

# WITHDRAWN: Establishment and Validation of a Novel Metabolism-related Gene Signature for Predicting Overall Survival of Patients with Breast Cancer

Jian Li

my-1j@163.com

Taian City Central Hospital <https://orcid.org/0000-0002-6633-4340>

Yang Liu

Taian City Central Hospital

Fei Liu

Taian City Central Hospital

Qiang Tian

Taian City Central Hospital

Baojiang Li

Taian City Central Hospital

Zongwang Zhang

Liaocheng People's Hospital

Guangwei Wei

Shandong University School of Medicine: Shandong University Cheeloo College of Medicine

---

## Research Article

**Keywords:** metabolic signature,tcga,geo,breast cancer TME

**Posted Date:** February 23rd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-222322/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

EDITORIAL NOTE:

The full text of this preprint has been withdrawn by the authors while they make corrections to the work. Therefore, the authors do not wish this work to be cited as a reference. Questions should be directed to the corresponding author.

# Abstract

It is well known that Breast cancer is a heterogeneous disease. Although the current recurrence and mortality rate have been greatly improved, many people still suffer relapse and metastasis. Metabolic reprogramming is currently considered to be a new hallmark of cancer. Therefore, in this study, we comprehensively analyzed the prognostic effect of metabolic-related gene signatures in breast cancer and its relationship with the immune microenvironment. We constructed a novel metabolic-related gene signature containing 6 genes to distinguish between high and low risk groups by univariate Cox regression and least absolute shrinkage and selection operator (LASSO) regression, and validated its robustness and accuracy through multiple databases.

The metabolic gene signature may be an independent risk factor for BC both in the training and the testing set, the nomogram has a moderately accurate performance, and the C index was 0.757 and 0.728 respectively. The signature can reveal metabolic characteristics based on gene set enrichment analysis and at the same time monitor the status of TME. This gene signature can be used as a promising independent prognostic marker for BC patients, and can indicate the current status of TME, providing more clues for exploring new diagnostic and treatment strategies.

## Introduction

Breast cancer (BC) has emerged as the highest incidence in women accounting for nearly 24.2% of female new cases worldwide and the leading cause of death according to the latest report of Global Cancer Statistics 2018(1). The incidence and mortality of breast cancer are rising year by year (2). Although big progress has been made in the treatment of breast cancer over the previous decades, the long-term prognosis for patients remains poor due to the high rate of recurrence and metastasis. Breast cancer is a heterogeneous disease consisting of five molecular subtypes and different treatments approaches are needed according to the type of biological molecule markers(3). Therefore looking for new targets for treatment of breast cancer and the development of diagnostic and prognosis markers is an urgent need and could provide early and effective treatment. Improved prognostic models will help to individualize treatment regimens for breast cancer patients(4).

Tumor cells have powerful characteristics of adapting to adversity and growing rapidly. This adaptability is achieved by changing the way of energy metabolism of tumor cells, called metabolic-related reprogramming(5). Metabolic-related reprogramming has become one of the ten hallmarks of tumors, which is also one of the four emerging characteristics of tumors(6). Changes in tumor cell metabolism enable tumor cells to obtain growth advantages, escape metabolic-related pressure, promote tumor cell immune escape, and enhance their invasion ability(7). Metastasis formation accounts for the majority of tumor-related deaths and consists of different steps, each of which is characterized by a unique adaptive phenotype of cancer cells. Metabolic reprogramming represents one of the main adaptive phenotypes utilized by cancer cells in all major steps of tumor and metastasis progression. However, the underlying

molecular mechanism and pathways of metabolic-related reprogramming are still unclear, especially in breast cancer.

To further investigate the prognostic role of these metabolic genes in BC, we conduct this study. In this study, we downloaded the data from the The Cancer Genome Atlas (TCGA) databases to identify the differentially expressed metabolism-related genes (DE-MRGs) in breast cancer and established prognosis molecular signatures and validated it through the Gene Expression Omnibus (GEO) database. The important prognostic role of metabolic-related genes (MRGs) in breast cancer has been elucidated.

## Materials And Methods

### Data collection and Identification of differentially expressed MRGs (DE-MRGs)

We downloaded the messenger RNA (mRNA) expression profiles (FPKM format) and the corresponding clinical information of breast cancer patients from TCGA database (<https://portal.gdc.cancer.gov>) concluding 113 normal breast and 1109 BC cases. Missing data was removed. The data were all female patients. The GSE20685 gene expression profile and clinical information was collected from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>), which contains a total of 327 BC cases. TCGA dataset were used as the training cohort and the geo dataset were used as the validation cohort. Samples within 30 days of follow-up data were deleted.

The MRGs were obtained from the downloads section of GSEA website (<http://software.broadinstitute.org/gsea/index.jsp>). GSEA v4.1.0 for Windows and c2.cp.kegg.v7.2.symbols.gmt were downloaded from the GSEA website for the further extraction analysis. Genes enriched in metabolism pathways of the KEGG database were utilized in this study as MRGs. In total, 948 MRGs were obtained, of which 861 MRGs were common between TCGA and GSE20685 dataset. The mRNA expression of common MRGs in the TCGA database and GEO database was extracted respectively. Then, an adjustment was given to adjust different mRNA expression levels of metabolic-related genes between TCGA and GEO databases by 'sva' package (8) in R software (version 4.0.2).

Genes were selected as candidate metabolic-related genes for meeting the following two conditions: (a) they have the same expression pattern in TCGA database and the GEO database; and (b) they were listed in the metabolic-associated pathways. The R package 'limma' (9) was used to screen the DE-MRGs in the TCGA database. The screening criteria were set as  $|\log FC| > 0.5$  and  $P\text{-value} < 0.05$ .

### Prediction of potential drugs for BC treatment by Cmap

The Cmap database (<http://www.broadinstitute.org/cMAP/>) is a computer simulation method to predict small molecular compounds for a specific disease. The DE-MRGs were divided into upregulated and downregulated genes group which were subsequently uploaded to the Cmap, and searches for small

molecule drugs that may treat BC. The higher negative connectivity score of drugs indicate greater suggestive of its possible treatment. Drugs with a score of  $\leq 0.75$  were considered as candidate drugs. Tomograph of these relative molecular drugs was searched in Pubchem database.

## GO and KEGG Analysis

The functions of the DE-MRGs were analyzed using the Gene Ontology (GO) tool and the Kyoto Encyclopedia of Genes and Genomes (KEGG) by using the “clusterProfiler” package(10). A P-value of  $< 0.05$  was considered significant.

## Construction and Assessment of the Metabolic Gene Signature and Survival Analysis

Univariate Cox regression analysis was used to identify prognostic associated MRGs in the TCGA database. Genes with  $P < 0.05$  were defined as candidate genes to further construct the least absolute shrinkage and selection operator (LASSO)-penalized Cox regression model. The LASSO regression was performed to construct the prognostic signature and avoid overfitting of this model by glmnet package(11). The risk score of every patient was calculated with this model. The formula of risk score was as follows:  $\text{risk score} = (\text{Coefficient}_{\text{mRNA1}} \times \text{expression of mRNA1}) + (\text{Coefficient}_{\text{mRNA2}} \times \text{expression of mRNA2}) + \dots + (\text{Coefficient}_{\text{mRNA}_n} \times \text{expression mRNA}_n)$ .

Based on the median of risk score, samples were classified into high-risk group and low-risk group. A Kaplan–Meier (K-M) survival curve was plotted to compare the predictive survival time between the two groups. The model was evaluated by the area under the curves (AUCs) of the receiver operator characteristic (ROC) curve. The risk score was also proved as an independent risk factor by univariate and multivariate Cox regression respectively. During the analyses, the TCGA cohort was used as a training set and the GSE20685 cohort was used as an external testing cohort.

## Construction and Validation of a Prognostic Nomogram

Based on the lasso Cox regression model results, a prognostic nomograms including clinical features and risk scores were constructed and assessed to predict 3 and 5 year survival rates in BC patients for TCGA and GEO cohorts respectively. The calibration curves was plotted to evaluate the prediction probabilities and fitness of the metabolic signature. The distinguishing ability of the nomograms was evaluated by the concordance index (C-index), time-Dependent ROC Curve and the area under the curve (AUC). Finally, a net benefit curve for patients was plotted to reflect the potential utility and evaluate the clinical value of this model by decision curve analysis (DCA).

## External Validation of the Prognostic Signature.

The genes included in the signature validated their expression significance in TIMER database(<https://cistrome.shinyapps.io/timer/>).The cBioportal database (<https://www.cbioportal.org/>) was used to investigate the overview of the alteration that occurred in BC for the novel metabolic-related genes.Moreover,The protein expression levels was validated in the Human Protein Atlas database (<https://www.proteinatlas.org/>) to visually compare the differentiation between tumor and control tissues.

## GO and KEGG analyses by GSEA

We utilized GSEA software(12)(version 4.1.0,permutations was set as 1000) to perform KEGG pathway analyses for the TCGA cohort.The potential molecular mechanism of genes in this signature was elucidated. Enrichment results satisfying a nominal P-value cutoff of  $<0.05$  with FDR  $<0.25$  were considered statistically significant.

## Clinical Application in Tumor Microenvironment.

The ESTIMATE scores for each sample was calculated and the comparison between high- and low-risk groups was made.The ESTIMATE scores contained immune score, stromal score,and tumor purity which respectively reflected the infiltration level of immune cells,the stromal content, and the estimated tumor purity(13).Furthermore,we calculated specific infiltration levels for 22 subtypes of immune cells through the CIBERSORT system to extend the utility of this metabolic signature(14).

## Statistical analysis

All the statistical analyses were conducted by R software v4.0.2.The Cox and LASSO regression were employed to screen the survival related variables.The survival curves were compared by the log-rank test. The differences for the independent samples were analyzed by the Wilcoxon rank-sum test.The coefficient of correlation was calculated by Pearson correlation analysis.  $p<0.05$  was considered statistically significant.

## Results

### Extraction of MRGs from the TCGA database and identification of DE-MRGs

We downloaded a total of 1222 cases from the TCGA database and extracted 927 MRGs from the TCGA gene expression matrix.After adjusting with GSE20685 expression profile,861 MRGs was finally selected

as candidate MRGs. After screening, we identified 315 DE-MRGs in the TCGA database under the screening criteria ( $|\log FC| > 0.5$  and the  $P\text{-value} < 0.05$ ), including 163 up-regulated MRGs and 152 down-regulated MRGs. The top 10 differentially up-regulated MRGs and the top 10 differentially down-regulated MRGs were demonstrated in Table 1. The heatmaps of these top differentially up-regulated and down-regulated MRGs were demonstrated in Figure 1.

Table I  
The top 10 up-regulated and top 10 down-regulated genes

Gene	ConMean	TreatMean	LogFC	P-value	fdr
Up-regulated gene					
NAT1	2.717125	40.79225	3.908142	3.60E-14	7.05E-14
TK1	2.818031	32.68337	3.535797	2.27E-60	1.81E-58
PAFAH1B3	5.771998	38.95801	2.754777	1.05E-61	1.08E-59
PYCR1	5.698108	32.29472	2.502744	5.44E-58	1.78E-56
RRM2	2.16075	11.76869	2.44535	9.89E-59	3.94E-57
DEGS2	3.049416	16.23833	2.412798	3.65E-25	1.07E-24
TYMP	7.227289	27.98339	1.953044	2.53E-39	1.53E-38
CA12	14.65247	54.9247	1.906311	1.81E-15	3.68E-15
CA9	1.263399	4.680852	1.889461	0.000527	0.000644
TSTA3	8.178872	29.50948	1.851205	1.94E-55	3.76E-54
Down-regulated gene					
GPD1	145.5579	3.467195	-5.39168	3.42E-55	6.30E-54
ADH1B	165.5105	4.995066	-5.05028	2.05E-58	7.74E-57
LPL	199.7108	7.804309	-4.6775	5.30E-58	1.78E-56
CA3	40.57587	2.127329	-4.25351	1.35E-56	3.12E-55
GPAM	50.65867	3.613939	-3.80917	6.57E-59	2.88E-57
AOC3	96.77993	7.335834	-3.72168	1.05E-57	3.26E-56
GPX3	234.1973	18.43828	-3.66695	5.10E-51	6.91E-50
PYGM	13.24036	1.343298	-3.30109	8.35E-60	4.61E-58
PCK1	11.45108	1.412414	-3.01925	1.04E-47	1.10E-46
MAOA	38.72915	5.103234	-2.92394	4.88E-56	1.06E-54

# The potential drug treatments for breast cancer

From the prediction of the Cmap dataset, 6 candidate drugs that scored  $\leq -0.75$  were considered as potential drugs for BC treatment. Strong negative correlation was found between BC and LY-294002, tanespimycin, apigenin, omeprazole, liothyronine, sirolimus. Strong positive correlation was found between BC and adiphenine, viomycin, Prestwick-692, genistein, isoflupredone and atractyloside (Table 2). These drugs might have therapeutic effects on BC. The tomographs of the top 3 associated molecule drugs were investigated in Pubchem database (Figure 2a-c).

Table II  
Strong negative and positive drug with BC in CMAP

rank	cmap name	mean	n	enrichment	p
2	LY-294002	-0.524	61	-0.387	0
3	tanespimycin	-0.452	62	-0.353	0
11	apigenin	-0.725	4	-0.866	0.00062
13	omeprazole	-0.725	4	-0.843	0.00105
19	liothyronine	-0.679	4	-0.833	0.00143
20	sirolimus	-0.384	44	-0.284	0.00144
1	adiphenine	0.821	5	0.937	0
4	viomycin	0.762	4	0.923	0.00004
5	Prestwick-692	0.687	4	0.902	0.0001
6	genistein	0.313	17	0.503	0.0002
7	isoflupredone	0.741	3	0.945	0.00022
8	atractyloside	0.636	5	0.824	0.00032

## Functional enrichment of DE-MRGs

Function annotation analyses of the 315 DE-MRGs were performed. A total of 791 GO terms (including 593 biological processes, 28 cellular components, and 170 molecular functions) were enriched.

The GO analysis includes three categories (biological process, molecular function, and cellular component), and the 10 significant enrichment terms were displayed for each category (Figure 3A). From the biological

process, it was observed that DEGs were predominantly associated with small molecule catabolic process, purine-containing compound metabolic process and nucleoside phosphate biosynthetic

process. For the cellular component, these DEGs were enriched in mitochondrial matrix, organelle outer membrane and outer membrane. In molecular function, these DEGs were associated with coenzyme binding, oxidoreductase activity, acting on CH-OH group of donors and oxidoreductase activity.

Furthermore, analysis of the KEGG pathways of these metabolic genes showed that 32 KEGG pathways were enriched, mainly including purine metabolism, Carbon metabolism, Fatty acid degradation, Tryptophan metabolism and Pyruvate metabolism (Fig. 3B)

## **Construction and validation of the prognostic metabolic gene signature**

Univariate Cox regression analysis and lasso-penalized Cox regression analysis (15) were performed to identify the prognosis related metabolic genes and construct the prognostic gene signature.  $P < 0.01$  in univariate Cox regression analysis were considered statistically significant. LASSO analysis identified six DE-MRGs (NT5E, PAICS, PFKL, PLA2G2D, QPRT and SHMT2) which were included in the classifier. The risk score =  $0.0697 \times \text{expression of NT5E} + 0.0182 \times \text{expression of PAICS} + 0.0220 \times \text{expression of PFKL} - 0.1127 \times \text{expression of PLA2G2D} + 0.0239 \times \text{expression of QPRT} + 0.0068 \times \text{expression of SHMT2}$ . All patients were assigned to high- ( $n = 519$ ) and low-risk ( $n = 520$ ) groups based on median risk scores. The AUC of ROC to the risk score was the best in both training set and testing set and the K-M survival analysis indicated significantly different survival time between the two groups (Figure 4). In both univariate and multivariate Cox regressions, the hazard ratio of risk score was maximal compared with other clinical features. The univariate Cox regression focused on the individual variable but may be affected by the confounding factors. The multivariate Cox regression avoided this limitation. These analyses complemented each other and indicated that the risk score could be a definitely independent risk factor for the prognosis of BC (Figure 5).

## **Constructing and validating a predictive nomogram in the TCGA and GEO cohort**

Nomograms were created for OS based on their common clinical traits. The prognostic nomogram for 3-, 5- and 10-year OS of the TCGA and GEO cohort were shown in Figure 6A,B. In order to identify the discriminating superiority of nomograms, various methods were used in this study, including calibration curves, C-index values and DCA curves. Calibration plots showed that the performance of the nomogram was best in predicting 3-, 5- and 10-year OS (Figure 6C,D). The C-index of the nomogram for the prediction of OS were 0.757 in training group, and 0.728 in validation group. The DCA curves showed some net benefit for predicting OS, especially for 3-, 5- and 10-year survival both in training group and validation group (Figure 6E,F).

## External Validation of the Prognostic Signature.

All these five genes have been confirmed that they were significantly differentially expressed between BC and control tissues in the TIMER2 database, with significant differences in expression (Figure 7), which is consistent with our results. Through the Human Protein Atlas Database (<https://www.proteinatlas.org/>), we compared the protein expression levels in breast cancer and normal tissues (Figure 8A). In the cBioportal database, the mutation rate of QPRT is the highest, 4% in the sample, and other genes also show changes (Figure 8B).

## Gene set enrichment analyses

GSEA was performed and found a great majority of the enriched pathways were metabolism related (such as KEGG\_PYRIMIDINE\_METABOLISM, KEGG\_GALACTOSE\_METABOLISM, KEGG\_PURINE\_METABOLISM, KEGG\_FRUCTOSE\_AND\_MANNANOSE\_METABOLISM, KEGG\_GLUTATHIONE\_METABOLISM) in the high-risk group. Besides, the metabolism-related pathways (such as KEGG\_ETHER\_LIPID\_METABOLISM, KEGG\_NITROGEN\_METABOLISM, KEGG\_GLYCEROPHOSPHOLIPID\_METABOLISM, KEGG\_INOSITOL\_PHOSPHATE\_METABOLISM) were enriched in the low-risk group, based on the criterion of a NOM p-value <0.05 (Figure 9).

## Correlation with Tumor Microenvironment.

Compared with the high-risk group, the Stromal score, Immune score and ESTIMATE score of the samples in the low-risk group were higher, and the tumor purity in the high-risk group was higher with significant differences (Figure 10).

The relationship was calculated between the signature and the level of immune cell infiltration through the CIBERSORT system. A total of 11 subtypes of immune cells (B cells naive, B cells memory, T cells CD8, T cells CD4 memory activated, T cells regulatory (Tregs), T cells gamma delta, NK cells activated, Macrophages M0, Macrophages M2, Dendritic cells activated, Neutrophils) have a clear correlation between infiltration level and risk score (Figure 11).

## Discussion

The incidence of breast cancer increases continuously in the world and has been a serious women's disease (16). Even with current significant progress in breast cancer treatment, 25% to 50% of breast cancers would eventually develop metastasis, leading to poor prognosis (17). The fields of personalized prognosis and precise prediction in medicine are rapidly growing ones (18). In the era of precision medicine, it is vital to include as much prognostic and predictive information as possible for decision-making (19).

In this study, we first performed a differential analysis of the metabolic genes in breast cancer and normal tissues in TCGA and GSE20605, and then performed enrichment analysis, and found DEGs mainly involved in various DNA, protein and fat metabolism processes. Further, through Cmap database, we mainly screened 6 kinds of drugs that may treat breast cancer, among which the first 3 kinds are LY-294002, tanespimycin, apigenin. LY294002 is a PI3K inhibitor that has been used to treat breast cancer in clinical practice. LY294002 has a significant inhibitory effect on the proliferation of TNBC cells, and LY294002 is more effective on TNBC cells lacking BRCA1(20). Tanespimycin is a quinone anti-tumor antibiotic, derived from the anti-tumor antibiotic geldanamycin. Tanespimycin binds to and inhibits the cytoplasmic chaperone function of heat shock protein 90 (HSP90). It has been reported in the literature that Tanespimycin has strong anti-proliferation activity against breast cancer, colorectal cancer and cervical cancer(21) and can reshape the immunosuppressive microenvironment of triple-negative breast cancer, which is conducive to checkpoint blocking immunotherapy(22). Acquired resistance to adriamycin is a major obstacle to triple negative breast cancer (TNBC) treatment, apigenin can enhance the cytotoxic effect of adriamycin on breast cancer cells(23).

It has been reported that the occurrence of tumors depends on the direct and indirect consequences of oncogenic mutations, namely the reprogramming of cell metabolism and has been proposed to be a core hallmark of cancer(24). Therefore, identifying new metabolic genes in cancer to predict the risk of cancer death has become a hot spot.

In this study, we performed LASSO regression analyses and identified a novel six metabolic prognosis-related genes signature including NT5E, PAICS, PFKL, PLA2G2D, QPRT and SHMT2 based on TCGA data set and verified its efficiency through the data downloaded from the GEO database (GSE20685). Among them, PLA2G2D was considered a protective factor while others were risk factors. It has satisfactory robustness and internal predictive ability in both data sets. According to the all patients' risk score results, the low-risk and high-risk patients are effectively stratified. The data from the GEO database further validated the efficacy, indicating that the predictive efficacy of these metabolism-related genes has strong prognostic value. Abnormal metabolic function in the tumor microenvironment may lead to various prognosis of patients, and genes related to metabolism can be used as prognostic indicators for tumors. Thus, we explored the correlation between signatures and TME to expand clinical applications and provide more clues for the choice of treatment strategies.

Previous studies have reported the study of genes in the model. NT5E, 5'-nucleotidase ecto, also known as CD73, is one of the key components of tumor immunosuppressive microenvironment formation. Studies have reported that the high expression of NT5E has a poor prognosis in breast and ovarian cancer, but has a protective effect in lung and gastric cancer. (25). Buisseret L et al reported that the expression of CD73 is related to poor prognosis of human TNBC and low anti-tumor immune function. Targeting CD73 may be a promising strategy to reshape the tumor microenvironment of this subtype(26). CD73 is highly expressed in breast cancer tissues, and increases with the increase of tumor grade and lymph node metastasis. CD73 is highly expressed in more malignant cells, and CD73 overexpression

promotes the proliferation of breast cancer cells in vivo and in vitro. It activates the akt/gsk-3 and  $\beta/\beta$ -catenin/cyclinD1 signaling pathways through mechanisms such as CD73 enzyme activity(27).

PAICS(phosphoribosamidoimidazole carboxylase and phosphoribosamidoimidazole succinylloxycarbonylacetamide synthetase) catalyze two basic steps in the de novo purine biosynthetic pathway and overexpressed in many cancers and may become a promising target for the development of cancer treatments(28). The depletion of PAICS largely eliminates the expansion of breast cancer, exemplifying a prognostic gene related to breast cancer activity(29). Knockdown of PAICS inhibits malignant proliferation of human breast cancer cell lines, These findings demonstrate that PAICS plays an essential role in breast cancer proliferation in vitro, which provides a new opportunity for discovering and identifying novel effective treatment strategies(30).PAICS is a therapeutic target for pancreatic cancer and can be targeted by small molecules(31).

Phosphofructokinase(PFK) is a key regulator of glycolysis and a key control point for glycolytic flux by catalyzing the phosphorylation of fructose 6 phosphate to fructose 1,6 bisphosphate. PFK is a tetramer composed of three subunits M (muscle), L (liver) and P (platelets), which are respectively encoded by three different genes PFKM, PFKL and PFKP (32).PFKL rs2073436C>G can predict the efficacy of NSCLC chemotherapy(33).Study reported A20 is an E3 ubiquitin ligase that interacts with the liver-type phosphofructokinase (PFKL) in liver cancer A20 to promote the degradation of PFKL, thereby inhibiting glycolysis of liver cancer cell lines(34). The PFKL/miR-128 axis regulates glycolysis by inhibiting AKT phosphorylation and predicts the low survival rate of lung cancer patients(35). PFKs are key genes responsible for glycolysis, and their expression is up-regulated in bladder cancer.Targeting this pathway can inhibit the growth of bladder cancer cells(36).

PLA2G2D, a secreted PLA subtype ,Quantitative PCR analysis of bovine PLA2G2D transcripts showed that their expression levels in breast samples were different between the dry period and lactation period, and their expression in liver tissue was polymorphic(37). Recent studies have shown that secreted phospholipase A2 (sPLA2s) represents an attractive potential tumor biomarker and a therapeutic target for various cancers, and the expression of PLA2G2D is reduced by up to 23 times. May be an ideal candidate for new biomarkers that influence colon cancer(38). In the context of cancer, the lack of PI2g2d leads to a significant reduction in the occurrence of skin cancer, which may be due to enhanced anti-tumor immunity. In summary, these results emphasize the general role of sPLA2-IIID as an immunosuppressive sPLA2, which allows the lipid balance of the microenvironment to enter an anti-inflammatory state, and exert beneficial or harmful effects according to different pathophysiological backgrounds in inflammation and cancer(39).

Quinolate phosphoribosyltransferase (QPRT) is a key enzyme in the de novo synthesis of NAD.Compared with normal tissues, DCTPP1 and QPRT are highly expressed in BC. The overexpression of DCTPP1 and QPRT is related to the slow progress of BC, which promotes the growth, migration and invasion of MCF7 and T47D cells, but inhibits cell apoptosis DSCAM-AS1 gene knockout reduced the expression of DCTPP1 and QPRT, and inhibited the growth, migration and invasion of estrogen receptor-

positive BC(40). Phosphoribosyl transferase (QPRT) is the terminal rate-limiting enzyme in KP Kynurenine pathway ,and its expression is down-regulated, Widespread dysregulation of KP in renal cell carcinoma is common and may lead to tumor immune escape, which is of great significance for effective targeted therapy of this key pathway(41).

Serine hydroxymethyltransferase 2 (SHMT2) is a protein-coding gene that regulates the production of glycine in mitochondria, which is an important intermediate in purine biosynthesis, It is a valuable marker in a variety of cancers, including intrahepatic cholangiocarcinoma, large B-cell lymphoma,Kidney Cancer and breast cancer(42).Studies have shown that SHMT2 expression is an independent prognostic factor for breast cancer(43).

GSEA analysis revealed that the difference in enrichment pathways between high-risk and low-risk groups.The results show that the two risk groups have significantly different metabolic characteristics.The pathway in the high-risk group is mainly related to DNA synthesis metabolism, and the pathway in the low-risk group is related to lipid metabolism.

Changes in tumor cell metabolism have caused a very large impact on the tumor microenvironment, causing corresponding changes in the pH value and metabolites of the microenvironment.In this study,we have not only clarified the metabolic characteristics and mechanisms under the conditions of different risk stratifications, but also predicted the metabolic microenvironment and prognostic characteristics through different immune scoring methods. Correlation analysis shows that our metabolic signature is statistically correlated with approximately half of immune cells,therefore,we can monitor the infiltration composition of immune cells and the degree of immune response through metabolic characteristics.Our metabolic signature can reflect the changes in TME from different aspects,and may provide clues for explaining immunological resistance and immunotherapy applications.

Our study has several limitations.Firstly,the included metabolic genes were from the TCGA and GSE20685 databases, and some metabolic genes may be missed.Secondly,although our signatures had been validated in multiple databases,the experiments are still needed to further confirm its clinical application value.Thirdly,our data was based on transcriptomics data, it was difficult to reflect the overall landscape of tumor metabolism.

## Conclusions

Based on the TCGA data set,we determined a new six-gene metabolic gene signature for BC prognosis prediction.Our signature may reflect the dysregulation of the metabolic microenvironment and provide potential biological markers for metabolic therapy and treatment response prediction.However,more functional experiments are needed to validate our signature.

## Declarations

## Data Availability

The datasets used and/or analyzed during the current study are available in The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>) and GSE20605 dataset from the Gene Expression Omnibus (GEO) database(<https://www.ncbi.nlm.nih.gov/geo/>). Other data comes from the online databases(TIMER database( <https://cistrome.shinyapps.io/timer/>),The cBioportal database (<https://www.cbioportal.org/>),the Human Protein Atlas database (<https://www.proteinatlas.org/>)).

## Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Funding

This study was funded by the Tai'an Science and Technology Development Plan (Guiding Plan) Fund(No.2018NS0222)

## Author Contribution

Z.W.Z. and G.W.W. designed the experiments.J.L. wrote this manuscript. J.L.,F.L. and Q.T. analyzed data.B.J.L.,G.W.W and Z.W.Z. reviewed and revised the manuscript.All authors read and approved the final manuscript.

## Abbreviations

TCGA,The Cancer Genome Atlas;BC,Breast cancer;TME,Tumor microenvironment;MRGs,metabolic-related genes;DE-MRGs,differentially expressed MRGs;AUC, Area under the curves;DCA, Decision curve analysis;FC,Fold change;FDR, False discovery rate;GEO, Gene Expression Omnibus;GSEA,Gene set enrichment analysis;KEGG, Kyoto Encyclopedia of Genes and Genomes;K-M, Kaplan–Meier;LASSO,Least absolute shrinkage and selection operator;ROC,Receiver operator characteristic.

## References

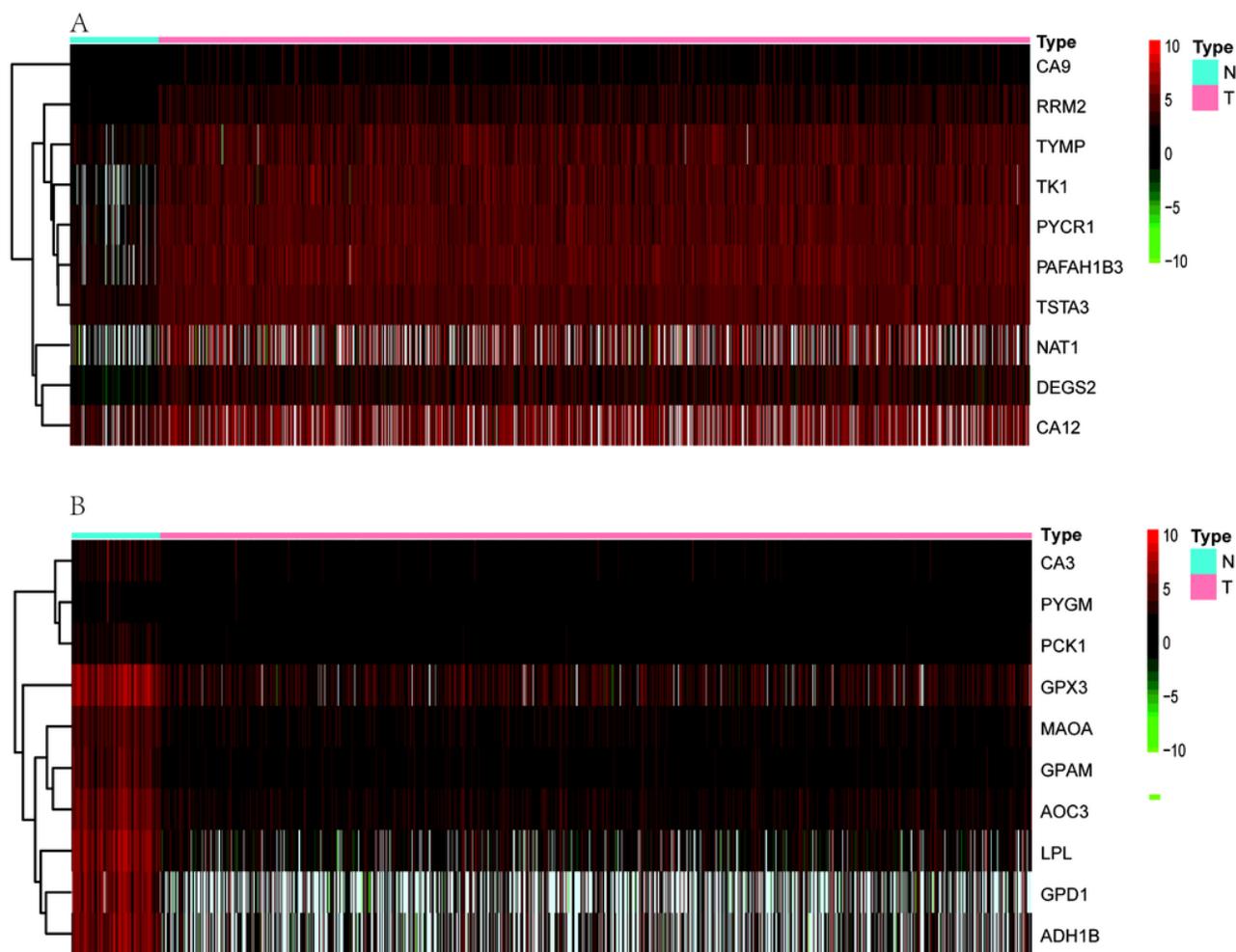
1. Freddie, Bray, Jacques, et al.: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. 2018.
2. Momenimovahed Z and Salehiniya H: Epidemiological characteristics of and risk factors for breast cancer in the world. Breast Cancer (Dove Med Press) 11: 151-164, 2019.
3. Goldhirsch A, Winer EP, Coates AS, et al.: Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early

- Breast Cancer 2011. 22: 1736-1747, 2011.
4. Koo TR, Eom KY, Kang EY, et al.: Prognostic value of the nodal ratio and ki-67 expression in breast cancer patients treated with postmastectomy radiotherapy. *J Breast Cancer* 16: 274-284, 2013.
  5. Yoshida, Go JJJoE and Cr CCR: Metabolic reprogramming: the emerging concept and associated therapeutic strategies. 34: 1-10, 2015.
  6. Douglas, Hanahan, and, Robert A and Cell WJ: Hallmarks of Cancer: The Next Generation. 2011.
  7. Lindsey A, Phillip K, Haik M, Christie-Lynn M and Michael DJe, 6,: Glycolysis and glutaminolysis cooperatively control T cell function by limiting metabolite supply to N-glycosylation. 62017.
  8. Leek JTJNAR: svaseq: removing batch effects and other unwanted noise from sequencing data. 0-0, 2014.
  9. Ritchie ME, Phipson B, Wu D, et al.: limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* 43: e47, 2015.
  10. Yu G, Wang LG, Han Y and He QYJO-aJoIB: clusterProfiler: an R package for comparing biological themes among gene clusters. 16: 284-287, 2012.
  11. Friedman JH, Hastie T and Tibshirani RJJJoSS: Regularization Paths for Generalized Linear Models via Coordinate Descent. 332010.
  12. Subramanian A, Tamayo P, Mootha V, et al.: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. 102: 15545-15550, 2005.
  13. Yoshihara K, Shahmoradgoli M, Martínez E, et al.: Inferring tumour purity and stromal and immune cell admixture from expression data. 42013.
  14. Newman AM, Liu CL, Green MR, et al.: Robust enumeration of cell subsets from tissue expression profiles. 12: 453-457, 2015.
  15. Tibshirani RJSiM: The lasso method for variable selection in the Cox model. 16: 385-395, 1997.
  16. Wang CCN, Li CY, Cai J-H, et al.: Identification of Prognostic Candidate Genes in Breast Cancer by Integrated Bioinformatic Analysis. *J Clin Med* 8: 1160, 2019.
  17. Chiang K-C, Yeh T-S, Chen S-C, et al.: The Vitamin D Analog, MART-10, Attenuates Triple Negative Breast Cancer Cells Metastatic Potential. *Int J Mol Sci* 17: 606, 2016.
  18. Ow GS and Kuznetsov VA: Big genomics and clinical data analytics strategies for precision cancer prognosis. *Scientific reports* 6: 36493-36493, 2016.
  19. Cheng B, Wang C, Zou B, et al.: A nomogram to predict outcomes of lung cancer patients after pneumonectomy based on 47 indicators. *Cancer Med* 9: 1430-1440, 2020.
  20. Guney Eskiler GJJop and pharmaceutical sciences : a publication of the Canadian Society for Pharmaceutical Sciences Scdsp: The Interaction of PI3K Inhibition with Homologous Recombination Repair in Triple Negative Breast Cancer Cells. 22: 599-611, 2019.
  21. El-Shafey H, Gomaa R, El-Messery S, Goda FJB and letters mc: Quinazoline Based HSP90 Inhibitors: Synthesis, Modeling Study and ADME Calculations Towards Breast Cancer Targeting. 30: 127281, 2020.

22. Liu Y, Qiu N, Shen L, et al.: Nanocarrier-mediated immunogenic chemotherapy for triple negative breast cancer. 323: 431-441, 2020.
23. Korga-Plewko A, Michalczyk M, Adamczuk G, et al.: Apigenin and Hesperidin Downregulate DNA Repair Genes in MCF-7 Breast Cancer Cells and Augment Doxorubicin Toxicity. 252020.
24. Jang M, Kim S, Lee JJE and Medicine M: Cancer cell metabolism: implications for therapeutic targets. 452013.
25. Jiang T, Xu X, Qiao M, et al.: Comprehensive evaluation of NT5E/CD73 expression and its prognostic significance in distinct types of cancers. 18: 267, 2018.
26. Buisseret L, Pommey S, Allard B, et al.: Clinical significance of CD73 in triple-negative breast cancer: multiplex analysis of a phase III clinical trial. 29: 1056-1062, 2018.
27. Yu J, Wang X, Lu Q, et al.: Extracellular 5'-nucleotidase (CD73) promotes human breast cancer cells growth through AKT/GSK-3 $\beta$ / $\beta$ -catenin/cyclinD1 signaling pathway. 142: 959-967, 2018.
28. Škerlová J, Unterlass J, Göttmann M, et al.: Crystal structures of human PAICS reveal substrate and product binding of an emerging cancer target. 295: 11656-11668, 2020.
29. Gallenne T, Ross K, Visser N, et al.: Systematic functional perturbations uncover a prognostic genetic network driving human breast cancer. 8: 20572-20587, 2017.
30. Meng M, Chen Y, Jia J, Li L and Yang SJB: Knockdown of PAICS inhibits malignant proliferation of human breast cancer cell lines. 51: 24, 2018.
31. Agarwal S, Chakravarthi B, Kim H, et al.: PAICS, a De Novo Purine Biosynthetic Enzyme, Is Overexpressed in Pancreatic Cancer and Is Involved in Its Progression. 13: 100776, 2020.
32. Wegener G and Krause UJBST: Different modes of activating phosphofructokinase, a key regulatory enzyme of glycolysis, in working vertebrate muscle. 30: 264, 2001.
33. Choi S, Jin C, Do S, et al.: Polymorphisms in Glycolysis-Related Genes Are Associated with Clinical Outcomes of Paclitaxel-Cisplatin Chemotherapy in Non-Small Cell Lung Cancer. 98: 468-477, 2020.
34. Feng Y, Zhang Y, Cai Y, et al.: A20 targets PFKL and glycolysis to inhibit the progression of hepatocellular carcinoma. 11: 89, 2020.
35. Yang J, Li J, Le Y, Zhou C, Zhang S and Gong ZJAJocr: PFKL/miR-128 axis regulates glycolysis by inhibiting AKT phosphorylation and predicts poor survival in lung cancer. 6: 473-485, 2016.
36. Sun C, Xiong D, Yan Y, Geng J, Liu M and Yao XJTljobm: Genetic alteration in phosphofructokinase family promotes growth of muscle-invasive bladder cancer. 31: e286-293, 2016.
37. Golik M, Cohen-Zinder M, Loor J, et al.: Accelerated expansion of group IID-like phospholipase A2 genes in *Bos taurus*. 87: 527-533, 2006.
38. Mounier C, Wendum D, Greenspan E, Fléjou J, Rosenberg D and Lambeau GJBjoc: Distinct expression pattern of the full set of secreted phospholipases A2 in human colorectal adenocarcinomas: sPLA2-III as a biomarker candidate. 98: 587-595, 2008.
39. Miki Y, Kidoguchi Y, Sato M, et al.: Dual Roles of Group IID Phospholipase A2 in Inflammation and Cancer. 291: 15588-15601, 2016.

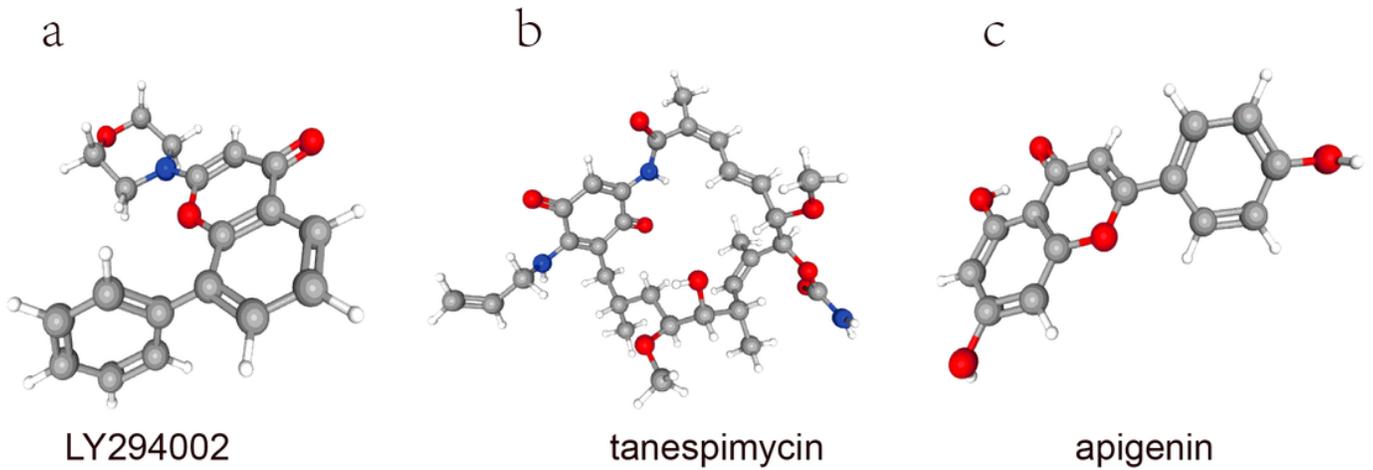
40. Yue Z, Shusheng J, Hongtao S, et al.: Silencing DSCAM-AS1 suppresses the growth and invasion of ER-positive breast cancer cells by downregulating both DCTPP1 and QPRT. 12: 14754-14774, 2020.
41. Hornigold N, Dunn K, Craven R, et al.: Dysregulation at multiple points of the kynurenine pathway is a ubiquitous feature of renal cancer: implications for tumour immune evasion. 123: 137-147, 2020.
42. Wang H, Chong T, Li B, Chen X and Zhen WJBr: Evaluating the clinical significance of SHMT2 and its co-expressed gene in human kidney cancer. 53: 46, 2020.
43. Bernhardt S, Bayerlová M, Vetter M, et al.: Proteomic profiling of breast cancer metabolism identifies SHMT2 and ASCT2 as prognostic factors. 19: 112, 2017.

## Figures



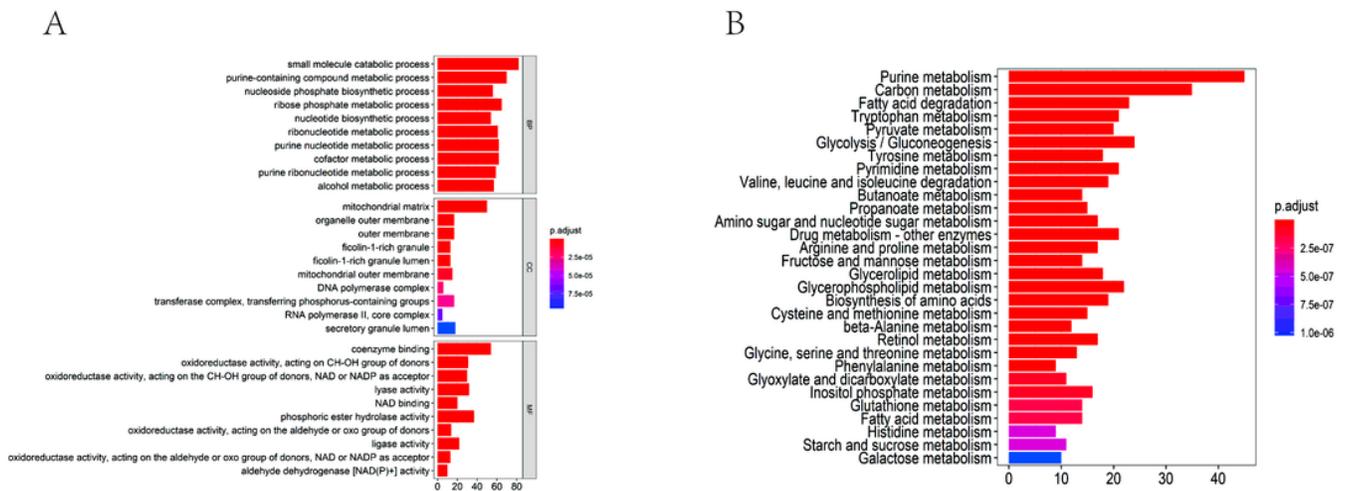
**Figure 1**

The heatmaps of differentially expressed genes. A, The heatmap of the top 10 up-regulated genes. B, The heatmap of the top 10 down-regulated genes



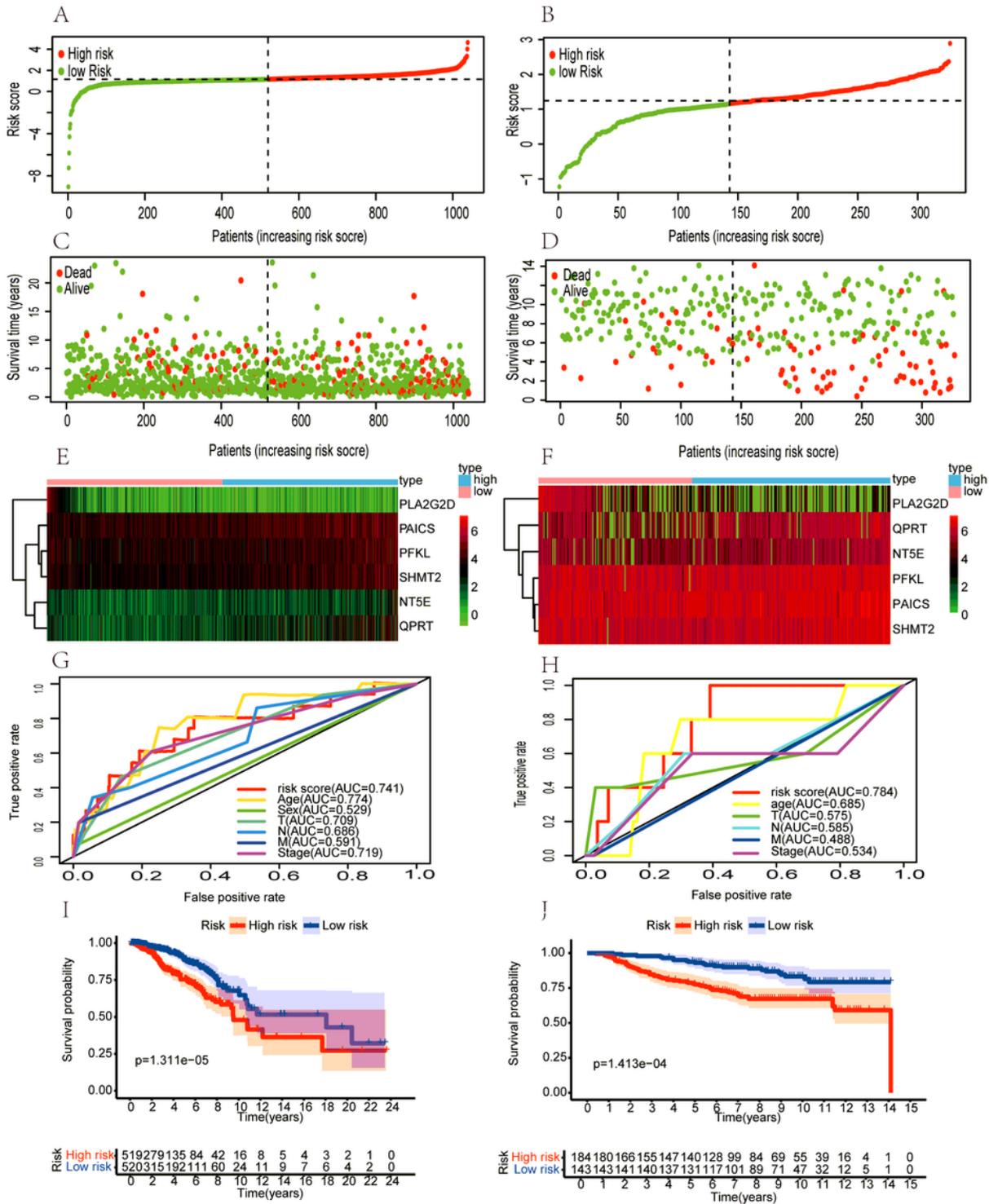
**Figure 2**

Top 3 molecule drugs. a LY294002, b tanespimycin, c apigenin



**Figure 3**

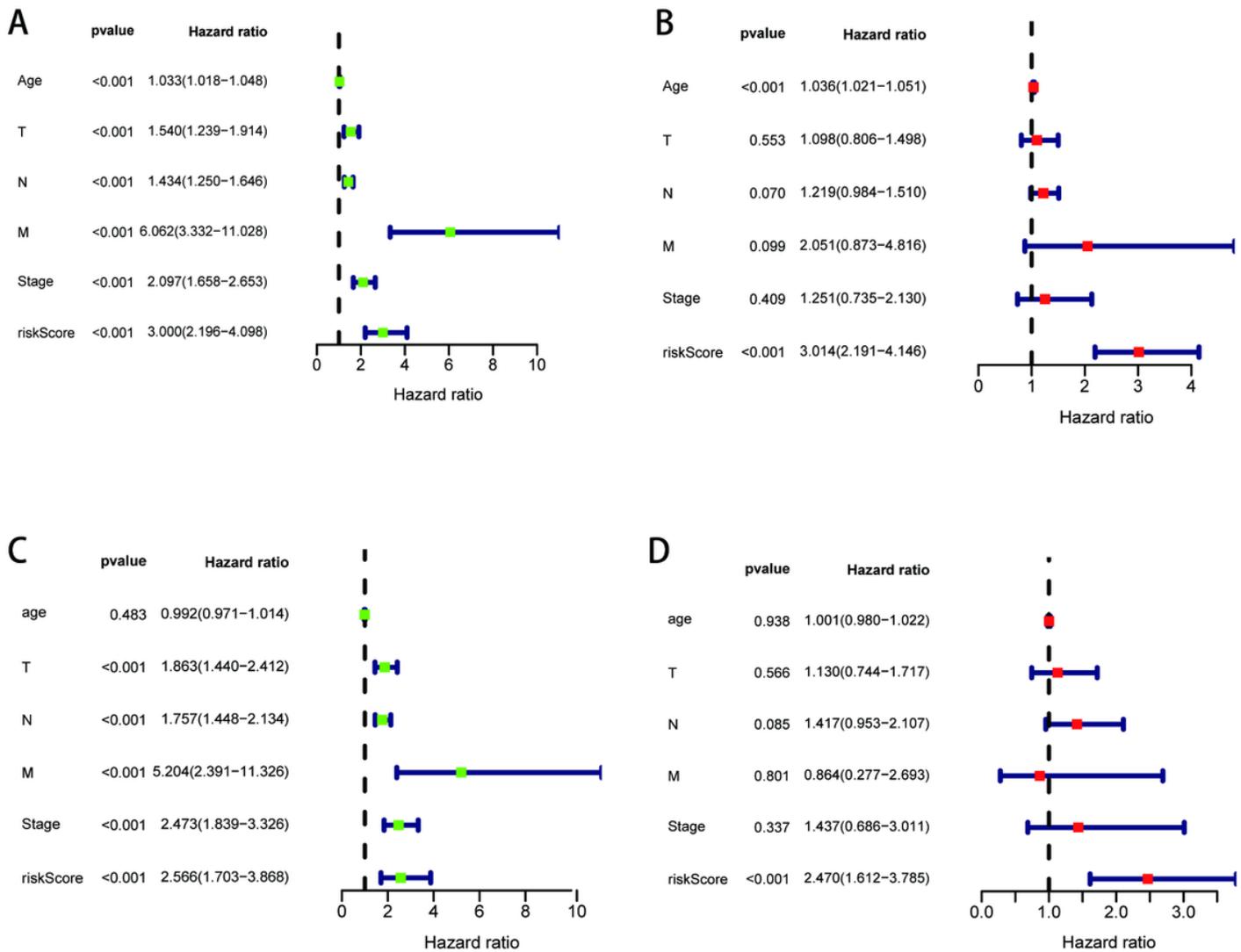
Gene Ontology (GO) and KEGG enrichment of DE-MRGs. A, GO analysis shows the biological processes, cellular component, and molecular functions involved in differential genes. b KEGG analysis shows significantly enriched pathways of DE-MRGs.



**Figure 4**

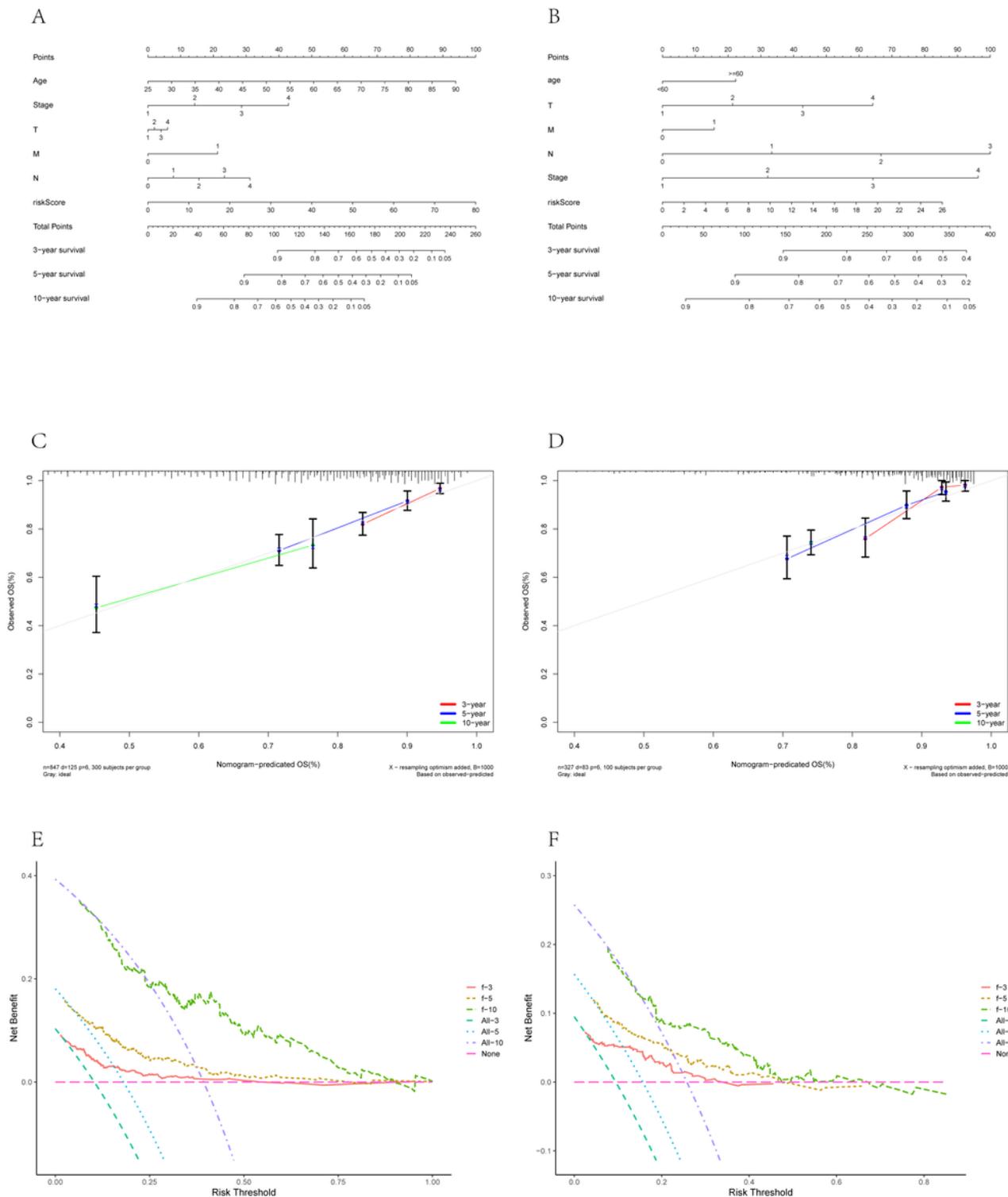
The performance of the signature in training set and testing set. The samples were classified into high- and low-risk group according to the median of risk score in the TCGA cohort. (A, B) The curve of risk score which represented the distribution of patients. (C, D) The dot plot which represented the survival status of patients. (E, F) The heatmap which represented the expression profiling of the 6 genes in the whole dataset. (G, H) The comparison of the AUC to the clinical features and risk scores. (I, J) Kaplan–Meier

survival curves for the prediction of prognostic outcomes based on the metabolic gene signature. In both training set and testing set, significant differences were observed in high- and low-risk groups. TCGA cohort (A,C,E,J,I). GEO cohort(B,D,F,H,J).



**Figure 5**

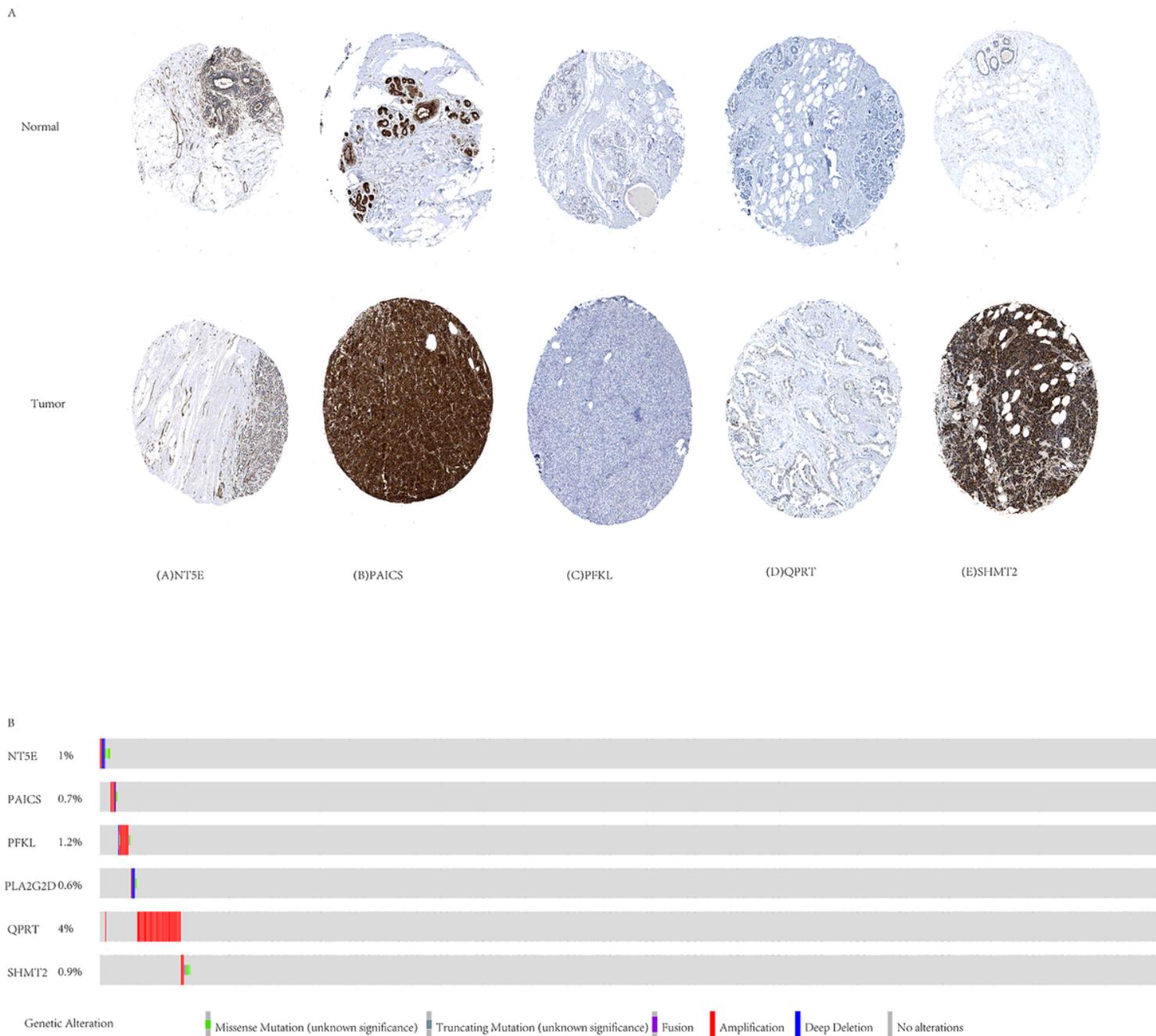
The forest plots of the univariate and multivariate Cox regression analysis in BC. The univariate Cox regression in the (A) TCGA cohort and (C) GEO cohort. The multivariate Cox regression in the (B) TCGA cohort and (D) GEO cohort.



**Figure 6**

Building and validation of the nomogram predicting overall survival for BC patients in the TCGA cohort and GSE20685 cohort. (A,B) The nomogram plot was built based on risk score, clinical parameters, and pathological stages for 3-, 5-, and 10-year overall survival in BC. (C,D) The calibration plot for internal validation of the nomogram for 3-, 5-, and 10-year overall survival in BC. (E,F) The DCA curves of the





**Figure 8**

Expression and genetic alterations of the six predictive genes. (A) The representative protein expression of the six genes in BC and normal liver tissue. Data were from the Human Protein Atlas (<http://www.proteinatlas.org>) database. Data of PLA2G2D were not found in the database. (B) Genetic alterations of the six genes in TCGA-BRCA. Data were from the cBioportal for Cancer Genomics (<http://www.cbioportal.org/>).

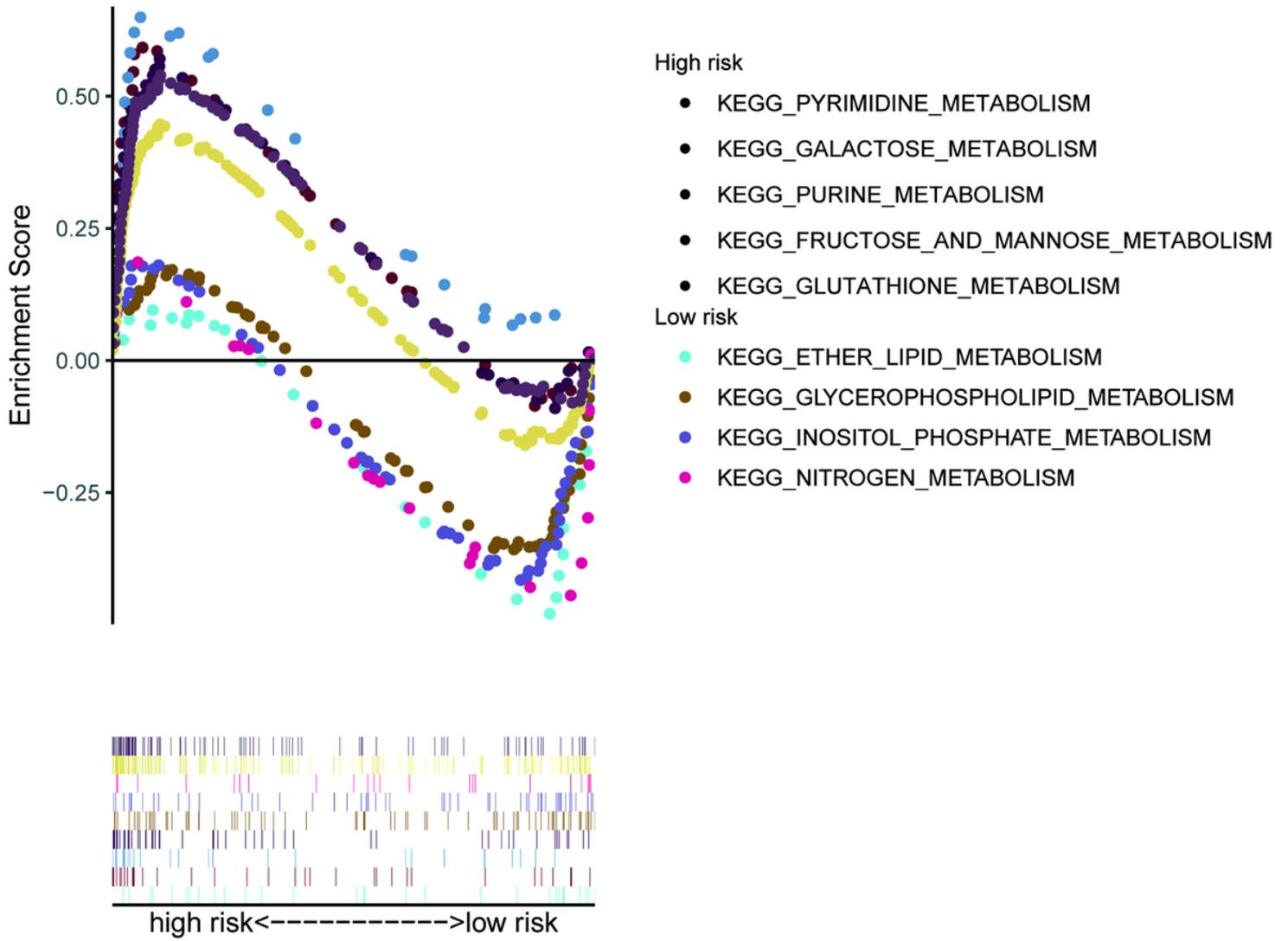


Figure 9

The top 5 significant enriched KEGG pathways in the TCGA cohort.

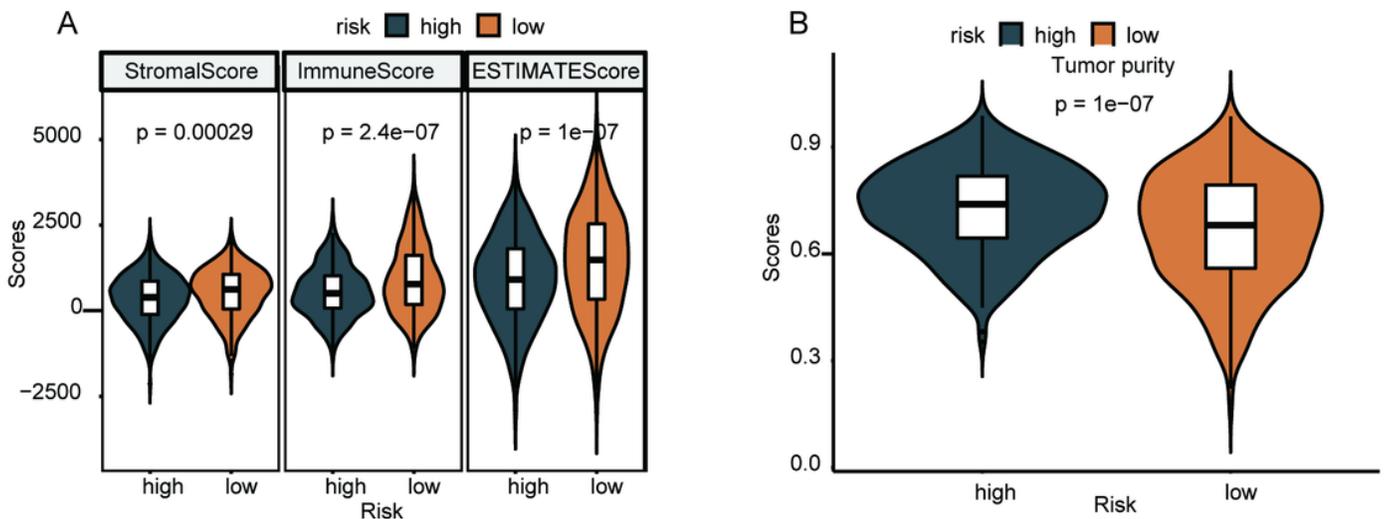
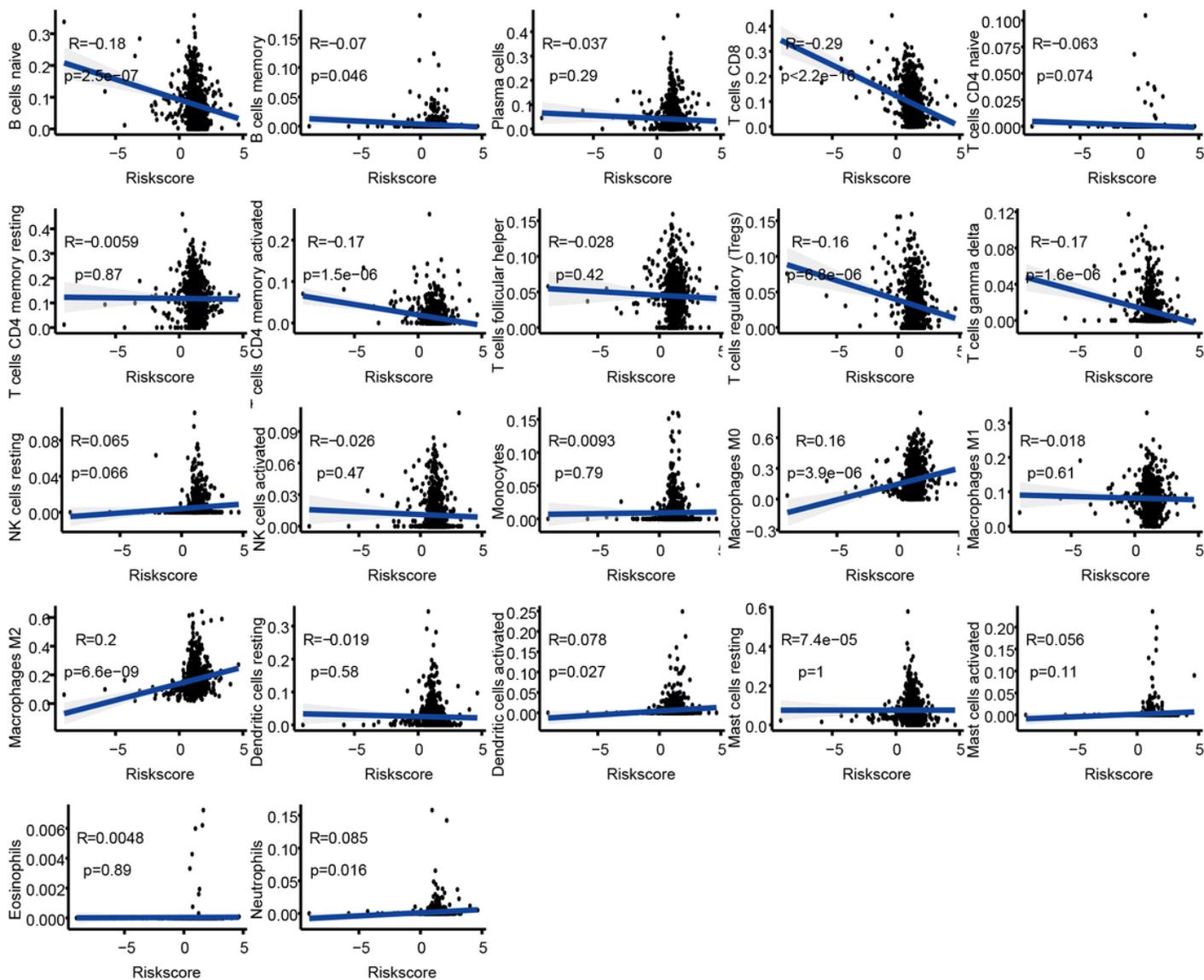


Figure 10

The comparison of ESTIMATE scores. The green part represented the low-risk group, and the yellow part represented the high-risk group. (A) The infiltration level of the immune cells, the stromal content, and the total scores. (b) The estimated tumor purity



**Figure 11**

The relationships between the metabolic signature and the infiltration levels of 22 subtypes of immune cells calculated by the CIBERSORT system.