Comprehensive Analysis of the Immune Microenvironment Reveals that CD52 is a Marker for Breast Cancer Treatment with Immune Checkpoint Inhibitors

Qiang Wang

First Affiliated Hospital of Harbin Medical University

Xuxu Liu

First Affiliated Hospital of Harbin Medical University

Pengfei Wang

First Affiliated Hospital of Harbin Medical University

Dankun Luo

First Affiliated Hospital of Harbin Medical University

Wenqi Gao

First Affiliated Hospital of Harbin Medical University

Yuanyuan Sun

First Affiliated Hospital of Harbin Medical University

Biao Ma (✉ mabiao@hrbmu.edu.cn)

The First Affiliated Hospital of Harbin Medical University, Harbin   https://orcid.org/0000-0002-7429-908X

Dongbo Xue

First Affiliated Hospital of Harbin Medical University   https://orcid.org/0000-0001-9012-4805

Hezheng Sun

Fourth Affiliated Hospital of Harbin Medical University

Primary research

Abstract

Background: Breast cancer (BC) is one of the most common tumors in women. Recent years, immune checkpoint inhibitors (ICIs) have brought good news to BC patients. Although significant achievements have been made through treatment with ICIs, some people who experience serious immune-related adverse events (IrAEs) are still insensitive to this approach. The response to ICI treatment depends on the type of tumor microenvironment (TME).

Methods:WGCNA (weighted gene co-expression network analysis), ESTIMATE algorithm, LASSO regression analysis, survival analysis, functional enrichment analysis are conducted to analyze the BC data in the TCGA database. Immunohistochemistry was used to verify the expression of CD52 in BC.

Results:WGCNA and **ESTIMATE algorithm** found that the CD52 is closely related to the immune microenvironment. CD52 highly expressed in various breast cancer subtypes, and patients with high expression of CD52 have longer survival time. Compared with the low-CD52 group, the high-CD52 group had more immune cell infiltration. TIMER database verification results showed that CD8+ T cells, activated memory CD4 T cells, memory B cells, γδ T cells, and Tregs were positively correlated with CD52 expression, while M2 macrophages were negatively correlated. CD52 can change the trend of TIC (CD8+ T) and tumor-associated macrophage (TAM) infiltration with respect to the survival time of breast cancer patients. Based on the expression of CD52, we explored the relationship between CD52 and the adaptive immune response (AIR). CD52 is a marker of AIR stratification in breast cancer patients. We constructed a CD52-related adaptive immune response gene signature (CD52rAIRGsig) which is an independent prognostic factor for breast cancer and related to genome instability and the immune cells infiltration in the TME. CD52 and CD52rAIRGsig were associated with PD-1 signaling and immune checkpoint inhibitor markers, which proves that patients with high CD52 expression and low risk of CD52rAIRGsig are more suitable for ICI treatment. We then screened chemotherapeutics for personalized medicine based on CD52rAIRGsig.

Conclusion: Therefore, we have discovered a new marker to guide the treatment and prognosis of breast cancer patients with ICIs. This provides a combined treatment strategy including different combinations of ICIs combined with chemotherapeutic drugs to treat breast cancer.

Background

Breast cancer is one of the most common tumors in women and seriously threatening women's health worldwide [1]. The abnormal expression of specific genes in breast cancer determines its occurrence, development and metastasis [2,3]. On the one hand, the tumor microenvironment fundamentally affects the gene expression of tumor tissues [4,5]. On the other hand, the latter can also affect the former [6]. The cellular milieu where a tumor is located is called the tumor microenvironment and consists of immune cells, mesenchymal cells, endothelial cells, inflammatory mediators and extracellular matrix (ECM) molecules [7,8]. Among these ingredients, immune cells and stromal cells are the two main types of

nontumor cells and contribute to the prognostic evaluation of tumor patients [7-9]. For example, Jia D et al. and Shah N et al. used an algorithm called ESTIMATE to predict the prognoses of patients with glioma and prostate cancer [9,10]. The tumor microenvironment has a dual effect on tumors. It can not only kill tumor cells but also may be used by tumor cells to induce immune escape and promote tumor growth [11,12].

In the tumor microenvironment (TME), tumor cells escape immune surveillance through mechanisms such as the expression of inhibitory receptors and their ligands, the expression of immunosuppressive cells and inhibitory cytokines, and cytotoxic T cell (TCL) and natural killer cell (NK cell) depletion [13,14]. These mechanisms enable tumor immune escape, leading to tumor occurrence and proliferation. Therefore, effectively preventing tumor immune escape is an important treatment method to prevent tumor occurrence and progression. In recent years, tumor immunotherapy has shown a high degree of efficacy against a variety of cancers, including refractory malignancies such as metastatic melanoma and advanced squamous non-small cell lung cancer (Nsclc) [15]. Among them, the development of immune checkpoint inhibitors (ICIs) is a revolutionary milestone in the field of immuno-oncology because the immune system is regulated by inhibitory receptors and ligands in the process of killing tumors. The goal of this therapy is not to activate the immune system to attack specific targets on tumor cells but to remove inhibitory pathways that prevent effective antitumor T cell responses. ICIs revive the antitumor immune response by blocking the immune checkpoint signaling pathway and promote immune-mediated tumor cell clearance. The classic immune checkpoint receptors mainly include CTLA-4 and PD-1. CTLA-4 is mainly expressed on CD4+ "helper" T cells, not on CD8+ "killer" T cells. It mainly reduces the magnitude of the T cell response by inhibiting CD28 co-stimulation. PD-1 is absent on resting naive and memory T cells and is expressed when the TCR is involved. PD-1 directly inhibits TCR-mediated effector functions and increases the migration of T cells in tissues [16]. Ipilimumab is an ICI targeting cytotoxic T lymphocyte antigen-4 (CTLA-4). It is used to treat patients with advanced melanoma [17]. The drug can effectively prevent the inactivation of T lymphocytes and promote the activation and proliferation of effector T cells. Nivolumab and pembrolizumab (anti-PD1) can target programmed cell death protein 1 (PD1) and are highly effective immune checkpoint inhibitors for the treatment of melanoma and non-small cell lung cancer [18,19]. Atezolizumab and avelumab can target programmed cell death 1 ligand (PD-L1) to treat various advanced malignant tumors. In addition to melanoma and non-small cell lung cancer, it can also be used for renal cell carcinoma, Hodgkin's lymphoma, and bladder cancer [20]. For cancer patients, this is certainly compelling. However, clinical trials have shown that ICIs are only useful for a small number of patients [21], most patients experience an insignificant or no effect at all, and the single application and combined application of ICIs will cause adverse consequences to patients. Therefore, exploring the predictive markers of immunotherapy is particularly critical for the personalized treatment of patients. This helps the direct treatment to the patients most likely to benefit and may provide the prognosis of the patient's treatment.

Method

Data source

The RNA expression profiles and the clinical data for 1,066 BC patients and 112 normal control patients were downloaded from the TCGA database (counts). We removed samples with incomplete clinical information. All expression profiles were annotated with the Ensemble reference database. The mRNA expression data of breast cancer (BRCA) in the validation data set (GSE42568) come from GEO (https://www.ncbi.nlm.nih.gov/geo/).

WGCNA and ESTIMATE algorithm

For WGCNA Network Construction and Module Identification, we used the WGCNA R package to construct the co-expression network. First, the samples were clustered to assess whether there were obvious outliers. Second, the automatic network construction function was used to construct a co-expression network. The R function pickSoftThreshold was used to calculate the soft thresholding power β, to which co-expression similarity is raised to calculate adjacency. Third, hierarchical clustering and the dynamic tree cut function were used to detect modules. Fourth, gene significance (GS) and module membership (MM) were calculated to relate modules to clinical traits. The corresponding module gene information was extracted for further analysis. The immune score and stromal score were obtained by the ESTIMATE algorithm.

Screening of differentially expressed RNAs

Differentially expressed mRNAs between high-score samples and low-score samples were detected with the "limma" package in R software. We defined RNAs with adjusted P values <0.05 and log|fold change| values >1 as differentially expressed RNAs. Heatmaps and volcano plots were visualized with the "ggplot2" and "pheatmap" packages of R software. The GEPIA database was used to analyze CD52, PLC and TPSD1 (http://gepia.cancer-pku.cn/detail.php).

Enrichment analysis

GSEA of CD52 was carried out with "GSEA4.1.0" (P<0.05 and FDR<25%). Matescape was used to perform GO enrichment and R-HSA pathway enrichment analysis on the differentially expressed genes of the two groups with high and low expression of CD52.

Overall survival curve

Kaplan-Meier plots were drawn to clarify the relationship between patient survival rates, and the log-rank test was used to test the overall survival rate of patients.

Infiltration of immune cells

The immune cell infiltration of each sample was calculated by the CIBERSORT method, and the corrplot package of R software was used to calculate the correlations between immune cells. Using the median expression of CD52 as the cutoff value, we classified the BC patients into a high-CD52 group and a low-CD52 group. The differences in immune cell infiltration between the high-CD52 group and the low-CD52 group were determined by the "limma" package of R software and visualized as a violin chart with the "vioplot" package. We also verified the results using the TIMER database (http://timer.cistrome.org/).

The distribution of CD52 in immune cells and the functional enrichment analysis of immune cells

The Tumor Immune Single Cell Hub (TISCH) database was used to select the BRCA_GSE114727 data set to analyze the distribution of CD52 in immune cells and perform a functional enrichment analysis of immune cells [15] (http://tisch.comp-genomics.org/home/).

Construction and Validation of the Prognostic Prediction Models

The univariate Cox model was used to screen potential CD52-related adaptive immune response genes with prognostic value (CD52rAIGs). A LASSO Cox regression model was further applied to narrow the range of prognostic AIGs. Second, a multivariate Cox regression model was used to screen out the CD52rAIGs that are most closely related to survival, and these genes were used to construct a risk model (CD52rAIGsig). Risk score = $\beta_1 X_1 + \beta_2 X_2 + \beta i X i$, where βi is the coefficient of each gene derived from the Cox regression and Xi is the expression level of each gene. The CD52rAIGsig risk score of the training set and the validation set was calculated according to the formula. The Kaplan-Meier method and ROC curve were used to analyze OS and evaluate the sensitivity and specificity of the signature model. Comprehensive age, sex, TNM classification and risk score were used to conduct univariate and multivariate Cox regression analyses to evaluate the independent prognostic effect of CD52rAIGsig.

Exploration of the significance of the model in clinical treatment

We download tumor RNA-seq (FPKM) from the Cancer Genome Atlas (TCGA) database. Convert PFKM data to TPM and normalize the data log2 (TPM+1), while keeping samples with clinical information recorded. We predicted the chemotherapeutic response for each sample based on the largest publicly available pharmacogenomics database [the Genomics of Drug Sensitivity in Cancer (GDSC), https://www.cancerrxgene.org/].The prediction process was conducted by R package "pRRophetic" where

the samples' half-maximal inhibitory concentration (IC50) was estimated by ridge regression and the prediction accuracy. All parameters were set by the default values with removal of the batch effect of combat and tissue type of allSoldTumours, and duplicate gene expression was summarized as mean value.AJCC recommends camptothecin, gemcitabine, pirubicin, paclitaxel, doxorubicin and other antitumor drugs (bleomycin, bortezomib, bryostatin.1, methotrexate, rapamycin, sunitinib, temsirolimus, tipifarnib, vorinostat）to treat BC.

Immunohistochemistry (IHC)

To evaluate the expression of CD52 in different molecular subtypes of breast cancer (luminal A, luminal B, triple-negative breast cancer (TNBC), Her-2 positive (HR negative), and Her-2 positive (HR positive)), we used IHC to detect CD52 (CD52Ab-DF8797, Affinity) molecular typing according to CSCO breast cancer guidelines. All specimens come from the Department of Pathology, the First Affiliated Hospital of Harbin Medical University.

Statistics

Genetic difference analysis used Wilcoxon test (Mann-Whitney); survival analysis adopted Log-rank test; all correlation was analyzed by Spearman correlation. The Wilcoxon test was used to compare the difference in IC50 between the high-risk group and the low-risk group. All data are expressed as the mean±standard deviation (x±s), *** p ⍰0.001, ** p ⍰0.01, * p ⍰0.05.

Result

WGCNA results

We clustered the samples and checked for any abnormal samples, and the results showed that there were no abnormal samples. We input the clinical traits (normal and tumor) to cluster again to obtain the sample cluster heat map (Figure 1A). To construct a WGCNA network, we first calculated the soft thresholding power β, to which the co-expression similarity is raised to calculate adjacency. We used the pickSoft Threshold function in WGCNA, which performs the analysis of network topology. The soft thresholding power β was set at 3 in the subsequent analysis because the scale independence reached 0.9 and had a relatively high average connectivity (Figure 1B). We constructed the gene network and identified modules associated with tumor using the one-step network construction function of the WGCNA R package. We constructed 10 gene co-expression modules (Figure 1C) and further analyzed the similarity between the modules (Figure 1D). Among them, MEblack and MEred are highly similar, so the two modules were merged. Correlation analysis showed that MEbrown has the strongest positive correlation with tumor modules(R=0.45, p<0.001). Therefore, we regard MEbrown as the key module related to tumors (Figure 1E). The MEbrown module includes 2,062 genes. Figure 1F showed that the

genes highly related to tumors are important elements in the modules that are significantly related to tumors.

Identification of key immune genes

We obtained the gene expression profiles and clinical information of all 1,066 BC patients from the TCGA database. Immune cells and stromal cells are important components of TME. Then, we calculated the immune scores (-2042.70-2076.56) and stromal scores (-1171.02-3632.68) using the ESTIMATE algorithm. We then explored the potential correlations between the overall survival rate and immune score and stromal score. Using the median immune score as the cutoff value, we classified the BC patients into a high-immune score group and a low-immune score group. The high-stromal score group and low-stromal score group were obtained in the same way. Our results showed that although the stromal score (P=0.655) (Figure 2A) had no significant relationship with the prognoses of BC patients, the high-immune score group had a higher survival rate (P=0.011) than the low-immune score group (P=0.011) (Figure 2B). This indicated that the immune response in TME of breast cancer is an important factor affecting the prognosis of patients. Hence, we focused on the differentially expressed genes between the high-immune score group and the low-immune score group. Difference analysis between high-immune score group and the low-immune score group showed that compared to low-immune score group, 730 genes were upregulated in high-immune score group(Figure 2C shows only the top 50 genes that were upregulated and downregulated in high immune score group).In order to screen out genes that are highly related to breast cancer and have important value in the immune microenvironment. We took the intersection of 2,062 genes in the MEbrown module and 730 immune-related genes. The results showed that CD52 and TPSD are related to the immune and tumor characteristics of breast cancer (Figure 2D). Difference analysis showed that, compared with normal breast tissue, CD52 was up-regulated in BC tissue, but TPSD did not show a significant difference (Figure 3A). In addition, we found through literature search that CD52 plays an important role in tumor immunotherapy (such as CAR T cell modification, lymphoma treatment, etc.). Therefore, we chose CD52 as the Key gene for subsequent analysis.

The expression of CD52 in the tumor microenvironment

There was no difference in the expression of TPSD in normal breast and breast cancer (Figure 3A). However, CD52 (P<0.001) was differentially expressed between tumor tissues and normal tissues (Figure 3B). Furthermore, we verified the expression of CD52 in paired samples, and the results showed that it was highly expressed in tumor tissues (Figure 3C). We further explored the expression of CD52 in each subtype and found that the expression of CD52 in each subtype of breast cancer was higher than that in normal tissues (Figure 3F). We further confirmed this result by immunohistochemical findings (CSCO breast cancer guideline molecular classification) (Figure 3G). Subsequently, using the median as the cutoff value, we classified the BC patients into a high-expression CD52 group and a low-expression CD52 group based on CD52 expression. As shown in Figure 3D, the overall survival rate of patients in the high-

expression groups of CD52 was significantly better than that of the low-expression groups, which indicates that CD52 is related to favorable prognosis for breast cancer patients. In addition, we performed GSEA on CD52 to explore the molecular functions of CD52. The results showed that CD52 is related to immune processes, cell apoptosis, cytokine and cytokine receptor interactions, the JAK-STAT signaling pathway, and the Toll-like receptor signaling pathway (Figure 3E).

Based on breast cancer single-cell transcriptome sequencing data, we found that CD52 was mainly highly expressed in immune cells but less expressed in stromal cells and milagnant cells (Supplementary 1A). Our study revealed that the high expression of CD52 in immune cells is related to worse TNM staging (Supplementary 1B). Further analysis revealed that CD52 was mainly expressed in CD4Tconv, CD8 T cells, CD8Tex, Tprolif and Treg cells (Supplementary 1C). Interestingly, in CD4Tconv, CD8 T cells, and CD8Tex, the expression of CD52 was related to clinicopathological staging, but there was no similar relationship in Tprolif and Treg cells (Supplementary 1D). We found that in the high pathological stage of breast cancer, CD52 is highly expressed in CD4Tconv, CD8T cells, and CD8Tex.

In breast cancer, the expression of CD52 is related to immune cell infiltration.

We quantified the composition of immune cells in the tumor microenvironment of each breast cancer sample and performed a correlation analysis between each immune cell type (Supplementary 2). Our results (Figure 4A) showed that compared with the low-CD52 group, the high-CD52 group had more naive B cells (P=0.001), memory B cells (P=0.001), CD8+ T cells (P≤0.001), resting memory CD4 T cells (P≤0.001), activated memory CD4 T cells (P≤0.001), follicular helper T cells (P≤0.001), Tregs (P≤0.001), γδ T cells (P≤0.001), resting NK cells (P≤0.001), M1 macrophages (P≤0.001), resting dendritic cells (P≤0.001) and activated mast cells (P≤0.001) but fewer plasma cells (P=0.015), M0 macrophages (P≤0.001), M2 macrophages (P≤0.001), resting mast cells (P=0.003) and neutrophils (P=0.003). Correlation analysis between CD52 expression and immune cell infiltration in the tumor microenvironment suggests that CD8+ T cells, activated memory CD4+ T cells, memory B cells, γδ T cells, Tregs and M2 macrophages have significant correlations with CD52 expression (Figure 4B). Among them, CD8+ T cells, activated memory CD4 T cells, memory B cells, γδ T cells, and Tregs were positively correlated with CD52 expression, while M2 macrophages were negatively correlated (Figure 4B). In view of the positive correlation between CD52 and a variety of T cells, we further analyzed the signaling pathways of T cells in tumor tissues (TICs). The results showed that the P53 signaling pathway was highly enriched in CD4 Tn, CD8T, CD8Tex, Tprolif and Treg cells (Figure 4C). In addition, we found that the IL6/JAK/STAT3 signaling pathway was slightly enriched in CD4 Tn, CD8T, CD8Tex, Tprolif and Treg cells (Figure 4C). The IL2/STAT5 signaling pathway and Wnt/β-catenin pathway showed almost no enrichment in any T cells (Figure 4C). Finally, we analyzed the upregulated and downregulated KEGG gene sets in CD4Tconv, CD8T, CD8Tex and Tprolif (Supplementary 3).

CD52 can change the trend of TIC (CD8+ T) and tumor-associated macrophage (TAM) infiltration with respect to the survival time of breast cancer patients.

The TIMER database and Kaplan-Meier Plotter analysis found that high expression of CD52 led to longer survival (Figure 5A and 5B), which is consistent with our previous results (Figure 3D). Analysis of the TIMER database showed that M2 macrophages and CD8+ T cells were related to survival prognosis (Figure 5C and 5E). The higher the level of M2 macrophage infiltration, the worse the prognosis, the higher the level of CD8+ T cell infiltration, and the longer the survival. We further analyzed the combined survival of CD52 cells with M2 macrophages and CD8+ T cells. Through further analysis, we found that the condition of high CD52 expression and low macrophage M2 infiltration correlated with the highest survival rate, while low CD52 expression and high macrophage M2 infiltration had the lowest survival rate (Figure 5D). Compared with curve 2 and curve 4, in the case of high macrophage M2 infiltration, high expression of the CD52 gene improved the survival rate (Figure 5D). Correspondingly, the condition of high CD52 expression and high CD8+ T cell infiltration yielded the highest survival rate, while low CD52 expression and low CD8+ T cell infiltration had the lowest survival rate (Figure 5F). In the case of low CD8+ T cell infiltration, high expression of the CD52 gene improved the survival rate (Figure 5F).

CD52 is a marker of AIR stratification in breast cancer patients.

We divided breast cancer patients into two groups with high and low expression of CD52 and then analyzed the differentially expressed genes between these two groups. Figure 6A shows the 10 most significantly upregulated genes and the 10 most downregulated genes (Figure 6A). We performed GO enrichment analysis of differentially expressed genes. We focused on the biological processes (BPs) involved in these differentially expressed genes. The results show that the biological processes enriched by differential genes are all related to the immune response (Figure 6B). Considering that CD52 is a protein anchored on the surface of T lymphocytes and is closely related to adaptive immunity, we focused on 155 genes involved in AIR. We explored the relationship between genes enriched in AIR and CD52. Surprisingly, all AIR genes were upregulated in the CD52 high expression group (Figure 6C). Spearman correlation typing shows that CD52 is positively correlated with AIR genes. This suggests that CD52 is a marker of AIR stratification in breast cancer patients.

Determination of the prognostic value of CD52rAIRGs in BC.

To identify the CD52-related adaptive immune response gene signature (CD52rAIRGsig) with prognostic value, 155 CD52rAIRGs were analyzed using CPHR to obtain 45 genes (Figure 6D). To prevent overfitting, the LASSO model was further used for processing (Figure 6E-F). Multivariate Cox regression analysis

identified 21 important CD52rAIRGs (TNFAIP8L2, CD70, SLAMF6, HLA. DPB1, PLA2G2D, LAPTM5, CD74, CD8A, RAC2, CCL19, HLA. DRA, CD2, IL12RB1, IRF1, CCR7, APBB1IP, JAK3, CCL5, IL7R, TIGIT, and TBX21, Table 1). We used these 21 genes to construct a prognostic gene signature. We calculated the risk value based on the recommended formula. Then, the patients in the training data set were divided into high-risk groups and low-risk groups according to the risk score cutoff (-0.218). The Kaplan-Meier survival curve showed that the survival time of the low-risk group was significantly longer than that of the high-risk group (p<0.001, Figure 7A). Figure 7C shows the risk scores of the patients in the training set, the distribution of patient survival status, and the gene expression profiles related to gene signatures. The area under the curve (AUC) value of the risk score was 0.764. The area under the curve (AUC) value of the pathological stage was 0.707 (T: AUC=0.699, M: AUC=0.601, N: AUC=0.628). The area under the curve (AUC) value of sex was 0.530. The area under the curve (AUC) value of age was 0.778 (Figure 7E). Univariate Cox regression analysis (Supplementary 4A) showed that the risk score was significantly correlated with OS (HR=3.225, 95% CI=2.434-4.273, P<0.001). Multivariate Cox regression analysis was performed on the basis of age, sex, pathological stage (T stage, N stage, and M stage), and risk score. The results showed that only the genetic signature risk score was an independent prognostic factor for breast cancer (Supplementary 4C). In the training data set, the 1-, 3-, and 5-year OS-predicted AUCs based on CD52rAIRGsig were 0.757, 0.670, and 0.713, respectively (Supplementary 4A). We used the GSE42568 data set to verify the performance of 21 gene signatures. We used the same method to calculate the risk score of each patient in the validation set and then used the risk score cutoff (-0.218) obtained in the training set to divide the patients into high- and low-risk groups. The results of the study show that in the validation set, compared with the high-risk group, low-risk patients had longer survival time (Figure 7B). Figure 7D shows the risk scores of patients in the verification set, the distribution of patient survival status, and gene expression profiles related to gene signatures. The area under the curve (AUC) values of the risk score and pathological grade were 0.640 and 0.758, respectively (Figure 7F). Univariate Cox regression analysis (Supplementary 4B) showed that the risk score was significantly correlated with OS (HR=2.018, 95% CI=1.025-3.947, P=0.042). Multivariate Cox regression analysis showed that the gene signature risk score was an independent prognostic factor for breast cancer (Supplementary 4D). In the verification data set, the 1-, 3-, and 5-year OS predicted AUCs based on CD52rAIRGsig were 0.640, 0.571, and 0.622, respectively (Supplementary 4F). In summary, CD52rAIRGsig exerts a positive effect in predicting the prognosis of breast cancer patients.

Table 1

multivariate COX regression analysis screened out 21 important CD52rAIRGs

Gene	Coef	HR	HR.95L	HR.95H	pvalue
TNFAIP8L2	0.079	1.080	0.999	1.168	0.023
CD70	-0.115	0.944	0.965	1.002	0.033
SLAMF6	0.053	1.002	1.001	1.004	0.004
HLA.DPB1	-0.002	0.998	0.996	0.999	0.002
PLA2G2D	-0.116	0.879	0.797	0.969	0.010
LAPTM5	0.002	1.001	0.967	1.014	0.024
CD74	-0.001	0.990	0.995	0.997	0.034
CD8A	0.026	1.058	1.031	1.025	0.000
RAC2	-0.033	0.820	0.643	1.046	0.011
CCL19	-0.002	1.002	1.001	1.004	0.002
HLA.DRA	0.001	1.018	1.011	1.025	0.000
CD2	0.019	1.080	0.999	1.168	0.032
IL12RB1	0.109	1.328	1.057	1.669	0.012
IRF1	-0.027	0.981	0.964	0.998	0.031
CCR7	0.019	1.070	0.979	1.148	0.012
APBB1IP	-0.039	0.879	0.786	0.968	0.003
JAK3	0.146	1.403	1.049	1.380	0.008
CCL5	-0.010	0.920	0.837	1.012	0.037
IL7R	-0.049	0.747	0.610	0.850	0.036
TIGIT	0.016	1.020	0.979	1.068	0.043
TBX21	0.311	1.428	1.157	1.669	0.015

Coef: Coefficient; HR:Hazard Ratio

CD52rAIRGsig is related to the clinical characteristics of breast cancer patients.

CD52rAIRGsig had a significant correlation with the clinical characteristics of breast cancer patients (Figure 8A). The results of the study show that the risk score is significantly higher in elderly breast cancer patients. The risk score is also significantly higher in patients with distant metastases. The genetic signatures of patients with lymph node metastasis are also applicable, and the risk scores of

patients in stages N2-3 are significantly higher than those of patients in stages N0-1. In the early stages (stages I-II) and late stages (stages III-IV) of breast cancer, the risk scores were also significantly different, and the risk scores in the late stages were significantly higher than those in the early stages.

CD52rAIRGsig is related to genome instability

It is well known that the occurrence of breast cancer is genetically determined. In recent years, breast cancer susceptibility gene testing has been widely used. Studies have confirmed that the two key factors associated with a high risk of breast cancer are mutations in BRCA1 and BRCA2. In addition, VCAN, BRIP1, CDH1, PIK3CA, LSP1, and DIRAS3 are also included. Importantly, we found that the expression of these breast cancer susceptibility genes was significantly increased in the high-risk group (Figure 8B).

CD52rAIRGsig is related to the infiltration of immune cells in the TME.

We also used the CIBERSORT method to quantify the immune cells in the tumor microenvironments of the high- and low-risk groups. The results showed that memory B lymphocytes, M1 macrophages, NK cells, CD8+ T cells, Tregs, and follicular helper T cells significantly increased infiltration in the low-risk group. This result is consistent with the CD52-High group; in fact, CD52 is highly expressed in the low-risk group. However, tumor-associated macrophages (M2 macrophages) were highly infiltrated in the high-risk group, consistent with the CD52-Low group (Figure 8C).

CD52 and CD52rAIRGsig are related to PD-1 signaling and ICI markers.

We performed R-HSA pathway enrichment analysis on the differentially expressed genes of the two groups with high and low expression of CD52 (Figure 9A). These differentially expressed genes are not only related to the adaptive immune system but are also closely related to PD-1 signaling. Figure 9B shows the PD-1 signaling pathway and the genes enriched in this pathway. In addition, ZAP-70 translocates to immunological synapses, phosphorylates CD3, and TCR zeta chains interact with PD-1 signaling. These three pathways regulate each other through the central genes of the network diagram. Interestingly, we found that the genes enriched in the PD-1 signaling pathway were all highly expressed in the CD52-High group (Figure 9C), including PD-1 (PDCD1) and PD-L1 (CD274). We speculate that CD52 has a potential connection with the PD-1 signaling pathway. We performed Spearman correlation analysis on CD52 and PD-1 signal-related genes, and the results showed that CD52 has a significant positive correlation with these genes (Figure 9D). This further confirmed our hypothesis. We further analyzed the relationship between CD52rAIRGsig and these genes, and the results showed that the expression of these genes in the low-risk group was significantly higher than that in the high-risk group

(Supplementary 5A). High expression of PD-1 signal-related genes will reduce the risk to patients. We performed OS analysis on these genes, and the results showed that high expression of these genes in patients prolonged survival time (Supplementary 6A). To determine the relationship between CD52rAIRGsig and ICI markers, in addition to PD-1 and PD-L1, we also studied other ICI markers (CTLA-4, LAG-3, HAVCR2, and TIGIT). PD-1, PD-L1, and CTLA-4 were low in the high-risk group and high in the low-risk group (Figure 10A). These results suggest that patients in the low-risk group are suitable for ICI treatment, while the high-risk group is not suitable for this therapy.

Screening of chemotherapeutics for personalized treatment based on CD52rAIRGsig.

In addition to ICI therapy, we tried to screen common chemotherapy drugs with obvious curative effects based on CD52rAIRGsig by GDSC database. We estimated the IC50 of each sample, and observed that the IC50 difference of 14 chemotherapy drugs between the high-risk group and the low-risk group was statistically significant. The antitumor drugs recommended by the AJCC, camptothecin, gemcitabine, pirubicin, and paclitaxel, are more effective in the low-risk group because the low-risk group has lower IC50 values than the high-risk group. However, doxorubicin (p=0.83) did not differ significantly between the two groups (Figure 11A). In addition, our research showed that the IC50 values of atezolizumab, bleomycin, bortezomib, bromostatin.1, methotrexate, rapamycin, sunitinib, temsirolimus, tipifarnib, and vorinostat chemotherapy drugs in the low-risk group were significantly lower than in the high-risk group (Figure 11B). This result suggests that the above chemotherapy drugs are suitable for low-risk patients yet less effective for high-risk patients.

Discussion

The immune response to ICIs is a complex process. There is increasing evidence that only a small percentage of cancer patients benefit from checkpoint inhibitors, and some patients receiving ICI treatment have experienced serious immune-related adverse events (IrAEs) [22]. IrAEs occur due to the suppression of immune checkpoints. These checkpoints cause physical barriers to resist normal immunity, leading to various local and systemic autoimmune reactions [22]. Therefore, the development of predictive biomarkers is essential to distinguish between responders and non-responders in order to avoid any adverse reactions. Predictive biomarkers can determine the outcome of a patient's treatment before treatment begins. These biomarkers should indicate whether the patient can benefit from a single checkpoint or whether a combination therapy is needed.

In previous studies, the biomarkers for ICI treatment were immunosuppressive receptors or their corresponding ligands, including classic PD-1 and CTLA-4. They can be expressed on immune cells such as T lymphocytes, B lymphocytes, NK cells, and monocytes [23]. These immunosuppressive receptors are upregulated when immune cells continue to activate, usually as a form of negative regulation of the immune response to prevent excessive damage to surrounding tissues. Once the inhibitory receptor is

activated, immune cells are suppressed and inactivated. This signaling mechanism is used by tumor cells to achieve tumor immune escape. Immune checkpoint inhibitors (ICIs) can block the interactions between inhibitory receptors and their ligands, thereby overcoming the inherent resistance of immune surveillance by initiating an antitumor immune response and stimulating T cells, B cells and other immune cell responses. The activity of immune cells is regulated by the complex signaling system produced by stimulating and inhibiting receptors [24]. The expression of inhibitory receptors is not only spatial but also temporal. The expression of inhibitory receptors increases with the enhancement of the TME immune response. As the inhibitory signal increases, the stimulus signal gradually weakens and tends to balance the immune system. If tumor cells continue to release inhibitory signals and cross the equilibrium point until the immune cells in the TME are completely inactivated, this is detrimental to ICI therapy. Such patients exhibit poor response to ICI treatment. If immune cells and the TME are in an immune response state before reaching the immune balance point, good results may be achieved with ICI therapy. Studies have confirmed that the response to ICI treatment is related to tumor infiltrating lymphocytes (TILs) and other immune cells in the TME [25]. The RNA expression study of patients receiving ipilimumab showed that compared with non-responding patients, patients benefiting from greater clinical efficacy have more genes involved in innate immunity and adaptive immunity. This indicates that the preimmune TME is important for ipilimumab treatment [26]. However, we lack markers to predict the immunogenicity of the TME.

WGCNA results in modules and genes that are significantly positively related to tumors (Figure 1). The ESTIMATE algorithm obtains the immune score of each sample, and the immune score is related to patient survival (Figure 2B). Immune-related genes were identified through high and low immune scores (Figure 2C). Finally, the immune gene CD52, which is significantly positively related to the tumor, was obtained by the intersection method (Figure 2D). Compared with normal breast tissues, CD52 was highly expressed in breast cancer tissues (Figure 3B), and consistent results were obtained from the paired cancer and paired adjacent breast tissues (Figure 3C). Breast cancer can be divided into different molecular subtypes: luminal A, luminal B, triple-negative breast cancer (TNBC), Her-2 positive (HR negative) and Her-2 positive (HR positive) [27]. We further explored the expression of CD52 in each subtype and found that the expression of CD52 in each subtype of breast cancer was higher than that in normal tissues (Figure 3E). We further confirmed this result by immunohistochemical findings [Figure 3F]. CD52 is a human leukocyte differentiation antigen of 12 amino acids anchored to glycosylphosphatidylinositol (GPI) [28]. To date, the regulatory mechanisms of CD52 expression have rarely been reported. In Flt3-ITD myeloid leukemia, Flt3-ITD enhances the expression of CD52 by phosphorylating STAT5 [29]. In breast cancer, GSEA showed that CD52 is related to immune/inflammatory pathways (Figure 3F), such as antigen presentation, the B cell receptor pathway, the T cell receptor pathway, NK cell-mediated cytotoxicity, the JAK-STAT pathway, and the TOLL-like receptor pathway [30].

CD52 is a very interesting molecule, but its molecular function is still unclear. Studies have shown that in T cells, CD52 is associated with the TCR to activate the intracellular signal of T cell activation and promote cell proliferation and differentiation [31]. CD52 is a stimulating antigen in the TME and

participates in the antitumor immune response of immune cells. In fact, in response to CD52-mediated activation signals, an anti-CD52 monoclonal antibody (alemtuzumab) is clinically used to treat chronic lymphocytic leukemia, relapsing-remitting multiple sclerosis, hormone-refractory acute graft-versus-host disease (GVHD) and improve organ transplantation [32-35]. The CD52 antibody Campath-1H can kill T cells in vivo through complement- and non-complement-mediated mechanisms [36,37]. Interestingly, CD52 can be shed from the cell membrane surface to become a soluble antigen under the action of phospholipase C (PLC). Soluble CD52 molecules interact with inhibitory sialic acid-bound immunoglobulin-like lectin 10 (Siglec10), thereby significantly inhibiting the proliferation and activation of T cells [38]. Soluble CD52/Siglec10 is an inhibitory signal in immune cells. CD52 on the membrane surface can provide co-stimulatory signals for the activation and proliferation of T cells, while soluble CD52 molecules can significantly inhibit the proliferation and activation of T cells. This is obviously a contradiction, and the specific function of CD52 in breast cancer is worthy of discussion. The release of soluble CD52 requires PLC. Although CD52 is highly expressed in breast cancer, PLC is significantly expressed at low levels (Supplementary 1E). We speculate that due to the lack of PLC in breast cancer tissues, CD52 may participate in immune regulation in the form of membrane surface molecules and manifest an antitumor effect. As shown in Figure 3D, the overall survival rate of patients in the high-expression groups of CD52 was significantly better than that of the low-expression groups, which indicates that CD52 is related to favorable prognosis in breast cancer patients.

CD52 is present during the final differentiation stage of all lymphocytes, as well as in eosinophils, monocytes/macrophages and DCs [39-42]. Based on breast cancer single-cell transcriptome sequencing data, CD52 is mainly expressed in immune cells and less expressed in stromal cells and malignant cells in the breast cancer TME (Supplementary 1A). Further analysis revealed that CD52 was mainly expressed in CD4Tconv, CD8 T cells, CD8Tex, Tprolif and Treg cells (Supplementary 1C), which is consistent with previous studies. To explore the relationship between CD52 and tumor-infiltrating immune cells (TIICs), we quantified the composition of immune cells in the tumor microenvironment of each breast cancer sample. Our results showed that compared with the low-CD52 group, the high-CD52 group had more TIICs (Figure 4A). Correlation analysis between CD52 expression and immune cell infiltration in the tumor microenvironment suggests that CD8+ T cells, activated memory CD4+ T cells, memory B cells, γδ T cells, Tregs and M2 macrophages have significant correlations with CD52 expression (Figure 4B). Among them, CD8+ T cells, activated memory CD4 T cells, memory B cells, γδ T cells, and Tregs were positively correlated with CD52 expression, while M2 macrophages were negatively correlated (Figure 4B). The high infiltration of TILs in various malignant tumors is an effective predictive biomarker of response to ICI therapy [43,44]. According to the infiltration of immune cells, the TME can be classified into three categories: immune-desert, immune-excluded and immune-inflamed [45]. High TIL infiltration is a biomarker of the immune-inflamed phenotype, which exhibits immune-mediated clearance of tumor cells. Obviously, breast cancer patients with high expression of CD52 have higher immunogenicity. In addition, we found that the higher the TNM stage was, the higher the expression of CD52 in immune cells (Supplementary 1B-D). According to previous results, CD52 and TIICs have a positive correlation,

suggesting that there will be more immune cell infiltration in high pathological stages. This shows that patients with advanced breast cancer are more immunogenic and more suitable for ICI treatment.

In addition, immune cells in the TME can also be used as predictors to predict the prognoses of tumor patients. In BC, an increase in cytotoxic T cells, γδT cells, Tfh cells, activated NK cells, memory B cells, memory CD4+ T cells and M1 macrophages implies a better prognosis, while an increase in M0 macrophages, M2 macrophages and neutrophils indicates a poor prognosis [46-53]. The positive and negative effects of high and low CD52 expression on the prognosis of breast cancer patients are consistent with the potential effects of different immune cell types. This further confirms the correlation between the expression of CD52 and the survival of BC patients. We further analyzed the combined survival with respect to CD52 in the context of M2 macrophages and CD8+ T cells. As expected, we found that patients with high CD52 expression and low macrophage M2 infiltration had the highest survival rate, while patients with low CD52 expression and high macrophage M2 infiltration had the lowest survival rate (Figure 5D). Compared with curve 2 and curve 4, in the case of high macrophage M2 infiltration, high expression of the CD52 gene improved the survival rate (Figure 5D). Correspondingly, the condition of high CD52 expression and high CD8+ T cell infiltration was associated with the highest survival rate, while low CD52 expression and low CD8+ T cell infiltration was related with the lowest survival rate (Figure 5F). In the case of low CD8+ T cell infiltration, high expression of the CD52 gene improved the survival rate (Figure 5F). This further proves that CD52 can regulate the TME to improve the prognosis of patients. In summary, CD52 can not only assess the TME but also predict the patient's prognosis in BC.

To explore the reasons for the high immunoreactivity of the CD52-High group, we conducted transcriptome difference analysis of the CD52-High and -Low groups (Figure 6A). GO enrichment analysis showed that the differentially expressed genes were all related to the immune response (Figure 6B). Considering that CD52 is anchored on the surface of lymphocytes and is closely related to adaptive immunity, we focused on 155 genes involved in the adaptive immune response (AIR). Interestingly, AIR genes (AIRGs) were all upregulated in the CD52-High group (Figure 6C). Spearman correlation analysis suggested that CD52 is positively correlated with AIRGs. Our results show that BC patients can be distinguished into two adaptive immune response states based on CD52 expression, which suggests that CD52 can be used as a biomarker for stratification of the adaptive immune response in BC (Figure 6C). All genes involved in the adaptive immune response in the CD52-High group were upregulated, including immunostimulatory genes and immunosuppressive genes. We then explored the prognostic value of CD52rAIRGs. We identified CD52rAIRGsig as having prognostic value through CPHR, the LASSO model and multivariate Cox regression analysis (Figure 6D-F, Table 1). The risk score based on CD52rAIRGsig was an independent predictor of breast cancer prognosis (Supplementary 4) and performed well in both the training data set and the verification data set [Figure 7]. CD52rAIRGsig is associated with the clinical characteristics of BC patients [Figure 8A], which further illustrates the accuracy of CD52rAIRGsig for predicting prognosis. Moreover, we found that CD52rAIRGsig is associated with immune cell infiltration. The results showed that memory B lymphocytes, M1 macrophages, NK cells, CD8+ T cells, Tregs, and follicular helper T cells significantly increased infiltration in the low-risk group, which is consistent with

the CD52-High group [Figure 8C]. However, tumor-associated macrophages (M2 macrophages) were highly infiltrated in the high-risk group, and other cells showed low infiltration, which is consistent with the CD52-Low group. Therefore, we defined the TME of patients with high CD52 expression and low CD52rAIRGsig risk score as the immune-inflamed type. At this time, the TME may not have reached immune balance. Patients with low CD52 expression and CD52rAIRGsig risk score are defined as immune-desert type, in which a small amount of immune cells infiltrate.

In addition, we found that genes associated with the PD-1 signaling pathway, including PD-1 (PDCD1) and PD-L1 (CD274) were all highly expressed in the CD52-High group (Figure 9C). We speculate that CD52 has a potential connection with the PD-1 signaling pathway. We performed Spearman correlation analysis on CD52 and PD-1 signaling-related genes, and the results showed that CD52 has a significant positive correlation with these genes (Figure 9D). To verify the relationship between PD-1 signal-related genes and CD52rAIRGsig, we analyzed the expression of these genes in the high- and low-risk groups. The results showed that these genes were significantly highly expressed in the low-risk group (Supplementary 5). To verify that the PD-1 signal-related genes are related to the low risk score of CD52rAIRGsig, we performed OS analysis on these genes, and the results showed that the high expression of these genes is associated with better prognosis in BC patients (Supplementary 5). Immune checkpoints and their ligands are frequently upregulated in a variety of malignant tumor TMEs. Key negative regulators of T cell activation include CTLA-4, PD-1, TIM-3, indoleamine 2,3-dioxygenase (IDO), V-domain Ig suppressor of T cell activation (VISTA), killer-cell immunoglobulin-like receptor (KIR), TIGIT, B and T lymphocyte attenuator (BTLA) and LAG-3 [54,55]. Several immune checkpoint inhibitors (ICIs) targeting immune checkpoints and their ligands have been developed and have shown exciting effects in some cancer patients. To date, blocking the interaction between cytotoxic T lymphocyte antigen 4 (CTLA-4) and CD80/86 and programmed death 1 (PD-1) and PD-L1/2 is the most effective strategy [56]. However, studies have shown that the therapeutic effect and clinical response of ICIs depend on the characteristics of the tumor. Recent studies have shown that overexpression of PD-L1 may not be an effective biomarker for predicting sensitivity to ICIs because tumors with low PD-L1 expression can also be sensitive to PD-1/PD-L1 [26]. Anti-CTLA-4 therapy is only applicable to some cancer patients. Therefore, there is an urgent need to identify predictive biomarkers for screening patients to receive ICI treatment. Our results showed that not only PD-1, PD-L1, and CTLA-4 but also LAG-3, HAVCR2, and TIGIT were highly expressed in the CD52-High group and CD52rAIRGsig low-risk group (Figure 10). This suggests that CD52 and CD52rAIRGsig may be good predictive biomarkers for the treatment of ICIs. BC patients with both high CD52 expression and low risk of CD52rAIRGsig have significantly increased expression of multiple immune checkpoints, which provides a variety of combinations to assess the sensitivity for ICI treatment. In fact, studies have shown that the combination of multiple ICIs can overcome the limitations of single ICI treatment [57]. In addition to ICI therapy, we attempted to screen common chemotherapy drugs with obvious curative effects based on CD52rAIRGsig. The antitumor drugs recommended by the AJCC, camptothecin, gemcitabine, pirubicin, and paclitaxel, are more effective in the low-risk group because the low-risk group has a lower IC50 than the high-risk group. However, doxorubicin (p=0.83) did not differ significantly between the two groups. In addition, our research showed

that the IC50 values of atezolizumab, bleomycin, bortezomib, bromostatin.1, methotrexate, rapamycin, sunitinib, temsirolimus, tipifarnib, and vorinostat chemotherapy drugs in the low-risk group were significantly lower than in the high-risk group (Figure 11). This result suggests that the above chemotherapy drugs are suitable for low-risk patients, but less effective for high-risk patients. Therefore, for patients with high CD52 expression and low risk of CD52rAIRGsig, we can not only choose different ICI combination therapies but may also combine the above conventional chemotherapy drugs.

With the deepening of research, we have an enhanced understanding of how the TME affects tumor development. Breast matrix accounts for 80% of the normal breast, including fibroblasts, endothelial cells, smooth muscle cells, immune cells, nerve cells, adipocytes, extracellular matrix (ECM) and other macromolecules that support epithelial cells [58]. Studies have reported genomic instability in the tumor-associated stroma, further confirming that the TME is a dynamic entity that promotes the occurrence of tumors through morphological and molecular changes[59-61]. Our results suggested that CD52 expression also differs with respect to high and low stromal scores (Figure2A), for which reason we explored the association between CD52rAIRGsig and genetic instability. Importantly, we found that these breast cancer susceptibility genes (BRCA1, VCAN, BRIP1, CDH1, PIK3CA, LSP1, DIRAS30) were significantly more highly expressed in the high-risk group (Figure 8B). This indicates that the high-risk group has a high genetic predisposition and risk of recurrence. Through CD52rAIRGsig, we can identify high-risk patients and provide personalized and precise treatment.

Conclusion

There is a dynamic interaction between immune cells and tumor cells, and the response to ICIs depends on the immune components of the TME. Our study found that CD52, which is highly expressed in luminal A, luminal B, HER-2 enriched, and triple-negative breast cancer, is a key component of the TME. The high expression of CD52 in the TME is not only associated with better prognosis but can also be used to evaluate TIICs in BC. The expression of CD52 is positively correlated with the infiltration of TICs, NK cells, memory B cells, and M1 macrophages in the TME of breast cancer patients but negatively associated with M2 macrophages. More importantly, CD52 can influence the trends of T cells in tumor tissues (TICs, CD8+ T cells) and tumor-associated macrophages (TAMs) in predicting the prognosis of BC patients. It is worth noting that CD52 is a biomarker of adaptive immune response stratification in BC patients. According to the gene expression of CD52, breast cancer patients can be divided into two adaptive immune response states. Considering the inaccuracy of a single biomarker for patient prognosis prediction and treatment guidance, we therefore identified the prognostic value of CD52-related AIR genes (CD52rAIRGs) in BC, and the results indicate that CD52rAIRGsig is an independent biomarker for predicting prognosis. CD52rAIRGsig is related to the clinical characteristics of BC patients. Further results show that CD52rAIRGsig is related to the infiltration of immune cells in the TME and the genetic susceptibility of breast cancer, which helps to screen patients with a high risk of recurrence. Our results showed that not only PD-1, PD-L1, and CTLA-4 but also LAG-3, HAVCR2, and TIGIT were highly expressed in the CD52-High group and CD52rAIRGsig low-risk group, which suggests that CD52 and CD52rAIRGsig may be effective biomarkers for sensitivity to ICI treatment. In addition, based on CD52rAIRGsig,

commonly used chemotherapeutics with obvious curative effects were screened. The combination of ICIs and chemotherapy is more beneficial to BC patients. Of course, there are deficiencies in our research, such as the conclusions we obtained through data analysis, which need to be verified and confirmed by a large number of experiments. We also gave some specific solutions: for clinical trials, we selected two cohorts with high CD52 expression and low CD52 expression, and obtained lesion samples in a reasonable way. Then we detected the infiltration of various immune cells in the two groups to evaluate the immunogenicity of TME. We can compare the expression of immune checkpoint receptors and ligands between the two groups, and compare the prognosis of the two groups of patients. We can give two groups of immune checkpoint inhibitors to compare the efficacy of the two groups.

Abbreviations

BC Breast cancer

TCGA The Cancer Genome Atlas database

OS overall survival

CPHR cox proportional hazards regression

PPI protein-protein interaction network

TISCH Tumor Immune Single Cell Hub

DEGs Differentially expressed genes

Tregs regulatory T cells

ESTIMATEEstimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data

ICIs immune checkpoint inhibitors

IrAE immune-related adverse events

TME tumor microenvironment

AIR adaptive immune response

CD52rAIRGsig CD52-related adaptive immune response gene signature

ECM extracellular matrix

TCL cytotoxic T cells

NK cellsnatural killer cells

Nsclc non-small cell lung cancer

CTLA-4 cytotoxic T lymphocyte antigen-4

PD1 programmed cell death protein 1

PD-L1 programmed cell death 1 ligand

GO Gene Ontology

KEGG Kyoto Encyclopedia of Gene and Genomes

CD52rAIRGs CD52-related adaptive immune response genes

IHC Immunohistochemistry

TICs T cells in tumor tissues

BP biological processes

AUC area under the curve

TIL tumor infiltrating lymphocytes

GPI glycosylphosphatidylinositol

PLC phospholipase C

Siglec10 sialic acid-bound immunoglobulin-like lectin 10

AIRGs AIR genes

TIICs tumor-infiltrating immune cells

TICs T cells in tumor tissues

TAM tumor-associated macrophages

Declarations

Ethics approval and consent to participate:

Breast cancer tissues were taken from patients undergoing surgical resection at the First Affiliated Hospital of Harbin Medical University(Harbin, China). The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the First Affiliated Hospital of Harbin Medical University. The study did not involve animals.

Informed Consent Statement:

Written informed consent has been obtained from the patients to publish this paper.

Author contributions:

Biao Ma and Dongbo Xue conceived the study; Qiang Wang designed and analyzed articles;Xuxu Liu wrote articles; Pengfei Wang and Dankun Luo edited the picture; Wenqi Gao and Yuanyuan Sun collected the data.

Conflict of Interest:

The authors declare no conflict of interest.

Data Availability Statement:

The data sets used and/or analyzed during the current study are available from the TCGA database (https://tcga-data.nci.nih.gov/, accessed on 11 January 2021) and GEO database (https:// www. ncbi. nlm. nih. gov/ geo/).

Author information

[1]Department of General Surgery, Laboratory of Hepatosplenic Surgery, Ministry of Education, The First Affiliated Hospital of Harbin Medical University, Harbin,

Heilongjiang Province, China.

[2]Department of General Surgery, The Fourth Affiliated Hospital of Harbin Medical University, Harbin, Helongjiang, China

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN

2012. *Int J Cancer* 2015, **136**(5):E359-386. doi:10.1002/ijc.29210

2. Ross JB, Huh D, Noble LB, Tavazoie SF. Identification of molecular determinants of primary and metastatic tumour re-initiation in breast cancer. *Nat Cell Biol* 2015, **17**(5):651-664. doi:10.1038/ncb3148

3. Zhang S, Fei F, Wang H, Gu Y, Li C, Wang X, Zhao Y, Li Y. Stromal immunoglobulin κC expression is associated with initiation of breast cancer in TA2 mice and human breast cancer. *Cancer Sci* 2018, **109**(6):1825-1833. doi:10.1111/cas.13620

4. Winslow S, Lindquist KE, Edsjö A, Larsson C. The expression pattern of matrix-producing tumor stroma is of prognostic importance in breast cancer. *BMC Cancer* 2016, **16**(1):841. doi:10.1186/s12885-016-2864-2

5. Galon J, Pagès F, Marincola FM, Thurin M, Trinchieri G, Fox BA, Gajewski TF, Ascierto PA. The immune score as a new possible approach for the classification of cancer. *J Transl Med* 2012, **10**:1. doi:10.1186/1479-5876-10-1

6. Zhao J, Li H, Min L, Han X, Shu P, Yang Y, Gan Q, Wang X, Wang H, Ruan Y, et al. High expression of tumor necrosis factor receptor-associated factor 2 promotes tumor metastasis and is associated with unfavorable prognosis in gastric cancer. *J Gastroenterol Hepatol* 2018, **33**(2):431-442. doi:10.1111/jgh.13818

7. Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* 2012, **21**(3):309-322. doi:10.1016/j.ccr.2012.02.022

8. Fane M, Weeraratna AT. How the ageing microenvironment influences tumour progression. *Nat Rev Cancer* 2020, **20**(2):89-106. doi:10.1038/s41568-019-0222-9

9. Jia D, Li S, Li D, Xue H, Yang D, Liu Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. *Aging (Albany NY)* 2018, **10**(4):592-605. doi:10.18632/aging.101415

10. Shah N, Wang P, Wongvipat J, Karthaus WR, Abida W, Armenia J, Rockowitz S, Drier Y, Bernstein BE, Long HW, et al. Regulation of the glucocorticoid receptor via a BET-dependent enhancer drives antiandrogen resistance in prostate cancer. *Elife* 2017, **6**. doi:10.7554/eLife.27861

11. Bates JP, Derakhshandeh R, Jones L, Webb TJ. Mechanisms of immune evasion in breast cancer. *BMC Cancer* 2018, **18**(1):556. doi:10.1186/s12885-018-4441-3

12. Steven A, Seliger B. The Role of Immune Escape and Immune Cell Infiltration in Breast Cancer. *Breast Care (Basel)* 2018, **13**(1):16-21. doi:10.1159/000486585

13. Thommen DS, Schumacher TN. T Cell Dysfunction in Cancer. *Cancer Cell* 2018, **33**(4):547-562. doi:10.1016/j.ccell.2018.03.012

14. Speiser DE, Ho PC, Verdeil G. Regulatory circuits of T cell function in cancer. *Nat Rev Immunol* 2016, **16**(10):599-611. doi:10.1038/nri.2016.80

15. Hoos A. Development of immuno-oncology drugs - from CTLA4 to PD1 to the next generations. *Nat Rev Drug Discov* 2016, **15**(4):235-247. doi:10.1038/nrd.2015.35

16. Honda T, Egen JG, Lämmermann T, Kastenmüller W, Torabi-Parizi P, Germain RN. Tuning of antigen sensitivity by T cell receptor-dependent negative feedback controls T cell effector function in inflamed tissues. *Immunity* 2014, **40**(2):235-247. doi:10.1016/j.immuni.2013.11.017

17. Robert C, Thomas L, Bondarenko I, O'Day S, Weber J, Garbe C, Lebbe C, Baurain JF, Testori A, Grob JJ, et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N Engl J Med* 2011, **364**(26):2517-2526. doi:10.1056/NEJMoa1104621

18. Paz-Ares L, Luft A, Vicente D, Tafreshi A, Gümüş M, Mazières J, Hermes B, Çay Şenler F, Csőszi T, Fülöp A, et al. Pembrolizumab plus Chemotherapy for Squamous Non-Small-Cell Lung Cancer. *N Engl J Med* 2018, **379**(21):2040-2051. doi:10.1056/NEJMoa1810865

19. Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Cowey CL, Lao CD, Schadendorf D, Dummer R, Smylie M, Rutkowski P, et al. Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. *N Engl J Med* 2015, **373**(1):23-34. doi:10.1056/NEJMoa1504030

20. Liu Z, Li M, Jiang Z, Wang X. A Comprehensive Immunologic Portrait of Triple-Negative Breast Cancer. *Transl Oncol* 2018, **11**(2):311-329. doi:10.1016/j.tranon.2018.01.011

21. Braun DA, Burke KP, Van Allen EM. Genomic Approaches to Understanding Response and Resistance to Immunotherapy. *Clin Cancer Res* 2016, **22**(23):5642-5650. doi:10.1158/1078-0432.Ccr-16-0066

22. Feng Y, Roy A, Masson E, Chen TT, Humphrey R, Weber JS. Exposure-response relationships of the efficacy and safety of ipilimumab in patients with advanced melanoma. *Clin Cancer Res* 2013, **19**(14):3977-3986. doi:10.1158/1078-0432.Ccr-12-3243

23. Zander H, Müller-Egert S, Zwiewka M, Groß S, van Zandbergen G, Engelbergs J. [Checkpoint inhibitors for cancer therapy]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2020, **63**(11):1322-1330. doi:10.1007/s00103-020-03221-9

24. Probst HC, McCoy K, Okazaki T, Honjo T, van den Broek M. Resting dendritic cells induce peripheral CD8+ T cell tolerance through PD-1 and CTLA-4. *Nat Immunol* 2005, **6**(3):280-286. doi:10.1038/ni1165

25. Cogdill AP, Andrews MC, Wargo JA. Hallmarks of response to immune checkpoint blockade. *Br J Cancer* 2017, **117**(1):1-7. doi:10.1038/bjc.2017.136

26. Darvin P, Toor SM, Sasidharan Nair V, Elkord E. Immune checkpoint inhibitors: recent progress and potential biomarkers. *Exp Mol Med* 2018, **50**(12):1-11. doi:10.1038/s12276-018-0191-1

27. Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. Breast cancer. *Lancet* 2021, **397**(10286):1750-1769. doi:10.1016/s0140-6736(20)32381-3

28. Treumann A, Lifely MR, Schneider P, Ferguson MA. Primary structure of CD52. *J Biol Chem* 1995, **270**(11):6088-6099. doi:10.1074/jbc.270.11.6088

29. Karnan S, Hanamura I, Ota A, Takasugi S, Nakamura A, Takahashi M, Uchino K, Murakami S, Wahiduzzaman M, Quang Vu L, et al. CD52 is a novel target for the treatment of FLT3-ITD-mutated myeloid leukemia. *Cell Death Discov* 2021, **7**(1):121. doi:10.1038/s41420-021-00446-8

30. Owen KL, Brockwell NK, Parker BS. JAK-STAT Signaling: A Double-Edged Sword of Immune Regulation and Cancer Progression. *Cancers (Basel)* 2019, **11**(12). doi:10.3390/cancers11122002

31. Rowan WC, Hale G, Tite JP, Brett SJ. Cross-linking of the CAMPATH-1 antigen (CD52) triggers activation of normal human T lymphocytes. *Int Immunol* 1995, **7**(1):69-77. doi:10.1093/intimm/7.1.69

32. Fox EJ, Wynn D, Coles AJ, Palmer J, Margolin DH. Alemtuzumab improves neurological functional systems in treatment-naive relapsing-remitting multiple sclerosis patients. *J Neurol Sci* 2016, **363**:188-194. doi:10.1016/j.jns.2016.02.025

33. Hui YM, Regan F, Willecombe M, Taube D. Use of non-irradiated blood components in Campath (alemtuzumab)-treated renal transplant patients. *Transfus Med* 2016, **26**(2):138-146. doi:10.1111/tme.12292

34. Schub N, Günther A, Schrauder A, Claviez A, Ehlert C, Gramatzki M, Repp R. Therapy of steroid-refractory acute GVHD with CD52 antibody alemtuzumab is effective. *Bone Marrow Transplant* 2011, **46**(1):143-147. doi:10.1038/bmt.2010.68

35. Garnock-Jones KP. Alemtuzumab: a review of its use in patients with relapsing multiple sclerosis. *Drugs* 2014, **74**(4):489-504. doi:10.1007/s40265-014-0195-7

36. Isaacs JD, Wing MG, Greenwood JD, Hazleman BL, Hale G, Waldmann H. A therapeutic human IgG4 monoclonal antibody that depletes target cells in humans. *Clin Exp Immunol* 1996, **106**(3):427-433. doi:10.1046/j.1365-2249.1996.d01-876.x

37. Riechmann L, Clark M, Waldmann H, Winter G. Reshaping human antibodies for therapy. *Nature* 1988, **332**(6162):323-327. doi:10.1038/332323a0

38. Zhao Y, Su H, Shen X, Du J, Zhang X, Zhao Y. The immunological function of CD52 and its targeting in organ transplantation. *Inflamm Res* 2017, **66**(7):571-578. doi:10.1007/s00011-017-1032-8

39. Buggins AG, Mufti GJ, Salisbury J, Codd J, Westwood N, Arno M, Fishlock K, Pagliuca A, Devereux S. Peripheral blood but not tissue dendritic cells express CD52 and are depleted by treatment with alemtuzumab. *Blood* 2002, **100**(5):1715-1720.

40. Ratzinger G, Reagan JL, Heller G, Busam KJ, Young JW. Differential CD52 expression by distinct myeloid dendritic cell subsets: implications for alemtuzumab activity at the level of antigen presentation in allogeneic graft-host interactions in transplantation. *Blood* 2003, **101**(4):1422-1429. doi:10.1182/blood-2002-04-1093

41. Gilleece MH, Dexter TM. Effect of Campath-1H antibody on human hematopoietic progenitors in vitro. *Blood* 1993, **82**(3):807-812.

42. Linn R, DuPont BR, Knight CB, Plaetke R, Leach RJ. Reassignment of the 92-kDa type IV collagenase gene (CLG4B) to human chromosome 20. *Cytogenet Cell Genet* 1996, **72**(2-3):159-161. doi:10.1159/000134175

43. Pagès F, Berger A, Camus M, Sanchez-Cabo F, Costes A, Molidor R, Mlecnik B, Kirilovsky A, Nilsson M, Damotte D, et al. Effector memory T cells, early metastasis, and survival in colorectal cancer. *N Engl J Med* 2005, **353**(25):2654-2666. doi:10.1056/NEJMoa051424

44. De Angulo G, Yuen C, Palla SL, Anderson PM, Zweidler-McKay PA. Absolute lymphocyte count is a novel prognostic indicator in ALL and AML: implications for risk stratification and future studies.

*Cancer* 2008, **112**(2):407-415. doi:10.1002/cncr.23168

45. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature* 2017, **541**(7637):321-330. doi:10.1038/nature21349

46. Shi JY, Gao Q, Wang ZC, Zhou J, Wang XY, Min ZH, Shi YH, Shi GM, Ding ZB, Ke AW, et al. Margin-infiltrating CD20(+) B cells display an atypical memory phenotype and correlate with favorable prognosis in hepatocellular carcinoma. *Clin Cancer Res* 2013, **19**(21):5994-6005. doi:10.1158/1078-0432.Ccr-12-3497

47. Bense RD, Sotiriou C, Piccart-Gebhart MJ, Haanen J, van Vugt M, de Vries EGE, Schröder CP, Fehrmann RSN. Relevance of Tumor-Infiltrating Immune Cell Composition and Functionality for Disease Outcome in Breast Cancer. *J Natl Cancer Inst* 2017, **109**(1). doi:10.1093/jnci/djw192

48. Verma C, Kaewkangsadan V, Eremin JM, Cowley GP, Ilyas M, El-Sheemy MA, Eremin O. Natural killer (NK) cell profiles in blood and tumour in women with large and locally advanced breast cancer (LLABC) and their contribution to a pathological complete response (PCR) in the tumour following neoadjuvant chemotherapy (NAC): differential restoration of blood profiles by NAC and surgery. *J Transl Med* 2015, **13**:180. doi:10.1186/s12967-015-0535-8

49. Zhang SC, Hu ZQ, Long JH, Zhu GM, Wang Y, Jia Y, Zhou J, Ouyang Y, Zeng Z. Clinical Implications of Tumor-Infiltrating Immune Cells in Breast Cancer. *J Cancer* 2019, **10**(24):6175-6184. doi:10.7150/jca.35901

50. Su S, Liao J, Liu J, Huang D, He C, Chen F, Yang L, Wu W, Chen J, Lin L, et al. Blocking the recruitment of naive CD4(+) T cells reverses immunosuppression in breast cancer. *Cell Res* 2017, **27**(4):461-482. doi:10.1038/cr.2017.34

51. Li Y, Shao Y, Bai L, Zhou X. Increased derived neutrophil-to-lymphocyte ratio and Breast Imaging-Reporting and Data System classification predict poor survival in patients with non-distant metastatic HER2+ breast cancer treated with neoadjuvant chemotherapy. *Cancer Manag Res* 2018, **10**:3841-3847. doi:10.2147/cmar.S174537

52. Ren K, Yin Y, He F, Shao Y, Wang S. Prognostic role of derived neutrophil-to-lymphocyte ratio in surgical triple-negative breast cancer. *Cancer Manag Res* 2018, **10**:4891-4898. doi:10.2147/cmar.S180695

53. Gu-Trantien C, Loi S, Garaud S, Equeter C, Libin M, de Wind A, Ravoet M, Le Buanec H, Sibille C, Manfouo-Foutsop G, et al. CD4 follicular helper T cell infiltration predicts breast cancer survival. *J Clin Invest* 2013, **123**(7):2873-2892. doi:10.1172/jci67428

54. Sharma P, Hu-Lieskovan S, Wargo JA, Ribas A. Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell* 2017, **168**(4):707-723. doi:10.1016/j.cell.2017.01.017

55. Anderson AC, Joller N, Kuchroo VK. Lag-3, Tim-3, and TIGIT: Co-inhibitory Receptors with Specialized Functions in Immune Regulation. *Immunity* 2016, **44**(5):989-1004. doi:10.1016/j.immuni.2016.05.001

56. Toor SM, Sasidharan Nair V, Decock J, Elkord E. Immune checkpoints in the tumor microenvironment. *Semin Cancer Biol* 2020, **65**:1-12. doi:10.1016/j.semcancer.2019.06.021

57. Sharma P, Allison JP. The future of immune checkpoint therapy. *Science* 2015, **348**(6230):56-61. doi:10.1126/science.aaa8172

58. Shekhar MP, Pauley R, Heppner G. Host microenvironment in breast cancer development: extracellular matrix-stromal cell contribution to neoplastic phenotype of epithelial cells in the breast. *Breast Cancer Res* 2003, **5**(3):130-135. doi:10.1186/bcr580

59. Weber F, Shen L, Fukino K, Patocs A, Mutter GL, Caldes T, Eng C. Total-genome analysis of BRCA1/2-related invasive carcinomas of the breast identifies tumor stroma as potential landscaper for neoplastic initiation. *Am J Hum Genet* 2006, **78**(6):961-972. doi:10.1086/504090

60. Moinfar F, Man YG, Arnould L, Bratthauer GL, Ratschek M, Tavassoli FA. Concurrent and independent genetic alterations in the stromal and epithelial cells of mammary carcinoma: implications for tumorigenesis. *Cancer Res* 2000, **60**(9):2562-2566.

61. Fukino K, Shen L, Patocs A, Mutter GL, Eng C. Genomic instability within tumor stroma and clinicopathological characteristics of sporadic primary invasive breast carcinoma. *Jama* 2007, **297**(19):2103-2111. doi:10.1001/jama.297.19.2103

Supplemental

Supplementary 6 is not available with this version.

Figures
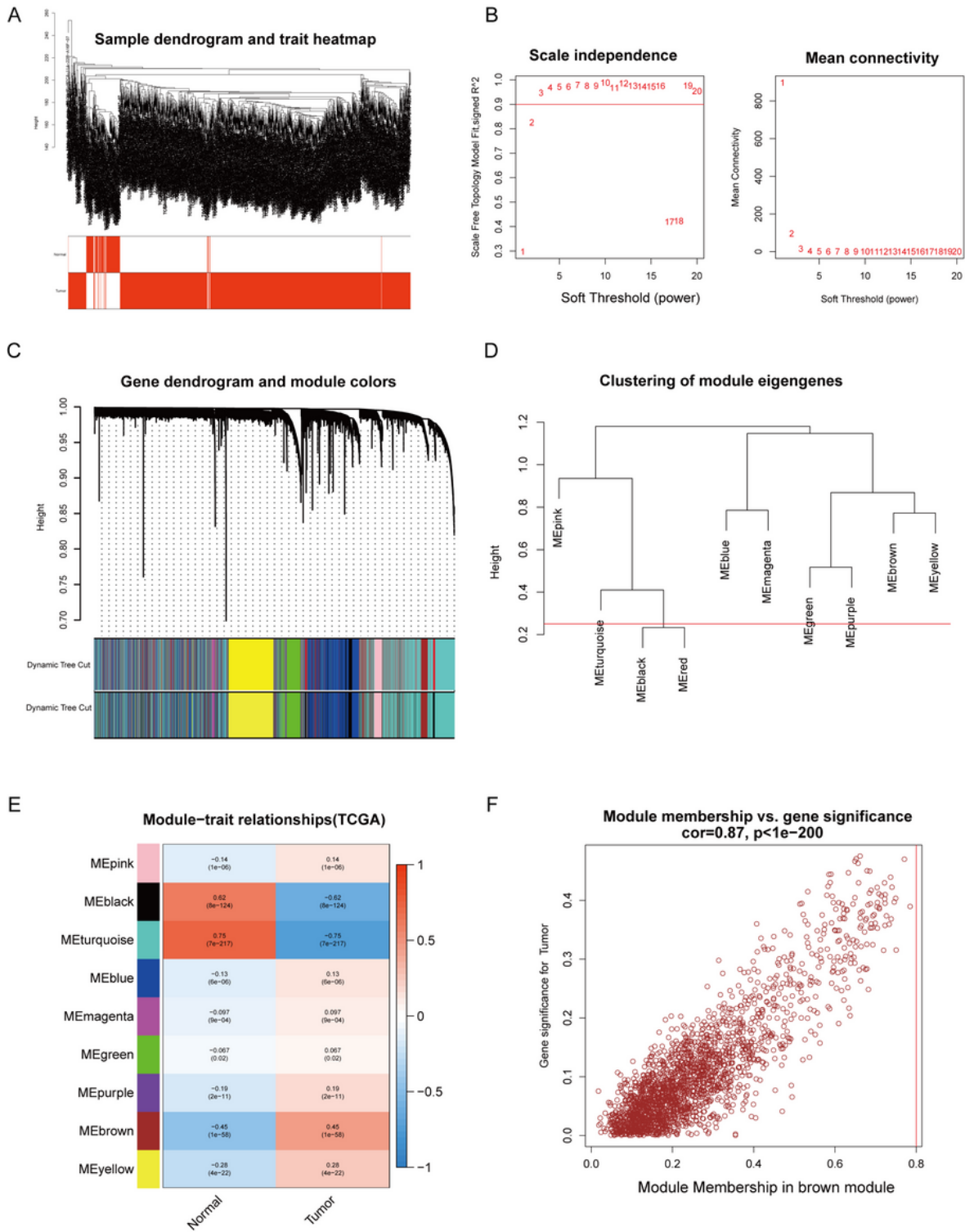
Figure 1

The relationship between immune score, stromal score and the transcriptome of breast cancer. A: Analysis of genetic differences between the high and low immune score groups. B: Genetic difference analysis between high and low stromal score groups. C: 203 genes upregulated in the high immune score group and the high stromal score group. D: Twenty-five genes were downregulated in the high immune score group and the high stromal score group. E: GO enrichment analysis of common differentially
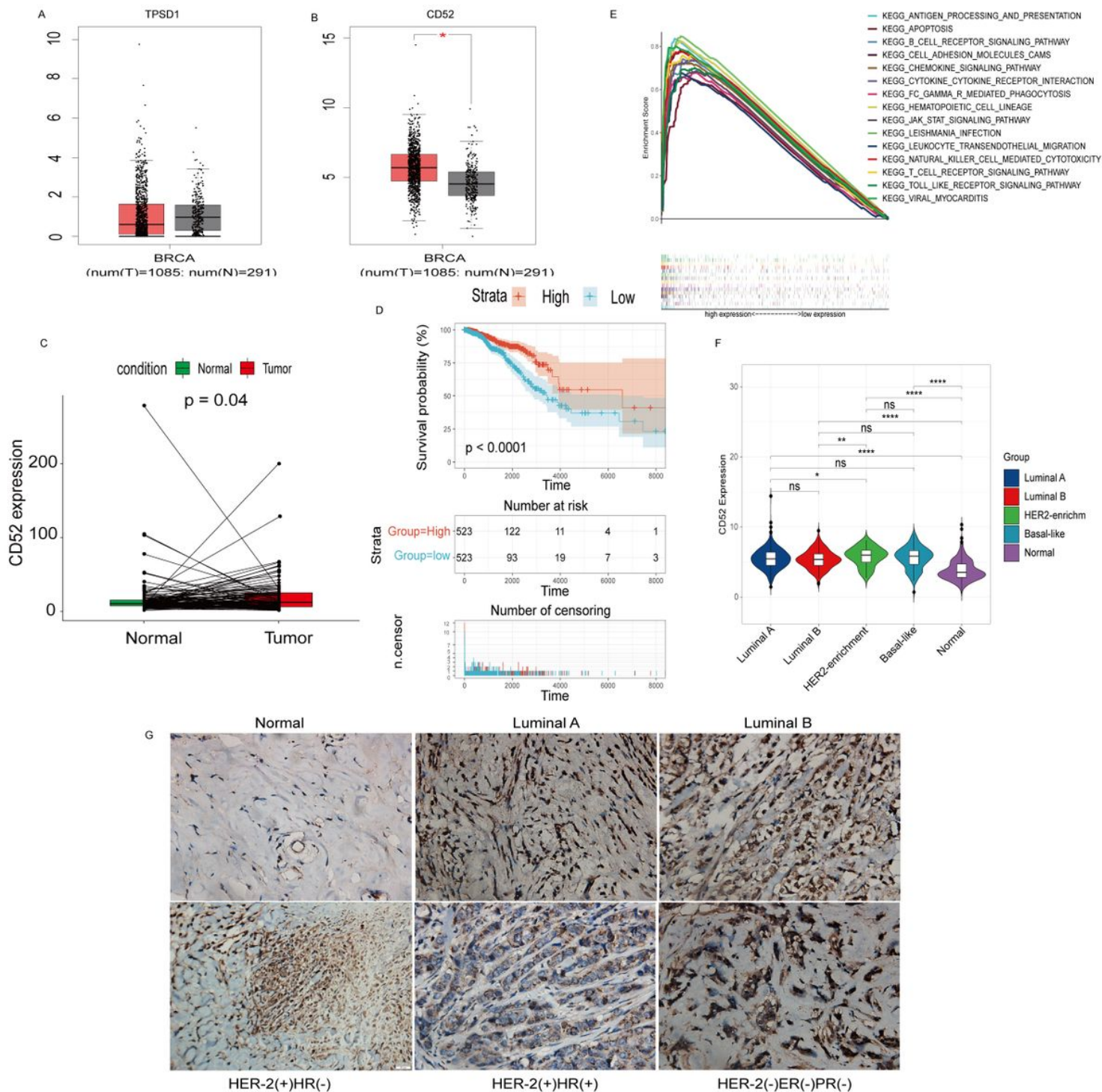
expressed genes between the high immune score group and the high matrix score group. F: KEGG enrichment analysis of common differentially expressed genes between the high immune score group and the high matrix score group.
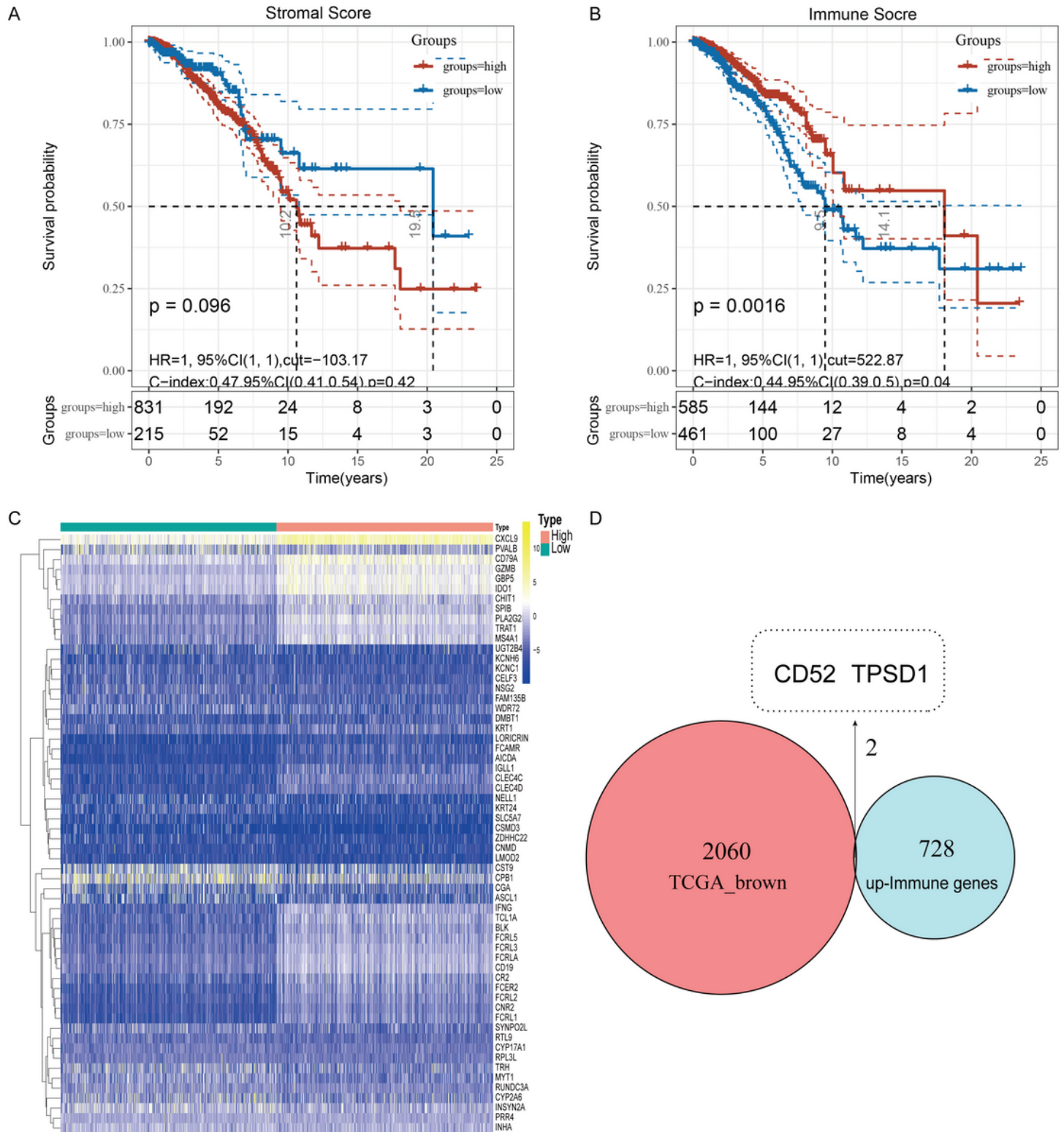
Figure 2

Screening of hub genes. B: PPI network of common differential genes. B: The top 30 hub genes calculated by the MCC algorithm.

Figure 3

The expression of CD52 in breast cancer. A: Venn diagram of the intersection of the hub gene and the prognostic gene. B and C: The expression of CD52 is significantly increased in breast cancer tissues. D: The higher the expression of the CD52 genes, the higher the overall survival rate of patients. E: GSEA of CD52. The results show that CD52 is related to immune processes, apoptosis, cytokine and cytokine receptor interactions, the JAK-STAT signaling pathway, and the TOLL-like receptor signaling pathway. F: The expression of CD52 in each subtype revealed that the expression of CD52 in each subtype of breast cancer was higher than that in normal tissues. G: Immunohistochemistry confirmed that compared with

normal breast tissue, CD52 was significantly increased in various molecular subtypes (CSCO breast cancer guideline molecular classification) (Wilcoxon test, ***p <0.001, **p <0.01, *p <0.05, NS denotes no significant difference).
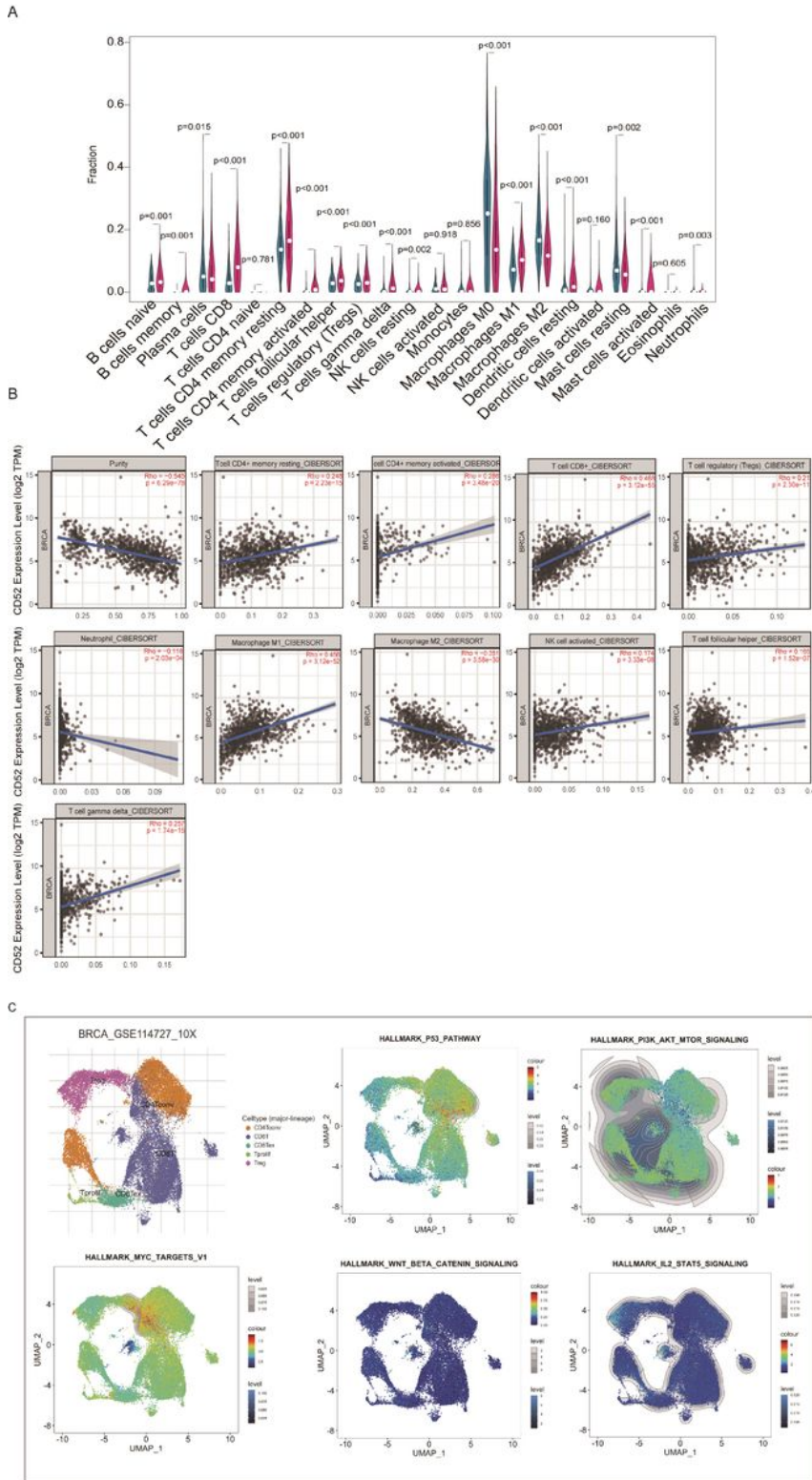
Figure 4

The relationship between CD52 and the immune microenvironment. A: The infiltration of various immune cells between the CD52 low expression group (green) and the CD52 high expression group (red). B: The

correlation between CD52 expression and immune cell infiltration. C: The P53 signaling pathway is highly enriched in CD4 Tn, CD8T, CD8Tex, Tprolif and Treg cells. The IL6/JAK/STAT3 signaling pathway was slightly enriched in CD4 Tn, CD8T, CD8Tex, Tprolif and Treg cells. The IL2/STAT5 signaling pathway and Wnt/β-catenin pathway showed no enrichment in almost any T cells. (Wilcoxon test, ***p <0.001, **p <0.01, *p <0.05, NS denotes no significant difference).
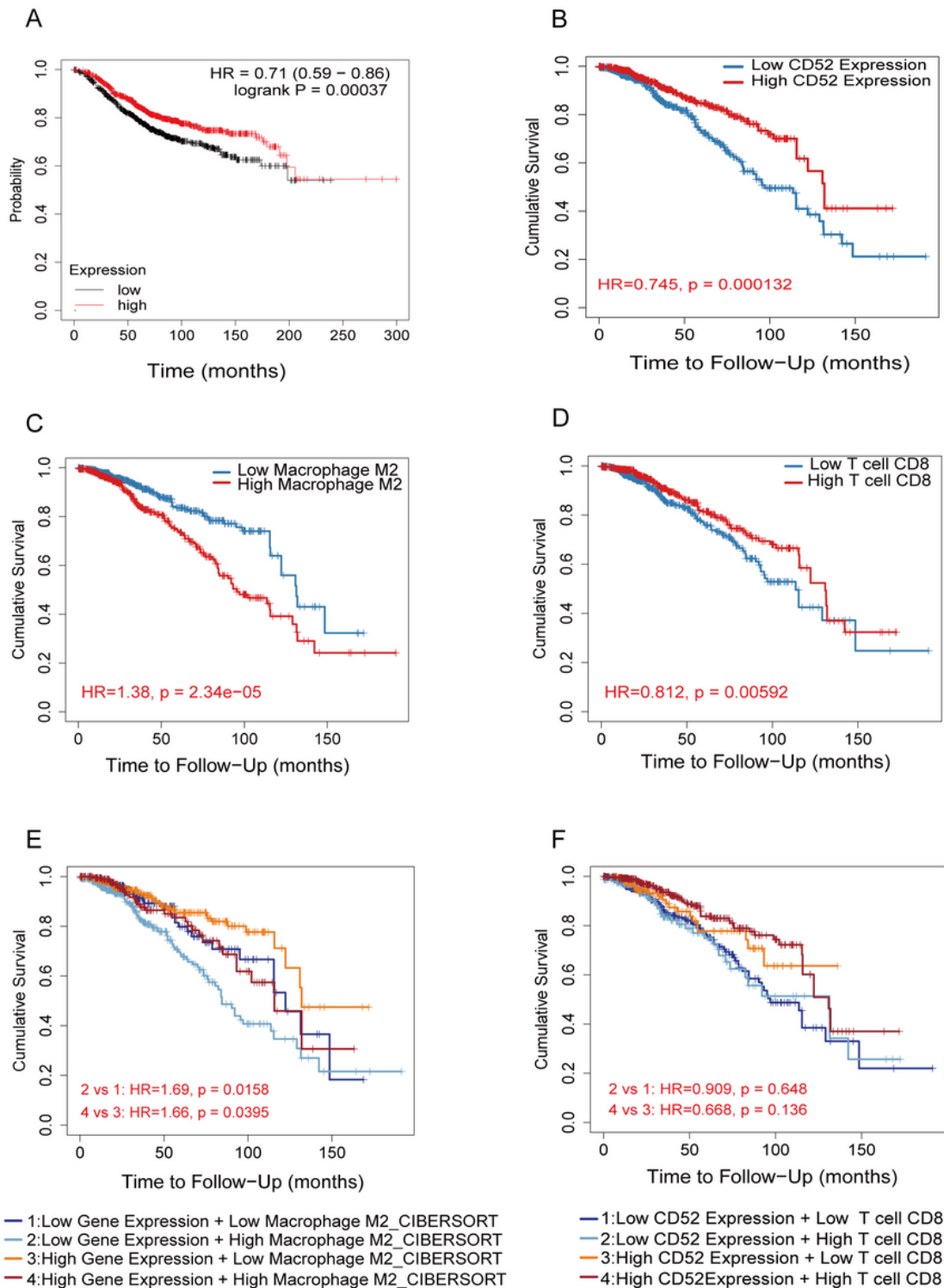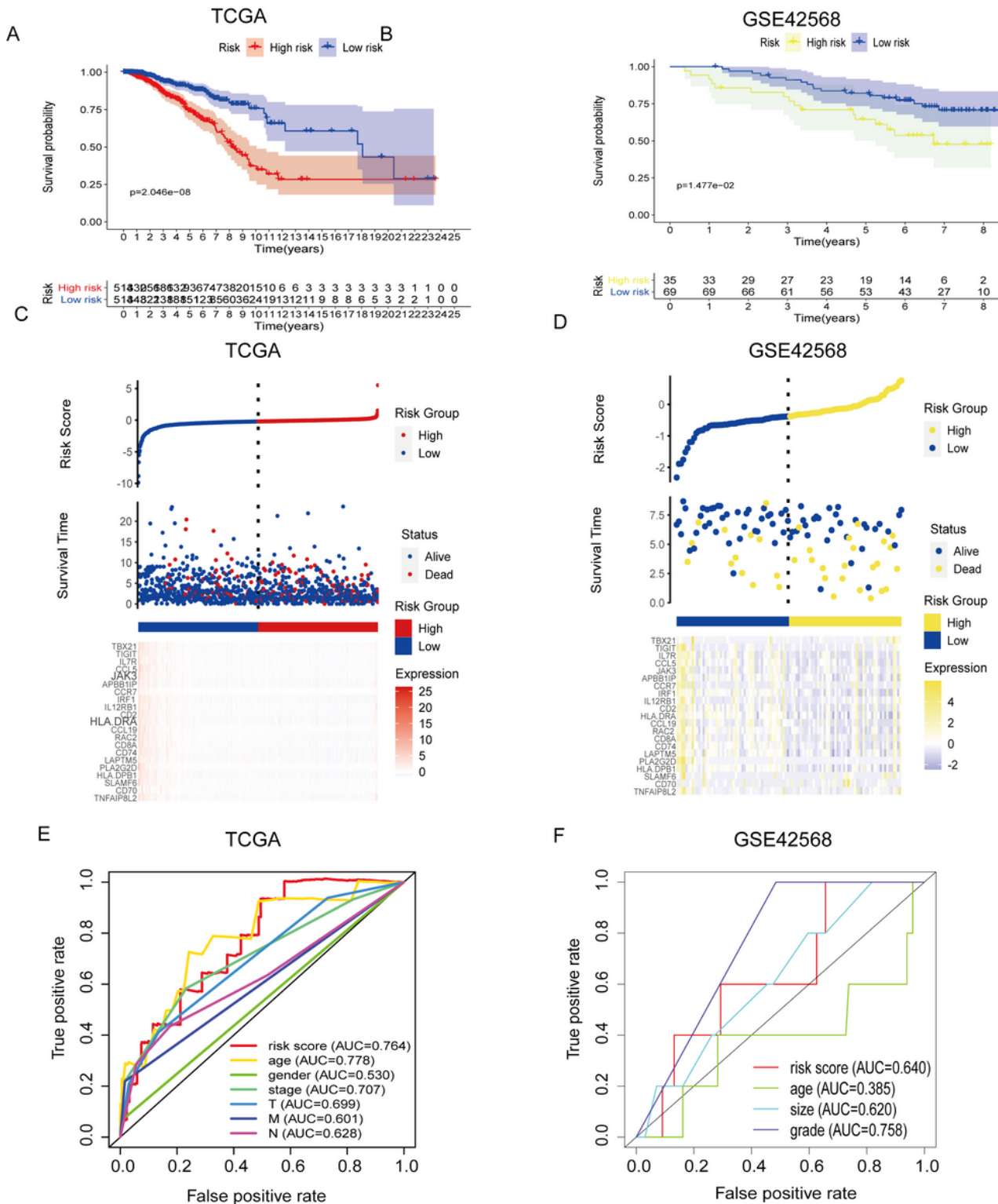

Figure 5

The relationship between CD52, M2 macrophages, CD8+ T cells and survival. A: TIMER database analysis revealed that high expression of CD52 led to longer survival. B: Kaplan-Meier Plotter analysis found that high expression of CD52 led to longer survival. C: The level of macrophage M2 infiltration is related to survival. D: CD52 combined with macrophage M2 survival analysis. E: The level of CD8+ T cell infiltration is related to survival. F: CD52 combined with T cell CD8+ survival analysis.
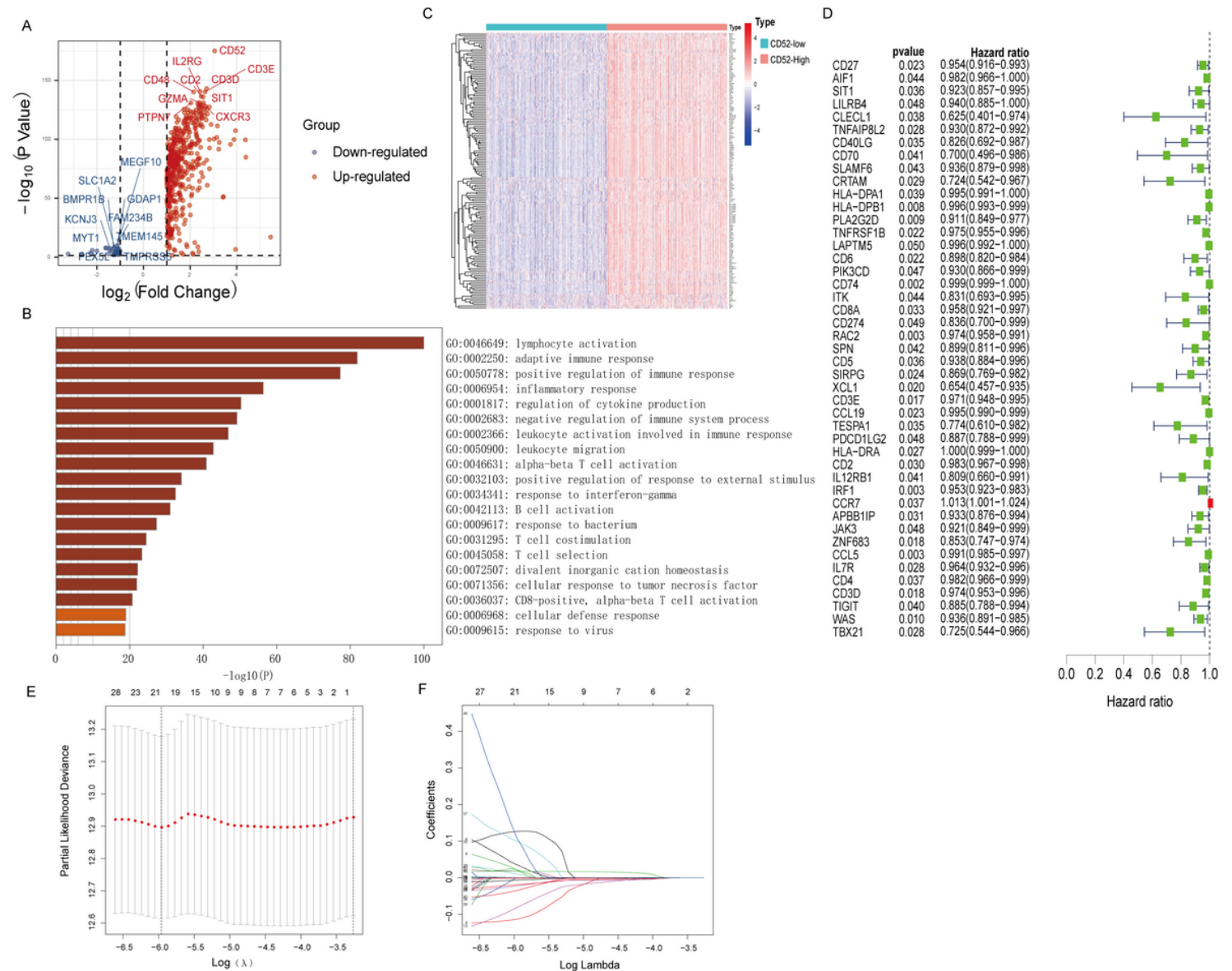
Figure 6

CD52 is a marker of AIR stratification in breast cancer patients. A: The figure shows the 10 most significantly upregulated genes and the 10 most downregulated genes. B: The results show that the biological processes enriched by differential genes are all related to the immune response. C: All genes enriched in the adaptive immune response were upregulated in the CD52 high expression group. D: To identify the CD52-related adaptive immune response gene signature (CD52-AIRGsig) with prognostic value, 155 CD52-AIRGs were analyzed using CPHR to obtain 45 genes. E-F: The 45 prognostic genes were

processed by the LASSO model to obtain 21 genes. The model was performed for 100 cycles to prevent overfitting.

Figure 7

Determination of the prognostic value of the AIR gene associated with CD52 (CD52-AIRGs) in BC. A-B: Kaplan−Meier estimates of OS according to the CD52-AIRGs in the training set (A) and validation set (B). C-D: The distribution of risk scores, survival overview and expression profiles of 21 genes in the training
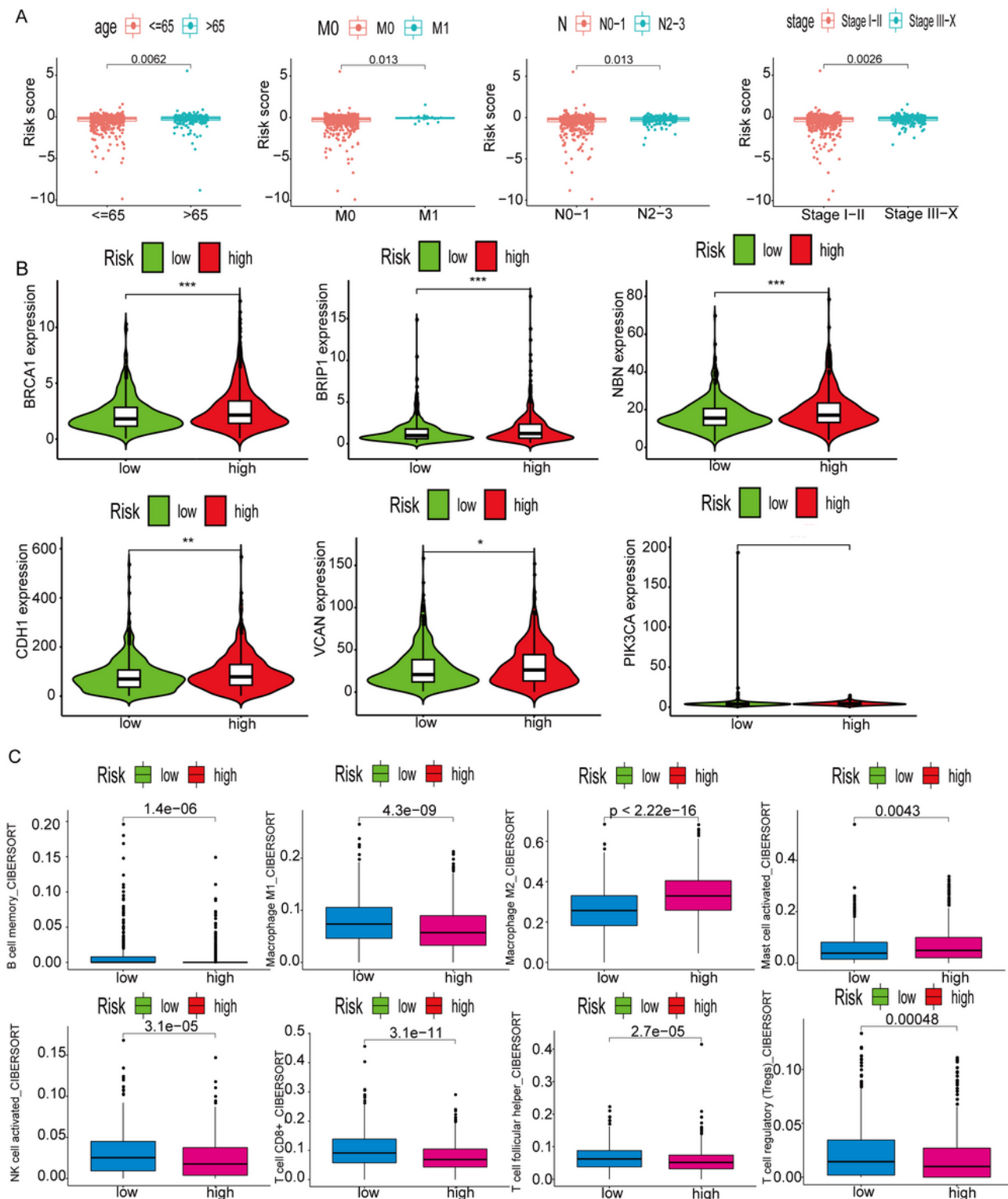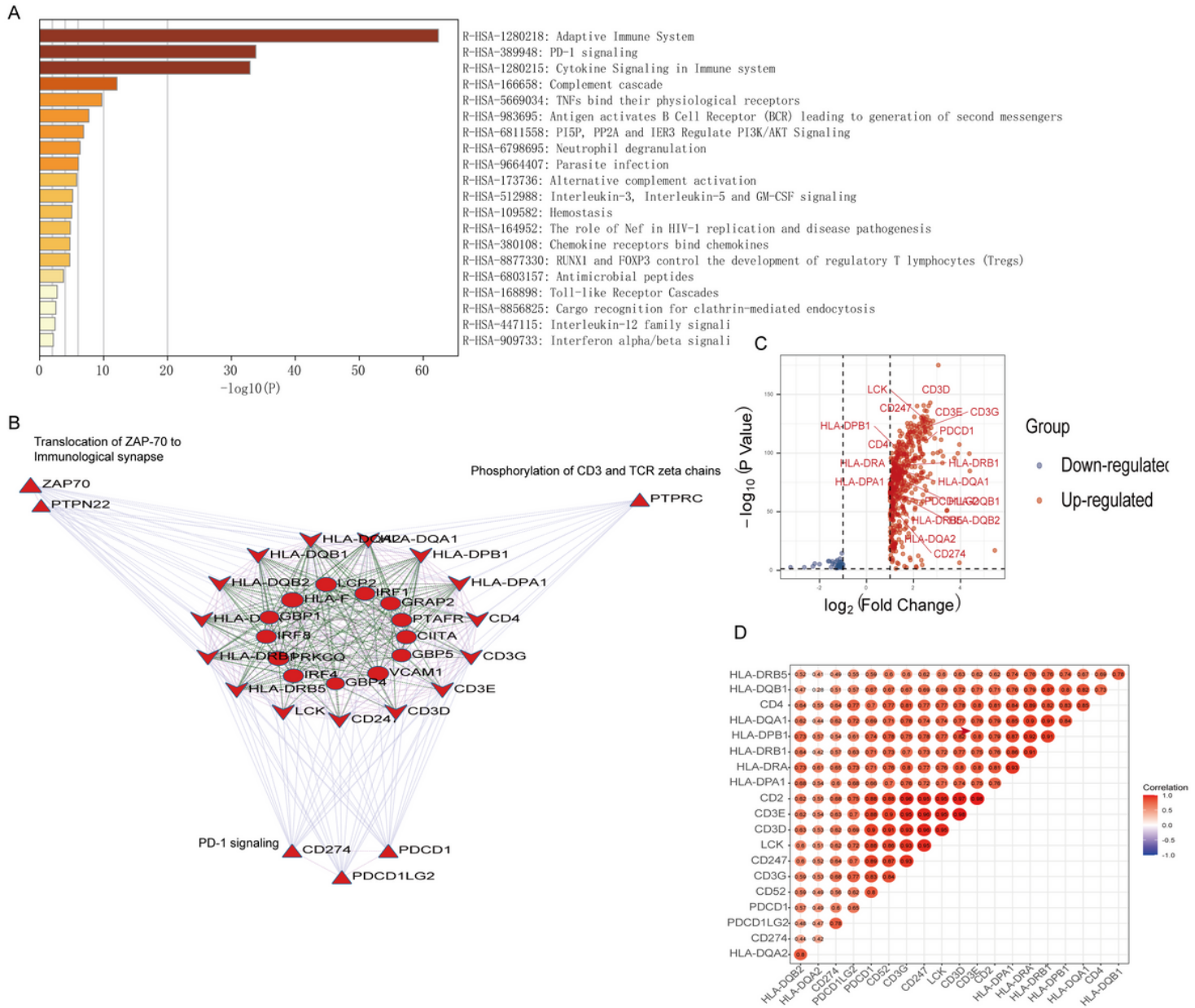
Figure 8

CD52-AIRGsig is related to the clinical features of breast cancer, genome instability and TME immune cell infiltration. A: The CD52-AIRGsig score is significantly increased in patients with advanced age, distant metastasis, stage N2-3 and stage III-IV disease. B: The expression of these breast cancer susceptibility

genes (BRCA1, VCAN, BRIP1, CDH1, PIK3CA, DIRAS3, and NBN) was significantly increased in the high-risk group. C: The results showed that memory B lymphocytes, M1 macrophages, NK cells, CD8+ T cells, Tregs, and follicular helper T cells significantly increased infiltration in the low-risk group. However, tumor-associated macrophages (M2 macrophages) were highly infiltrated in the high-risk group, consistent with the CD52-Low group. (Wilcoxon test, ***p <0.001, **p <0.01, *p <0.05, NS denotes no significant difference).

Figure 9

CD52 is related to PD-1 signaling. A: R-HSA pathway enrichment analysis of the differentially expressed genes of the two groups with high and low expression of CD52. B: The figure shows the PD-1 signaling pathway and the genes enriched in this pathway. C: The genes enriched in the PD-1 signaling pathway were all highly expressed in the CD52-High group. D: We performed Spearman correlation analysis on
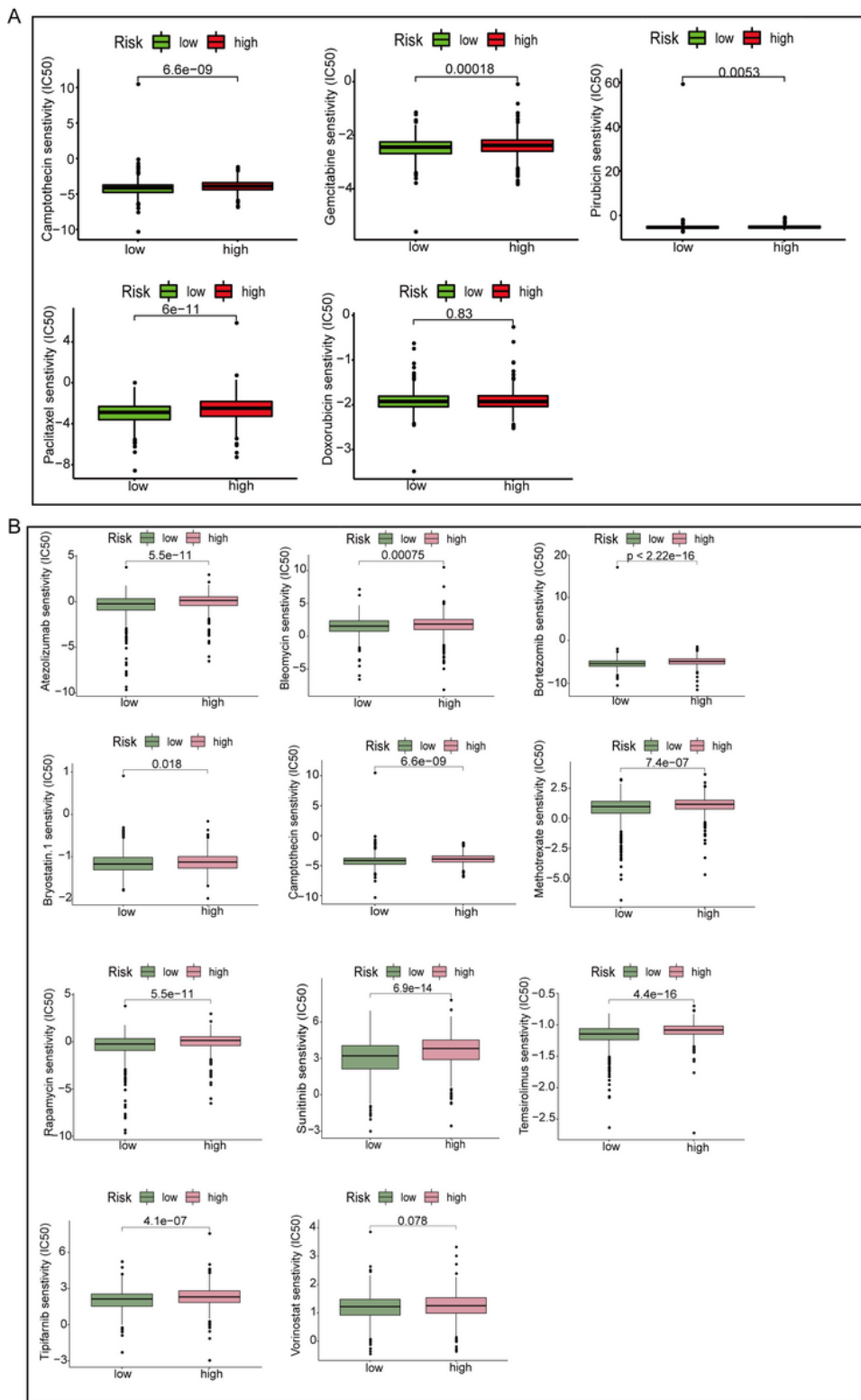
CD52 and PD-1 signal-related genes, and the results showed that CD52 has a significant positive correlation with these genes (p⬚0.05⬚r⬚0.4).
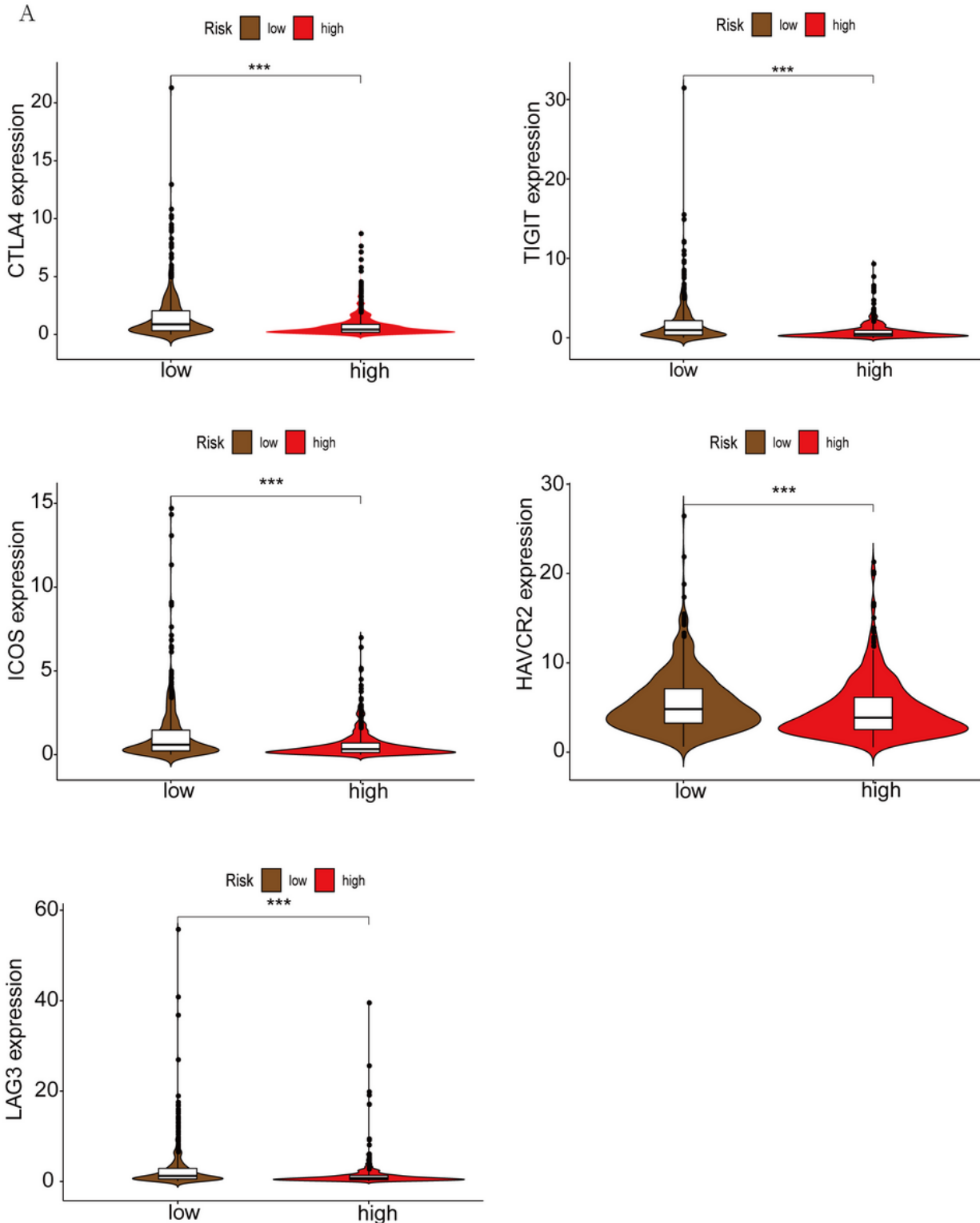
Figure 10

CD52-AIRGsig is related to ICI markers. ICI markers (CTLA-4, LAG-3, HAVCR2, TIGIT, ICOS) were more highly expressed in the low-risk group. (Wilcoxon test, ***p <0.001, **p <0.01, *p <0.05, NS denotes no significant difference).

Figure 11

Screening of chemotherapeutics for personalized treatment based on CD52-AIRGsig. Research showed that the IC50 values of atezolizumab, bleomycin, bortezomib, bromostatin.1, camptothecin, methotrexate, rapamycin, sunitinib, temsirolimus, tipifarnib, and vorinostat chemotherapy drugs in the low-risk group were significantly lower than in the high-risk group. (Wilcoxon test, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, NS denotes no significant difference).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfigure1.tif
- Supplementaryfigure2.jpg
- Supplementaryfigure3.jpg
- Supplementaryfigure4.tif
- Supplementaryfigure5.tif