terms for each section of GO and KEGG pathway were shown in Fig. 2B -2E (sorted by -log10 of Q value). GO functions indicated that these genes are mainly enriched in protein binding, immune system process and immune response and membrane part (Fig. 2B-2D). Moreover, cytokine-cytokine receptor interaction, chemokine signaling pathway were also obtained from KEGG pathways analysis (Fig. 2E).

Screening and functional analysis of genes related to TNBC prognosis

In order to screen genes associated with prognosis of TNBC, we performed survival analysis of all the DEGs, among which 171 genes were significantly related to DFS ($p < 0.05$) (Supplementary Table 3). The survival curves of six genes with the lowest p values from the 171 genes were shown in Fig. 3A, including SH2D1A, CST7, GPR18, LCP2, CLIC2 and ITK.

To reveal the potential role of DEGs related to DFS of TNBC, we performed PPI analysis of the 171 genes above via STRING (Fig. 3B). The core proteins included CD2, SELL, CCR5, IL10RA, and LCP2.

Subsequently, the GO enrichment and KEGG pathway analyses on the genes mined by survival analysis were carried out. The data showed these genes are mainly enriched in the TME and immune-related pathway (Figs. 3C-3F).

We integrated the protein interaction of the 171 DFS-related genes from the PPI network module via MCODE analysis, the top four of the eight modules obtained were selected for further investigation. The protein interaction of these four modules were shown in Fig. 4, with the core nodes having higher degrees. Module 1 contained a total of 24 nodes and 245 edges. Among them, SELL, ITGAL, CD8A, CD52, and CD2 have been confirmed as cell adhesion markers, which are involved in a series of important physiological and pathological processes, such as immune response, tumor metastasis, and wound healing (hsa04514) [18–20]. Module 2 contained a total of 20 nodes and 70 edges. Among them, C1QB, HLA-DRA, C3, and HLA-DPA1 have been confirmed to be associated with *S. aureus* infection and systemic lupus erythematosus, indicating that these genes are closely related to immune response (hsa05150, hsa05322) [21]. Module 3 contained a total of 20 nodes and 64 edges. Among them, CASP1, GBP4, and GBP5 have been confirmed to be related to Nod-like receptor signaling pathway, which is an important way for eukaryotes to recognize pathogens (hsa04621) [22]. Module 4 contained a total of 12 nodes and 28 edges, four (CD4, CCR5, CD3D, and ITK) of which have been reported to be closely related to immune response (hsa04658, hsa04060, hsa04660) [23, 24]. As shown in Fig. 4, SELL, ITGAL, CD69, and HLA-DRA had higher node degrees, indicating that they might be important immune microenvironment-related genes of TNBC. Among these 171 DFS-related genes, 11 genes have been reported to be significantly related with breast cancer prognosis, including CD3D, CD8A, CORO1A, GZMB, LCK, TRBC1, HLA-DRA, ACSL5, EOMES, IRF4, IRF8. In addition, CD3D,CD8A□CORO1A□GZMB□EOMES and IRF8 have been reported to be related with the prognosis of TNBC [25–29].The association of the remaining genes with TNBC prognosis, including CD247, CD3E, LAX1, LPXN, PRKCB, and SIRPG, have not been reported. These genes may be potential immune microenvironment-related prognostic markers of TNBC.

Relationships Between Immune Microenvironment-related Genes And Tnbc Patient Prognosis

DFS-related genes (n = 171) were further analyzed in KM-plotter website (http://kmplot.com/analysis/index.php?p=service&cancer=breast) and the expression levels of 36 genes were significantly correlated with DFS in E-MTAB-365 dataset ($p < 0.01$). Detailed information of the 36 genes was provided in Table 2. Additionally, GSE58812 dataset with clinical information of TNBC samples were downloaded from the GEO database and were used to verify the correlation of the above 36 genes with prognosis of TNBC samples. These genes were further analyzed by univariate Cox regression using the R-package “survival”. The detailed information of univariate log-rank was provided in Supplementary Table 4 and only 14 genes were significantly correlated with DFS. Kaplan–Meier survival analysis showed that patients with higher expression of these genes have better disease-free survival (Fig. 5). The detailed information of the 14 genes was provided in Table 3. Among these genes, SELL, GZMB, IL2RB, LCP2, and CD8A were involved in Module 1, CASP1 and TRIM22 were involved in Module 3,

EOMES and ITK were involved in Module 4. These genes may be potential genes for the poor prognosis of TNBC and may provide therapeutic value for TNBC in the future.

Table 2
Genes significantly associated with disease-free urvival in the GSE21653 dataset and verified by KM plotter

Categories	Gene symbols
T cell receptor complex	**CD247**, *CD3D, *CD8A
immunological synapse	**CD3E**, *CORO1A, *GZMB, LCK
plasma membrane	**LAX1**, **LPXN**, **PRKCB**, **SIRPG**, TRBC1, **TRAC**
cell surface	**CD19**, **CD27**, HLA-DRA, **IL2RB**, **SELL**
cell-cell junction	**CRTAM**, **ITK**, **LCP2**
membrane part	ACSL5, **BIRC3**, **CLIC2**
cytoplasmic vesicle	**AOAH**, **CTSS**, **GNLY**, **GPR18**
cell part	**CASP1**, **CST7**, *EOMES, IRF4, *IRF8, **MZB1**, **SH2D1A**, **TRIM22**

The genes in normal font are reported genes related to breast cancer prognosis; the genes marked with an asterisk (*) are reported genes related to TNBC prognosis; the genes in bold font have not been reported to be related to breast cancer prognosis.

Table 3
Genes significantly associated with survival in the GSE58812 dataset and verified by R package

Gene symbol	Gene ID	Biological Process (GO)
BIRC3	330	GO:0001959 regulation of cytokine-mediated signaling pathway
CASP1	834	GO:0065009 regulation of molecular function
CD8A	925	GO:0007165 signal transduction
CLIC2	1193	GO:0003824 catalytic activity
EOMES	8320	GO:0010467 gene expression
GNLY	10578	GO:0050896 response to stimulus
GZMB	3002	GO:0007165 signal transduction
ITK	3702	GO:0071704 organic substance metabolic process
LCP2	3937	GO:0035592 establishment of protein localization to extracellular region
SELL	6402	GO:0010033 response to organic substance
CRTAM	56253	GO:0008104 protein localization
IL2RB	3560	GO:0019899 enzyme binding
TRBC1	28639	GO:0050776 regulation of immune response
TRIM22	10346	GO:0009894 regulation of catabolic process

Verification Of Immune Microenvironment-related Genes In Tnbc

To further confirm the reliability of these immune microenvironment-related genes in TNBC prognosis, we verified the expression patterns of some genes. Firstly, we used immunohistochemical staining to detect the expression of 14 genes in normal mammary tissues and breast cancer samples from The Human Protein Atlas website. Results showed that BIRC3, CASP1, CD8A, EOMES and TRIM22 were significantly downregulated in breast carcinoma samples (Figs. 6A and 6B). Meanwhile, the relative mRNA expression levels of these 14 genes were detected by real-time RT-PCR (RT-qPCR) in MCF-10A, MDA-MB-231 and Hs 578T cells. The results showed that BIRC3, CASP1, CLIC2, EOMES, GZMB, IL2RB, and TRIM22 were notably downregulated in MDA-MB-231 and Hs 578T cells compared with MCF-10A cells (Fig. 6C). All the results showed that these immune microenvironment-related genes might be good prognostic biomarkers of TNBC.

Multivariate Cox-LASSO regression assay was used to further validate the significance of the 14 genes. BIRC3 (hazard ratio, 0.68; 95% CI, 0.43–1.1; p = 0.1), CD8A (hazard ratio, 0.89, 95% CI, 0.67–1.2; p =

0.439), GNLY (hazard ratio, 0.98; 95% CI, 0.73–1.3; p = 0.895), TRIM22 (hazard ratio, 0.72; 95% CI, 0.44–1.2; p = 0.195) were further screened out which might have tight associations with prognosis (Fig. 7A). Subsequently, the time-dependent receiver operating characteristic (ROC) curve of these four genes for the prognosis was completed. The predictive accuracy for one-year survival rate was remarkable, yielding an AUC value of 0.95. In general, these four immune microenvironment-related genes have a greatly importance in TNBC prognosis.

Discussion

In this paper, we used the ESTIMATE algorithm [11] to calculate the Estimate/immune/stromal scores of gene expression data of TNBC samples from GEO database GSE21653. These samples were divided into high- and low-score groups based on the ESTIMATE scores. By analyzing the differential expression genes between the high- and low-score groups, we acquired 278 DEGs. After functional enrichment analysis of upregulated and downregulated genes, many genes were found to be related to TME and immunity. The results were consistent with previous studies, which reported the important role of immune cells and stromal cells in TME [12, 30–32]. A total of 171 core genes were extracted from the survival analysis and analyzed by PPI network analysis and modules. Functional enrichment analysis of these genes indicated that they were related to immune response. We then performed a KM plotter analysis of the 171 DFS-related genes, further analyzed them, and identified 36 genes significantly associated with DFS, which could be used as TNBC prognosis biomarkers. These 36 genes were reanalyzed by univariate Cox regression analysis in GSE58812 dataset. Only 14 genes of 36 were significantly related to DFS of TNBC. Subsequent immunohistochemical staining and RT-qPCR verification revealed that these genes were likely to be prognostic markers for TNBC. Finally, Cox LASSO regression assay revealed that BIRC3, CD8A, GNLY and TRIM22, might be potential immune microenvironment-related prognostic markers of TNBC.

BIRC3 (baculoviral IAP repeat containing 3) participates in immunity activities by regulating NF-κB signaling and other inflammatory signals. It also acts as an E3 ubiquitin protein ligase in the TME in mice. The TNFa-TNFR2-BIRC3-TRAF1 signaling pathway has been shown to promote metastases in mice, and the activation of this pathway is associated with the poor prognosis of gastrointestinal stromal tumor patients [33].

CD8A (CD8 antigen) is a cell surface glycoprotein on most cytotoxic T lymphocytes, which mediates effective cell-to-cell interactions in the immune system. It functions as a coreceptor with T cell receptors on T lymphocytes and recognizes antigens displayed by antigen-presenting cells in class I MHC molecules. As previously reported that, CD8A was predictive for increased pathologic complete response (pCR) in the neoadjuvant GeparSixto trial [19]. The CD8A gene is associated with an improved outcome in several public breast cancer datasets. High expression of CD8A is related with a better relapse free survival in TNBC [26]. Our results further supports the role of CD8A in TNBC.

GNLY (granulysin) is part of the saposin-like protein (SAPLIP) family and is located in the cytotoxic granules of T cells and released after antigen stimulation. GNLY may tempt ER stress-mediated apoptosis [34]. It is correlated to the ability of NK-extracellular vesicles (EV) to induce cytotoxicity [35]. Additionally, serum GNLY may be a potential biomarker for nasopharyngeal carcinoma, the early stage of colorectal adenocarcinoma and muscle-invasive bladder cancer [36–38]. The function of GNLY in TNBC immune microenvironment and immunotherapies deserves further research.

TRIM22 (Stimulated Trans-Acting Factor of 50 kDa, Staf-50) is a E3 ubiquitin-ligase and a member of the C-IV group of tripartite motif (TRIM) family, which is strongly induced by interferon stimulation and takes part in innate immunity of cells. Besides the antiviral effects, TRIM22 is a potential therapeutic target and prognosis marker for NSCLC [39]. We will further explore the expression and function of TRIM22 in TNBC.

Our results were due to the publicly available data and ESTIMATE algorithm. Multiple datasets were used to verify that the selected genes were reliable and common. The method “CIBERSORT” was used to analyze microarray data and not for TCGA RNA-seq data [40]. The method “TIMER” has limited sample size and relevance for distinguishing the location of immune cells in the stroma or tumor or capturing tumor cell heterogeneity [41]. The ESTIMATE algorithm is applicable in microarray expression data sets, new microarray and RNA-seq-based transcriptome profiles. The predictive ability of this method has been verified in large independent data sets. This method provides a good indication of biomarkers for tumor prognosis, provides a basis for the next step of research, links the tumor microenvironment and tumor prognosis together, and contributes to the better resolution of tumor prognosis.

This study has certain limitations. The ROC curve analysis shows that these 4 genes can predict the prognostic effect of 1 year, but there is a limitation when considering 3 or 5 years or long-term prognosis, and other predictive markers may be needed.

In conclusion, based on the ESTIMATE algorithm, we systematically analyzed the expression profile of TNBC and screened out some genes related to TME and prognosis. These genes may be potential prognostic markers of the TNBC immune microenvironment. Future work will be focused on comprehensively exploring the pathogenic mechanism of these genes in the immune microenvironment, and the potential associations between TME and TNBC prognosis. Our findings have greatly contributed to a better understanding of the complex regulatory network of TNBC and of immune and stromal cell-related genes on TNBC progression. These findings may provide new promising biomarkers for TNBC therapy.

Declarations

Ethics approval and consent to participate

The study was authorized by the Ethics Committee of Capital Medical University, and the written informed consent was signed by each patient.

Consent for publication

All authors give consent for the publication of the manuscript in Cancer Cell International.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 81902960 (to WH)and 81672726 (to HY)], and the Natural Science Foundation of Beijing (7204241 to WH).

Authors' contributions

JG, HY and WH contributed conception and designed the research study. DF, WL and GL collected and organised the database. JW, LZ and YF performed data analysis. YR and XT performed the validation experiments. DF and JG wrote the manuscript. HY and WH revised the manuscript. All authors read and finally approved the manuscript.

Acknowledgements

We would like to thank Editage (www.editage.cn) for English language editing.

References

1. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. CA: a cancer journal for clinicians. 2016; 66: 115-132.

2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018; 68: 394-424.

3. Brown M, Tsodikov A, Bauer KR, Parise CA, Caggiano V. The role of human epidermal growth factor receptor 2 in the survival of women with estrogen and progesterone receptor-negative, invasive breast cancer: the California Cancer Registry, 1999-2004. Cancer. 2008; 112: 737-747.

4. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P, Narod SA. Triple-negative breast cancer: clinical features and patterns of recurrence. Clinical cancer research : an official journal of the American Association for Cancer Research. 2007; 13: 4429-4434.

5. Fulford LG, Reis-Filho JS, Ryder K, Jones C, Gillett CE, Hanby A, Easton D, Lakhani SR. Basal-like grade III invasive ductal carcinoma of the breast: patterns of metastasis and long-term survival. Breast cancer research : BCR. 2007; 9: R4.

6. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490: 61-70.

7. Pitt JM, Marabelle A, Eggermont A, Soria JC, Kroemer G, Zitvogel L. Targeting the tumor microenvironment: removing obstruction to anticancer immune responses and immunotherapy. Annals of oncology : official journal of the European Society for Medical Oncology. 2016; 27: 1482-1492.

8. Sharma P, Allison JP. The future of immune checkpoint therapy. Science (New York, NY). 2015; 348: 56-61.

9. Francis DM, Thomas SN. Progress and opportunities for enhancing the delivery and efficacy of checkpoint inhibitors for cancer immunotherapy. Advanced drug delivery reviews. 2017; 114: 33-42.

10. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. Nature reviews Clinical oncology. 2016; 13: 674-690.

11. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nature communications. 2013; 4: 2612.

12. Jia D, Li S, Li D, Xue H, Yang D, Liu Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. Aging. 2018; 10: 592-605.

13. Yang S, Liu T, Nan H, Wang Y, Chen H, Zhang X, Zhang Y, Shen B, Qian P, Xu S, Sui J, Liang G. Comprehensive analysis of prognostic immune-related genes in the tumor microenvironment of cutaneous melanoma. J Cell Physiol. 2020; 235: 1025-1035.

14. Nagy A, Lanczky A, Menyhart O, Gyorffy B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. Sci Rep. 2018; 8: 9227.

15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research. 2015; 43: e47.

16. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic acids research. 2015; 43: D447-452.

17. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003; 13: 2498-2504.

18. Harland KL, Day EB, Apte SH, Russ BE, Doherty PC, Turner SJ, Kelso A. Epigenetic plasticity of Cd8a locus during CD8(+) T-cell development and effector differentiation and reprogramming. Nat Commun. 2014; 5: 3547.

19. Denkert C, von Minckwitz G, Brase JC, Sinn BV, Gade S, Kronenwett R, Pfitzner BM, Salat C, Loi S, Schmitt WD, Schem C, Fisch K, Darb-Esfahani S et al. 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. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2015; 33: 983-991.

20. Balada E, Castro-Marrero J, Felip L, Ordi-Ros J, Vilardell-Tarres M. Clinical and serological findings associated with the expression of ITGAL, PRF1, and CD70 in systemic lupus erythematosus. Clin Exp Rheumatol. 2014; 32: 113-116.

21. Niehrs A, Garcia-Beltran WF, Norman PJ, Watson GM, Holzemer A, Chapel A, Richert L, Pommerening-Roser A, Korner C, Ozawa M, Martrus G, Rossjohn J, Lee JH et al. A subset of HLA-DP molecules serve as ligands for the natural cytotoxicity receptor NKp44. Nat Immunol. 2019; 20: 1129-1137.

22. Wang W, Nguyen LT, Burlak C, Chegini F, Guo F, Chataway T, Ju S, Fisher OS, Miller DW, Datta D, Wu F, Wu CX, Landeru A et al. Caspase-1 causes truncation and aggregation of the Parkinson's disease-associated protein alpha-synuclein. Proc Natl Acad Sci U S A. 2016; 113: 9587-9592.

23. Jiao X, Nawab O, Patel T, Kossenkov AV, Halama N, Jaeger D, Pestell RG. Recent Advances Targeting CCR5 for Cancer and Its Role in Immuno-Oncology. Cancer Res. 2019; 79: 4801-4807.

24. Huang W, Solouki S, Koylass N, Zheng SG, August A. ITK signalling via the Ras/IRF4 pathway regulates the development and function of Tr1 cells. Nat Commun. 2017; 8: 15871.

25. Klintman M, Buus R, Cheang MC, Sheri A, Smith IE, Dowsett M. Changes in Expression of Genes Representing Key Biologic Processes after Neoadjuvant Chemotherapy in Breast Cancer, and Prognostic Implications in Residual Disease. Clinical cancer research : an official journal of the American Association for Cancer Research. 2016; 22: 2405-2416.

26. Araujo JM, Gomez AC, Aguilar A, Salgado R, Balko JM, Bravo L, Doimi F, Bretel D, Morante Z, Flores C, Gomez HL, Pinto JA. Effect of CCL5 expression in the recruitment of immune cells in triple negative breast cancer. Scientific reports. 2018; 8: 4899.

27. He Y, Jiang Z, Chen C, Wang X. Classification of triple-negative breast cancers based on Immunogenomic profiling. Journal of experimental & clinical cancer research : CR. 2018; 37: 327.

28. Maeda T, Hiraki M, Jin C, Rajabi H, Tagde A, Alam M, Bouillez A, Hu X, Suzuki Y, Miyo M, Hata T, Hinohara K, Kufe D. MUC1-C Induces PD-L1 and Immune Evasion in Triple-Negative Breast Cancer. Cancer research. 2018; 78: 205-215.

29. Wang S, Beeghly-Fadiel A, Cai Q, Cai H, Guo X, Shi L, Wu J, Ye F, Qiu Q, Zheng Y, Zheng W, Bao PP, Shu XO. Gene expression in triple-negative breast cancer in relation to survival. Breast cancer research and treatment. 2018; 171: 199-207.

30. Guo S, Deng CX. Effect of Stromal Cells in Tumor Microenvironment on Metastasis Initiation. International journal of biological sciences. 2018; 14: 2083-2093.

31. Bussard KM, Mutkus L, Stumpf K, Gomez-Manzano C, Marini FC. Tumor-associated stromal cells as key contributors to the tumor microenvironment. Breast cancer research : BCR. 2016; 18: 84.

32. Choi Y, Kim JW, Nam KH, Han SH, Kim JW, Ahn SH, Park DJ, Lee KW, Lee HS, Kim HH. Systemic inflammation is associated with the density of immune cells in the tumor microenvironment of gastric cancer. Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association. 2017; 20: 602-611.

33. Ivagnes A, Messaoudene M, Stoll G, Routy B, Fluckiger A, Yamazaki T, Iribarren K, Duong CPM, Fend L, Caignard A, Cremer I, LeCesne A, Adam J et al. TNFR2/BIRC3-TRAF1 signaling pathway as a novel NK cell immune checkpoint in cancer. Oncoimmunology. 2018; 7: e1386826.

34. Saini RV, Wilson C, Finn MW, Wang T, Krensky AM, Clayberger C. Granulysin delivered by cytotoxic cells damages endoplasmic reticulum and activates caspase-7 in target cells. Journal of immunology (Baltimore, Md : 1950). 2011; 186: 3497-3504.

35. Wu CH, Li J, Li L, Sun J, Fabbri M, Wayne AS, Seeger RC, Jong AY. Extracellular vesicles derived from natural killer cells use multiple cytotoxic proteins and killing mechanisms to target cancer cells. Journal of extracellular vesicles. 2019; 8: 1588538.

36. Lin J, Huang Y, Zhang L, Tang W, Li X, Wang X, Liu W. Evaluation of serum granulysin as a potential biomarker for nasopharyngeal carcinoma. Clinica chimica acta; international journal of clinical chemistry. 2016; 454: 72-76.

37. Janikowska G, Janikowski T, Pyka-Pająk A, Mazurek U, Janikowski M, Gonciarz M, Lorenc Z. Potential biomarkers for the early diagnosis of colorectal adenocarcinoma - transcriptomic analysis of four clinical stages. Cancer biomarkers : section A of Disease markers. 2018; 22: 89-99.

38. Jiang W, Zhu D, Wang C, Zhu Y. An immune relevant signature for predicting prognoses and immunotherapeutic responses in patients with muscle-invasive bladder cancer (MIBC). Cancer medicine. 2020; 9: 2774-2790.

39. Liu L, Zhou XM, Yang FF, Miao Y, Yin Y, Hu XJ, Hou G, Wang QY, Kang J. TRIM22 confers poor prognosis and promotes epithelial-mesenchymal transition through regulation of AKT/GSK3beta/beta-catenin signaling in non-small cell lung cancer. Oncotarget. 2017; 8: 62069-62080.

40. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD, Diehn M, West RB, Plevritis SK et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nature medicine. 2015; 21: 938-945.

41. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S, Signoretti S, Liu JS, Liu XS. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome biology. 2016; 17: 174.

Supplementary Files

Supplemental tables are not available with this version.

Figures

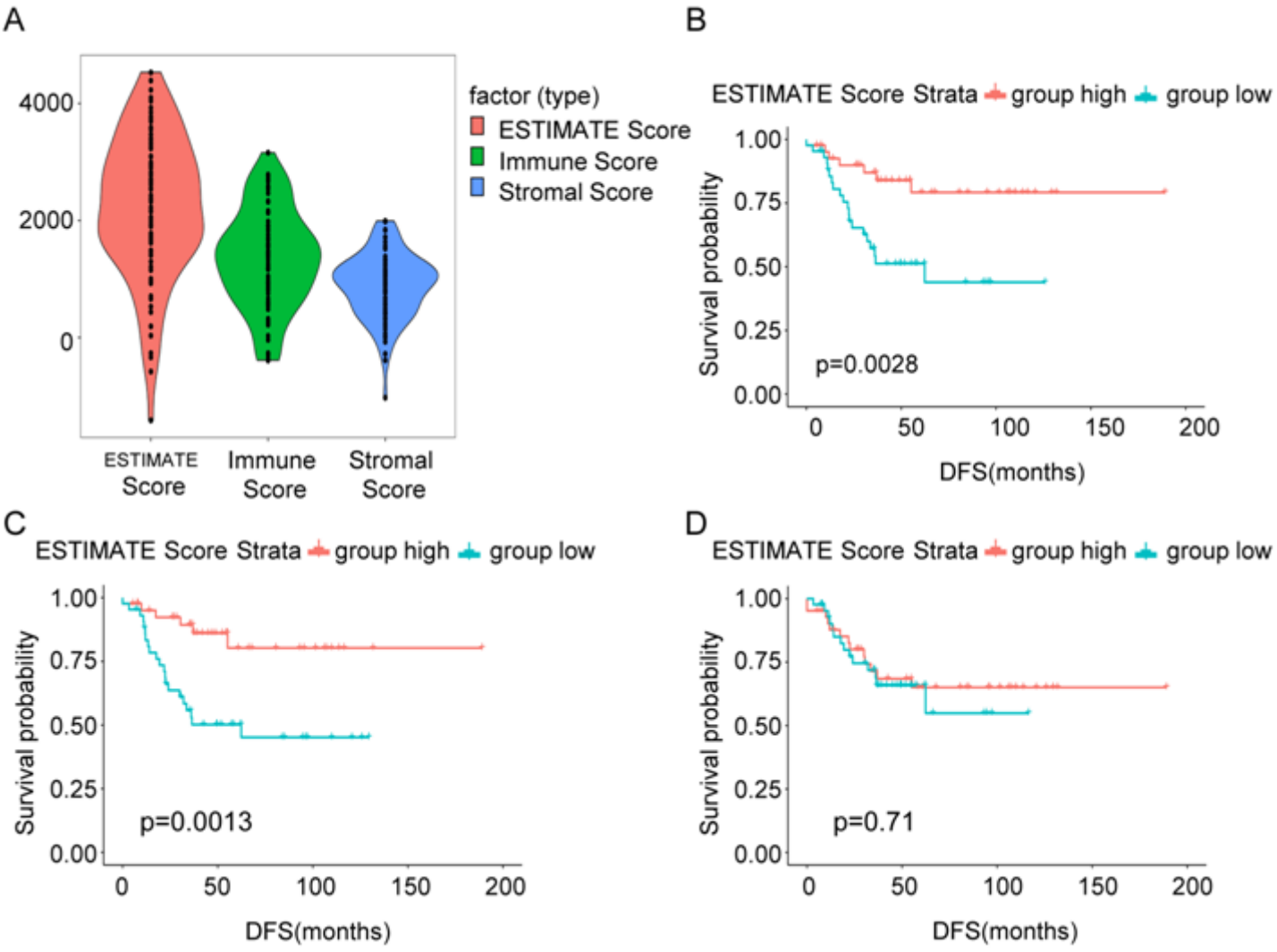

Figure 1

ESTIMATE and immune scores are closely related to the survival of TNBC. (A) Distribution of ESTIMATE, immune, and stromal scores of TNBC samples. The violin plot shows significant correlations between the TNBC samples and the ESTIMATE, immune, and stromal scores.(B) The three scores of TNBC samples were divided into high- and low-scoring groups (median as a standard). Survival analysis was performed using the clinical follow-up data corresponding to each sample. As shown in the figure, the ESTIMATE score was significantly correlated with the DFS of TNBC samples in the GEO database (p<0.05).

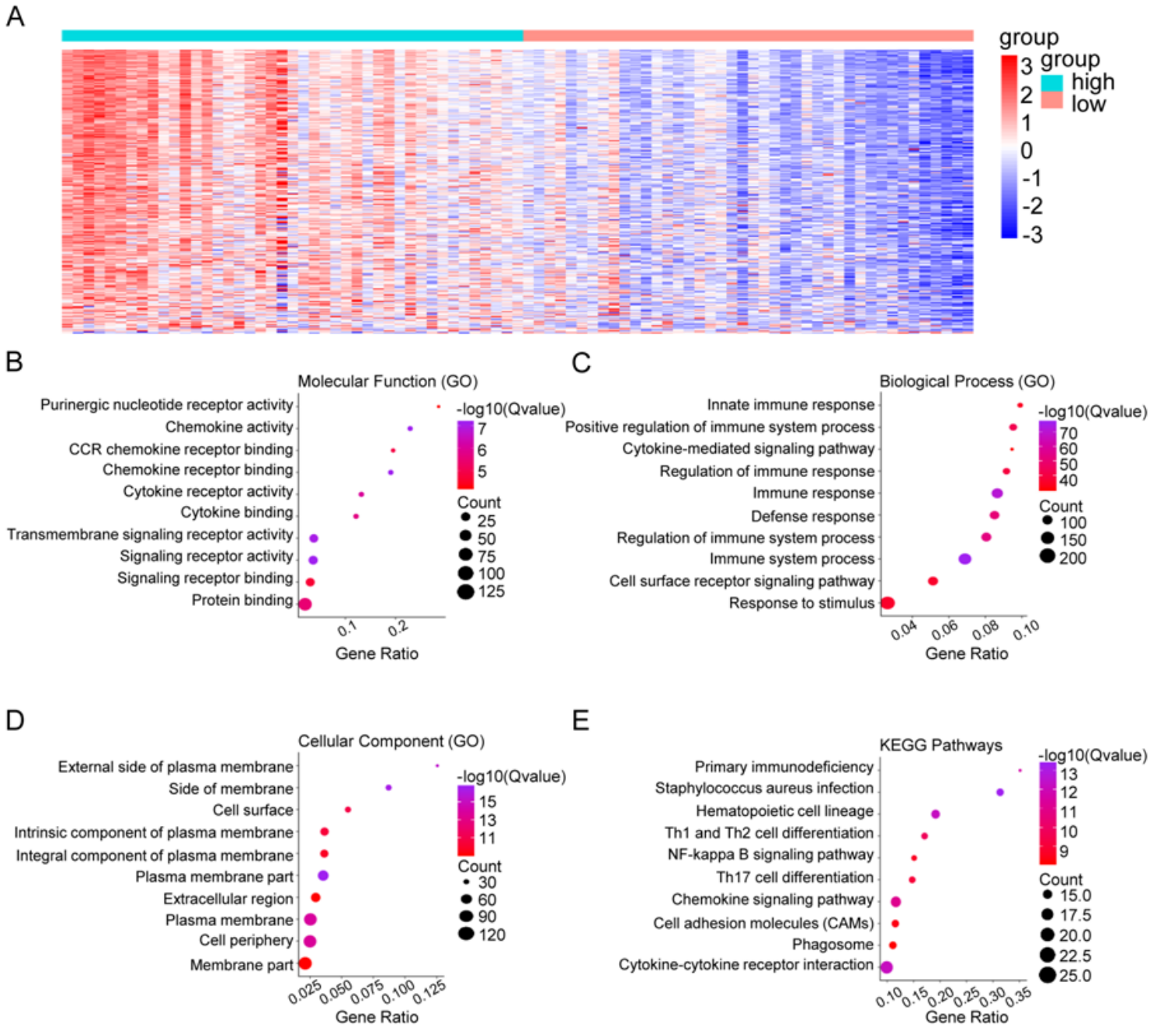

Figure 2

Identification of differentially expressed genesbased on the TNBC ESTIMATE score.(A) Heatmap representation of DEGs (|log2FC| >1, p<0.05). Each row represents a gene and each column, a sample. The samples are sorted according to the ESTIMATE score from high to low and from left to right. The

blue group on the left is a sample from the ESTIMATE high score group, and the right pink group is a sample from the ESTIMATE low score group. The genes were ranked according to the p value of the differential expression analysis from low to high. Red represents high gene expression and blue represents low gene expression. The darker the red or blue color, the greater the degree of difference in gene expression. (B-E) Top ten enrichment results for GO annotation and KEGG pathway enrichment on all 278 DEGs using the STRING database. (B) Molecular function. (C) Biological process. (D) Cellular component. (E) KEGG pathways.

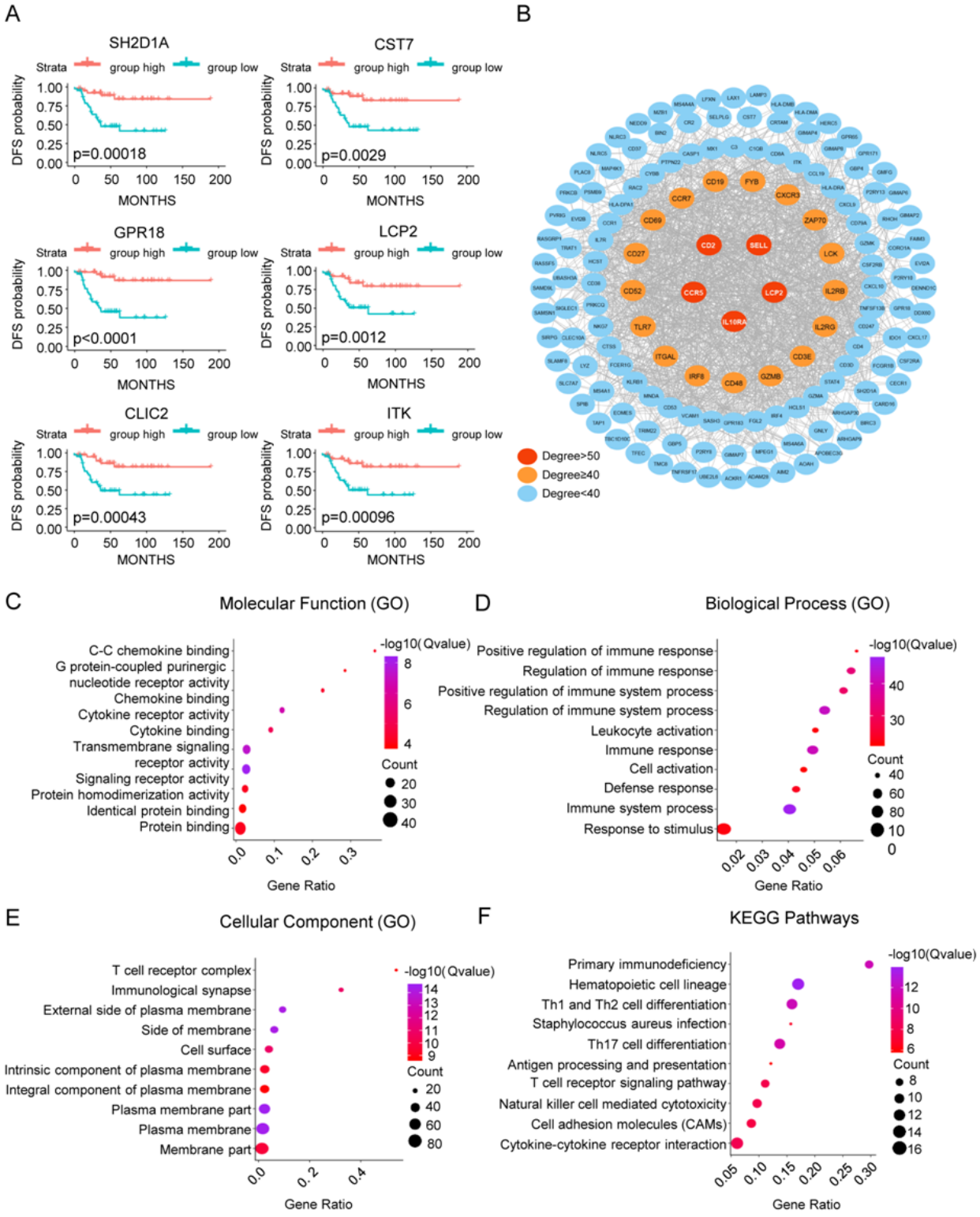

Figure 3

Functional analysis of DEGs related to TNBC prognosis. (A) Survival curves for 6 representative genes of 171 differentially expressed genes (DEGs) significantly associated with DFS ($p<0.05$). (B) A PPI network of 171 DFS-related genes was constructed using the STRING tool and it contained a total of 145 nodes and 1438 edges. The color of a node in the PPI network reflects the log (FC) value and the size of the node represents the degree. The thickness of the edge reflects the comprehensive score of the degree of interaction between the nodes. (C-F) Top ten enrichment results for GO annotation and KEGG pathway enrichment of the PPI network using the STRING database. (C) Molecular function. (D) Biological process. (E) Cellular component. (F) KEGG pathways.

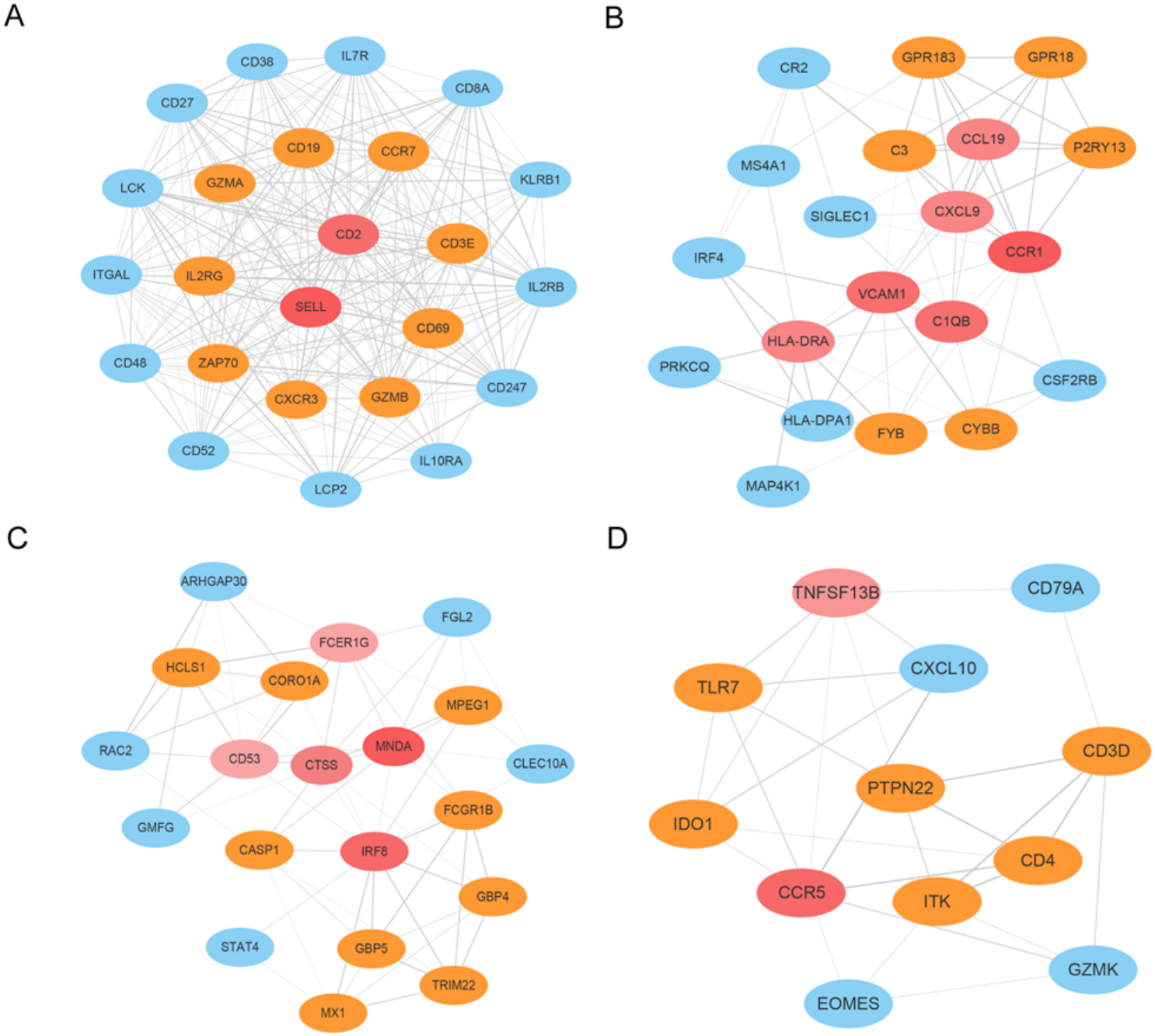

Figure 4

Associations among the 171 DFS-related genes in TNBC prognosis. (A-D) Module analysis of PPI network. The color of a node in the PPI network reflects the log (FC) value and the size of the node represents the degree. The thickness of the edge reflects the comprehensive score of the degree of interaction between the nodes. (A) Module 1 contains 24 nodes and 245 edges. (B) Module 2 contains 20 nodes and 70 edges. (C) Module 3 contains 20 nodes and 64 edges. (D) Module 4 contains 12 nodes and 28 edges.

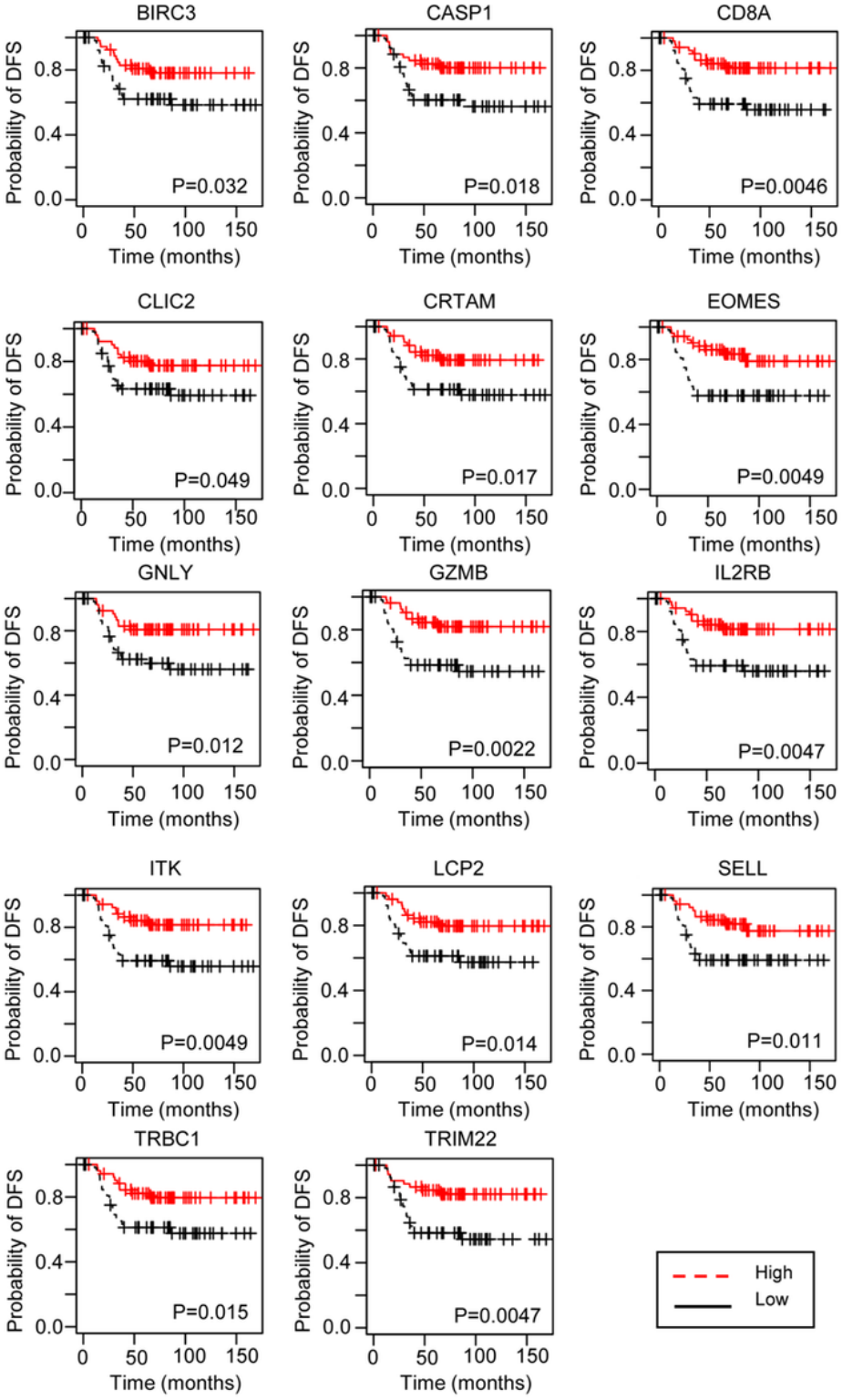

Figure 5

Relationships between immune microenvironment-related genes and TNBC patient prognosis. Kaplan–Meier survival analysis of the relationship between survival time and 14 genes expression signatures by R package survival. These 14 genes are likely to be prognostic markers for the immune microenvironment of TNBC.

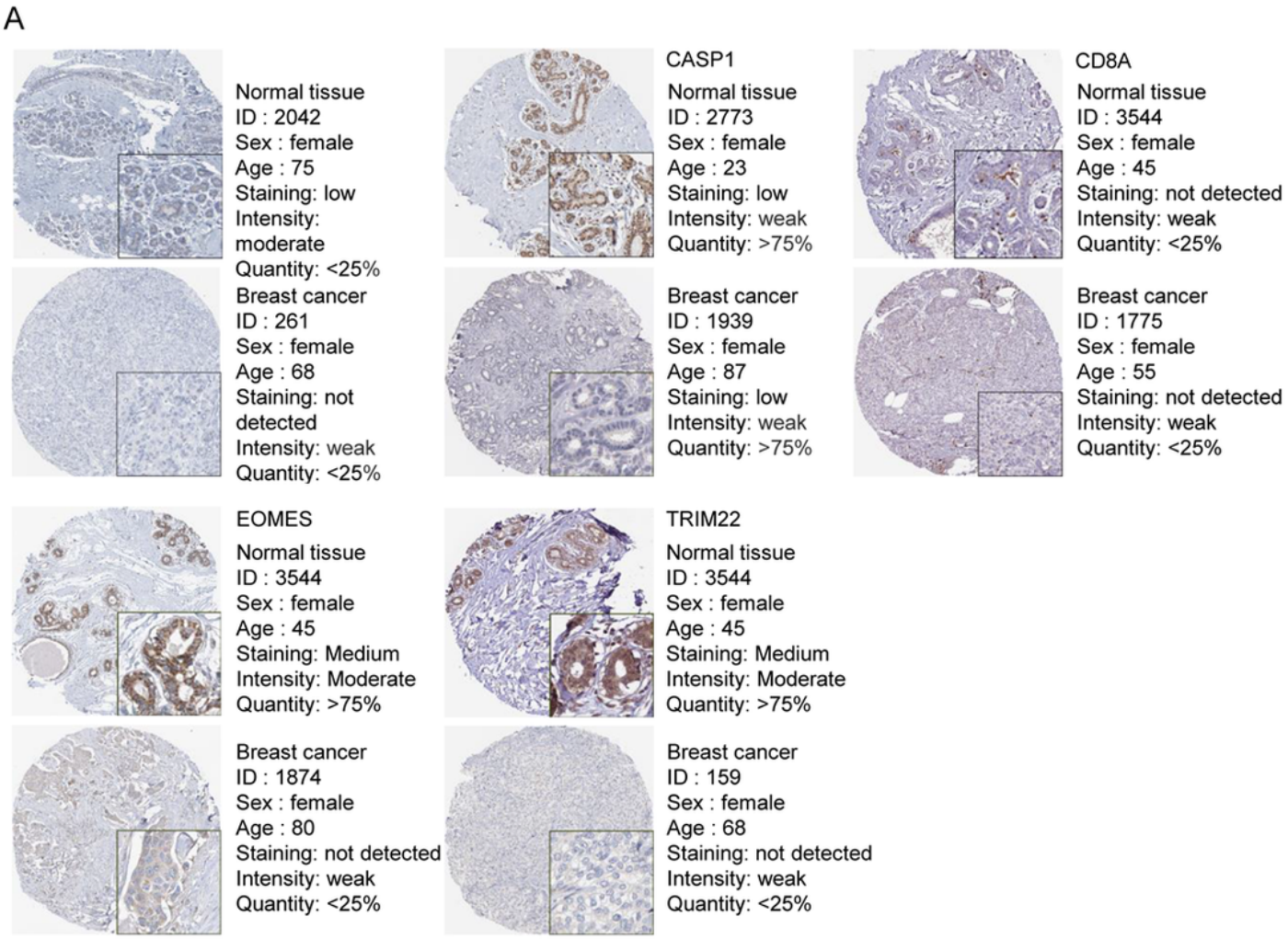

Figure 6

Verification of immune microenvironment-related genes in TNBC. (A) Immunohistochemical staining of BIRC3, CASP1, CD8A, EOMES and TRIM22 in normal breast tissues and breast carcinoma samples. (B) RNA expression of genes with prognostic values between normal mammary tissues and TNBC in The Human Protein Atlas database. num(N), Normal sample size; num(T), Tumor sample size.***p<0.001 by a two-tailed unpaired t test. (C) The expression levels of BIRC3, CASP1, CLIC2, EOMES, GZMB, IL2RB, and TRIM22 were measured by RT-qPCR in MCF-10A, MDA-MB-231 and Hs 578T cells. The mRNA levels were normalized to GAPDH. Error bars represent the means ± SDs of three independent experiments (**p<0.01, ***p<0.001, two-tailed unpaired t test).

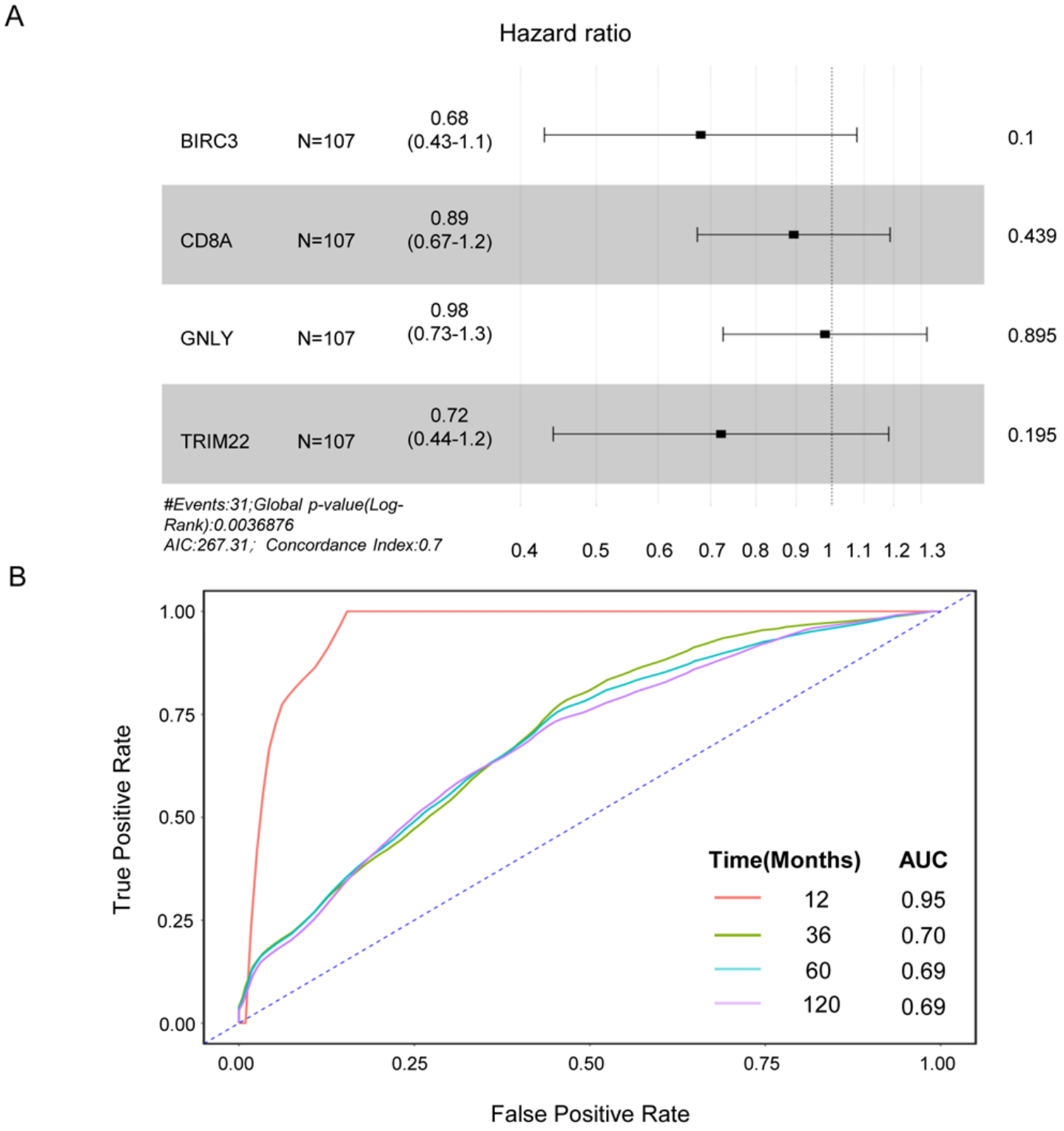

Figure 7

Construction and validationof immune microenvironment-related gene panel associated with the prognosis of TNBC. (A) Multivariate Cox-LASSOregression analyses of the association between immune microenvironment-related genes and disease-free survival. (B) ROC curves displayed the predictive power of the four immune microenvironment-related genes for the time-dependent disease-free survival rate.