ORIGINAL PAPER

A CD8+ T Cell-Related Genes Expression Signature Predicts Prognosis and the Efficacy of Immunotherapy in Breast Cancer

Lian-hua Lv[1] · Jia-rong Lu[2] · Tao Zhao[2] · Jing-li Liu[3] · Hai-qi Liang[2]

Abstract
Immunotherapy has been applied to patients with breast cancer. However, only part of patients benefits from the current immunotherapy. Accurate prediction of individual response to immunotherapy can be beneficial for breast cancer management. CD8+ T cells are the main force of anti-tumor immunity. This study aimed to establish a CD8+ T cell-related gene expression signature for prediction of breast cancer prognostic and immunotherapy efficacy. RNA-seq transcriptomic data was the basics of this research. Weighted gene co-expression network analysis (WGCNA) and the least absolute shrinkage and selection operator (LASSO) Cox regression analysis established the prognostic signature. We identified 290 CD8+ T cell-related genes in the training set and established a risk-score model based on 8-genes panel (*SOCS1*, *IL10*, *CAMK4*, *CXCL13*, *KIR2DS4*, *TESPA1*, *CD70* and *ICAM4*). Subsequently, univariate Cox regression analysis suggested that high risk-score was a risk factor for breast cancer (HR = 3.1, 95%CI 2.0–4.8, P < 0.001). In tumor microenvironment, high-risk tumors present decreased tumor infiltrating CD8+ T cells and increased M2 macrophages. The low-risk patients may benefit more from immune checkpoint blockade immunotherapy than the high-risk patients. Moreover, breast tumors which sensitive to immune checkpoint inhibitor (ICI) showed higher *IL10* expression.

Keywords Breast cancer · Immunotherapy response · Immune checkpoint blockade · Prognosis · T cell

Abbreviations

TCGA-BRCA	The Cancer Genome Atlas- breast cancer
ssGSEA	Single-sample gene set enrichment analysis
WGCNA	Weighted gene co-expression network analysis
OS	Overall survival
LASSO	Least absolute shrinkage and selection operator
KEGG	Kyoto Encyclopedia of Genes and Genomes
ROC	Receiver operating characteristic
AUC	Area under the ROC curve
ICI	Immune checkpoints inhibitors
IPS	Immunophenoscore
CR	Complete response
PR	Partial response
SD	Stable disease
PD	Progressive disease

Lian-hua Lv and Jia-rong Lu contributed equally to this work.

✉ Hai-qi Liang
liahiki@163.com

1 The Second Clinical Medical College, Guangxi Medical University, Nanning 530021, Guangxi, China

2 The First Clinical Medical College, Guangxi Medical University, Nanning 530021, Guangxi, China

3 Nursing College, Guangxi Medical University, Nanning 530021, Guangxi, China

Background

Breast cancer is the most common malignant tumor in women, and the incidence rate is increasing by about 0.5% per year [1]. According to Global cancer statistics 2018, there are 2,088,849 new cases of breast cancer worldwide each year, and 626,679 related deaths [2]. With the treatment improved in recent years, the 5-year survival rate of breast cancer in the United States has reached 90%, however, once distant metastasis occurs, the survival rate drops to 26% [1, 3]. Current treatment methods cannot cure metastatic advanced breast cancer, and the pros and cons of treatment need to be measured individually to obtain the best outcome [4]. Immunotherapy has been used in advanced breast cancer

patients who are resistant to first-line therapy and has shown clinical benefits, which is a promising new therapy [5].

The tumor immune microenvironment is a key part in the development and progression of tumors [6]. Large clinical trial has shown that the infiltration of CD8[+] T cells predicts a better prognosis for breast cancer [7]. However, there is significant heterogeneity in the condition of tumor-infiltrating T cells, and T cell function is a window to assess the immune microenvironment [8]. T cells have immune checkpoints such as cytotoxic T-lymphocyte antigen 4 (CTLA-4), which directly inhibit cellular immune response [9]. Immune checkpoint inhibitor (ICI) is currently used in immunotherapy to induce anti-tumor immunity by blocking CTLA-4 and PD-1/PD-L1 [10]. The patient selection and efficacy evaluation system for immune checkpoint blockade therapy is still incomplete, and tumor CD8[+] T cells are one of the important reference elements.

Previous study has used immunohistochemical (IHC) to evaluate the relationship between tumor-infiltrating CD8[+] T cells level and breast cancer prognosis [11]. With the advent of the era of high-throughput sequencing, the acquisition of tumor gene expression data has become convenient. Gene transcription and epigenetics were proved to play a crucial role in the differentiation and functional activation of CD8[+] T cells [5]. Xie et al. reported an immune-related gene prognosis model of breast cancer, but it had no directly correlation with tumor-infiltrating CD8[+] T cells and did not be tested the relationship with immune checkpoints expression [12]. CD8[+] T cell-related genes have shown potential for the prognosis of breast cancer. Therefore, we expected to predict the prognosis of breast cancer and evaluate the tumor immune status through the expression of CD8[+] T cell-related genes.

This study employed gene transcription data to construct a prognostic signature of breast cancer. We used training set and validation set to investigated the relationship between the signature and clinical characteristics of breast cancer. The risk-score was further used to predicted the efficacy of immune checkpoint blockade response. In addition, gene functions and biological pathways were further explored.

Methods

Data Collection

The training set came from The Cancer Genome Atlas-breast cancer (TCGA-BRCA) database. The RNA transcription data and clinical information of 1097 cases of tumor and 121 cases of normal breast tissues were downloaded through the UCSC Xena website (https://xenabrowser.net) [13]. The gene transcription data is converted using the formula log2 (expression value + 1) and corrected by

RSEM. The validation sets included GSE20685, GSE7390, GSE16446, GSE69031, GSE21653, GSE25066, GSE26971, GSE75688, GSE124821 [14–22] was downloaded from the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/). IMvigor210 data has been available in IMvigor210CoreBiologies R package and is freely available in website [23]. Detail of above dataset was described in Supplementary Table 1.

Identify CD8[+] T Cell-Related Genes in the Training Set

The ESTIMATE algorithm calculated immune-score through transcription profile analysis based on the specific gene expression of tumor [24]. We calculated an immune-score for each individual TCGA-BRCA sample. Samples with an immune-score above the upper quartile or below the lower quartile were divided into high- and low- immune score group, respectively. The immune-related genes were defined as the differential expressed genes between high- and low- immune score group samples. We chose immune-score only above the upper quartile or below the lower quartile because of greater differences in immune infiltration status may exist between the two groups. Thus, the differential expressed genes may be more related to the tumor immune microenvironment. In the WGCNA and later analyses, all the 1097 tumor samples were included. Next, single-sample gene set enrichment analysis (ssGSEA) [25] employed the expression of gene markers to evaluate the relative levels of tumor infiltrating immune cells in training set. Weighted gene co-expression network analysis (WGCNA) [26] is a systems biology method to find highly related gene modules [26]. We used WGCNA to identify CD8[+] T cell-related gene modules in breast cancer. The method details of ssGSEA and WGCNA were described in Supplementary Method.

Construction of Prognostic Signature Based on CD8[+] T Cell-Related Genes

In order to make the signature reflect the patient's prognosis, univariate Cox regression analysis was used to screen CD8[+] T cell-related genes that affect overall survival (OS). When acquiring superabundant prognostic-related genes, the least absolute shrinkage and selection operator (LASSO) Cox regression analysis would reduce the dimensionality of the model to avoid the risk of overfitting. The prognostic signature was established by the LASSO Cox regression fitting model, which provided key prognostic genes and corresponding regression coefficients, and verified by tenfold crossover. A risk-score model is constructed by formula $riskscore = \sum_{i=1}^{n}(Coefficient_i \times Expression\ of\ gene_i)$. In individual validation sets, multivariate Cox regression

analysis was used to calculate the coefficient of risk-score model based on the same gene panel. The median risk score was used to divide the patients into high- and low-risk groups in both training and validation sets.

Prognostic Ability of the Risk Score Model

Based on the risk score model, the risk score was calculated for each patient in the training set. Kaplan–Meier curve and log-rank test were applied to verify the prognostic value of the risk score model. Time-dependent receiver operating characteristic (ROC) curves assessed the accuracy of risk score for predicting breast cancer OS at 3, 5 and 10 years. Subsequently, univariate and multivariate Cox regression analysis verified whether the risk score was an independent prognostic indicator. At the same time, a nomogram was generated for clinical application. In addition, Kaplan–Meier analysis was performed on all the validation dataset.

The Risk-Score Model Predicts Immunotherapy Response

We first explored the correlation of risk score and expression of immune checkpoints in TCGA dataset. Subsequently, the CIBERSORT score of 22 immune infiltration cells and ssGSEA score of 28 immune infiltration cells were calculated in TCGA dataset. The detail process to access CIBERSORT [27] score was also described in Supplementary Method. Due to traditional whole genome sequencing was limited in the evaluation of gene panel expression in differential cell types, we introduce single-cell sequencing dataset GSE75688 to detect the expression of gene panel in breast cancer tumor microenvironment. We further obtained the corresponding anti-CTLA-4 and anti-PD-1/PD-L1 predictive response values of TCGA breast cancer patients from The Cancer Immunome Atlas (https://tcia.at/) [28]. The Cancer Immunome Atlas provides results of comprehensive immunogenomic analyses for 20 solid cancers from TCGA and other data sources. Each sample is given an immunophenoscore (IPS) in specific ICI therapy by scoring tumor immunogenicity from 0 to 10 that can predict the efficacy of ICI therapy. We then compared the IPS between high- and low- risk groups breast cancer. At the same time, the immunotherapy response prediction ability of the risk-score model was verified in IMvigor210 and GSE124821 dataset. IMvigor210 dataset contains transcriptome data of clinical trials that applied atezolizumab in locally advanced and metastatic urothelial carcinoma [29]. And GSE124821 includes bulk mRNA-seq data of anti-CTLA-4/anti-PD1 -treated and pre-treated from sensitive and resistant mouse mammary tumor samples [22].

Gene Functional Enrichment Analysis

In order to understand the biological functions related to the prognostic signature, differentially expressed genes between the high- and low- risk groups were searched in the whole TCGA-BRCA transcription set. Following, the differentially expressed genes were used to enrich the biological pathways of the Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The enriched Gene Ontology terms and KEGG pathways may reveal the activation difference of biological process in difference risk groups.

Statistical Analysis

Statistical analyses were performed using R software (version 3.6.1) (https://www.r-project.org/) and Stata (version 15.0). R packages involved in this study were concluded in Supplementary Method. When $P < 0.05$, the analysis results can be considered as statistically significant.

Results

Sample Characteristics and Immune-Score

The baseline information of the TCGA-BRCA cohort is summarized in Table 1, and the clinical information of the GSE20685 cohort is provided in Supplementary Table 2. The flowchart of signature construction and verification was presented in Fig. 1. We first calculated an immune-score for each tumor sample in TCGA-BRCA dataset (Fig. 2A, Supplementary Table 3). Then, 1557 differentially expressed genes between the high- and low- immune-score groups were identified as immune-related genes for subsequent genetic screening (Fig. 2B, Supplementary Table 4). The ssGSEA assessed the abundance of tumor infiltrating immune cells in TCGA-BRCA samples (Fig. 2C).

Gene screening and Prognostic Signature Establishment

WGCNA analysis was performed on 1557 immune-related genes and the optimal soft threshold ($β = 2$) were chosen (Fig. 2D). The genes clustered into 6 different modules (Fig. 2E). The turquoise module clustered the most genes and had the highest correlation with tumor-infiltrating immune cells. Supplementary Table 5 provided the genes in the turquoise module. Correlation coefficient of the turquoise module and activated CD8$^+$ T cells was 0.85 (Fig. 2F). The turquoise module genes with both correlation coefficient and significance greater than 0.6 were defined as CD8$^+$ T cell-related genes in breast cancer (n = 290, Fig. 2G). Subsequently, univariate Cox regression analysis identified 21

Table 1 Baseline information of training set TCGA cohort

Characteristics		Total (n = 1097)	Low-risk (n = 548)	High-risk (n = 549)	P (χ^2 test)
Age	< 60	446 (40.7%)	256 (46.7%)	190 (34.6%)	< 0.001*
	≥ 60	372 (33.9%)	153 (27.9%)	219 (39.9%)	
	NA	279 (25.4%)			
AJCC stage	I-II	579 (52.8%)	303 (55.3%)	276 (50.3%)	0.004*
	III-IV	190 (17.3%)	81 (14.8%)	119 (21.7%)	
	NA	328 (29.9%)			
T stage	T1-T2	679 (61.9%)	354 (64.6%)	325 (59.2%)	0.080
	T3-T4	109 (9.9%)	47 (8.6%)	62 (11.3%)	
	NA	309 (28.2%)			
N stage	N0-N1	649 (59.2%)	333 (60.8%)	316 (57.6%)	0.051
	N2-N3	142 (12.9%)	60 (10.9%)	82 (14.9%)	
	NA	306 (27.9%)			
M stage	M0	771 (70.3%)	386 (70.4%)	385 (70.1%)	0.287
	M1	14 (1.3%)	5 (0.9%)	9 (1.6%)	
	NA	312 (28.4%)			
ER status	Positive	599 (54.6%)	287 (52.4%)	312 (56.8%)	0.062
	Negative	179 (16.3%)	100 (18.2%)	79 (14.4%)	
	NA	319 (29.1%)			
PR status	Positive	521 (47.5%)	253 (46.2%)	268 (48.8%)	0.320
	Negative	254 (23.2%)	133 (24.3%)	121 (22.0%)	
	NA	322 (29.4%)			
HER2 status	Positive	130 (11.9%)	56 (10.2%)	74 (13.5%)	0.134
	Negative	431 (39.3%)	218 (39.8%)	213 (38.8%)	
	NA	536 (48.9%)			

T tumor, *N* Node, *M* Metastasis, *ER* estrogen receptor, *PR* progesterone receptor

* *P* < 0.05

prognostic-related genes (Fig. 3A). Further, LASSO Cox regression analysis constructed the prognostic signature that included 8 CD8$^+$ T cell-related genes panel (Fig. 3B, Table 2). The formula was as follows: risk score = (-0.121)×*SOCS1* + 0.520×*IL10* + (-0.191)×*CAMK4* + (-0.013)×*CXCL13* + (-0.122)×*KIR2DS4* + (-0.127)×*TESPA1* + (-0.123)×*CD70* + (-0.022)×*ICAM4*. Risk score of patients in training and validation set was calculated and the patients were divided into high- and low- risk groups according to the median risk score. The risk-score of TCGA-BRCA and corresponding 8-genes panel expression were exhibited (Fig. 3C). Interestingly, risk-score of basal-like subtype patients was lower than Luminal B subtype (Fig. 3D). Results also showed that risk-score of patients in AJCC stage III/IV was higher than in stage I/II (Fig. 3E).

Prognostic Ability of the Risk Score Model

Kaplan–Meier analysis proved that patients with high-risk had worse survival probability in TCGA-BRCA dataset (P < 0.001, Fig. 4A), and the result in GSE20685 was similar (P = 0.002, Fig. 4B). The high-risk patients also had poorer OS, disease free survival (DFS), or metastatic free interval

(MFI) in the validation set (Supplementary Fig. 1). *IL10* is the heaviest weighed gene in the risk-score model. It is also a risk factor of breast cancer patients (HR = 1.261, 95%CI 1.035–1.430, Fig. 4C), but no significant was seen in the subgroup analysis (Supplementary Fig. 2). Time-dependent ROC curve analysis showed that area under the ROC curve (AUC) of the risk score model predicting 3, 5, and 10-year OS of TCGA breast cancer patients were 0.736, 0.676, and 0.761, respectively (Fig. 4D). Meanwhile, AUC of the risk score model predicting 3, 5, and 10-year OS of GSE20685 cohort were 0.818, 0.781, and 0.817, respectively (Fig. 4E). Besides, AUC of TCGA-BRCA molecular subtypes was displayed (Fig. 4F). The chi-square test showed that the elderly (age > 60) and AJCC stage III/IV patients had a higher proportion in the high-risk group (Table 1). Multivariate Cox regression analysis indicated that risk-score was an independent risk factor for breast cancer (Table 3). Further, a nomogram included indexes age, AJCC stage, HER2 status, PR status, and risk-score was drawn for the clinical prognostic application, and the calibration curve estimated the efficiency of nomogram prediction (Fig. 4G, H). The C-index was 0.80, which means a relative high accuracy of the nomogram.
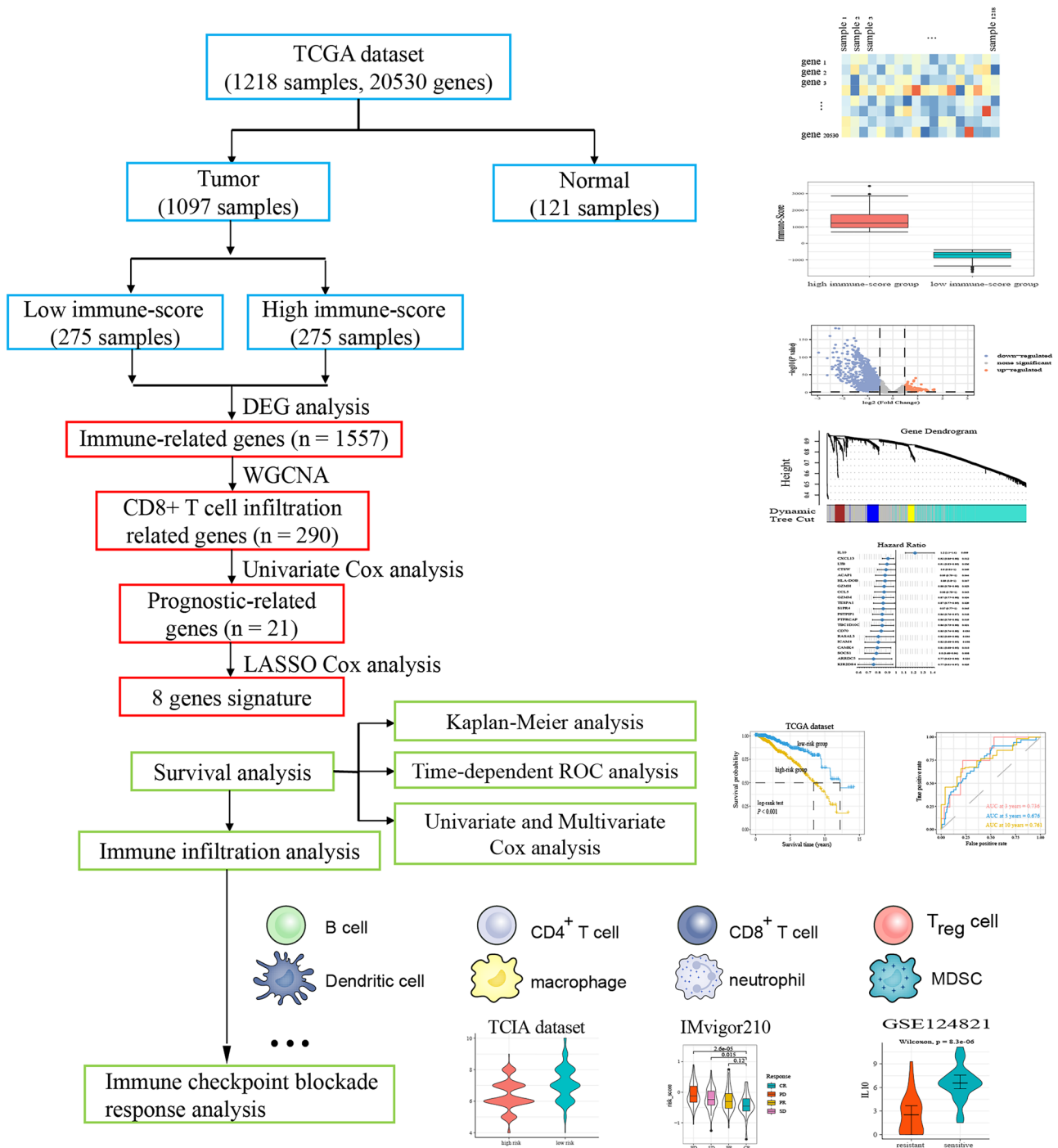
Fig. 1 The flowchart of signature construction and verification

The Risk Score Associated with Immune Infiltration and Efficacy of Immunotherapy

Single-cell RNA sequencing dataset GSE75688 showed that *CAMK4*, *CXCL13*, and *TESPA1* were higher expressed in T cells, but *IL10* was lower expressed in T cells (Fig. 5A-H). Both CIBERSORT and ssGSEA method proved that high-risk tumors had decreased tumor-infiltrating CD8$^+$ T cells (Supplementary Fig. 3). High-risk tumors also had decreased tumor-infiltrating M1 macrophages but increased M2 macrophages (Supplementary Fig. 3A), which suggested the immunosuppression microenvironment in high-risk tumors. Bulk data analysis showed that the risk score was negatively correlated with expression of the immune
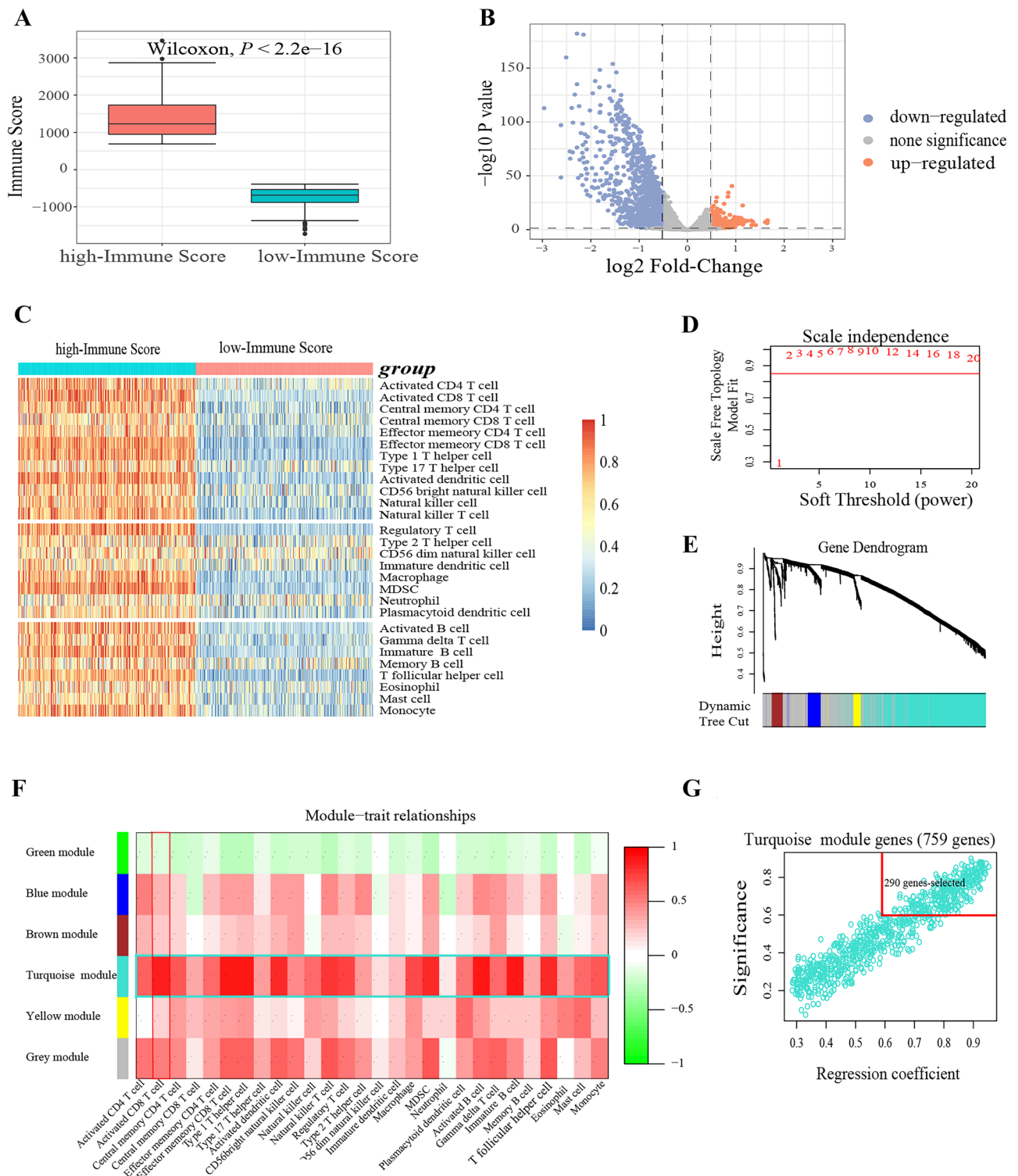
Fig. 2 Selection of CD8+ T cells-related genes. **A**: Immune-score of TGGA-BRCA samples. **B**: Differentially expressed genes between high- and low- immune-score groups TGGA-BRCA. **C**: 28 types tumor infiltrating immune cells in high- and low- immune-score groups TGGA-BRCA. **D-G** WGCNA identified the CD8+ T cells-related genes. **D**: selection of the optimal soft threshold (β = 2). **E**: 6 gene clusters in the gene dendrogram. **F**: the correlation coefficient between 6 gene modules and 28 kinds tumor-infiltrating immune cells. **G**: the regression coefficient of turquoise module genes expression and tumor-infiltrating CD8 + T cells level
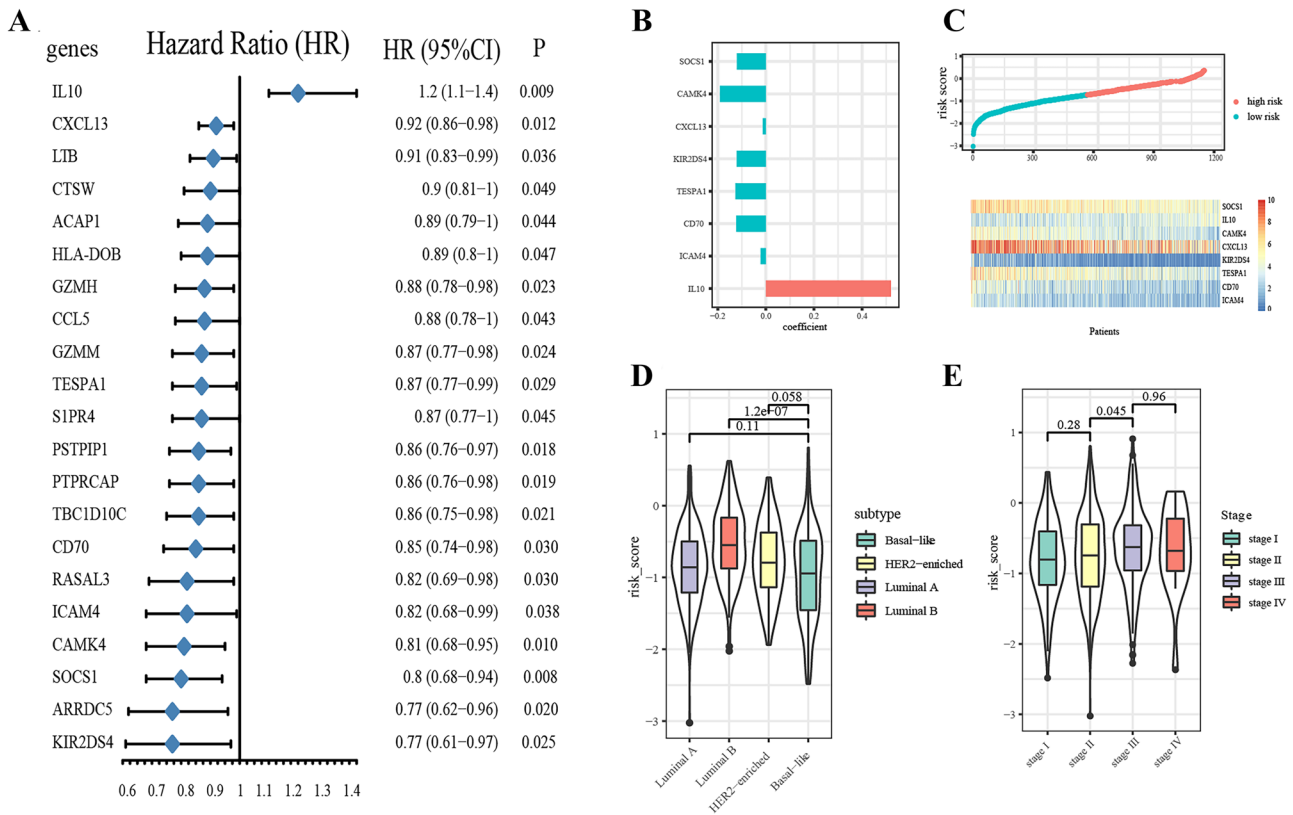
Fig. 3 Construction of CD8$^+$ T cells-related genes expression signature. **A**: Univariate Cox regression analysis result of 21 survival-related genes in the study. **B**: Coefficient of the 8 genes in the signature. **C**: Risk scores distribution and heatmap of the 8 signature component genes expression. **D**: Risk scores of TGGA-BRCA molecular subtypes. **E**: Risk scores of TGGA-BRCA in different AJCC stages

checkpoints such as PDCD1 (PD-1), CD274 (PD-L1), CTLA4, IDO1 and TIGIT, as well as immune stimulators including CD27, CD40, CD48, ICOS and CXCR4 (Fig. 6A). In the treatment response prediction of TCGA-BRCA patients to ICI, the anti-CTLA-4 and anti-PD-1/PD-L1 IPS were lower in high-risk group, that may mean lack of sensitivity to ICI in high-risk breast cancer patients (Fig. 6B). In addition, the risk-score model was validated in IMvigor210 dataset and the results indicated that urothelial carcinoma

patients who achieved complete response (CR) to PD-L1 blockade had lower pre-treatment risk-score than the stable disease (SD) and progressive disease (PD) patients (Fig. 6C). Meanwhile, patients who received CR may expressed higher *IL10* before PD-L1 blockade than the partial response (PR) and PD patients (Fig. 6D). In the bulk mRNA-seq dataset GSE124821 of breast cancer mouse models, we also found tumors that sensitive to PD-1 and CTLA4 blockade had higher *IL10* expression in pre-treatment (Fig. 6E).

Table 2 Signature component genes and the corresponding coefficients

| Gene symbol	Full name	Coefficient
SOCS1	Suppressor of cytokine signaling 1	-0.121
IL10	Interleukin 10	0.520
CAMK4	Calcium/calmodulin-dependent protein kinase IV	-0.191
CXCL13	Chemokine (C-X-C motif) ligand 13	-0.013
KIR2DS4	Killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 4	-0.122
TESPA1[†]	Thymocyte expressed, positive selection associated 1	-0.127
CD70	CD70 molecule	-0.123
ICAM4	Intercellular adhesion molecule 4	-0.022

[†]Also known as KIAA0748
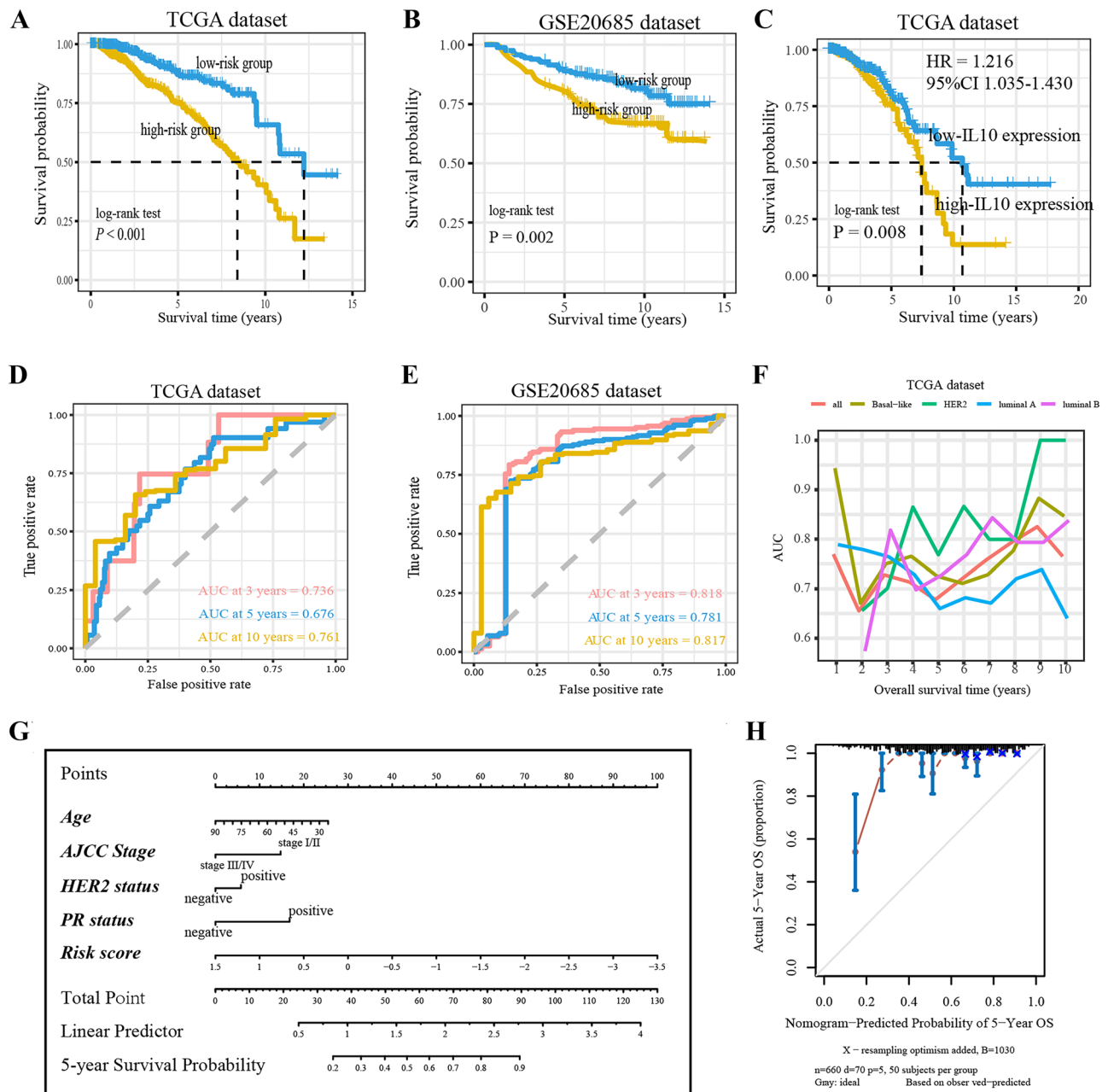
Fig. 4 Survival analyses results. **A-B** Kaplan–Meier plot analyzed overall survival (OS) probability of high- and low- risk breast cancer in **A**: TCGA-BRCA; **B**: GSE20685. **C**: OS probability of high- and low- IL10 expression groups in TCGA-BRCA. **D-E** Time-dependent ROC curve analysis predicted 3, 5 and 10 years OS in **D**: TCGA-BRCA; **E**: GSE20685. **F**: AUC of time-dependent ROC curve analysis OS in TCGA-BRCA molecular subtypes. **G**: The nomogram for the clinical use of the signature. **H:** A calibration curve examined the nomogram

Biological Pathways Associated with Risk Model

Gene functional enrichment analysis showed that risk-score were linked to some Gene Ontology and KEGG terms that related to tumor immune and signal transduction, such as "T cell activation", "JAK-STAT cascade", "chemokine signaling pathway", "protein kinase B signaling", "JNK cascade", "PI3K signaling pathway", "Wnt signaling pathway", "Ras

signaling pathway", "B cell receptor signaling pathway" and "PD-L1 expression and PD-1 checkpoint pathway in cancer" (Fig. 7A, B).

Pan-Cancer Analysis

The CD8 + T cell-related genes expression signature was further tested in TGGA pan-cancer dataset. The risk-score

Table 3 Univariate and multivariate Cox regression analyses of the signature and clinicopathological characteristics of OS

Characteristics		Univariate analysis		Multivariate analysis	
		HR (95%CI)	P	HR (95%CI)	P
Age		1.07 (1.02–1.12)	0.002*	1.020 (1.000–1.040)	0.048*
AJCC stage	III-IV VS. I-II	1.71 (1.11–2.69)	0.029*	2.258 (1.358–3.755)	0.002*
ER status	Negative VS. Positive	1.32 (0.87–2.04)	0.300	1.258 (0.595–2.660)	0.548
PR status	Negative VS. Positive	1.41 (0.94–2.23)	0.097	2.458 (1.221–4.950)	0.012*
HER2 status	Negative VS. Positive	0.94 (0.48–1.82)	0.864	2.283 (1.025–5.084)	0.043*
Risk score		3.82 (2.61–5.73)	<0.001*	4.382 (2.636–7.283)	<0.001*

ER estrogen receptor, *PR* progesterone receptor, *HR* hazard ratio, *95% CI* 95% confidence interval

* *P* < 0.05

models were additionally constructed in 8 types of malignant tumors: bladder urothelial carcinoma, cervical squamous cell carcinoma and endocervical adenocarcinoma, head and neck squamous cell carcinoma, lung adenocarcinoma, ovarian serous cystadenocarcinoma, stomach adenocarcinoma, liver hepatocellular carcinoma, colon adenocarcinoma. Kaplan–Meier analysis proved that survival probability was significant statistical difference between high- and low- risk patients among the 8 types tumors (Supplementary Fig. 4). Except for stomach adenocarcinoma, the IPSs were also lower in high-risk group (Supplementary Fig. 5). However, IL10 expression was no longer a risk-factor in pan-cancer meta-analysis (Supplementary Fig. 6).
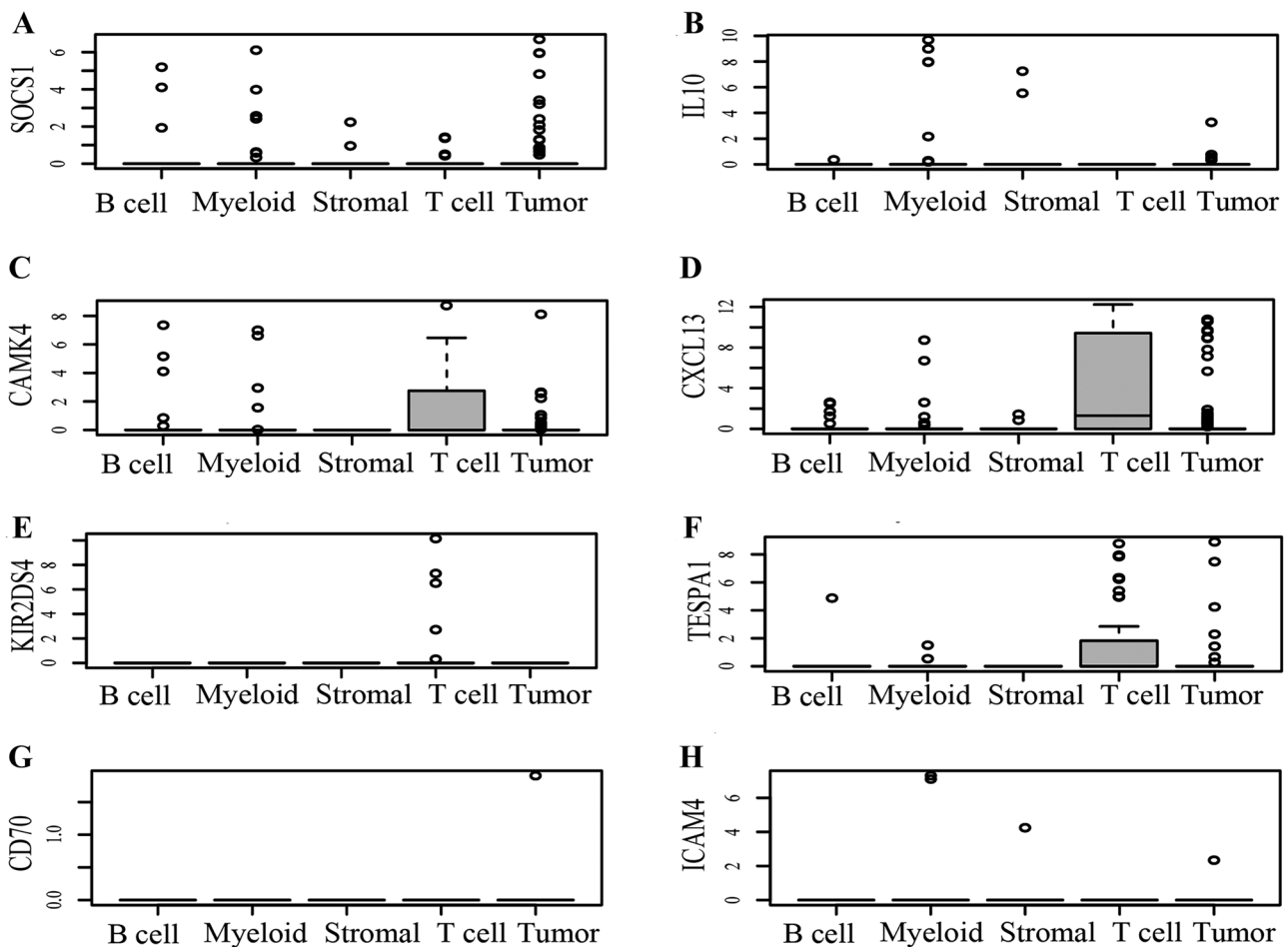

Fig. 5 Single-cell RNA sequencing dataset GSE75688 showed the expression of 8 signature component genes in T cell
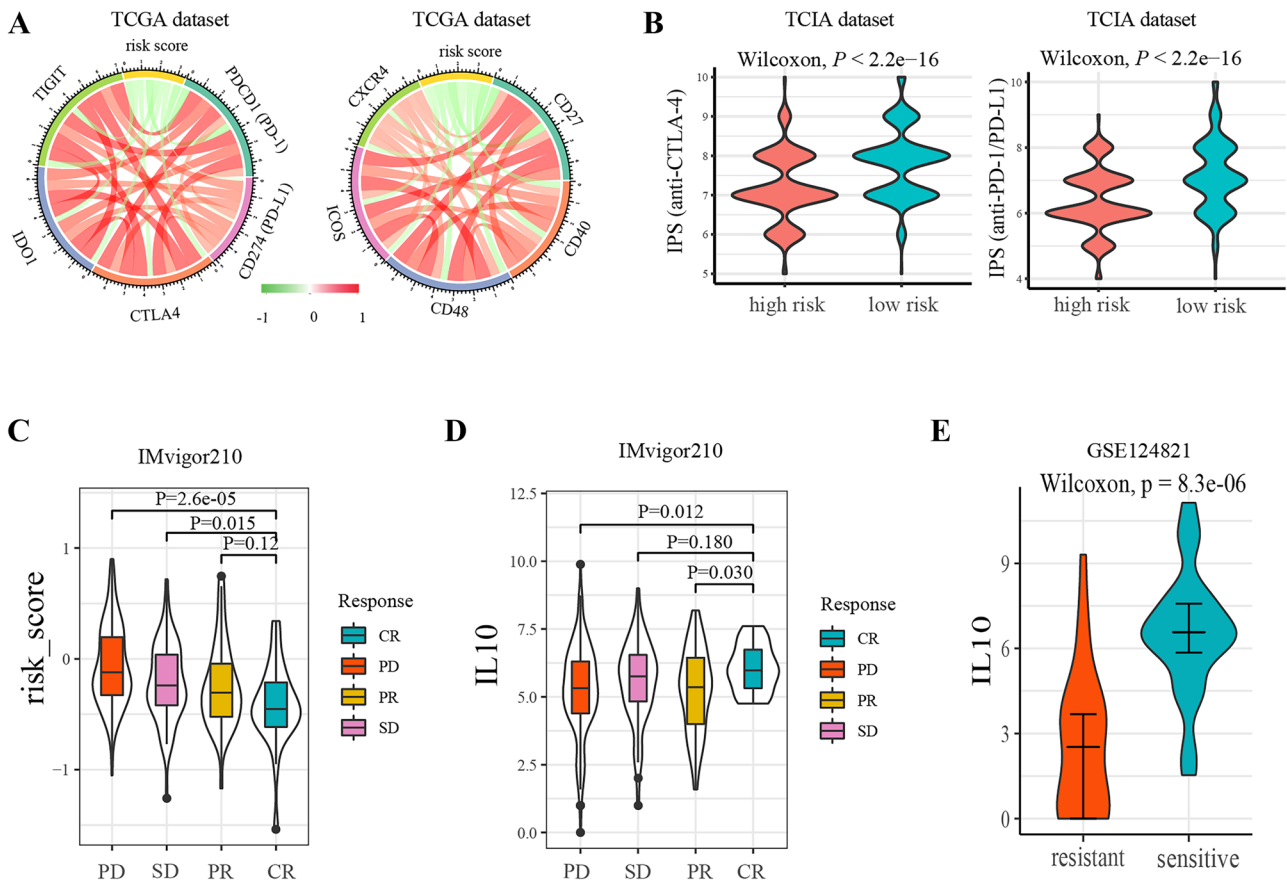
Fig. 6 Immunotherapy response prediction. **A**: The correlation of risk score and immune molecular expression. **B**: IPS predicted the relative response of anti-CTLA-4 and anti-PD-1/PD-L1 ICI therapy. **C**: Risk-score of different response group IMvigor210 clinic trial patients. **D**: Pre-treatment IL10 expression of different response group IMvigor210 clinic trial patients. **E**: Pre-treatment IL10 expression in the PD-1/CTLA4 block-sensitive group in mouse mammary tumors. *CR*, complete response; *PR*, partial response; *SD*, stable disease; *PD*, progressive disease

Discussion

In the present study, we constructed a CD8[+] T cell-related genes prognostic signature based on TCGA-BRCA dataset.

The construction of risk-score model is a process of hub-genes selection. We first used ESTIMATE to defined the 1557 immune-related genes from the whole gene expression profile with 20,530 genes. Then we performed WGCNA to
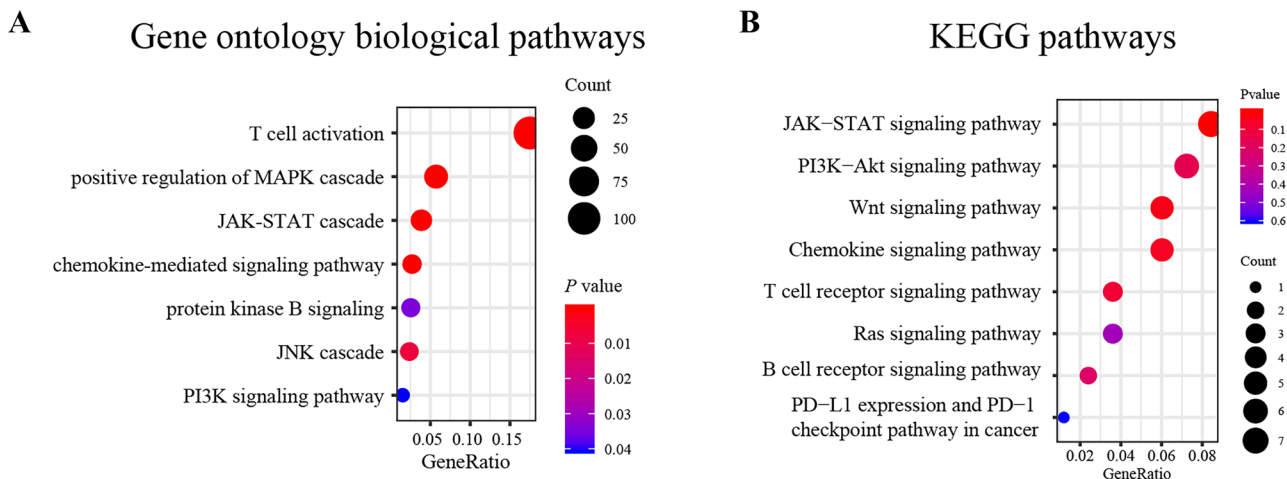

Fig. 7 The signature related biological pathways. **A**: gene ontology terms. **B**: KEGG biological pathways

obtain 290 CD8$^+$ T cell infiltration related genes for prognosis model establishment. Further, univariate Cox regression analysis identified 21 prognostic-related genes. Finally, LASSO Cox regression analysis punished the regression coefficient of 21 candidate genes and fitting the risk model. Totally 8 significant survival-related genes were comprised in the risk model. Among the 8 hub genes, *SOCS1* has been found to be a "non-classical" checkpoint blocker that negatively regulates cytokine receptor signaling and plays a key role in CD8$^+$ T cell response [30]. Deng et al. reported that IL-10 activated STAT3 and acted as an immunosuppressive molecule in triple-negative breast cancer [31]. Existing study has shown that *CAMK4* can activate transcription factors of immune cells such as T cells to enhance the immune response [32]. In recent years, *TESPA1* has been found to be a specific regulator on TCR signaling [33]. In addition, *ICAM4* is involved in antigen recognition of monocytes and macrophages in innate immunity [34]. Single-cell RNA sequencing data showed that *CAMK4*, *CXCL13* and *TESPA1* were highly expressed in T cells, while the other 5 genes, including *IL10*, were low expressed in T cells.

High-risk defined by the risk-score model indicated a poor prognosis for breast cancer. Since time-dependent ROC curve analysis considers both the follow-up time and death events [35], it was used to evaluate the prognostic prediction accuracy of the signature. In both training set and validation set, the risk-score was able to distinguish patients with differential survival outcomes, including overall survival, disease free survival and metastatic free survival. The risk-score, age, AJCC stages, HER2 status and PR status were independent risk factors for breast cancer in this study.

Pathological stage is an important means for clinicians to evaluate the survival of cancer patients, but tumor immunity is also a key factor affecting the prognosis [36]. There are already some immune-score models that complement the existing TNM stages [37, 38]. In addition to predicting prognosis, the signature we constructed has the potential to reflect the efficacy of immune checkpoint blockade. The immunotherapy response is close related to the expression of tumor immune checkpoints [39]. So, we first explored the correlation between risk score and immune checkpoints expression. Interestingly, the risk-score negatively correlated with the expression of immune checkpoints and immune stimulators in tumor.

Further analysis suggested the high-risk tumors presented lower level of tumor-infiltrating CD8$^+$ T cells. CD8$^+$ T cells are the main anti-tumor cells, and the application of ICI can enhance the activity of CD8$^+$ T cells [40]. Therefore, our CD8$^+$ T cell infiltration-related signature can predict the prognosis and may reflect immunotherapy response of breast cancer. High-risk tumors had also less tumor-infiltrating M1 macrophages but more M2 macrophages. Previous study demonstrated that both innate and adaptive immune cells in breast cancer showed immunosuppressive gene expression characteristics [21]. The decrease of M1/M2 ratio is the damage to anti-tumor immunity, M1 macrophages have anti-tumor effect, and the infiltration of M2 macrophages is linked to the poor prognosis of tumor [41]. At the same time, M2 macrophages can be induced by IL10 [42], which suggested that our signature is associated with the infiltration of macrophages.

In March 2019, atezolizumab became the first ICI to be approved for breast cancer, delivering improved outcomes and showing promising application potential [43]. However, there was still no large-scale transcriptome data of breast cancer immunotherapy. We used IPS that calculated by TCGA RNA-seq to substitute the efficacy of immune checkpoint blockade. Charoentong et al. have demonstrated that IPS can reflect cancer patients' relative response to anti-PD-1/PD-L1 and anti-CTLA-4 treatment [28]. We found the low-risk TCGA-BRCA patients achieved higher anti-CTLA-4 and anti-PD-1/PD-L1 IPS. It may guide the selection of immune checkpoint blockade therapy for breast cancer patients. Furthermore, the IMvigor210 dataset was involved in this study and it confirmed that the anti-PD-L1 CR group had lower risk-score. Additionally, the pre-treatment *IL10* expression was higher in human patients who achieved CR to atezolizumab, and higher pre-treatment *IL10* expression was found in the PD-1/CTLA4 block-sensitive group in mouse mammary tumors.

Our study established a CD8$^+$ T cell-related gene expression prognostic signature of breast cancer. Although some approval results have achieved, our study still has shortcomings. Firstly, this is a retrospective study based on the online database that still needs to be verified by prospective studies. Secondly, the IPS was used to detect the efficacy of immunotherapy for breast cancer, which may be different from the real world. We will continue to improve the research to obtain further conclusions.

Conclusions

This study established a breast cancer prognostic signature based on 8 CD8$^+$ T cell-related genes. A high-risk indicates an independent risk factor for breast cancer prognosis. Also, high-risk tumors present decreased tumor infiltrating CD8$^+$ T cells and increased M2 macrophages. The low-risk patients may benefit more from immune checkpoint blockade immunotherapy than the high-risk patients. In addition, breast cancer with enhanced *IL10* expression also has an adverse prognosis, but it is more sensitive to ICI. The results provide a new perspective on the prognosis and immunotherapy of breast cancer.

Declarations

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Consent for Publication All authors approved the publication of this manuscript.

Competing Interests The authors declare that they have no competing interests.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA Cancer J Clin. 2021;71:7–33.
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 Cancer J Clin. 2018;68:394–424.
3. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. Breast cancer statistics, 2017, racial disparity in mortality by state. CA Cancer J Clin. 2017;67:439–48.
4. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, et al. Breast cancer Nat Rev Dis Primers. 2019;5:66.
5. Reading JL, Gálvez-Cancino F, Swanton C, Lladser A, Peggs KS, Quezada SA. The function and dysfunction of memory CD8(+) T cells in tumor immunity. Immunol Rev. 2018;283:194–212.
6. Baxevanis CN, Fortis SP, Perez SA. The balance between breast cancer and the immune system: Challenges for prognosis and clinical benefit from immunotherapies. Semin Cancer Biol. 2019.
7. Ali HR, Provenzano E, Dawson SJ, Blows FM, Liu B, Shah M, et al. Association between CD8+ T-cell infiltration and breast cancer survival in 12,439 patients. Ann Oncol. 2014;25:1536–43.
8. van der Leun AM, Thommen DS, Schumacher TN. CD8(+) T cell states in human cancer: insights from single-cell analysis. Nat Rev Cancer. 2020;20:218–32.
9. Chambers CA, Kuhns MS, Egen JG, Allison JP. CTLA-4-mediated inhibition in regulation of T cell responses: mechanisms and manipulation in tumor immunotherapy. Annu Rev Immunol. 2001;19:565–94.
10. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. Science. 2018;359:1350–5.
11. Catacchio I, Silvestris N, Scarpi E, Schirosi L, Scattone A, Mangia A. Intratumoral, rather than stromal, CD8+ T cells could be a potential negative prognostic marker in invasive breast cancer patients. Transl Oncol. 2019;12:585–95.
12. Xie P, Ma Y, Yu S, An R, He J, Zhang H. Development of an Immune-Related Prognostic Signature in Breast Cancer. Front Genet. 2019;10:1390.
13. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12:996–1006.
14. Kao KJ, Chang KM, Hsu HC, Huang AT. Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. BMC Cancer. 2011;11:143.
15. Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B, et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin Cancer Res. 2007;13:3207–14.
16. Desmedt C, Di Leo A, de Azambuja E, Larsimont D, Haibe-Kains B, Selleslags J, et al. Multifactorial approach to predicting resistance to anthracyclines. J Clin Oncol. 2011;29:1578–86.
17. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell. 2006;10:529–41.
18. Sabatier R, Finetti P, Cervera N, Lambaudie E, Esterni B, Mamessier E, et al. A gene expression signature identifies two prognostic subgroups of basal breast cancer. Breast Cancer Res Treat. 2011;126:407–20.
19. Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A, et al. A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. JAMA. 2011;305:1873–81.
20. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. Clin Cancer Res. 2011;17:6012–20.
21. Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun. 2017;8:15081.
22. Hollern DP, Xu N, Thennavan A, Glodowski C, Garcia-Recio S, Mott KR, et al. B Cells and T Follicular Helper Cells Mediate Response to Checkpoint Inhibitors in High Mutation Burden Mouse Models of Breast Cancer. Cell. 2019;179:1191-206.e21.
23. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature. 2018;554:544–8.
24. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612.
25. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–50.
26. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
27. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12:453–7.
28. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. 2017;18:248–62.

29. Balar AV, Galsky MD, Rosenberg JE, Powles T, Petrylak DP, Bellmunt J, et al. Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. Lancet. 2017;389:67–76.

30. Ilangumaran S, Bobbala D, Ramanathan S. SOCS1: Regulator of T Cells in Autoimmunity and Cancer. Curr Top Microbiol Immunol. 2017;410:159–89.

31. Deng XX, Jiao YN, Hao HF, Xue D, Bai CC, Han SY. Taraxacum mongolicum extract inhibited malignant phenotype of triple-negative breast cancer cells in tumor-associated macrophages microenvironment through suppressing IL-10 / STAT3 / PD-L1 signaling pathways. J Ethnopharmacol. 2021;274:113978.

32. Koga T, Kawakami A. The role of CaMK4 in immune responses. Mod Rheumatol. 2018;28:211–4.

33. Hensel O, Burow P, Mages S, Wienke A, Kraya T, Zierz S. Increased Blood Flow Velocity in Middle Cerebral Artery and Headache Upon Ingestion of Ice Water. Front Neurol. 2019;10:677.

34. Ihanus E, Uotila LM, Toivanen A, Varis M, Gahmberg CG. Red-cell ICAM-4 is a ligand for the monocyte/macrophage integrin CD11c/CD18: characterization of the binding sites on ICAM-4. Blood. 2007;109:802–10.

35. Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. BMC Med Res Methodol. 2017;17:53.

36. Taube JM, Galon J, Sholl LM, Rodig SJ, Cottrell TR, Giraldo NA, et al. Implications of the tumor immune microenvironment for staging and therapeutics. Mod Pathol. 2018;31:214–34.

37. Hao D, Liu J, Chen M, Li J, Wang L, Li X, et al. Immunogenomic Analyses of Advanced Serous Ovarian Cancer Reveal Immune Score is a Strong Prognostic Factor and an Indicator of Chemosensitivity. Clin Cancer Res. 2018;24:3560–71.

38. Sui S, An X, Xu C, Li Z, Hua Y, Huang G, et al. An immune cell infiltration-based immune score model predicts prognosis and chemotherapy effects in breast cancer. Theranostics. 2020;10:11938–49.

39. Hu FF, Liu CJ, Liu LL, Zhang Q, Guo AY. Expression profile of immune checkpoint genes and their roles in predicting immunotherapy response. Brief Bioinform. 2020.

40. Petroni G, Buqué A, Zitvogel L, Kroemer G, Galluzzi L. Immunomodulation by targeted anticancer agents. Cancer Cell. 2021;39:310–45.

41. Gabrilovich DI, Ostrand-Rosenberg S, Bronte V. Coordinated regulation of myeloid cells by tumours. Nat Rev Immunol. 2012;12:253–68.

42. Funes SC, Rios M, Escobar-Vera J, Kalergis AM. Implications of macrophage polarization in autoimmunity. Immunology. 2018;154:186–95.

43. Schmid P, Rugo HS, Adams S, Schneeweiss A, Barrios CH, Iwata H, et al. Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Oncol. 2020;21:44–59.