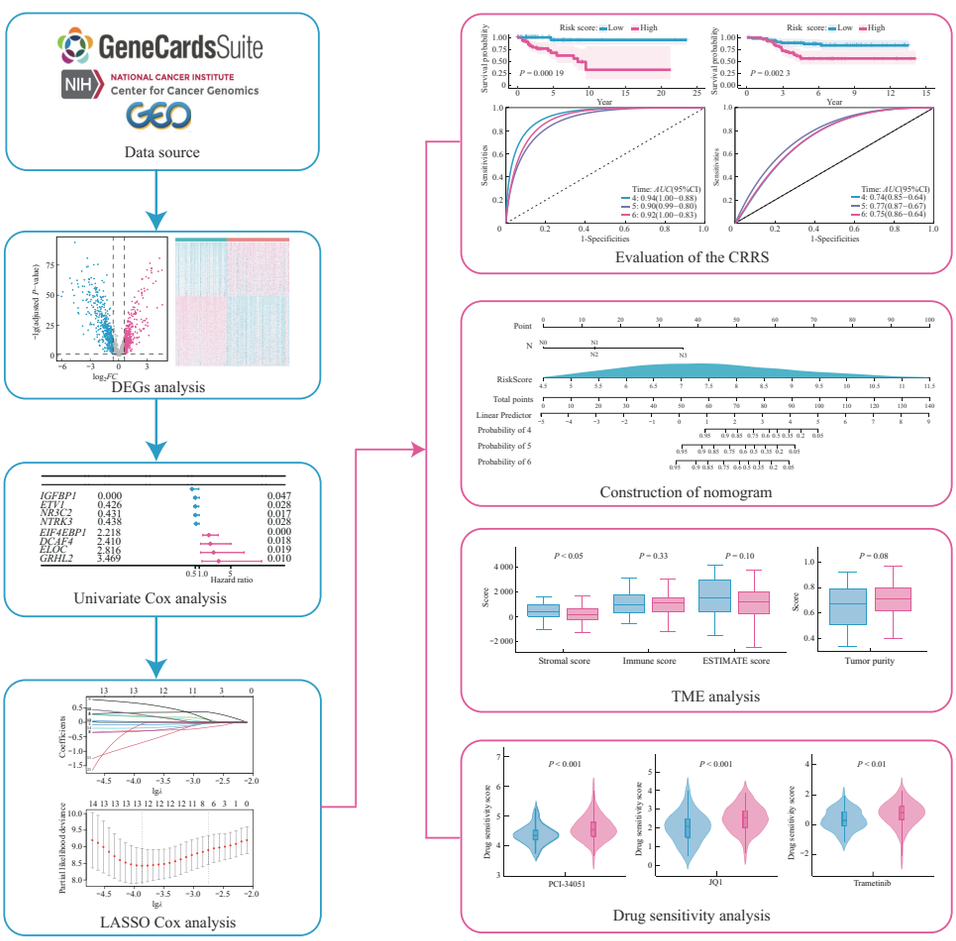
PIBB 生物化学与生物物理进展
Progress in Biochemistry and Biophysics
2024, 51(10): 2741~2756
www.pibb.ac.cn

A Prognostic Model Based on Colony Stimulating Factors-related Genes in Triple-negative Breast Cancer*

GUO Yu-Xuan[1)], WANG Zhi-Yu[1)], XIAO Pei-Yao[1)], ZHENG Chan-Juan[1)], FU Shu-Jun[1)], HE Guang-Chun[1)], LONG Jun[4)], WANG Jie[3)], DENG Xi-Yun[1)]**, WANG Yi-An[1,2)]**

([1)]*Key Laboratory of Model Animals and Stem Cell Biology in Hunan Province, School of Medicine, Hunan Normal University, Changsha* 410013, China;
[2)]*Engineering Research Center of Reproduction and Translational Medicine of Hunan Province, Changsha* 410013, China;
[3)]*Department of Pathology, Institute of Oncology & Diagnostic Pathology Center, The School of Basic Medical Sciences, Fujian Medical University, Fuzhou* 350122, China;
[4)]*Shenzhen Key Laboratory of Energy Electrocatalytic Materials, Shenzhen Key Laboratory of Polymer Science and Technology, Guangdong Provincial Key Laboratory of New Energy Materials Service Safety, College of Materials Science and Engineering, Shenzhen University, Shenzhen* 518055, China)

Graphical abstract



* This work was supported by grants from The National Natural Science Foundation of China (82303126, 82303069, 82173374, 82103342), Natural Science Foundation of Hunan Province (2024JJ6325, 2024JJ5287), and Natural Science Foundation of Fujian Province (2023J05038).
** Corresponding author.
WANG Yi-An. Tel: 86-731-88912415, E-mail: wangyian@hunnu.edu.cn
DENG Xi-Yun. Tel: 86-731-88912415, E-mail: dengxiyunmed@hunnu.edu.cn
Received: June 30, 2024 Accepted: September 2, 2024

Abstract **Objective** Triple-negative breast cancer (TNBC) is the breast cancer subtype with the worst prognosis, and lacks effective therapeutic targets. Colony stimulating factors (CSFs) are cytokines that can regulate the production of blood cells and stimulate the growth and development of immune cells, playing an important role in the malignant progression of TNBC. This article aims to construct a novel prognostic model based on the expression of colony stimulating factors-related genes (CRGs), and analyze the sensitivity of TNBC patients to immunotherapy and drug therapy. **Methods** We downloaded CRGs from public databases and screened for differentially expressed CRGs between normal and TNBC tissues in the TCGA-BRCA database. Through LASSO Cox regression analysis, we constructed a prognostic model and stratified TNBC patients into high-risk and low-risk groups based on the colony stimulating factors-related genes risk score (CRRS). We further analyzed the correlation between CRRS and patient prognosis, clinical features, tumor microenvironment (TME) in both high-risk and low-risk groups, and evaluated the relationship between CRRS and sensitivity to immunotherapy and drug therapy. **Results** We identified 842 differentially expressed CRGs in breast cancer tissues of TNBC patients and selected 13 CRGs for constructing the prognostic model. Kaplan-Meier survival curves, time-dependent receiver operating characteristic curves, and other analyses confirmed that TNBC patients with high CRRS had shorter overall survival, and the predictive ability of CRRS prognostic model was further validated using the GEO dataset. Nomogram combining clinical features confirmed that CRRS was an independent factor for the prognosis of TNBC patients. Moreover, patients in the high-risk group had lower levels of immune infiltration in the TME and were sensitive to chemotherapeutic drugs such as 5-fluorouracil, ipatasertib, and paclitaxel. **Conclusion** We have developed a CRRS-based prognostic model composed of 13 differentially expressed CRGs, which may serve as a useful tool for predicting the prognosis of TNBC patients and guiding clinical treatment. Moreover, the key genes within this model may represent potential molecular targets for future therapies of TNBC.

Key words triple-negative breast cancer, colony stimulating factors, prognostic model, tumor microenvironment, drug sensitivity
DOI：10.16476/j.pibb.2024.0281

According to the latest statistics from the International Agency for Research on Cancer (IARC), breast cancer is the second most common cancer and the fourth leading cause of cancer death globally, and it is also the most prevalent cancer type among women, both in terms of incidence and mortality[1]. As social economy develops, the incidence and mortality rates of breast cancer continue to rise annually, placing a significant burden on the healthcare system[2]. Breast cancer is classified into several subtypes, including luminal A, luminal B, HER2-enriched, and triple-negative, based on the molecular characteristics of the patient's tumor tissue. Among them, triple-negative breast cancer (TNBC) is characterized by the absence of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2)[3]. Due to the lack of these molecular targets, TNBC patients are insensitive to commonly used endocrine therapy or HER2-targeted therapy used for other subtypes of breast cancer. Additionally, TNBC often exhibits higher invasiveness and metastatic potential, tending to recur and resulting in poor prognosis for patients[4]. With the advancement of omics technology, researchers have discovered that TNBC is also a heterogeneous subtype of breast cancer. Therefore, it is crucial to classify TNBC patients based on their genetic characteristics, analyze prognostic features, and explore suitable treatment modalities[5-6].

Colony stimulating factors (CSFs) are a group of cytokines that stimulate the proliferation and differentiation of hematopoietic stem cells. Depending on the targeted cells, they are classified into granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), granulocyte-macrophage colony stimulating factor (GM-CSF), stem cell factor (SCF), erythropoietin (EPO), thrombopoietin (TPO), and so on[7]. CSFs can promote the development and maturation of red blood cells, platelets, macrophages, neutrophils, dendritic cells, *etc*., thus playing a significant role in immune response and immune regulation in the body[8]. In the occurrence and development of TNBC, G-CSF can promote the migration and invasion of TNBC cells[9], M-CSF can promote the accumulation of macrophages in the tumor microenvironment (TME) and enhance the invasiveness of cancer cells[10], GM-CSF can recruit myeloid-derived suppressor cells to shape an immunosuppressive microenvironment, promoting tumor growth and metastasis[11], and EPO is highly expressed in tumor tissues of TNBC patients and is correlated with prognosis[12]. Apart from their

significant role in the malignant progression of tumors, researchers have also focused on the therapeutic potential of CSFs. Injecting GM-CSF into tumors can make cancer cells sensitive to programmed death-1 (PD-1) treatment[13]. The use of PARP inhibitors combined with CSF-1R blocking antibodies can significantly enhance the anti-tumor immune response in TNBC[14]. The combination of GM-CSF and an HER2-targeted peptide vaccine has certain therapeutic efficacy in the treatment of TNBC patients[15-17]. Researchers have been focusing on the molecular mechanisms of CSFs in the pathogenesis of TNBC and designing treatment methods based on this. However, up to now, there has been no prognostic model based on CSFs available for guiding clinical treatment of TNBC.

This study aimed at developing a novel prognostic model related to CSFs in TNBC and validating its predictive capability for TNBC patient outcomes through comprehensive bioinformatics analysis of TCGA-BRCA and GEO databases. Furthermore, the study will explore the role of this model in assessing the TME and evaluate patient sensitivity to drugs based on colony stimulating factors-related genes risk score (CRRS), thereby providing insights for clinical treatment of TNBC patients.

1 Materials and methods

1.1 Acquisition and processing of TNBC datasets

We downloaded the dataset of breast cancer patient samples and normal breast samples from The Cancer Genome Atlas (TCGA) database with a retrieval date of 12 December 2023, and retrieved relevant clinical information from UCSC Xena (https://xena.ucsc.edu/). We extracted the Transcripts Per Kilobase per Million (TPM) data, removed low-expression genes, and normalized the sample expressions. We screened out TNBC patients based on the status of "breast_carcinoma_estrogen_receptor_status", "breast_carcinoma_progesterone_receptor_status", and "lab_proc_her2_neu_immunohistochemistry_receptor_status" of the samples. A total of 120 TNBC patient samples and 99 normal samples were included in the TCGA training set.

We downloaded the GSE58812 independent cohort from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) as an external dataset. After normalization of the sample expressions, we ultimately obtained 107 TNBC patient samples to serve as an external test set.

1.2 Screening of the differentially expressed colony stimulating factors−related genes

We downloaded colony stimulating factors-related genes (CRGs) from the GeneCards database (version 5.20) (https://www.genecards.org/) using the keyword "Colony Stimulating Factor" retrieving 7 180 genes. We further extracted proteins with a Relevance score>10 for subsequent analysis. We utilized the "limma" R package to perform differential expression analysis on these genes between TNBC patient samples and normal breast samples in TCGA. Genes with $|\log_2 FC| > 0.585$ and adjusted P-value<0.05 were considered as differentially expressed genes (DEGs).

The selection of DEGs for the high-risk and low-risk groups was also conducted using the "limma" R package, with the screening criteria set as $|\log_2 FC| > 0.585$ and adjusted P-value<0.05.

1.3 Functional enrichment analysis of DEGs

For the functional enrichment analysis of DEGs, we used the "org.Hs.eg.db" R package for gene ontology (GO) annotation and the Kyoto Encyclopedia of Genes and Genomes (KEGG) REST API (https://www.kegg.jp/kegg/rest/keggapi.html) to obtain the latest gene annotations for KEGG pathway, which served as the background. We mapped the genes to this background set. Enrichment analysis was performed using the "clusterProfiler" R package to obtain GO and KEGG enrichment results.

1.4 Construction and validation of a CSFs−related prognostic model

We performed univariate Cox regression analysis on gene expression and overall survival (OS) for TNBC patients in TCGA-BRCA. Using the "survival" and "ggforest" R packages, we obtained and visualized the univariate Cox results. Genes with P-values<0.05 were considered statistically significant for patient survival, and 21 survival-related DEGs were identified for further analysis.

Using the LASSO Cox regression model from the "glmnet" R package, we narrowed down the candidate genes and established a prognostic model. Ultimately, 13 genes and their coefficients were retained, and the penalty parameter λ was determined

based on the minimum criteria. Risk scores were calculated for the TCGA expression data using the risk score formula. Based on the median risk score value, TNBC patients in TCGA-BRCA were divided into high-risk and low-risk groups. The survival time between the two groups was then compared using Kaplan-Meier survival curve analysis. The predictive ability of the model in the training set was validated using 4-year, 5-year, and 6-year time-dependent receiver operating characteristic (ROC) curve analysis with the "survival" and "timeROC" R packages. Additionally, we used the external independent cohort GSE58812 as a validation set.

1.5 Construction of nomogram

We extracted TNBC patients with complete clinical data and performed univariate and multivariate Cox regression analyses on the clinical stage, T stage, N stage, and Risk Score of these patients. We utilized the "survival" and "ggfores" R packages to obtain and visualize the results of the univariate and multivariate Cox regression analyses. Features with *P*-values<0.05 were considered as independent prognostic factors.

Using the "rms" R package, we integrated data such as survival time, survival status, N stage, and Risk Score to construct nomogram plots with 4-year, 5-year, and 6-year OS as endpoints. We further developed a model to predict patients' survival time and assessed the ability of the nomogram plots to predict patient prognosis using calibration curves and ROC curves. Additionally, we compared the prognostic evaluation capability of the nomogram score with other clinical features using ROC curves.

1.6 Assessment of the tumor microenvironment in the high−risk and low−risk population groups

We calculated the stromal and immune scores, as well as tumor purity, for TNBC patients in TCGA-BRCA using the "estimate" R package. Subsequently, we employed the "CIBERSORT" R package to determine the proportions of 22 immune cell types in tumor samples, and the "ssGSEA" R package to evaluate the infiltration levels of 28 immune cell types in tumor samples. Furthermore, we analyzed the expression of 17 immune checkpoint genes using Student's *t*-test.

1.7 Prediction of drug sensitivity in TNBC patients

We used the "oncoPredict" R package to calculate the sensitivity of TNBC samples to 198 commonly used chemotherapeutic drugs, and then employed Student's *t*-test to compare drug sensitivity between the high-risk and low-risk groups.

1.8 Statistical analysis

All statistical analyses were performed using R software (version 4.3.1). Differences between two groups were analyzed using Student's *t*-test, and *P*-values<0.05 was considered statistically significant.

2 Results

2.1 Identification and functional analysis of CRGs in TNBC

First, we downloaded colony stimulating CRGs from the GeneCards database, retrieving a total of 7 180 proteins. We then extracted 1 869 molecules with a Relevance score>10 for subsequent analysis (Table S1). Using the "limma" R package, we identified 842 DEGs in normal breast and TNBC tissues from TCGA-BRCA. A volcano plot was used to visualize these DEGs, showing that compared to normal breast tissues, 368 genes were upregulated and 474 genes downregulated in TNBC (Figure 1a). A heatmap further revealed significant differences in the expression of CRGs between normal and TNBC tissues (Figure 1b). To analyze the biological significance of these DEGs, we performed functional enrichment analysis. GO enrichment analysis showed that DEGs were primarily clustered in the cellular response to chemical stimulus, the cell surface, and the transcription regulatory region DNA binding (Figure 1c). KEGG pathway analysis demonstrated that DEGs were primarily enriched in the PI3K-Akt signaling pathway, the MAPK signaling pathway, and the cytokine-cytokine receptor interaction (Figure 1d). These results suggest that CRGs may play an important role in the progression of TNBC.

2.2 Construction and validation of a CRGs−related prognostic model for TNBC

To explore the value of CRGs in the prognosis of TNBC, we collated the survival information of TNBC patients and performed univariate Cox regression analysis on 842 differentially expressed CRGs, screening for 21 genes closely related to the prognosis of TNBC patients (Figure 2a). Subsequently, the LASSO Cox regression algorithm was used for dimensionality reduction screening, and the optimal coefficients were calculated, resulting in the

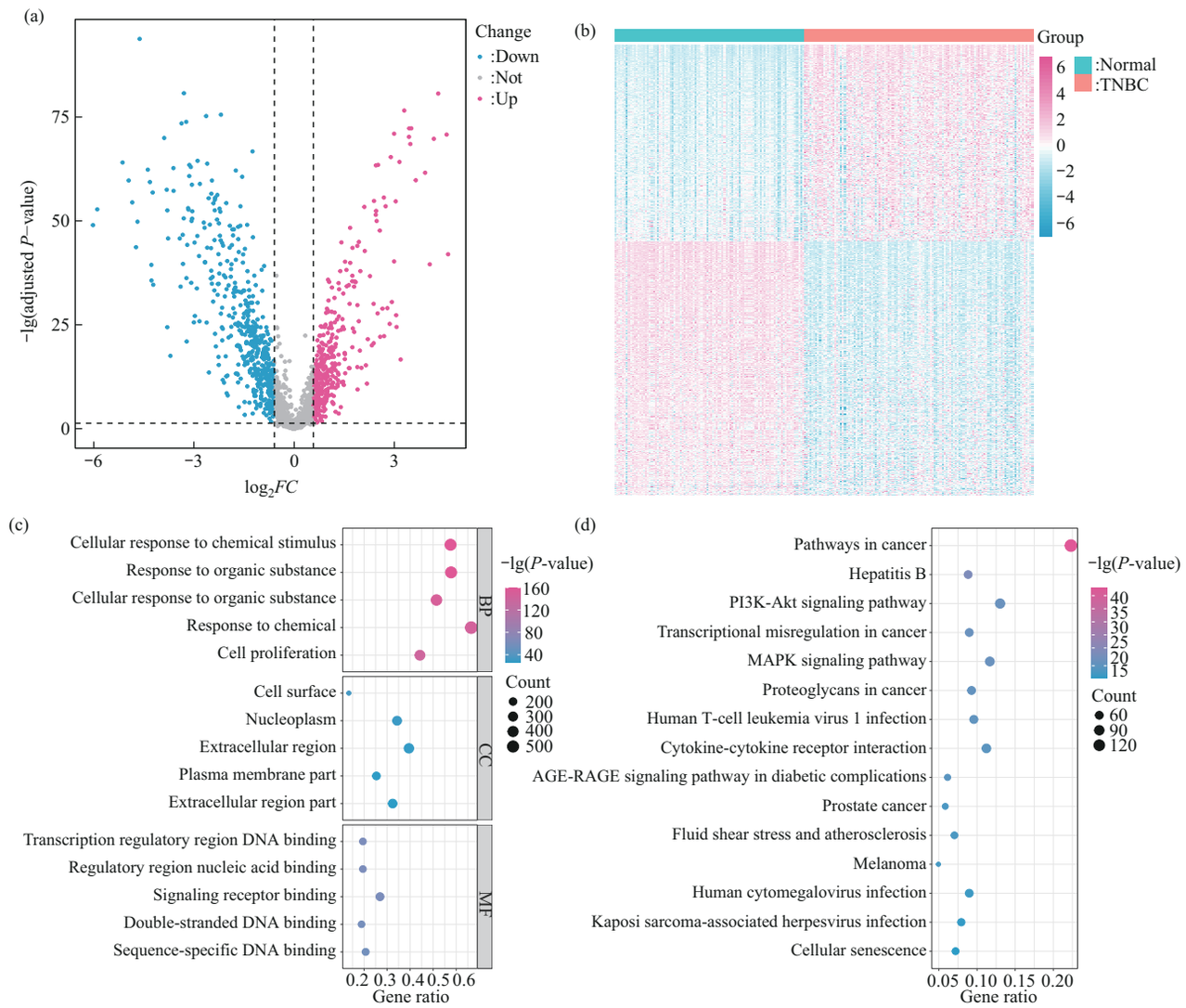


Fig. 1　Identification and functional analysis of colony stimulating factors−related genes in TNBC

(a) A volcano plot displaying the differential expression of CRGs in normal breast tissues and TNBC tissues from TCGA-BRCA. Red represents upregulated genes, blue represents downregulated genes, and gray represents genes with no statistically significant changes. Adjusted *P*-value<0.05, $|\log_2 FC|>0.585$. (b) A heatmap showing the DEGs related to CRGs. (c, d) Functional enrichment analysis of DEGs performed using GO (c) and KEGG (d) analyses.

construction of a prognostic model utilizing 13 genes (Figure 2b). Based on the expression levels of candidate genes, the CRRS was calculated: CRRS=0.342×*EIF4EBP1* − 0.288×*RASGRP1*+0.194×*GRHL2* − 0.088×*S100B*+0.211×*DCAF4* − 0.306×*TWIST1*+0.642×*HMOX1* − 0.701×*ETV1*+0.268×*KLF10*+0.017×*SOCS2* − 0.196×*GZMB*+0.255×*KRT18* − 0.056×*IGFBP1*.

Subsequently, scores were assigned to each TNBC patient in TCGA-BRCA based on the risk calculation formula, and patients were divided into high-risk and low-risk groups using the median score. Kaplan-Meier survival curves revealed that the high-risk group had a lower OS rate and worse prognosis compared to the low-risk group (Figure 2c). Additionally, the number of deaths in the high-risk group was higher than that in the low-risk group (Figure 2e). The area under the curve (*AUC*) of the receiver operating characteristic (ROC) curve at 4- , 5- , and 6-years was 0.94, 0.90, and 0.92, respectively (Figure 2g). We used GSE58812 from the GEO database for external validation, assigning risk

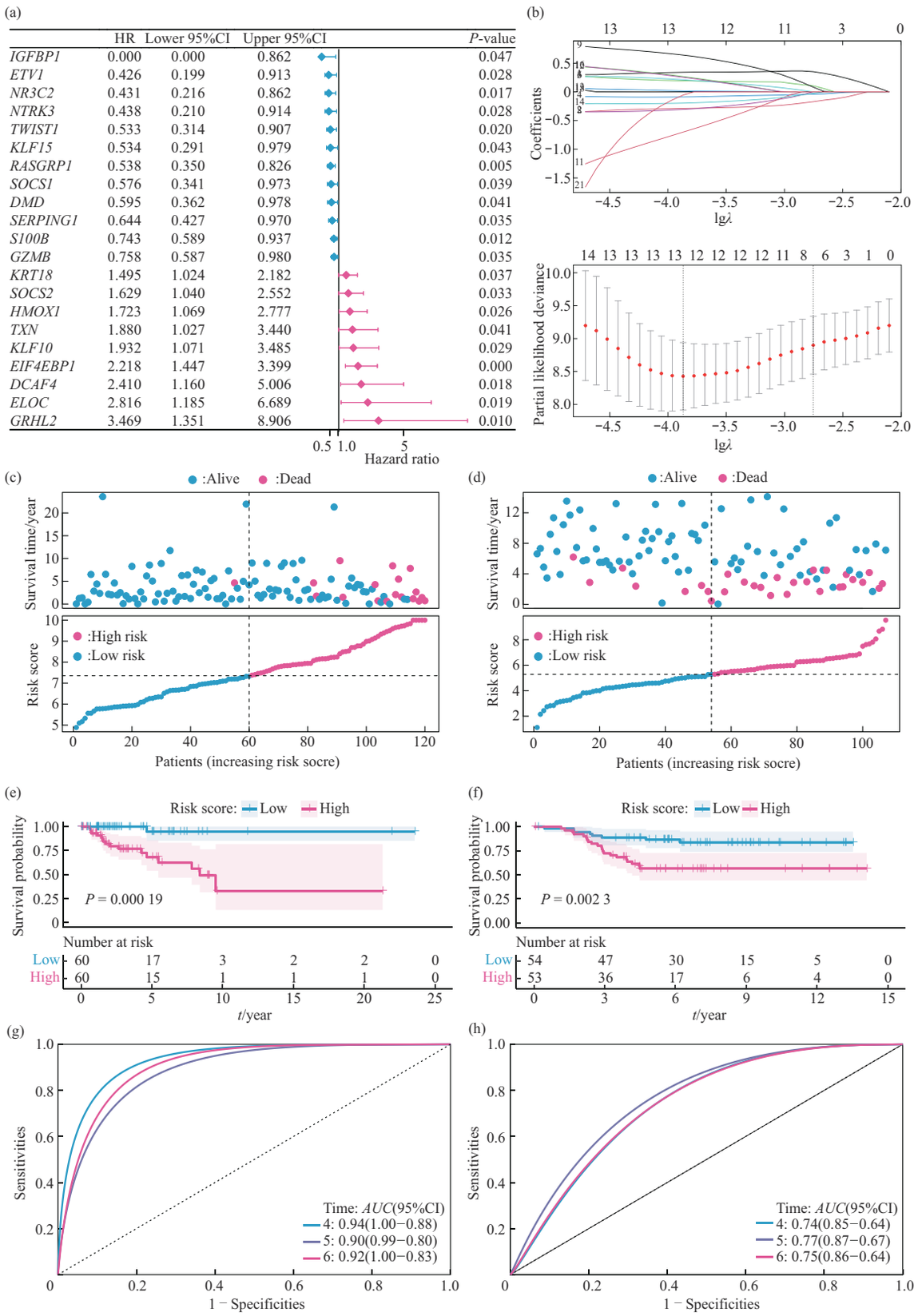


Fig. 2　Construction and validation of a CRGs−related prognostic model for TNBC

(a) Univariate Cox regression analysis of DEGs in the prognosis of TNBC. (b) LASSO Cox regression analysis to identify the DEGs for constructing the prognostic model. (c, d) The differences in overall survival between the high-risk and low-risk groups of TNBC patients in the TCGA-BRCA database (c) and the GEO58812 database (d) using Kaplan-Meier survival curves. (e, f) Distribution of Risk Score and survival status of TNBC patients in the TCGA-BRCA database (e) and the GEO58812 database (f). (g, h) The prognostic value of CRRS in the TCGA-BRCA database (g) and the GEO58812 database (h) using ROC curves.

scores to 107 TNBC patients and grouping them based on the median score. The prognostic analysis yielded consistent results with the training set, showing worse prognosis in the high-risk group. The *AUC* of the ROC curve for 4-, 5-, and 6-years was 0.74, 0.77, and 0.75, respectively (Figure 2d, f, h). These results demonstrate the accuracy and reliability of the prognostic model constructed using CRGs in TNBC.

We further analyzed the expression profiles of the CRGs involved in the prognostic model, and we found that, compared with normal breast tissues in the TCGA-BRCA cohort, *EIF4EBP1*, *RASGRP1*, *GRHL2*, *DCAF4*, *HMOX1*, *GZMB*, and *KRT18* were upregulated in TNBC, while *S100B*, *TWIST1*, *ETV1*, *KLF10*, *SOCS2*, and *IGFBP1* were downregulated in TNBC (Figure 3a). Subsequently, we used Kaplan-Meier survival curves to analyze the role of these molecules in the prognosis of TNBC, and the results revealed that high expression of *EIF4EBP1*, *GRHL2*, *DCAF4*, *HMOX1*, *KLF10*, *SOCS2*, and *KRT18* was associated with poor prognosis in TNBC patients (Figure 3b), while low expression of *RASGRP1*, *S100B*, *TWIST1*, *ETV1*, *GZMB*, and *IGFBP1* was related to poor prognosis in TNBC patients (Figure 3c).

2.3 The construction and validation of a nomogram for the CRGs−related prognostic model

To further analyze the clinical value of this prognostic model, we first investigated the expression differences of risk scores in TNM and clinical stages among TNBC patients in the TCGA-BRCA database. We found that patients with higher clinical stage, T stage, and N stage exhibited higher risk scores (Figure 4a−c). Subsequently, we explored whether CRRS and clinical characteristics were independent prognostic factors for patients through univariate and multivariate Cox analyses. The results of univariate Cox regression analysis indicated that clinical stage, T stage, N stage, and CRRS were significant prognostic risk factors (Figure 4d). Multivariate Cox regression analysis revealed that N stage and CRRS were independent prognostic risk factors (Figure 4e). Therefore, we chose N stage and CRRS to further establish a prognostic nomogram to predict the survival rate of TNBC patients (Figure 4f). The predictive ability of the nomogram was analyzed using time-dependent ROC curves, and we found that the *AUC* for the 4-, 5-, and 6-years was 0.97, 0.94, and 0.96, respectively (Figure 4g). The calibration curves demonstrated the predictive accuracy of the model for the 4-, 5-, and 6-years (Figure 4h). Then, we analyzed the predictive differences between the nomogram and other parameters, showing that the *AUC* of the nomogram was 0.944, which was higher than the predictive abilities of CRRS (*AUC*=0.922) and N stage (*AUC*=0.735) (Figure 4i). These results suggest that CRRS is an independent prognostic factor that can accurately assess the prognosis of TNBC patients.

2.4 Functional analysis of the CSFs−related prognostic model

To explore the reasons for the different prognoses of TNBC patients with varying CRRS, we used the “limma” R package to screen for DEGs between the high-risk and low-risk groups. Based on the criteria of adjusted *P*-value<0.05 and $|\log_2 FC| > 0.585$, we identified 39 upregulated genes and 27 downregulated genes in the TNBC cohort (Figure 5a). We further displayed the expression of the DEGs between the high-risk and low-risk groups using a heatmap (Figure 5b). Subsequently, GO enrichment analysis and KEGG pathway analysis were performed on the aforementioned DEGs. The GO enrichment analysis demonstrated that the DEGs were mainly related to the cellular developmental process, the vesicle, and the laminin binding (Figure 5c). The KEGG pathway analysis revealed that the DEGs were primarily associated with the pentose phosphate pathway, the AMPK signaling pathway, and the PI3K-Akt signaling pathway (Figure 5d).

2.5 Differences in the tumor microenvironment between the high−risk and low−risk groups of TNBC patients

TME contains a large number of infiltrating immune cells, with each immune cell type performing a unique regulatory function, which may be an important reason for the different prognoses of TNBC patients. We used the ESTIMATE algorithm to analyze the stromal score, immune score, ESTIMATE score, and tumor purity of the high-risk and low-risk groups in the TNBC cohort of TCGA-BRCA. Compared to the low-risk group, the high-risk group had a lower stromal score and an increasing trend in tumor purity (Figure 6a, b). We then used the CIBERSORT algorithm to explore the differences in

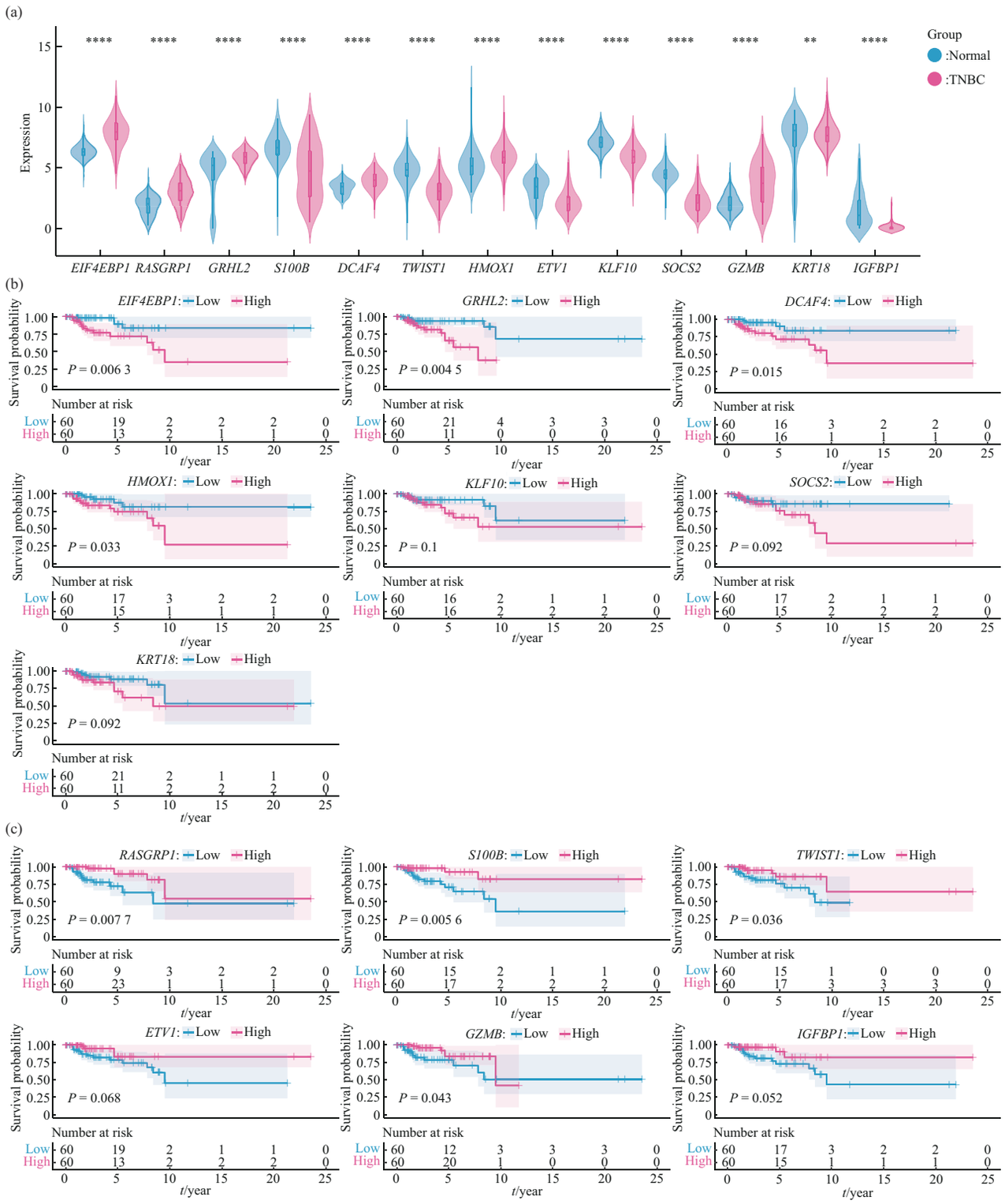


Fig. 3　The expression profiles of CRGs in TNBC and their association with prognosis

(a) Expression differences of *EIF4EBP1*, *RASGRP1*, *GRHL2*, *S100B*, *DCAF4*, *TWIST1*, *HMOX1*, *ETV1*, *KLF10*, *SOCS2*, *GZMB*, *KRT18*, and *IGFBP1* between TNBC cancer tissues and normal breast tissues in the TCGA-BRCA database. ***P*<0.01, *****P*<0.000 1. (b, c) Kaplan-Meier survival curves were used to analyze molecules that are unfavorable (b) or favorable (c) for the prognosis of TNBC patients.

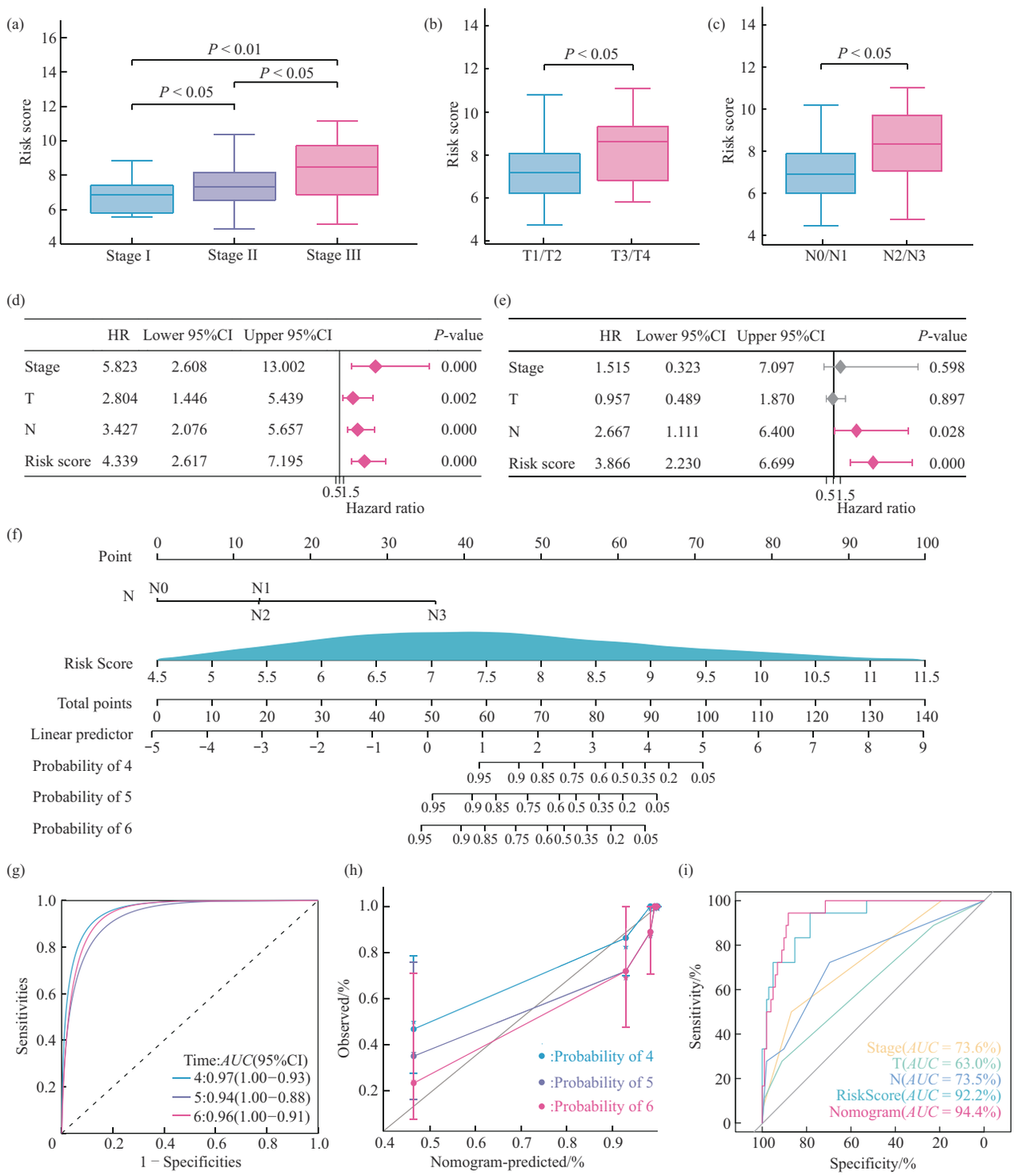


Fig. 4　Construction and validation of a nomogram for the CRGs−related prognostic model

(a−c) The risk scores of TNBC patients in the TCGA-BRCA database with the progression of clinical stage (a), T stage (b), and N stage (c). (d, e) Univariate (d) and multivariate (e) Cox regression analyses of clinicopathological characteristics and CRRS in the TNBC cohort from the TCGA-BRCA database. (f) A nomogram to demonstrate the ability of N stage and CRRS to evaluate prognosis. (g) ROC curves were used to predict the predictive ability of the nomogram at 4-, 5-, and 6-years. (h) Calibration curves displaying the 4-year, 5-year, and 6-year OS probabilities in the TCGA cohort. (i) ROC curves demonstrating the predictive ability of the nomogram, risk score, and clinical characteristics for patient prognosis.

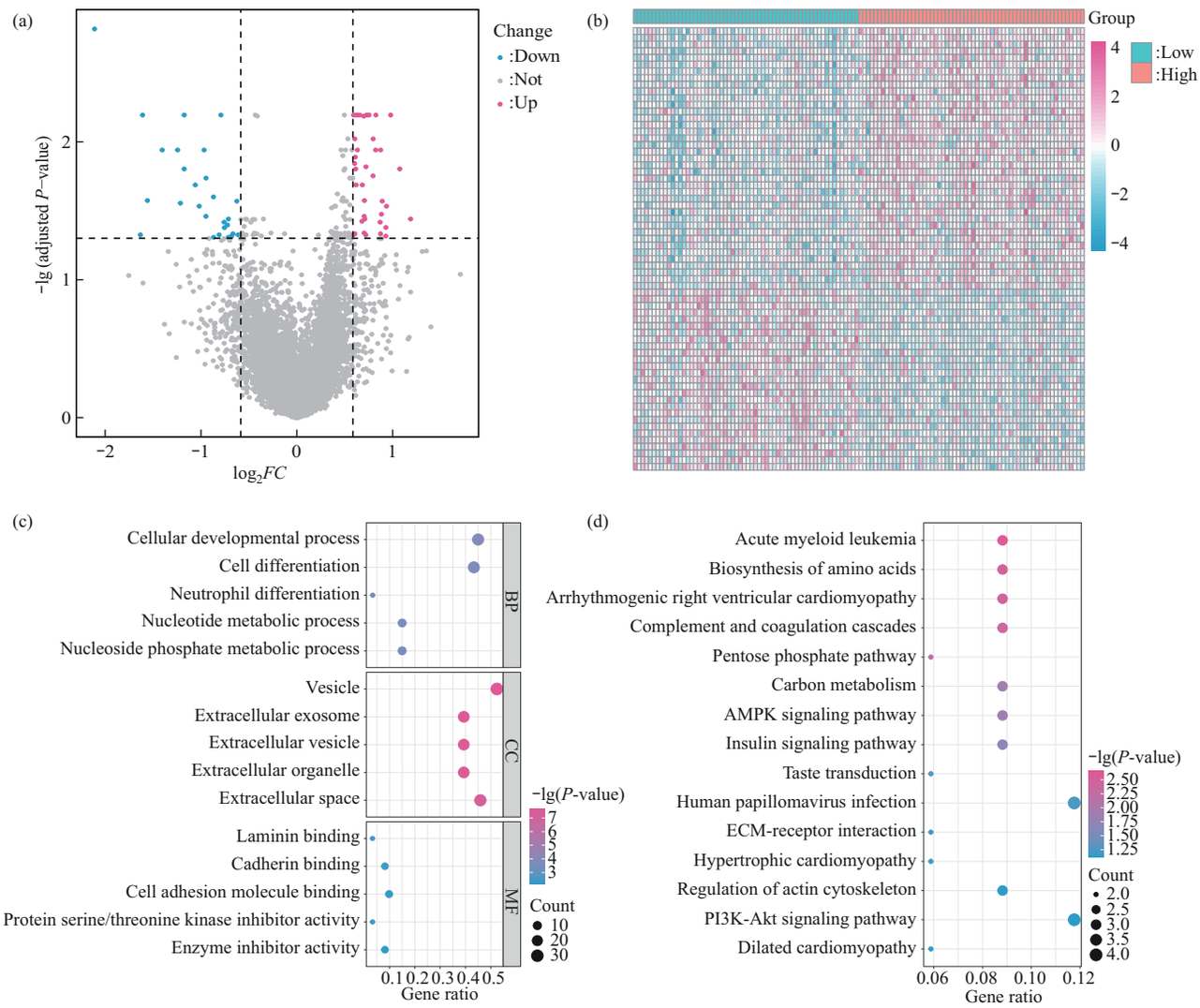


Fig .5　Functional analysis of DEGs between the high−risk and low−risk groups in the TCGA−BRCA TNBC cohort

(a) A volcano plot displaying the DEGs between the high-risk and low-risk groups in the TNBC cohort of TCGA-BRCA. (b) A heatmap showing the DEGs between the high-risk and low-risk groups. (c, d) GO functional enrichment (c) and KEGG pathway (d) analyses of the DEGs between the high-risk and low-risk groups.

the composition of immune cells in the TME, and we found that B naive cells were enriched in the low-risk group (Figure 6c). Subsequently, we used the ssGSEA algorithm to analyze the differences in immune cells infiltration between the two groups, and we found that the high-risk group had more infiltration of CD56dim natural killer cells, while the low-risk group had more infiltration of activated B cells, immature B cells, eosinophils, mast cells, plasmacytoid dendritic cells, and type 2 helper cells (Figure 6d). Finally, we analyzed the expression of some immune checkpoint molecules in the high-risk and low-risk groups[18], and we found that *CEACAM1* and *VTCN1* molecules were upregulated in the low-risk group (Figure 6e). The above results indicate that after grouping TNBC patients into high-risk and low-risk groups based on CRRS, there are certain differences in immune cell infiltration between the two groups, which may be potential targets for future precision therapy.

2.6　Drug sensitivity in the high−risk and low−risk groups of TNBC patients

To analyze the appropriate treatment options for TNBC patients and how this model can guide drug administration, we calculated the sensitivity scores (IC_{50}-like values) of TNBC patients to each drug. We found that patients in the high-risk group were more

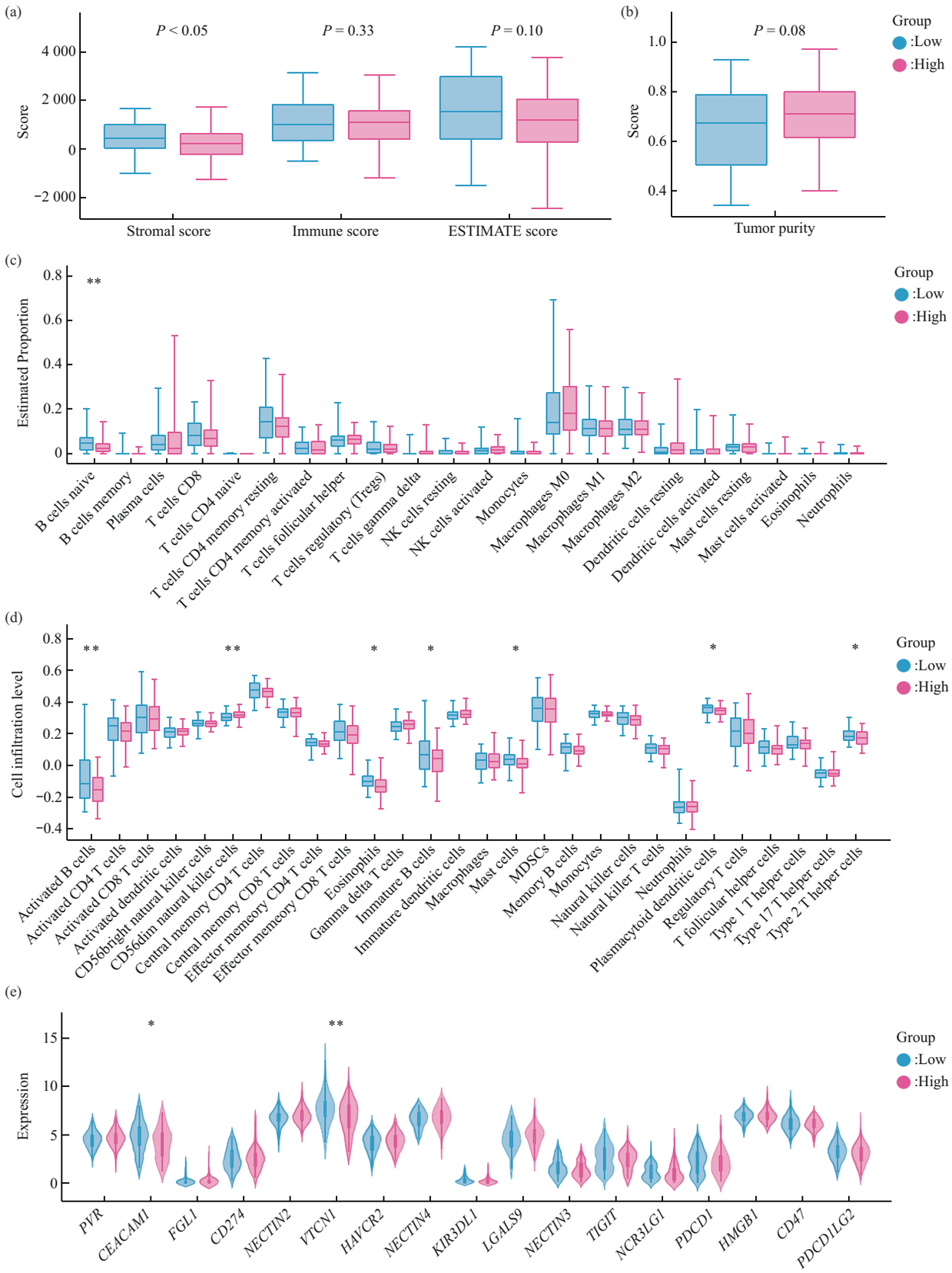


Fig. 6　Differences in the TME between the high−risk and low−risk groups of TNBC patients

(a, b) The ESTIMATE algorithm used to score the TME (a) and tumor purity (b) in the high-risk and low-risk groups. (c) The CIBERSORT algorithm employed to analyze differences in the abundance of 22 immune cell types between the high-risk and low-risk groups. (d) The ssGSEA algorithm used to analyze differences in the degrees of immune cell infiltration between the high-risk and low-risk groups. (e) Expression differences of immune checkpoints between the high-risk and low-risk groups. **P*<0.05; ***P*<0.01.

sensitive to 5-fluorouracil, ipatasertib, and paclitaxel (Figure 7a), while patients in the low-risk group were more sensitive to PCI-34051, JQ1, and trametinib (Figure 7b). These results suggest that TNBC patients can be scored based on CRRS, and individual targeted therapy can be administered with reference to the scoring results.

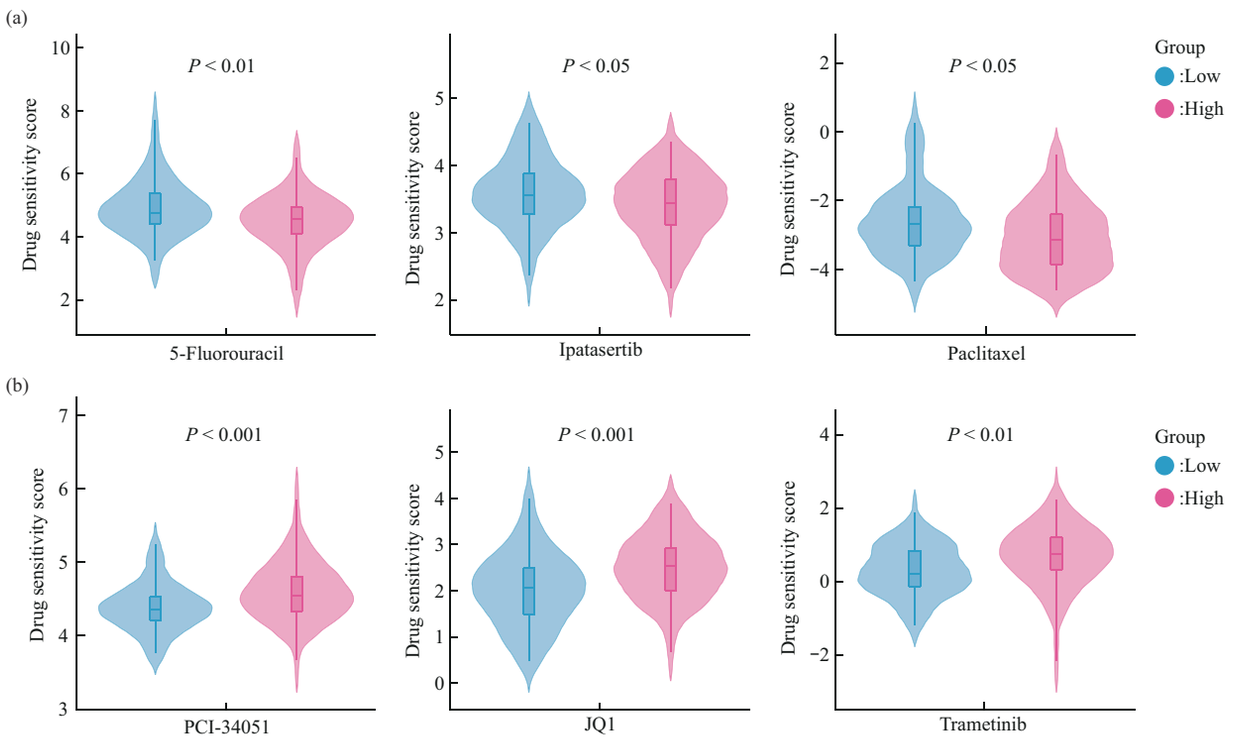


Fig. 7 Drug sensitivity in high/low-risk groups of TNBC patients

(a) Sensitivity scores of the high-risk and low-risk groups of patients to 5-fluorouracil, ipatasertib, and paclitaxel. (b) Sensitivity scores of the high-risk and low-risk groups of patients to PCI-34051, JQ1, and trametinib.

3 Discussion

TNBC is currently considered the most challenging subtype of breast cancer for clinical treatment, primarily due to the significant heterogeneity among patients, which hampers the selection of appropriate treatment methods. Therefore, it is crucial to categorize and perform targeted therapy based on patients' molecular characteristics[5]. With the advancement of sequencing technology, researchers have gained a deeper understanding of the heterogeneity among TNBC patients. By integrating clinical information and bioinformatics analysis, researchers can now predict the prognosis, immune infiltration, and drug sensitivity of individual patients[19-20]. However, the current gene prognostic models for TNBC are not entirely satisfactory, necessitating the development of a novel classification method to better guide clinical treatment.

CSFs are cytokines that accumulate significantly in the TME of TNBC patients and play an important role in TNBC tumorigenesis[21]. In this study, we extracted CRGs with differential expression in 120 TNBC patients and 99 normal people from TCGA-BRCA data. The results of GO and KEGG functional enrichment analysis suggest that CRGs may affect the growth and metastasis of TNBC by influencing gene transcription and regulating signaling pathways such as PI3K-Akt and MAPK. Using univariate Cox regression analysis and LASSO analysis, we constructed a prognostic model based on the CSFs-related genes, including *EIF4EBP1*, *RASGRP1*, *GRHL2*, *S100B*, *DCAF4*, *TWIST1*, *HMOX1*, *ETV1*, *KLF10*, *SOCS2*, *GZMB*, *KRT18*, and *IGFBP1*. The prognostic ability of this model was validated using TNBC cohorts from TCGA and GEO databases, demonstrating that the model can predict the

prognosis of TNBC patients. Among them, eukaryotic translation initiation factor 4E-binding protein 1 (*EIF4EBP1*) participates in regulating the translation process of mRNA, which is upregulated in TNBC and promotes the metastasis and proliferation of cancer cells[22-23]. RAS guanyl releasing protein 1 (*RASGRP1*) is highly expressed in TNBC, but Kaplan-Meier analysis results suggest that high expression of RASGRP1 leads to better prognosis, indicating that RASGRP1 plays a complex regulatory role in TNBC[24-25]. Grainyhead-like transcription factor 2 (*GRHL2*) is a transcription factor closely related to EMT, which is upregulated in TNBC and correlated to poor prognosis in patients[26]. S100 calcium binding protein B (*S100B*) is a potential diagnostic and prognostic biomarker for TNBC[27-28]. DDB1 and CUL4 associated factor 4 (*DCAF4*) is highly expressed in TNBC patients and is associated with poor prognosis, but there are currently few studies on DCAF4 in TNBC. Twist family bHLH transcription factor 1 (*TWIST1*) is a key transcription factor in the EMT process, promoting the metastasis, angiogenesis, stemness maintenance, and chemotherapy resistance of TNBC[29-31]. The TCGA-BRCA database shows that TWIST1 is expressed at low levels in TNBC patients, but immunohistochemical experiment results suggest that TWIST1 is upregulated in TNBC cancer tissues, which may be due to post-translational modifications leading to inconsistent expression levels between its mRNA and protein[30]. Heme oxygenase 1 (*HMOX1*) promotes ferroptosis in TNBC cells, affecting the effectiveness of immunotherapy[32-33]. ETS variant transcription factor 1 (*ETV1*) may affect the growth and invasion of TNBC, but the specific molecular mechanism is unclear[34-35]. KLF transcription factor 10 (*KLF10*) is involved in cell differentiation, but there are currently few studies on it in TNBC[36]. Suppressor of cytokine signaling 2 (*SOCS2*) serves as a tumor suppressor in breast cancer, and the study on it in TNBC is rare[37]. Granzyme B (*GZMB*) is mainly secreted by cytotoxic T cells, which can induce the death of cancer cells in the microenvironment, and TNBC patients with high expression of GZMB have better prognosis[38-39]. Keratin 18 (*KRT18*) is a marker for TNBC cell differentiation[40]. Insulin-like growth factor binding protein 1 (*IGFBP1*) regulates angiogenesis and radiosensitivity in TNBC[41-42]. Interestingly, we observed that the expression of some CRGs in TNBC patients is inconsistent with their prognostic outcomes. This could be attributed to the complex regulation mechanisms governing gene expression, encompassing epigenetic regulation, genetic background, environmental factors, and more. Consequently, the relationship between gene expression and tumor prognosis is not a simple linear one but rather a result of the intricate interplay of multiple factors. Therefore, the research on the expression and function of the above-mentioned molecules in TNBC is limited, this study provides a direction for further exploration.

The analysis results of DEGs between the high-risk and low-risk groups suggest that the different prognostic outcomes of the two patient groups may be closely related to intracellular or intercellular signal transduction and cellular metabolic status. In addition, the low-risk group has more immune cell infiltration, which may suppress tumor immune evasion and inhibit the progression of TNBC. Regarding changes in immune checkpoint molecules, we only observed upregulated expression of *CEACAM1* and *VTCN1* in the low-risk group. These changes indicate that, although they can serve as therapeutic targets, CRGs may regulate tumor progression in patients through non-immunomodulatory means. Currently, commonly used drugs for TNBC treatment include doxorubicin[43], paclitaxel[44], and cisplatin[45], but they have not achieved the expected therapeutic efficacy. Therefore, we further analyzed the drug sensitivity of high-risk and low-risk groups. High-risk patients are more sensitive to DNA synthesis inhibitor (5-fluorouracil), AKT enzyme inhibitor (ipatasertib), and microtubule depolymerization inhibitor (paclitaxel). Low-risk patients are more sensitive to HDAC8 inhibitor (PCI-34051), BET bromodomain inhibitor (JQ1), and MEK inhibitor (trametinib). The possible reason is that the changes in signaling pathways are different between the two groups. The drug sensitivity analysis results in this study may provide insights for the selection of targeted drugs for TNBC patients and for the development of new targeted therapeutic drugs. In the future, the specific molecular mechanism of CRGs regulating the progression of TNBC needs to be investigated.

4 Conclusion

In this study, we constructed a prognostic model composed of 13 CRGs, which effectively predicted

the prognosis of TNBC patients in TCGA and GEO cohorts. We also explored the changes in signaling pathways and immune infiltration levels in different risk groups, and predicted the sensitivity of patients to various drugs, thus providing guidance for clinical treatment of TNBC.

Supplementary Available online (http://www.pibb.ac.cn, http://www.cnki.net):
PIBB_20240281_Table_S1.xlsx

参 考 文 献

[1] Bray F, Laversanne M, Sung H, *et al*. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, 2024, **74**(3): 229-263

[2] Britt K L, Cuzick J, Phillips K A. Key steps for effective breast cancer prevention. Nat Rev Cancer, 2020, **20**(8): 417-436

[3] Farshid G, Walters D. Molecular subtypes of screen-detected breast cancer. Breast Cancer Res Treat, 2018, **172**(1): 191-199

[4] Yin L, Duan J J, Bian X W, *et al*. Triple-negative breast cancer molecular subtyping and treatment progress. Breast Cancer Res, 2020, **22**(1): 61

[5] Gong Y, Ji P, Yang Y S, *et al*. Metabolic-pathway-based subtyping of triple-negative breast cancer reveals potential therapeutic targets. Cell Metab, 2021, **33**(1): 51-64.e9

[6] Li Y, Zhang H, Merkher Y, *et al*. Recent advances in therapeutic strategies for triple-negative breast cancer. J Hematol Oncol, 2022, **15**(1): 121

[7] Bath P M W, Sprigg N, England T. Colony stimulating factors (including erythropoietin, granulocyte colony stimulating factor and analogues) for stroke. Cochrane Database Syst Rev, 2013(6): CD005207

[8] Metcalf D. The colony-stimulating factors and cancer. Cancer Immunol Res, 2013, **1**(6): 351-356

[9] Mei X, Ouyang H, Zhang H, *et al*. Scutellarin suppresses the metastasis of triple-negative breast cancer *via* targeting TNFα/TNFR2-RUNX1-triggered G-CSF expression in endothelial cells. Biochem Pharmacol, 2023, **217**: 115808

[10] Lu X, Yang R, Zhang L, *et al*. Macrophage colony-stimulating factor mediates the recruitment of macrophages in triple negative breast cancer. Int J Biol Sci, 2019, **15**(13): 2859-2871

[11] Li W, Tanikawa T, Kryczek I, *et al*. Aerobic glycolysis controls myeloid-derived suppressor cells and tumor immunity *via* a specific CEBPB isoform in triple-negative breast cancer. Cell Metab, 2018, **28**(1): 87-103.e6

[12] Badowska-Kozakiewicz A M, Budzik M P, Liszcz A, *et al*. Clinicopathological factors associated with novel prognostic markers for patients with triple negative breast cancer. Arch Med Sci, 2019, **15**(6): 1433-1442

[13] Mihalik N E, Steinberger K J, Stevens A M, *et al*. Dose-specific intratumoral GM-CSF modulates breast tumor oxygenation and antitumor immunity. J Immunol, 2023, **211**(10): 1589-1604

[14] Mehta A K, Cheney E M, Hartl C A, *et al*. Targeting immunosuppressive macrophages overcomes PARP inhibitor resistance in BRCA1-associated triple-negative breast cancer. Nat Cancer, 2021, **2**(1): 66-82

[15] Mittendorf E A, Ardavanis A, Symanowski J, *et al*. Primary analysis of a prospective, randomized, single-blinded phase II trial evaluating the HER2 peptide AE37 vaccine in breast cancer patients to prevent recurrence. Ann Oncol, 2016, **27**(7): 1241-1248

[16] Clifton G T, Hale, Vreeland T J, *et al*. Results of a randomized phase IIb trial of nelipepimut-S + trastuzumab versus trastuzumab to prevent recurrences in patients with high-risk HER2 low-expressing breast cancer. Clin Cancer Res, 2020, **26**(11): 2515-2523

[17] Chick R C, Clifton G T, Hale D F, *et al*. Subgroup analysis of nelipepimut-S plus GM-CSF combined with trastuzumab versus trastuzumab alone to prevent recurrences in patients with high-risk, HER2 low-expressing breast cancer. Clin Immunol, 2021, **225**: 108679

[18] Wang J, Wang Y, Jiang X, *et al*. Unleashing the power of immune checkpoints: post-translational modification of novel molecules and clinical applications. Cancer Lett, 2024, **588**: 216758

[19] Zou Y, Xie J, Zheng S, *et al*. Leveraging diverse cell-death patterns to predict the prognosis and drug sensitivity of triple-negative breast cancer patients after surgery. Int J Surg, 2022, **107**: 106936

[20] Liu Z, Ding M, Qiu P, *et al*. Natural killer cell-related prognostic risk model predicts prognosis and treatment outcomes in triple-negative breast cancer. Front Immunol, 2023, **14**: 1200282

[21] Liu L, Wu Y, Zhang C, *et al*. Cancer-associated adipocyte-derived G-CSF promotes breast cancer malignancy *via* Stat3 signaling. J Mol Cell Biol, 2020, **12**(9): 723-737

[22] Yan C, Liu Q, Jia R. Construction and validation of a prognostic risk model for triple-negative breast cancer based on autophagy-related genes. Front Oncol, 2022, **12**: 829045

[23] Vittori C, Jeansonne D, Yousefi H, *et al*. Mechanisms of miR-3189-3p-mediated inhibition of c-MYC translation in triple negative breast cancer. Cancer Cell Int, 2022, **22**(1): 204

[24] Wang S, Beeghly-Fadiel A, Cai Q, *et al*. Gene expression in triple-negative breast cancer in relation to survival. Breast Cancer Res Treat, 2018, **171**(1): 199-207

[25] Chou C W, Huang Y M, Chang Y J, *et al*. Identified the novel resistant biomarkers for taxane-based therapy for triple-negative breast cancer. Int J Med Sci, 2021, **18**(12): 2521-2531

[26] Mooney S M, Talebian V, Jolly M K, *et al*. The GRHL2/ZEB feedback loop——a key axis in the regulation of EMT in breast cancer. J Cell Biochem, 2017, **118**(9): 2559-2570

[27] Liu Y, Teng L, Fu S, *et al*. Highly heterogeneous-related genes of triple-negative breast cancer: potential diagnostic and prognostic biomarkers. BMC Cancer, 2021, **21**(1): 644

[28] Thalor A, Joon H K, Singh G, *et al*. Machine learning assisted analysis of breast cancer gene expression profiles reveals novel

potential prognostic biomarkers for triple-negative breast cancer. Comput Struct Biotechnol J, 2022, **20**: 1618-1631

[29] Wang S C, Sun H L, Hsu Y H, *et al*. α-Linolenic acid inhibits the migration of human triple-negative breast cancer cells by attenuating Twist1 expression and suppressing Twist1-mediated epithelial-mesenchymal transition. Biochem Pharmacol, 2020, **180**: 114152

[30] Guan T, Li M, Song Y, *et al*. Phosphorylation of USP29 by CDK1 governs TWIST1 stability and oncogenic functions. Adv Sci, 2023, **10**(11): e2205873

[31] Zhang D, Sun B, Zhao X, *et al*. Twist1 expression induced by sunitinib accelerates tumor cell vasculogenic mimicry by increasing the population of $CD133^+$ cells in triple-negative breast cancer. Mol Cancer, 2014, **13**: 207

[32] Wang Z, Li R, Hou N, *et al*. PRMT5 reduces immunotherapy efficacy in triple-negative breast cancer by methylating KEAP1 and inhibiting ferroptosis. J Immunother Cancer, 2023, **11**(6): e006890

[33] He G, Zhang Y, Feng Y, *et al*. $SBFI_26$ induces triple-negative breast cancer cells ferroptosis *via* lipid peroxidation. J Cell Mol Med, 2024, **28**(7): e18212

[34] Li J, Lai Y, Ma J, *et al*. MiR-17-5p suppresses cell proliferation and invasion by targeting ETV1 in triple-negative breast cancer. BMC Cancer, 2017, **17**(1): 745

[35] Kollareddy M, Martinez L A. Distinct classes of flavonoids and epigallocatechin gallate, polyphenol affects an oncogenic mutant p53 protein, cell growth and invasion in a TNBC breast cancer cell line. Cells, 2021, **10**(4): 797

[36] Subramaniam M, Hawse J R, Rajamannan N M, *et al*. Functional role of KLF10 in multiple disease processes. Biofactors, 2010, **36**(1): 8-18

[37] Wang J L, Ji W W, Huang A L, *et al*. CEBPA restrains the malignant progression of breast cancer by prompting the transcription of $SOCS_2$. Mol Biotechnol, 2024. DOI: 10.1007/s12033-024-01189-4

[38] Wang S, Chang C W, Huang J, *et al*. Gasdermin C sensitizes tumor cells to PARP inhibitor therapy in cancer models. J Clin Invest, 2024, **134**(1): e166841

[39] Mizoguchi K, Kawaji H, Kai M, *et al*. Granzyme B expression in the tumor microenvironment as a prognostic biomarker for patients with triple-negative breast cancer. Cancers, 2023, **15**(18): 4456

[40] Abu Bakar N F A B, Yeo Z L, Hussin F, *et al*. Synergistic effects of combined cisplatin and *Clinacanthus nutans* extract on triple negative breast cancer cells. J Taibah Univ Med Sci, 2023, **18**(6): 1220-1236

[41] Nassar E, Hassan N, El-Ghonaimy E A, *et al*. Syndecan-1 promotes angiogenesis in triple-negative breast cancer through the prognostically relevant tissue factor pathway and additional angiogenic routes. Cancers, 2021, **13**(10): 2318

[42] Krautschneider S L, Troschel F M, Vadillo E, *et al*. Enzymatic digestion of cell-surface heparan sulfate alters the radiation response in triple-negative breast cancer cells. Arch Med Res, 2022, **53**(8): 826-839

[43] Zeng W, Luo Y, Gan D, *et al*. Advances in Doxorubicin-based nano-drug delivery system in triple negative breast cancer. Front Bioeng Biotechnol, 2023, **11**: 1271420

[44] Mittendorf E A, Zhang H, Barrios C H, *et al*. Neoadjuvant atezolizumab in combination with sequential nab-paclitaxel and anthracycline-based chemotherapy versus placebo and chemotherapy in patients with early-stage triple-negative breast cancer (IMpassion031): a randomised, double-blind, phase 3 trial. Lancet, 2020, **396**(10257): 1090-1100

[45] Yan H, Luo B, Wu X, *et al*. Cisplatin induces pyroptosis *via* activation of MEG3/NLRP3/caspase-1/GSDMD pathway in triple-negative breast cancer. Int J Biol Sci, 2021, **17**(10): 2606-2621

基于集落刺激因子相关基因开发三阴性乳腺癌的预后模型*

郭雨轩[1)] 王智禹[1)] 肖沛瑶[1)] 郑婵娟[1)] 付淑君[1)] 贺光春[1)] 龙 俊[4)]
汪 婕[3)] 邓锡云[1)**] 王忆安[1,2)**]

([1)] 湖南师范大学医学院，模式动物与干细胞生物学湖南省重点实验室，长沙 410013；
[2)] 湖南省生殖与转化医学工程技术研究中心，长沙 410013；
[3)] 福建医科大学基础医学院肿瘤研究所病理学系及诊断病理中心，福州 350122；
[4)] 深圳大学材料科学与工程学院，深圳市能源电催化材料重点实验室，深圳市高分子科学与技术重点实验室，广东省新能源材料服务安全重点实验室，深圳518055）

摘要 目的 三阴性乳腺癌（TNBC）是目前预后最差的乳腺癌亚型，并且缺乏有效的治疗靶点。集落刺激因子（CSFs）是一类能调节血细胞生成和刺激免疫细胞生长发育的细胞因子，在TNBC的恶性进展中发挥重要作用。本文旨在根据集落刺激因子相关基因（CRGs）的表达情况，构建一种新的预后模型，并且分析TNBC对免疫治疗和药物治疗的敏感性。**方法** 从公共数据库中下载CRGs，在TCGA-BRCA数据库中筛选在正常和TNBC组织中表达差异的CRGs。通LASSO Cox回归分析确定用于构建模型的差异基因并建立CRGs风险评分（CRRS）。进一步在高低风险组中分析了CRRS与患者预后、临床特征、肿瘤微环境之间的相关性，并评估了CRRS与免疫治疗和药物敏感性之间的关系。**结果** 鉴定了842个在TNBC患者乳腺癌组织中存在表达差异的CRGs，并确立了13个CRGs用于构建预后模型。Kaplan-Meier生存曲线、时间依赖性受试者工作特征曲线等证实了高CRRS的TNBC患者总生存期更短，并且在GEO数据库中进一步证实了CRRS预后模型的预测能力。结合临床特征的列线图证实CRRS是TNBC患者预后的独立因素。并且高风险组患者肿瘤微环境中免疫浸润水平较低，对部分化疗药物敏感。**结论** 本文开发了由13个DEGs组成的CRRS模型，该模型可能成为预测TNBC患者预后和指导临床治疗的有用工具，并且其中的关键基因可能是未来治疗的潜在分子靶标。

关键词 三阴性乳腺癌，集落刺激因子，预后模型，肿瘤微环境，药物敏感性
中图分类号 R736，R318 **DOI**：10.16476/j.pibb.2024.0281

* 国家自然科学基金（82303126，82303069，82173374，82103342），湖南省自然科学基金（2024JJ6325，2024JJ5287）和福建省自然科学基金（2023J05038）资助。
** 通讯联系人。
王忆安 Tel：0731-88912415，E-mail：wangyian@hunnu.edu.cn
邓锡云 Tel：0731-88912415，E-mail：dengxiyunmed@hunnu.edu.cn
收稿日期：2024-06-30，接受日期：2024-09-02