

Original Article

Development of a metabolism-related signature for predicting prognosis, immune infiltration and immunotherapy response in breast cancer

Chunzhen Li^{1*}, Yijie Tao^{1*}, Yining Chen^{2*}, Yunyang Wu¹, Yixian He¹, Shulei Yin¹, Sheng Xu¹, Yizhi Yu¹

¹National Key Laboratory of Medical Immunology and Institute of Immunology, Naval Medical University, Shanghai 200433, China; ²Faculty of Health Sciences and Engineering, University of Shanghai for Science and Technology, Shanghai 200433, China. *Equal contributors and co-first authors.

Received September 23, 2022; Accepted November 27, 2022; Epub December 15, 2022; Published December 30, 2022

Abstract: Breast cancer (BRCA) is the most commonly diagnosed cancer and among the top causes of cancer deaths globally. The abnormality of the metabolic process is an important characteristic that distinguishes cancer cells from normal cells. Currently, there are few metabolic molecular models to evaluate the prognosis and treatment response of BRCA patients. By analyzing RNA-seq data of BRCA samples from public databases via bioinformatic approaches, we developed a prognostic signature based on seven metabolic genes (PLA2G2D, GNP-NAT1, QPRT, SHMT2, PAICS, NT5E and PLPP2). Low-risk patients showed better overall survival in all five cohorts (TCGA cohort, two external validation cohorts and two internal validation cohorts). There was a higher proportion of tumor-infiltrating CD8⁺ T cells, CD4⁺ memory resting T cells, gamma delta T cells and resting dendritic cells and a lower proportion of M0 and M2 macrophages in the low-risk group. Low-risk patients also showed higher ESTIMATE scores, higher immune function scores, higher Immunophenoscores (IPS) and checkpoint expression, lower stemness scores, lower TIDE (Tumor Immune Dysfunction and Exclusion) scores and IC50 values for several chemotherapeutic agents, suggesting that low-risk patients could respond more favorably to immunotherapy and chemotherapy. Two real-world patient cohorts receiving anti-PD-1 therapy were applied for validating the predictive results. Molecular subtypes identified based on these seven genes also showed different immune characteristics. Immunohistochemical data obtained from the human protein atlas database demonstrated the protein expression of signature genes. This research may contribute to the identification of metabolic targets for BRCA and the optimization of risk stratification and personalized treatment for BRCA patients.

Keywords: Metabolism-related genes, prognostic signature, breast cancer, immune infiltration, bioinformatics methods

Introduction

Breast cancer (BRCA) is a highly malignant female tumor and its incidence ranked first among cancers in 2020 worldwide [1]. Traditionally, breast cancer can be classified into Luminal A, Luminal B, HER2 (human epidermal growth factor receptor 2)-enriched and Basal-like type according to the status of specific receptors. The classification helps to assess the pathological features and prognosis of patients more accurately and thus adopt more precise treatments [2]. Although clinical treatment methods including surgery, hormonal therapy, adjuvant chemotherapy and radiother-

apy for breast cancer have made great progress in recent years, problems such as recurrence, metastasis, drug resistance and low response rates to immunotherapy still limit the further improvement of the survival rate of BRCA patients [3]. Therefore, it is of great significance to find novel molecular targets with prognostic and therapeutic value to help with the stratification, prognosis evaluation and individualized treatment of patients.

With advances in high-throughput sequencing, multi-omics and spatial biology, researchers now master more tools to discover key molecules regulating tumor heterogeneity and micro-

A metabolism-related prognostic signature for breast cancer patients

environment. Metabolic reprogramming is one of the characteristic biological processes of cancer cells [4]. Various metabolic processes, such as glycolysis, the tricarboxylic acid cycle, amino acid, nucleotide and lipid metabolism, are altered to varying degrees compared with normal cells. Oncogenes play a role in promoting metabolic dysregulation, which in turn provides favorable conditions for the survival of cancer cells [5]. Metabolites could also directly or indirectly affect the signal transduction involved in the growth and migration of cancer cells. Changes in nucleotide metabolism were shown to be correlated with chemotherapeutic drug resistance and the progression of breast cancer [6].

In addition, altered tumor metabolism interacts with other cells in the tumor microenvironment (TME), such as fibroblasts and immune cells, and shapes an environment conducive to the proliferation and immune evasion of tumors [7]. More and more studies have focused on the interaction between the metabolic processes of cancer cells and their microenvironment. For example, enhanced glycolytic activity in cancer cells produces large amounts of lactate and acidifies the microenvironment, which can inhibit the function of effector T cells by affecting their proliferation and IFN- γ production [8]. Cancer cells could also compete for glutamine with immune cells. On the one hand, limited glutamine inhibited immune cell proliferation; on the other hand, low glutamine level in the microenvironment could also promote the transformation of CD4⁺ T cells into Tregs, shaping the immunosuppressive environment, accelerating the exhaustion of effector T cells and undermining the effectiveness of immunotherapy [9, 10]. In breast cancer models, targeting the glutamate metabolism could also facilitate M1-like polarization of macrophages and reduce the recruitment of myeloid-derived suppressor cells (MDSC) [11]. And inhibiting the glycolysis could reduce the expression of granulocyte-macrophage colony-stimulating factor (GM-CSF) and tumor granulocyte colony-stimulating factor (G-CSF), decrease MDSC infiltration, and further enhance the anti-tumor immunity [12]. Additionally, a glycolytic molecular subtype of triple-negative breast cancer (TNBC) characterized by upregulation of carbohydrate and nucleotide metabolism was newly identi-

fied, and inhibition of LDH in this subtype significantly improved the efficacy of anti-PD-1 immunotherapy. The role of altered metabolism on the tumor microenvironment and anti-tumor immunity is still gaining considerable attention [13]. However, metabolic biomarkers or molecular models for predicting the immunotherapy response of breast cancer patients are still lacking. Therefore, targeting the metabolic process of cancer cells and studying the interaction with components of the microenvironment, such as immune cells, is of great significance to help us discover key molecules with prognostic and therapeutic value and further improve the effectiveness of immunotherapy.

In this study, we established a robust prognostic signature based on 7 metabolism-related genes for breast cancer, which performed well in stratifying the risk of patients, distinguishing the characteristics of tumor immune microenvironment and immune function, and predicting the sensitivity to immunotherapy and chemotherapy. These results may help provide insights into the discovery of novel molecular biomarkers for breast cancer, as well as the selection of individualized treatment for breast cancer patients.

Methods

Data acquisition and preprocessing

The general workflow of this study is shown in **Figure 1**. The RNA sequencing data and corresponding clinical information of female patients with breast cancer were downloaded from The Cancer Genome Atlas database (TCGA, <https://portal.gdc.cancer.gov>). We obtained 113 normal tissue samples and 1109 tumor samples from the TCGA database for differential analysis. 1023 tumor samples were used for subsequent prognostic analysis because 86 patients with less than 60 days of follow-up were excluded. Then we used the TCGA cohort (N = 1023) as a training cohort, from which we also randomly selected two groups of patients (2:1 ratio, N = 682 and N = 341) as internal validation cohorts. Two expression array datasets of BRCA patients serving as external validation cohorts, GSE21653 (N = 241) and GSE20685 (N = 327) and their corresponding clinical information were obtained from the Gene Ex-

A metabolism-related prognostic signature for breast cancer patients

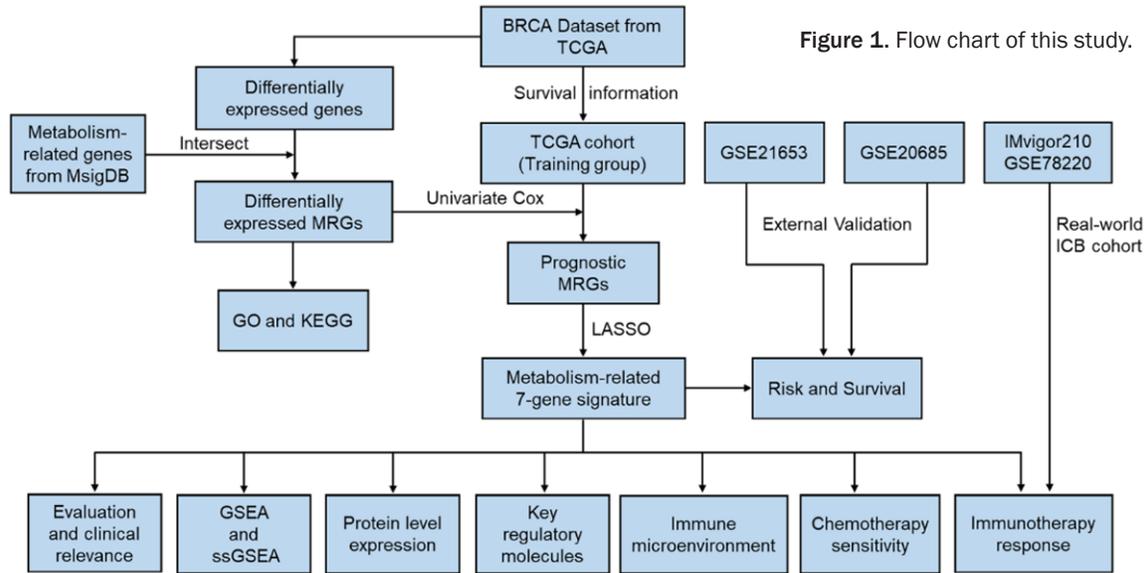


Figure 1. Flow chart of this study.

pression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>). When evaluating the prognostic value and clinical relevance of the signature, samples with incomplete TNM staging data were also excluded.

RNA-sequencing and clinical response data of a real-world cohort (IMvigor210, $n = 348$) of patients with urothelial carcinoma receiving anti-PD-1 therapy were obtained from the R package “IMvigor210CoreBiologies” [14]. Dataset GSE78220 ($n = 28$) containing transcriptome and response data of a cohort of metastatic melanoma patients undergoing anti-PD-1 therapy was downloaded from the GEO database, and one sample with incomplete clinical information was excluded [15]. These two cohorts were applied for validating the role of the signature in predicting the immunotherapy response.

Identification of metabolism-related genes and development of the metabolism-related signature

We used the Perl script to extract the metabolism-related genes (MRGs) from the pathway dataset “c2.cp.kegg.v7.4.symbols.gmt”, which was downloaded from the MsigDB (<https://www.gsea-msigdb.org>). Differentially expressed metabolic genes were screened using the R package “Limma”, with parameters set as $|\log_2(\text{FoldChange})| > 0.585$ and $\text{FDR} < 0.01$

[16]. Then the R package “ClusterProfiler” was applied to carry out the Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses to explore the function of these differentially expressed MRGs [17].

Univariate Cox regression analysis was performed to screen the prognostic MRGs. Differentially expressed MRGs satisfying $P < 0.01$ would be considered as prognostic MRGs. Next, the least absolute shrinkage and selection operator (LASSO) regression analysis was applied to control the number of prognostic MRGs and prevent the overfitting using the R package “glmnet”. Finally, we adopted multivariate Cox regression analysis to confirm the MRGs for model construction and output their hazard values and coefficients. Then the risk score of each patient could be calculated using this formula:

$$\text{Risk Score} = \sum_{i=1}^n \text{Exp}(i) \times \text{Coef}(i)$$

(N represents the number of MRGs used for signature construction. $\text{Exp}(i)$ and $\text{Coef}(i)$ represent the expression level and coefficient of each gene, respectively).

Patients in the TCGA cohort (the training cohort) were divided into high- and low-risk groups based on the median risk score. We used the R package “survival” to draw the Kaplan-Meier (K-M) curve to show differences in survival

A metabolism-related prognostic signature for breast cancer patients

between high- and low-risk groups. “Pheatmap” package was applied to visualize the patient survival status, risk score distribution and gene expression in the risk group. Time-dependent receiver operating characteristic (ROC) curves were drawn using the R package “survivalROC” to evaluate the predictive efficiency of the signature. We utilized two external independent cohorts (GSE21653 cohort and GSE20685 cohort) and two internal cohorts to validate the prognostic signature. Furthermore, univariate and multivariate regression analyses were performed to assess the independence of the signature as well as other clinical factors in predicting the overall survival of patients.

Protein expression of signature MRGs

Apart from analyses at the RNA level, we also explored protein abundance of the signature MRGs in normal breast and BRCA tissues through immunohistochemical data obtained from the HPA (The Human Protein Atlas) database (<https://www.proteinatlas.org/>) [18]. Immunohistochemical images of samples with well-defined tissue and using the same antibody may be selected.

Gene set enrichment analysis

Firstly, we applied the R package “Limma” to obtain the differentially expressed genes between the high-risk and low-risk groups. Then the gene set enrichment analysis was performed using the “org.Hs.eg.db” and “clusterProfiler” packages to explore the biological functional differences between the high-risk and low-risk tumors. The top 5 enriched GO terms or KEGG pathways were visualized.

Evaluation of tumor-infiltrating immune cells (TIICs) and immune function status

Multiple algorithms for estimating the tumor-infiltrating immune cells were used in this section. We obtained the immune cell infiltration data of breast cancer samples in TCGA using the TIMER database (<http://timer.cistrome.org/>) and analyzed the correlation between our risk score and the abundance of tumor-infiltrating cells estimated by three different algorithms (xCELL, MCPcounter and CIBERSORT) [19]. Then we used box plots to show differences in TIICs estimated by CIBERSORT between

the high- and low-risk groups. The violin plot was used to visualize the differences in micro-environment scores between the two groups. By running the “GSVA” package, we analyzed the characteristics of immune function in different risk tumors.

Correlation between the signature and clinicopathological variables

To evaluate the relationship between our risk signature and other clinicopathological indicators, we adopted the “ComplexHeatmap” package to show the distribution of patients with different clinicopathological features between risk groups. We also divided patients into subgroups based on different clinicopathological information to further confirm whether the signature is still valuable for each subgroup. The expression levels of MRGs used for signature construction in each clinicopathological subgroup were also presented via R software.

Evaluation of patient response to immunotherapy and chemotherapy

Firstly, we assessed the expression levels of common immune checkpoints and some costimulatory or coinhibitory molecules in two risk groups. Then we utilized two approaches, Immunophenoscore (IPS) and TIDE (Tumor Immune Dysfunction and Exclusion) algorithms, to analyze the immunogenicity and immunoreactivity characteristics of tumors, and further explored the performance of the risk signature in predicting the response of patients to ICB therapy. IPS scores of patients were downloaded from the TCIA (The Cancer Immunome Atlas) database (<https://tcia.at/home>) and TIDE scores were calculated using the TIDE database (<http://tide.dfci.harvard.edu/>) [20]. After predictive analyses, two real-world cohorts receiving anti-PD-1 therapy, IMvigor210 and GSE78220, were applied to validate the performance of the signature. The responses to chemotherapeutic or targeted drugs including methotrexate, lapatinib and others were predicted and analyzed based on the GDSC database (the Genomics of Drug Sensitivity in Cancer, <https://www.cancerrxgene.org/>). Then the half-maximal inhibitory concentration (IC50) values were compared using the R package “pRRophetic” [21].

A metabolism-related prognostic signature for breast cancer patients

Molecular subtype exploration

To further study whether breast cancer tumors could be classified based on these signature MRGs, we performed the “ConsensusClusterPlus” algorithm on R software. In addition, differences between subgroups in overall survival, immune microenvironment, immune function, and immunotherapeutic response were also investigated using the same method as above.

Statistical analysis

Data processing and analyses were implemented by R (version: 4.0.3) and Perl software (version: 5.26.1). Multiple R packages (limma, survival, GSVA, ClusterProfiler, ggplot2, etc.) were adopted in this study. A subset of data such as IPS scores and TIDE scores were calculated by online databases and further processed and visualized by R software. The Student's t-test was used to evaluate continuous variables, whereas the χ^2 test was used to compare categorical variables. For all statistical analyses, a *p*-value less than 0.05 was regarded as statistical significance.

Results

Screening and protein expression analysis of prognostic MRGs for signature construction

RNA-seq data containing 113 normal samples and 1109 breast cancer samples were downloaded from the TCGA database. After intersecting with metabolism-related genes (MRGs) and performing differential analysis, 267 differentially expressed MRGs (DE-MRGs) were identified (**Figure 2A, 2B**), which were mainly enriched in the oxidoreductase, lyase and coenzyme binding function, the membrane and mitochondria components, the small molecule catabolic metabolism and nucleotide-related metabolism bioprocesses or pathways (**Figure 2C, 2D**). After excluding samples with less than 60 days of follow-up, 1023 BRCA samples were retained in the TCGA cohort, which served as the training group. Combined with the overall survival information of patients in the training group, we screened the 7 most statistically significant MRGs with prognostic value using the univariate Cox regression. The hazard ratios (HR) and 95% confidence interval (95% CI) of 7

prognostic MRGs are shown in **Table 1**. Accordingly, PLA2G2D was a protective factor with a HR < 1, while the other six were risky ones (HR > 1, *P* < 0.01). We confirmed their coefficients for signature construction by LASSO regression (**Figure 2E, 2F**). Among these seven prognostic MRGs, except for NT5E, which was down-regulated, the other 6 MRGs were up-regulated (**Figure 2B**).

To further confirm the expression characteristics of 7 MRGs at the protein level, we obtained the immunohistochemical data from the HPA database for these MRGs except PLA2G2D. Consistent with their mRNA expression (**Figure 2B**), the protein levels of GNPAT1, QPRT, PLPP2, SHMT2 and PAICS were higher in breast cancer tissues than in normal tissues (**Figure S1**). But for NT5E, no obvious difference in its protein level between normal tissue and tumor tissue was observed (**Figure S1**).

Development, validation and clinical relevance evaluation of the prognostic signature

The coefficients of seven signature MRGs are shown in **Table 1**. Then the risk score for each patient in the TCGA cohort (N = 1023) can be calculated using this formula: Risk score = $-0.124668 \times \text{Exp (PLA2G2D)} + 0.012631 \times \text{Exp (GNPNAT1)} + 0.025496 \times \text{Exp (QPRT)} + 0.011223 \times \text{Exp (SHMT2)} + 0.026805 \times \text{Exp (PLPP2)} + 0.077629 \times \text{Exp (NT5E)} + 0.018399 \times \text{Exp (PAICS)}$. Patients in the TCGA cohort were divided into the low-risk group (N = 512) and high-risk group (N = 511) according to the median value of risk scores, which would also be taken as the cutoff for external validating groups. The Kaplan-Meier curve showed that high-risk patients had worse overall survival in the TCGA cohort (*P* < 0.001) (**Figure 3A**). Moreover, poorer outcomes of high-risk patients were verified in two external validation cohorts, the GSE21653 cohort (n = 241) (*P* < 0.01) and the GSE20685 cohort (n = 327) (*P* < 0.001) (**Figure 3B, 3C**). Consistent results were also achieved in two internal validation cohorts (*P* < 0.01) (**Figure S3A**). **Figure 3D-F** showed risk scores and survival status of patients and expression of the 7 MRGs in different risk groups. The 7 MRGs were highly expressed mainly in the tumors of high-risk patients, except for PLA2G2D, which may indicate that PLA2G2D

A metabolism-related prognostic signature for breast cancer patients

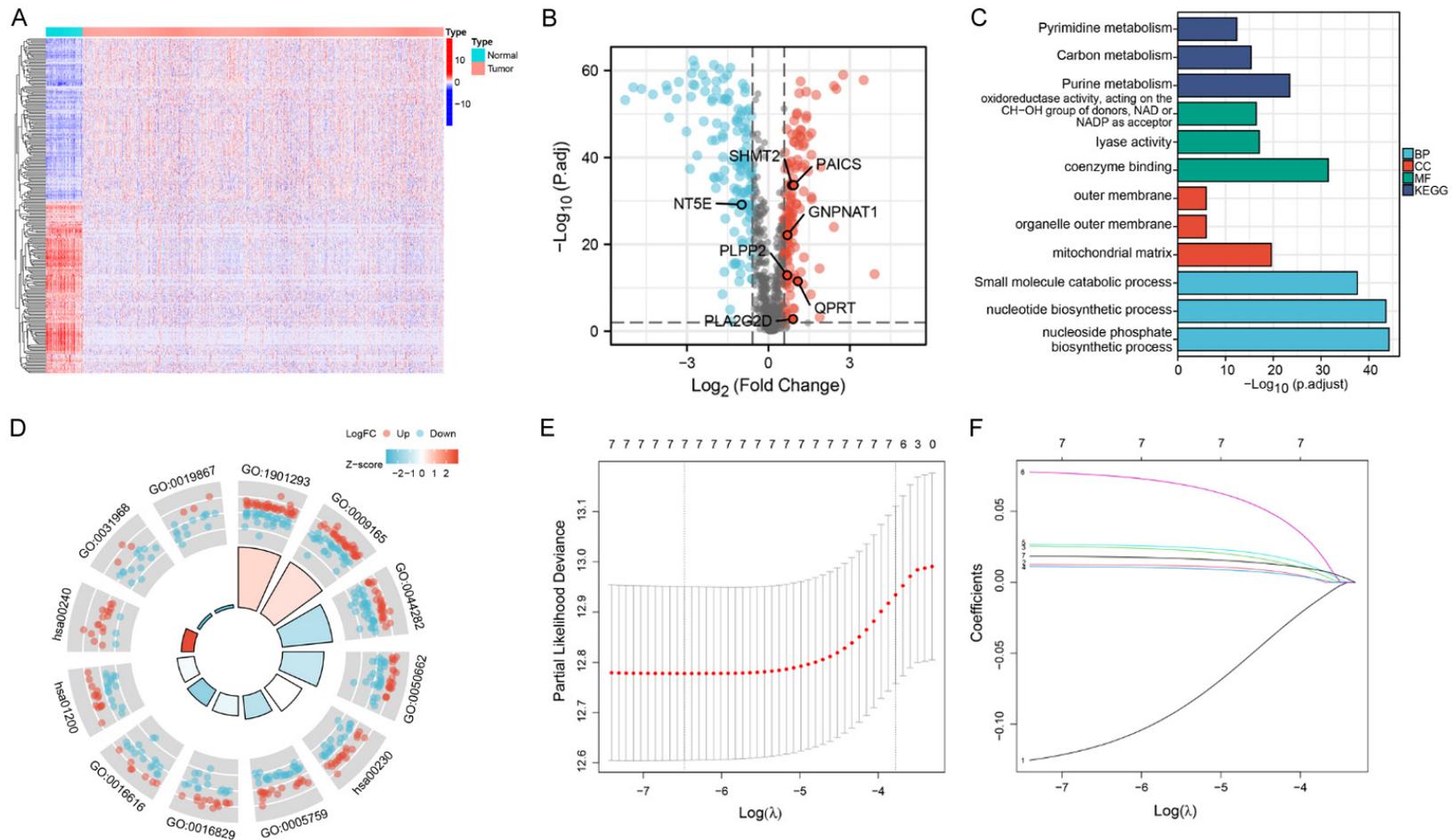


Figure 2. Screening of prognostic MRGs in breast cancer for signature construction. (A, B) Heatmap and volcano map of the expression of MRGs in normal and tumor samples. Red grids or dots represent high-expression MRGs, and blue grids or dots represent low-expression MRGs. Signature MRGs have been labeled in the figure. (C, D) GO and KEGG analyses of differentially expressed MRGs. In (C), the categories of enrichment analysis and the corresponding bar colors are shown in the legend (BP, Biological Process; CC, Cellular Component; MF, Molecular Function; KEGG, the Kyoto Encyclopedia of Genes and Genomes). (E, F) LASSO analysis determined the lambda parameter and coefficients of signature MRGs.

A metabolism-related prognostic signature for breast cancer patients

Table 1. Hazard ratio and coefficients of signature MRGs

id	Description	HR	HR. 95L	HR. 95H	P value	Coefficient
PLA2G2D	Phospholipase A2 Group IID	0.89759	0.83035	0.97027	0.00654	-0.124668
GNPNAT1	Glucosamine-Phosphate N-Acetyltransferase 1	1.02507	1.00645	1.04402	0.00810	0.012631
QPRT	Quinolate Phosphoribosyltransferase	1.02965	1.00925	1.05048	0.00422	0.025496
SHMT2	Serine Hydroxymethyltransferase 2	1.02144	1.00636	1.03674	0.00518	0.011224
PLPP2	Phospholipid Phosphatase 2	1.03105	1.01072	1.05179	0.00262	0.026805
NT5E	5'-Nucleotidase Ecto	1.05447	1.01650	1.09386	0.00459	0.077629
PAICS	Phosphoribosylaminoimidazole Carboxylase And Phosphoribosylaminoimidazolesuccinocarboxamide Synthase	1.01909	1.00877	1.02951	0.00027	0.018399

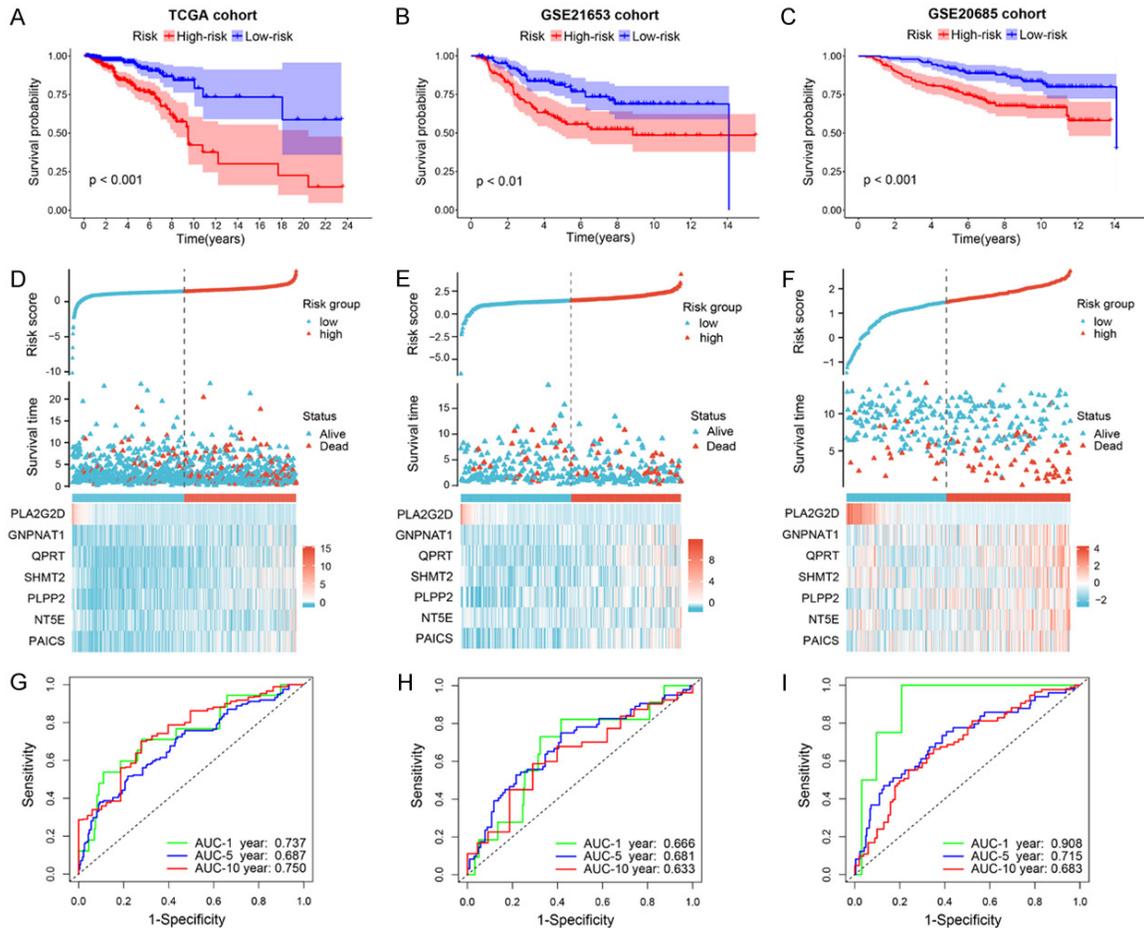


Figure 3. Construction, validation and evaluation of the metabolism-related prognostic signature. A-C. Kaplan-Meier curves of overall survival between high- and low-risk groups in the TCGA cohort (N = 1023), GSE21653 cohort (N = 241) and GSE20685 cohort (N = 327), respectively. D-F. Risk score distribution, survival status of patients and heatmaps of expression profiles of 7 MRGs involved in the signature. G-I. Time-dependent ROC curves for assessing the prognostic efficacy of the signature.

was a favorable factor (Figure 3D-F). Then the time-dependent ROC curves were applied to evaluate the prediction sensitivity of the prognostic signature; the 1-, 5- and 10-year area under the curve (AUC) values were 0.737, 0.687 and 0.750 in the TCGA cohort (Figure

3G). In external validating groups, the 1-, 5- and 10-year AUC values of two GEO cohorts were 0.666, 0.681, 0.633 and 0.908, 0.715 and 0.683, respectively (Figure 3H, 3I). The AUC of the signature can be around 0.7 at three time points in the TCGA cohort, but mostly

A metabolism-related prognostic signature for breast cancer patients

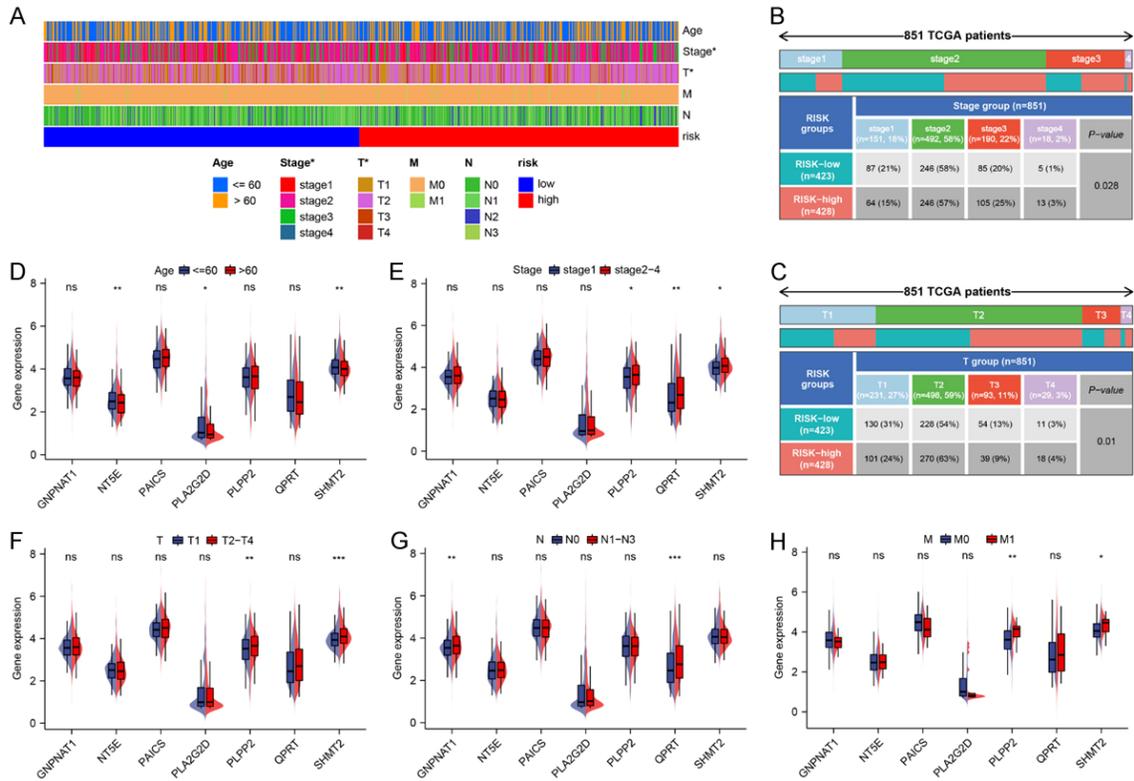


Figure 4. Clinicopathological relevance of the prognostic signature. (A-C) Heat maps of patients with different clinical or pathological characteristics in the high- and low-risk groups. The association between the expression of 7 MRGs and clinical factors including age (D), stage (E), T-stage (F), N-stage (G) and M-stage (H). In this figure, grouping variables including stage, T (tumor), N (node), and M (metastasis) were derived from the AJCC (American Joint Committee on Cancer) breast cancer staging system (ns: not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

between 0.65-0.7 in the GSE21653 cohort. In the GSE20685 cohort, the AUC values at 1- and 5-year were also still satisfactory, showing a relatively reliable performance of the signature.

To assess the association between the prognostic signature and clinicopathological factors, we presented the distribution of patients with different clinicopathological features in the low- and high-risk groups using heatmaps (Figure 4A). There were more stage-1 patients in the low-risk group and more stage-3 or stage-4 patients in the high-risk group ($P = 0.028$) (Figure 4B). Patients with T1 and T3 were enriched in the low-risk group and those with T2 and T4 in the high-risk group ($P = 0.01$) (Figure 4C). In addition, we also analyzed the relationship between the expression of signature MRGs and clinicopathological factors. The expression levels of NT5E, PLA2G2D and SHMT2 differed between the two groups separated by the age of 60 (all $P < 0.05$) (Figure 4D).

With the increase of clinicopathological stage, T stage and M stage, the expression of PLPP2 and SHMT2 elevated significantly (all $P < 0.05$) (Figure 4E, 4F, 4H). The expression of QPRT also increased in the T2-T4 group, while the statistical difference was not significant ($P = 0.08$) (Figure 4F). Consistent with GNPAT1, QPRT expressed higher in the group with a higher N stage (all $P < 0.01$) (Figure 4G). Moreover, QPRT expression showed a certain increasing tendency with the advancing of the clinicopathological stage and N stage (Figure S4B, S4C). The changing trend of SHMT2 expression with clinicopathological stage and T stage also showed a similar feature (Figure S4A, S4C). It can be found that signature MRGs such as PLPP2, SHMT2 and QPRT are correlated with clinicopathological factors to a certain extent.

Furthermore, univariate and multivariate Cox regression analyses showed that the signature could be regarded as a prognostic indicator independent of other clinicopathological fac-

A metabolism-related prognostic signature for breast cancer patients

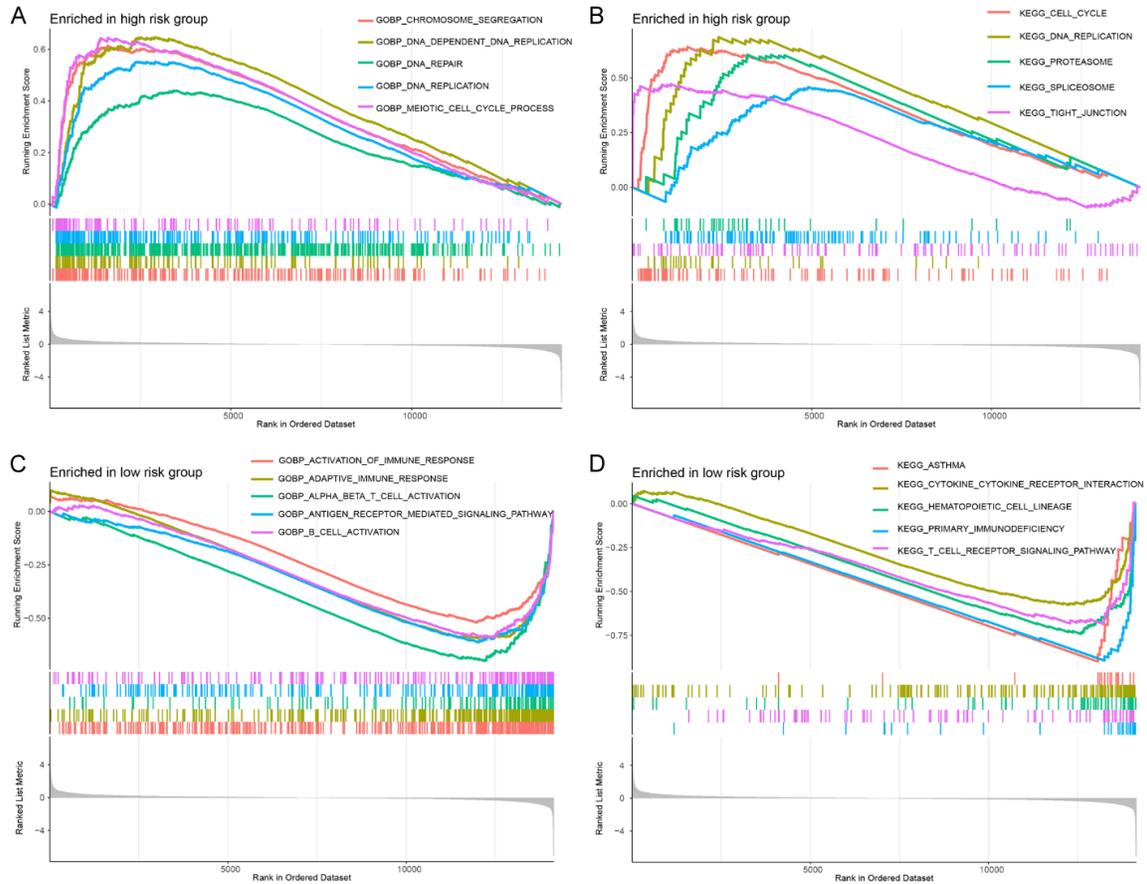


Figure 5. Gene Set Enrichment Analysis (GSEA). Significantly enriched biological processes (A) and KEGG pathways (B) in the high-risk group. Significantly enriched biological processes (C) and KEGG pathways (D) in the low-risk group.

tors (Figures S2A, S3B). In combination with the risk score and the age, a clinicopathological factor capable of independently indicating prognosis, a nomogram was developed to predict patient survival (Figure S2B). Decision curve analyses (DCA) indicated that the metabolism-related signature and integrated nomogram did better than other clinical factors including age and stage in predicting 5- and 10-year survival, especially the 10-year survival (Figure S2C).

To further explore whether the signature could distinguish survival outcomes of patient subgroups with different clinicopathological features, patients were regrouped according to clinicopathological features and survival analyses were performed. The signature still showed good risk stratification effect in patients with Luminal A subtype and Basal subtype ($P < 0.05$) (Figure S3C). The overall survival was shorter for high-risk patients in Her2-enriched subgroup and Luminal B subgroup, yet not sta-

tistically significant ($P = 0.07$ and $P = 0.235$) (Figure S3C). In other clinicopathological subgroups of the TCGA cohort, high-risk patients still demonstrated significantly worse survival outcomes ($P < 0.01$) (Figure S3C). Among the clinicopathological subgroups of the GSE206-85 cohort, only the T3-T4 subgroup showed no significant difference due to the limited number of patients ($P = 0.26$), and the overall survival differences between high- and low-risk patients in the remaining subgroups were all significant ($P < 0.05$) (Figure S4D).

Gene set enrichment analysis (GSEA)

To identify biological processes associated with the metabolism-related risk signature, we conducted GSEA. The biological processes mainly enriched in the high-risk group included chromosome segregation, DNA replication and repair and cell cycle-related processes (Figure 5A). While in the low-risk group, biological processes that were significantly enriched were

A metabolism-related prognostic signature for breast cancer patients

immunity-related processes including activation of the immune response, adaptive immunity and alpha-beta T cell activation and so on (**Figure 5C**). For KEGG pathways, there were also some cell cycle- and division-related ones enriched in the high-risk group, as well as some immune and cytokine-related pathways enriched in the low-risk group (**Figure 5B, 5D**). These results suggest that differences in the prognosis of patients with different risks may be strongly linked to immune responses and immune regulation, and therefore we may need to characterize the immune microenvironment of different risk tumors.

Analysis of tumor immune microenvironment and immune function

Four algorithms (ESTIMATE, xCELL, MCPcounter and CIBERSORT) were utilized to analyze the characteristics of tumor-infiltrating immune cells between different risk groups. **Figure 6A** showed the relationship between risk score and immune cell abundance analyzed by different algorithms. It can be found that risk score has a significant negative correlation with the proportion of tumor-infiltrating CD8⁺ T cells, B cells, myeloid dendritic cells, etc. A positive correlation between risk scores and the proportions of M0 macrophages, M2 macrophages and CD4⁺ Th2 cells was also observed (**Figure 6A**). High-risk group had lower Stromal, Immune, and ESTIMATE scores than the low-risk group and these three scores all negatively correlated with the risk score (**Figure 6A, 6C**). From the results calculated by the CIBERSORT algorithm we found that high-risk tumors had a lower proportion of naïve B cells, CD8⁺ T cells, CD4⁺ memory resting T cells, gamma delta T cells, resting dendritic cells and resting mast cells, and a higher proportion of M0 macrophages and M2 macrophages (**Figure 6B**). There were significant differences in stemness scores (RNAss) between high- and low-risk groups, and a significant positive correlation between risk score and RNAss was observed (**Figure 6D, 6E**).

Furthermore, the relationship between the expression of 7 signature MRGs and the abundance of immune cells was shown in **Figure 6F**. As we can notice, PLA2G2D showed a strong positive correlation with CD8⁺ T cells, M1 mac-

rophages and CD4⁺ memory activated T cells and so on, and a strong negative correlation with M0 macrophages and M2 macrophages, indicating its potential anti-tumor effect (**Figure 6F**). NT5E expression positively correlated with the infiltration of CD4⁺ memory T cells, resting NK cells, neutrophils and M2 macrophages, and negatively correlated with the infiltration of plasma cells, activated NK cells, follicular helper T cells and CD8⁺ T cells, which were similar to those characteristics of GNPAT1 (**Figure 6F**). In addition, the expression of PLPP2 and SHMT2 had a similar effect on immune cell infiltration, which was reflected in that they were both positively correlated with regulatory T cells, follicular helper T cells, activated NK cells and M0 macrophages, and both negatively correlated with CD4⁺ memory resting T cells, resting mast cells and resting dendritic cells (**Figure 6F**). Differences in immune function between the low- and high-risk groups were shown in **Figure 6G**. Most of the immune function scores such as the CD8⁺ T cell, cytolytic activity, and type-II IFN response activity scores were higher in the low-risk group than in the high-risk group, except for some functional scores that were not statistically different, which may indicate that the immune function of low-risk patients was more active, consistent with the GSEA results above.

Moreover, the expression levels of immune checkpoint genes, m6A mRNA methylation regulator genes and main mismatch repair (MMR) genes were compared. Most of the immune checkpoints including PDCD1 (PD-1), CTLA4, CD274 (PD-L1), HAVCR2 (TIM-3) and ICOS etc., expressed higher in the low-risk group, but CD276 (B7-H3) expressed lower ($P < 0.01$) (**Figure 7A**). In the low-risk group, expression levels of crucial MMR genes including MSH2, MSH6 and EPCAM were lower, and expression levels of MLH3, MLH1, and MSH3 were higher ($P < 0.05$) (**Figure 7B**). The risk score was negatively correlated with the expression of PDCD1, CTLA4 and CD274, while positively correlated with the expression of CD276 (all $P < 0.001$) (**Figure S5**). In addition, **Figure 7C** showed significantly decreased expression levels of YTHDF3, YTHDF1, HNRNPC, HNRNPA2B1, FMR1, and RBM15 in the low-risk group, which are key regulatory molecules of m6A mRNA methylation. The other molecules, ZC3H13, YTHDC2,

A metabolism-related prognostic signature for breast cancer patients

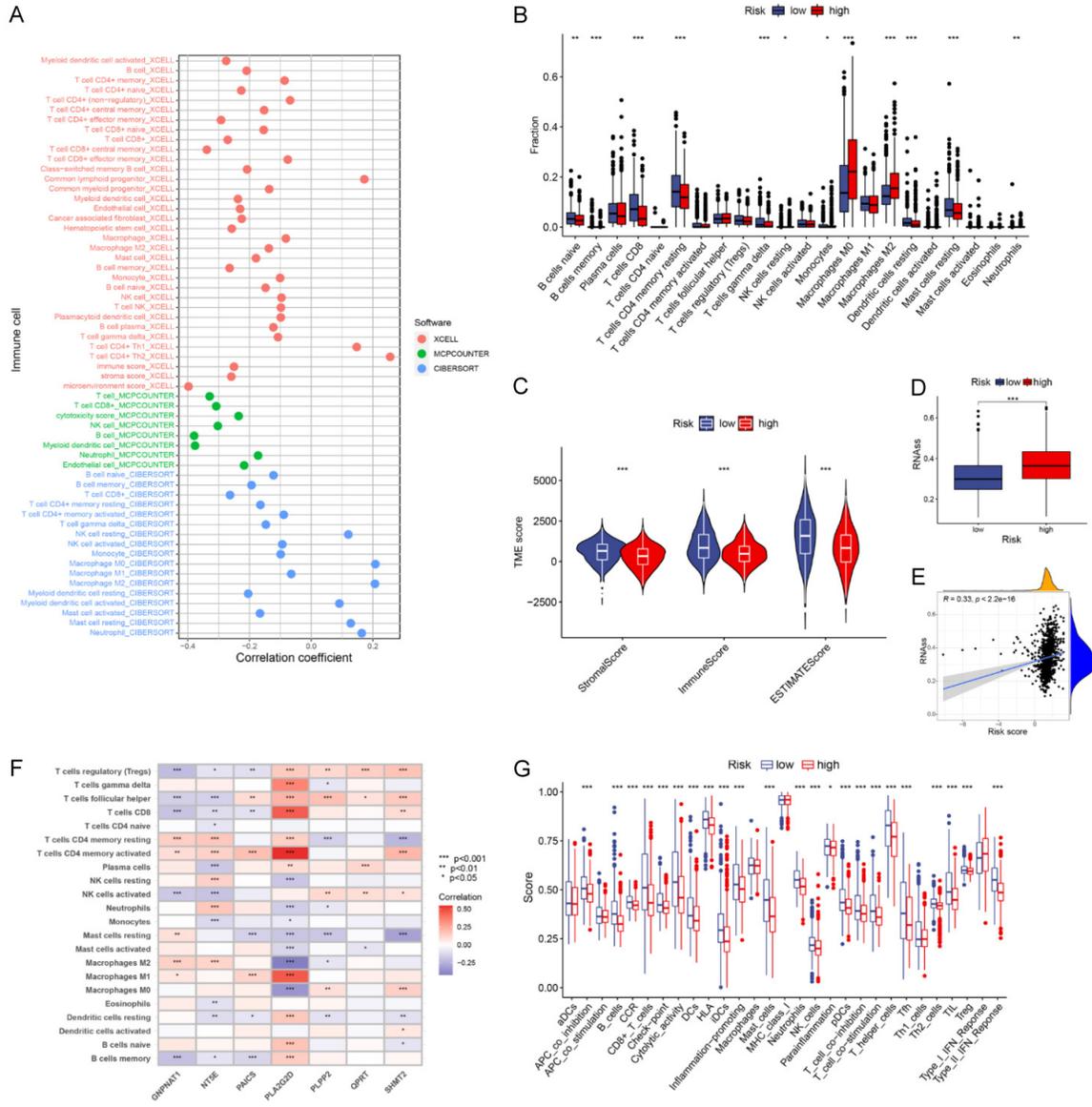


Figure 6. Characteristics of tumor immune microenvironment and immune function indifferent risk groups. **A.** The correlation between risk score and the tumor-infiltrating immune cells estimated by three algorithms (xCELL, MCPcounter and CIBERSORT). **B.** Differences in the TIICs calculated by CIBERSORT between two groups. **C.** Comparison of tumor microenvironment scores calculated by ESTIMATE between two groups. **D, E.** The correlation between risk score and stemness score. **F.** The correlation between the TIICs and expression of MRGs involved in the signature. **G.** Comparison of the immune function scores calculated by ssGSEA between different risk groups (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

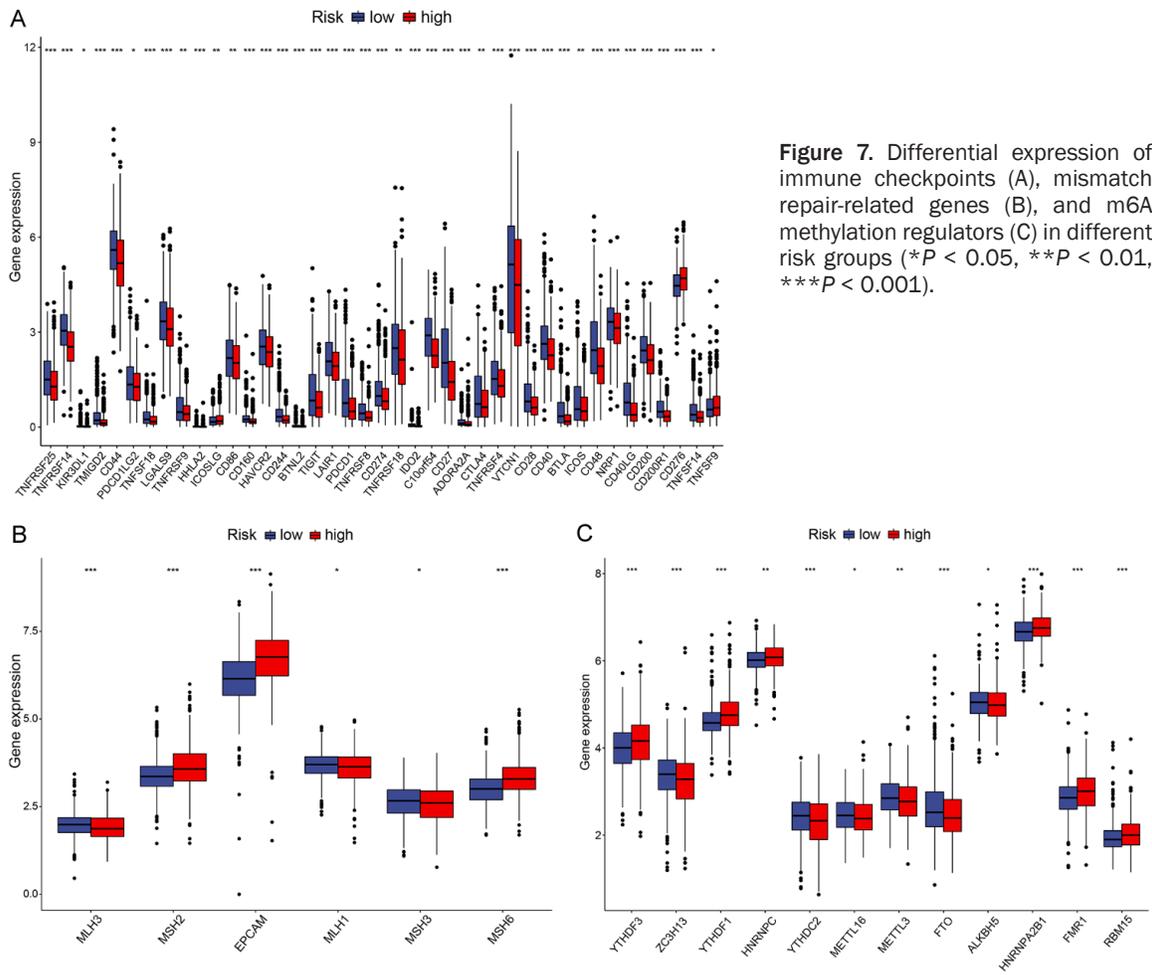
METTL16, METTL3, FTO, and ALKBH5, expressed higher in the low-risk group ($P < 0.05$) (Figure 7C).

Immunotherapy response prediction and validation

Immune subtype information of cancer samples was obtained from the TCGA database.

There were 6 immune subtypes reported in 2018, wound healing (Immune C1), IFN- γ dominant (Immune C2), inflammatory (Immune C3), lymphocyte depleted (Immune C4), immunologically quiet (Immune C5) and TGF- β dominant (Immune C6) subtypes [22]. The proportions of immune subtypes differed between the high- and low-risk groups, and there was no Immune C5 subtype in our cohort. The propor-

A metabolism-related prognostic signature for breast cancer patients



tion of C3 subtypes was lower, while the proportion of C1 and C2 subtypes, which are those with higher proliferative activity, angiogenic activity and macrophage polarization, was higher in high-risk patients (Figure 8A). Multiple approaches were utilized to evaluate the value of the signature in predicting immunotherapy response. Firstly, the expression of PD-1, PD-L1 and CTLA-4 was higher in the low-risk group (Figure 7A). Secondly, the TIDE scoring system calculated the dysfunction score, exclusion score, MSI score, TIDE score and predicted response of each patient. The high-risk group had lower dysfunction scores and higher TIDE scores, suggesting that immune escape may be more likely to occur in high-risk patients (Figure 8C-F). By comparing the distribution of patients responding to immunotherapy between two groups, we found that the proportion of patients with response to immune checkpoint blockade (ICB) therapy in the low-risk

group (19%, 80/851) was higher than that in the high-risk group (11%, 47/851) ($P = 0.002$) (Figure 8B). To further confirm this finding, we also conducted the Immunophenoscores (IPs) analysis. Consistently, patients in the low-risk group showed higher IPs scores, regardless of whether they were regrouped by CTLA-4 and PD-1 expression, suggesting that low-risk patients were more sensitive to immunotherapy ($P < 0.001$) (Figure 8G).

Since the above analyses were predictive, we still wanted to determine whether our signature could be used to distinguish responders and non-responders in the real-world immunotherapy cohort. Due to the lack of immunotherapy cohort data of BRCA patients, we used ICB therapy data of the urethral epithelial cancer (IMvigor210) and melanoma (GSE78220) for validation. High-risk patients from the IMvigor210 cohort showed worse overall survival

A metabolism-related prognostic signature for breast cancer patients

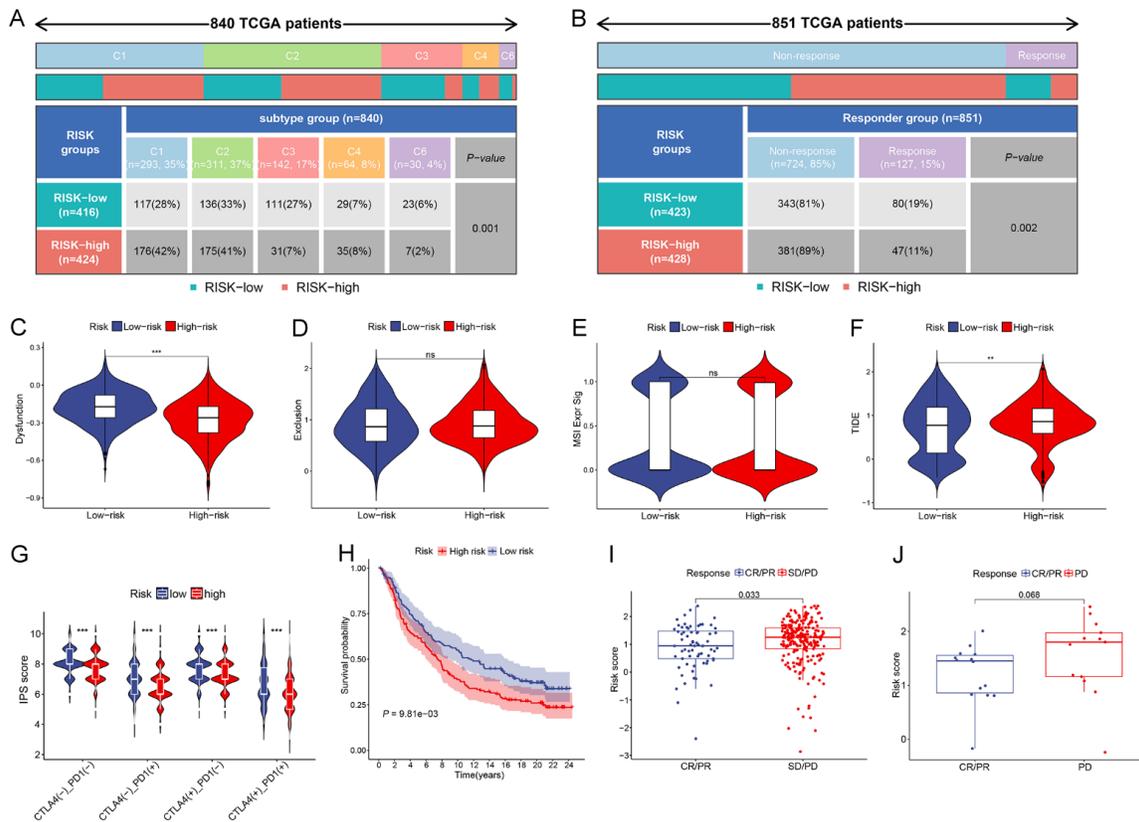


Figure 8. The value of the signature in differentiating immune subtypes and immunotherapy responses. (A) The distribution of patients with different immune subtypes in the high- and low-risk groups. (B) The predicted response of patients to immune checkpoint blockade (ICB) therapy in the high- and low-risk groups; (C-G) Comparison of dysfunction score (C), exclusion score (D), MSI signature score (E), TIDE score (F) and IPS score (G) between the two groups; (H) Risk stratification and survival analysis of external cohort, IMvigor210; (I, J) Patients responding to ICB therapy had lower risk scores in the IMvigor210 cohort and GSE78220 cohort (ns: not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

than low-risk patients (**Figure 8H**). Patients with disease progression or stabilization (PD/SD) had higher risk scores than patients with complete or partial response (CR/PR) ($P = 0.033$) (**Figure 8I**). In the melanoma patient cohort, though the difference was not statistically significant ($P = 0.068$), the risk score of the CR/PR group was lower, consistent with our predictive results above (**Figure 8J**). These analyses confirmed the value of our signature in predicting the response of patients to immunotherapy.

Chemotherapeutic sensitivity prediction

We analyzed the sensitivity to specific chemotherapy or targeted agents by calculating and comparing their IC50 values. We found that low-risk patients had a decreased IC50 of Lapatinib, Gefitinib, Axitinib, Bosutinib, Imatinib, Cytarabine, Methotrexate, Bexarotene, Br-

yostatin.1, Vinorelbine, Temsirolimus and Eleclomol, and an increased IC50 of Sorafenib, Mitomycin.C, Thapsigargin and S.Trityl.L.cysteine than the high-risk group ($P < 0.05$) (**Figure 9A-P**), which demonstrated that low-risk patients might be more sensitive to the first 12 drugs including Lapatinib, Gefitinib and Axitinib and so on. High-risk patients were less sensitive to immunotherapy and targeted agents such as Lapatinib, but those latter including Sorafenib, Mitomycin.C, etc., may be more effective.

Identification of molecular subtypes based on signature MRGs

We further used the 7 MRGs in the signature to explore whether these 7 genes could be used for molecular subtype classifying of patients. Patients were clearly divided into two groups

A metabolism-related prognostic signature for breast cancer patients

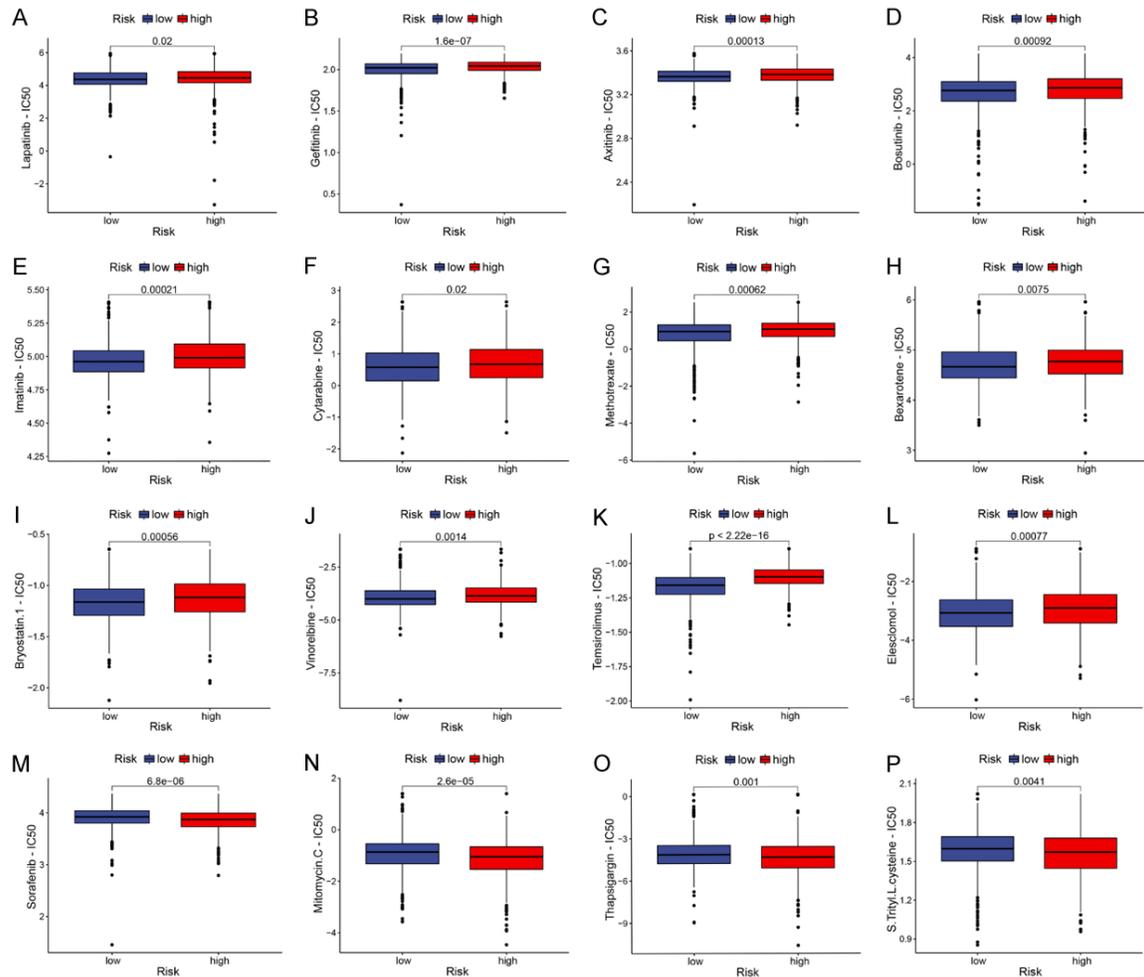


Figure 9. Prediction of the sensitivity to chemotherapeutic drugs and targeted drugs. Low-risk patients showed higher sensitivity to Lapatinib (A), Gefitinib (B), Axitinib (C), Bosutinib (D), Imatinib (E), Cytarabine (F), Methotrexate (G), Bexarotene (H), Bryostatins.1 (I), Vinorelbine (J), Temsirrolimus (K), Elesclomol (L), and lower sensitivity to Sorafenib (M), Mitomycin.C (N), Thapsigargin (O) and S.Trityl.L.cysteine (P).

after consensus clustering based on the 7 MRGs (Figure 10A, 10B). The two molecular subtypes showed different overall survival prognosis and the cluster-A survived better ($P < 0.001$) (Figure 10C). There were significant differences in the characteristics of infiltrating immune cells and immune function scores between the two molecular types of tumors (Figure 10D, 10E). Subtype-B patients had lower dysfunction, exclusion, and TIDE scores and a higher proportion of immunotherapy responders (Figure 10F-J). These results further demonstrated the potential of these seven metabolic molecules for signature developing as new prognostic indicators and intervention targets.

Discussion

Metabolic reprogramming is regarded as an important hallmark of cancer because it directly affects the growth and survival of cancer cells by generating phenotypes more adaptable to the environment [6]. Nowadays, the crosstalk between the metabolic reprogramming and the tumor microenvironment has also gained much attention [23]. The metabolic adaptations of cancer cells contribute to the shaping of the immunosuppressive tumor microenvironment, thus promoting the vascularization and immune escape. For example, increased glycolysis activity can lead to a decreased glucose available in the microenvi-

A metabolism-related prognostic signature for breast cancer patients

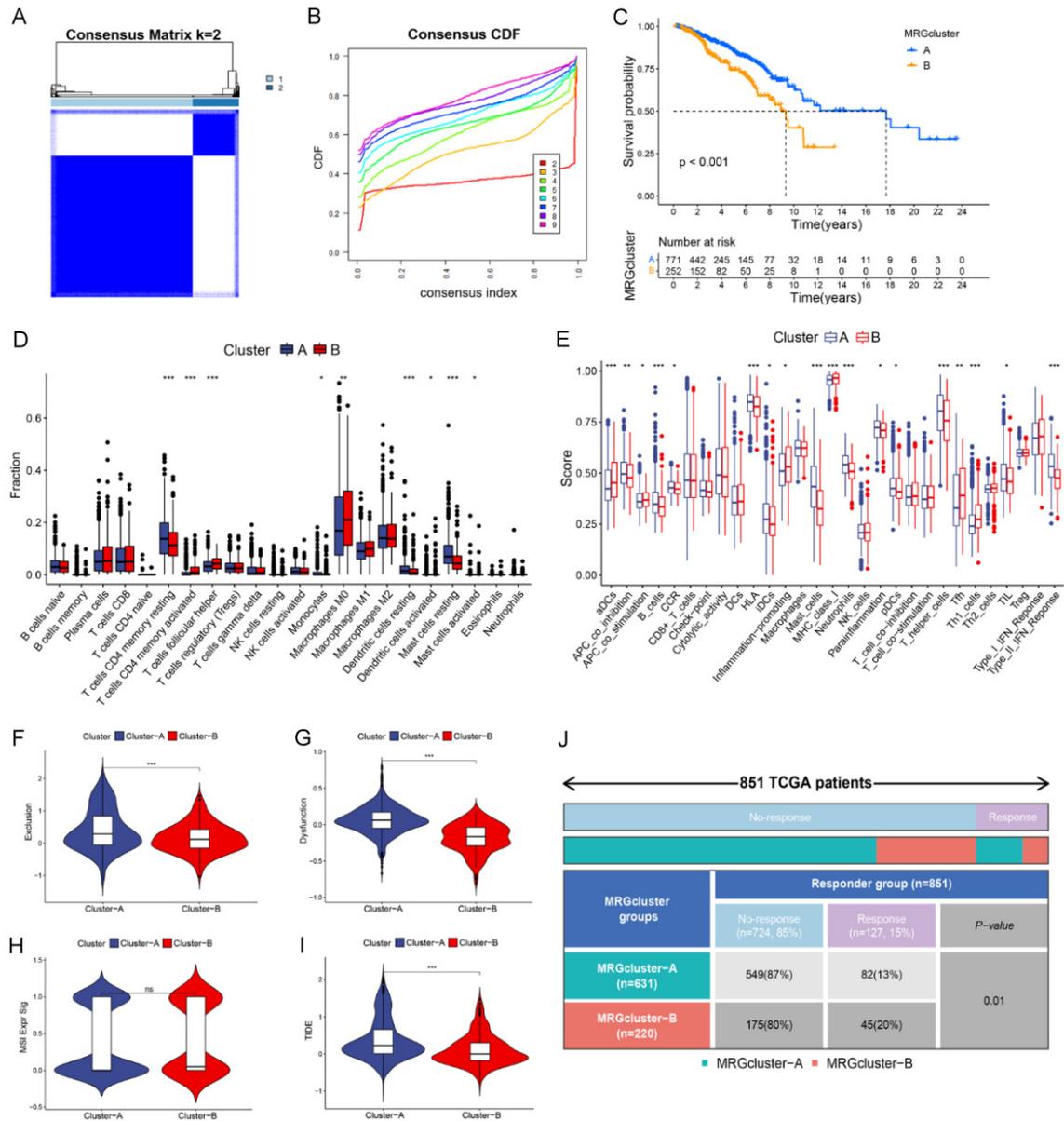


Figure 10. Identifying subtypes based on 7 MRGs in the signature. (A) Consensus matrix for $k = 2$. (B) Selection of CDF. (C) Kaplan-Meier curves of OS in subtypes. (D, E) Differences in tumor-infiltrating immune cells and immune function scores between subtypes. Comparison of dysfunction score (F), exclusion score (G), MSI score (H) and TIDE score (I) between subtypes. (J) The distribution of responders to ICB therapy in subgroups (ns: not significant, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$).

ronment, limiting the function and altering the apoptosis sensitivity of effector T cells. Low glucose could also induce the expression of genes such as FOXP3 that promote the differentiation of T cells into Tregs [24]. Not only the glucose metabolism, other molecules such as fatty acid, amino acid and nucleotide metabolism and small molecule metabolites are closely associated with the tumorigenesis, progres-

sion and crosstalk with the microenvironment of breast cancer [10]. For instance, acetyl-CoA, a molecule derived from glycolysis and glutamine metabolism, could promote mevalonate (MVA) biosynthesis to produce metabolites that sustain the survival and rapid growth of various cancer cells [25]. Up-regulated cholesterol synthesis promotes immune surveillance evasion and causes chemotherapy resistance [26].

A metabolism-related prognostic signature for breast cancer patients

However, the value of metabolic molecules in predicting prognosis, assessing drug sensitivity and even acting as potential therapeutic targets remains to be further explored.

In this study, a robust prognostic signature based on seven metabolic genes (PLA2G2D, GNPAT1, QPRT, PLPP2, NT5E, SHMT2, PAICS) was constructed for breast cancer patients. High-risk patients distinguished by the signature showed worse survival than low-risk patients in the TCGA cohort and two external independent cohorts. There were more patients with lower tumor stage or T-stage in the low-risk group, indicating the good clinicopathological relevance of the signature. ROC curves, DCA, clinicopathological subgroup survival curves and independence analyses showed that the metabolic signature performed well and could act as an independent prognosis indicator. Though some of genes involved in this signature have been reported to express low or high in tumors, we found differences in their expression with clinicopathological features and their relationship with immune cell infiltration and used immunohistochemical data from the HPA database to validate their protein levels.

PLA2G2D, phospholipase A2 Group IID, a kind of lipid metabolism-associated protein, could be secreted and involved in inflammatory responses via hydrolyzing membrane phospholipids and releasing unsaturated fatty acids [27]. In this analysis, PLA2G2D was identified as a low-risk gene, consistent with reports from other researchers [28-30]. Liu et al. reported that there were more tumor-infiltrating CD8⁺ T cells inside the tumor of cervical squamous cell carcinoma with high expression of PLA2G2D, and patients with higher PLA2G2D expression were more likely to benefit from immunotherapy [30]. Consistently, our analysis found that low-risk patients had higher PLA2G2D expression and were more sensitive to immunotherapy.

GNPNAT1, glucosamine-phosphate N-acetyltransferase 1, a key enzyme involved in the branching processes of glucose metabolism, the hexosamine biosynthetic pathway, was found to be up-regulated in breast cancer, lung cancer and prostate cancer tissues and associated with poor prognosis [31-33]. GNPAT1

expression was positively correlated with the infiltration of CD4⁺ memory T cells and M2 macrophages, and negatively correlated with the infiltration of Tregs, CD8⁺ T cells and memory B cells in breast cancer. Similarly, its correlation with B cells and CD4⁺ T cells was also found in lung adenocarcinoma [32, 34].

Quinolate phosphoribosyltransferase (QPRT), a rate-limiting enzyme regulating the generation of nicotinamide adenine dinucleotide (NAD⁺), participated in the cell cycle and essential metabolism processes of cancer cells [35, 36]. The high expression of QPRT was associated with poor prognosis and pathological progression of breast cancer patients in the present and previous studies [37]. A positive correlation between the QPRT expression and the abundance of infiltrating Tregs, plasma cells, activated NK cells and follicular helper T cells was found in this study.

PLPP2 (phospholipid phosphatase 2), also known as PPAP2C or LPP2, was a lipid-metabolism gene participating in regulating cell cycle and proliferation [38]. PLPP2, along with its family members PLPP1 and PLPP3, could dephosphorylate extracellular lysophosphatidic acid and sphingosine-1-phosphate on the cell surface [39]. And PLPP2 played a role in regulating cell proliferation by influencing the entry into the S-phase of the cell cycle [40]. High expression of PLPP2 was found in various cancers including breast, lung, ovarian and clear cell renal cell carcinoma, and was significantly associated with poor prognosis and progression of pathological stage in breast and clear cell renal cell carcinoma, which was partly verified in this study [38, 41].

NT5E encodes CD73, an important 5' exonuclease for the synthesis and homeostasis of extracellular adenosine levels [42]. CD73 has been found correlated with prognosis in breast, liver, pancreatic, and colon cancers and is expected to be a therapeutic target [43-47]. Adenosine could play a potent immunosuppressive role in the tumor microenvironment to inhibit the activity of various immune cells, such as the antigen-presenting process, T cell proliferation and cytokine production [48]. Recently, researchers found that tumor-derived adenosine could also act synergistically with GM-CSF released by

A metabolism-related prognostic signature for breast cancer patients

activated macrophages to promote the proliferation of immunosuppressive macrophages in hepatocellular carcinoma (HCC) [49]. Moreover, the interaction between CD39⁺ macrophages and CD73⁺ HCC cells, accompanied by elevated adenosine, could induce the resistance to ICB therapy in HCC [44]. And CD73 could promote the progression and metastasis of cancer cells and act on a specific fibroblast population leading to immunosuppression [46, 47].

Serine Hydroxymethyltransferase 2 (SHMT2) is a key enzyme participating in amino acid and one-carbon metabolism, redox homeostasis and nucleotide synthesis. We found the association between the expression of SHMT2 and poor prognosis, advanced stage and metastasis in breast cancer, which was partly consistent with previous studies [50, 51]. Breast cancer cells with overexpression of SHMT2 presented resistance to lapatinib and paclitaxel [52, 53]. In this study, we observed a strong correlation between the SHMT2 expression and the infiltration of resting mast cells, M0 macrophages, CD4⁺ memory T cells, follicular helper T cells and Tregs.

PAICS (Phosphoribosylaminoimidazole Carboxylase and Phosphoribosylaminoimidazolesuccinocarboxamide Synthase) is an important metabolic enzyme in the *de novo* purine synthesis. It was found to be differentially expressed and associated with prognosis in multiple cancers, including breast cancer, lung cancer and others [54-57]. PAICS participates in the regulation of cell growth, epithelial-mesenchymal transition (EMT) and migration in cancers [58]. A significant association between PAICS and immune cells such as M1 macrophages and activated CD4⁺ memory T cells was found in this study.

The microenvironmental characteristics of tumors in high-risk and low-risk groups were significantly different. There were more infiltrating M0 and M2 macrophages and fewer CD8⁺ T cells, resting memory CD4⁺ T cells and gamma delta T cells and so on inside the high-risk tumors. Tumor-associated macrophages (TAMs), especially the M2 phenotype, were found to play key roles in promoting the invasion and metastasis of cancer cells, facilitating the angiogenesis and restricting the function of

CD8⁺ T cells in the microenvironment of breast cancer [59, 60]. An increased number of M0 macrophages were correlated with a decreased overall survival, while the M1 phenotype could promote the cytotoxic response by attracting T lymphocytes [61]. High intratumoral density of CD8⁺ cytotoxic T lymphocytes is a marker of favorable prognosis in the majority of cancers including breast cancer, with the exception of Hodgkin lymphoma and clear cell renal cell carcinoma and so on [62]. The prognostic effect of CD4⁺ T cells is less than that of CD8⁺ T cells because of the considerable heterogeneity of subpopulations and different functions of each lineage. Researchers found that a higher proportion of resting memory CD4⁺ T cells and $\gamma\delta$ T cells was associated with a better prognosis in breast cancer [63]. For memory B cells and naïve B cells, studies have reported their favorable prognostic value and association with neoadjuvant chemotherapy response in breast cancer [64-66]. These findings above support our results regarding the characteristics of tumor-infiltrating immune cells in different risk groups.

When exploring the performance of the signature in helping with clinical decision making, sensitivity to immunotherapy and chemotherapy of patients with different risk was analyzed via multiple approaches. Expression of most checkpoints such as PD-1, PD-L1, CTLA-4 and TIM-3 was higher in the low-risk group, except for CD276. Patients in the low-risk group showed lower TIDE scores and higher IPS scores regardless of the PD-1 or CTLA-4 status, suggesting that low-risk tumors were less likely to develop immune evasion and more immunogenic, thus low-risk patients were more likely to benefit from immunotherapy. Moreover, analyses of real-world cohorts of patients receiving immunotherapy (IMvigor210 and GSE78220) also confirmed the predictive performance in evaluating the responsiveness to immunotherapy. Low-risk patients also showed higher sensitivities to various chemotherapeutics and targeted drugs including Cytarabine, Lapatinib and Bosutinib and so on. Elesclomol, which had a lower IC50 value in the low-risk group, is an oxidative stress inducer that can induce apoptosis of cancer cells. It has recently been reported as a promising copper carrier for anticancer therapy recently [67]. But high-risk patients could respond well to drugs such as

A metabolism-related prognostic signature for breast cancer patients

Sorafenib and Mitomycin C, which may be suitable candidates.

With an improved understanding of the immune microenvironment of breast cancer, a growing number of novel immunotherapy targets such as CD73 (NT5E), LAG3 and TIGIT, as well as new combination therapy, have also been studied [23, 68]. Our study showed that low-risk patients with more immune effector cells in their tumor microenvironment and relatively active immune function could respond better to immunotherapy and chemotherapy, which may partly explain their improved prognosis. Besides, molecular typing based on these 7 metabolic genes can also distinguish the prognosis and immune characteristics of patients.

The study is expected to provide insights into the exploration of metabolic molecular targets with prognostic and therapeutic values, as well as individualized treatment for patients with BRCA. Of course, our research still has some shortcomings. Firstly, this study was a bioinformatic analysis dependent on databases and algorithms, and lacked experimental and functional validation *in vivo* and *in vitro*. Secondly, the analysis was mainly based on RNA data, and the integration of multi-omics methods was inadequate. Additionally, this was a retrospective study, and we need to conduct more prospective clinical observations when conditions permit.

Conclusion

In conclusion, we have constructed a prognostic signature based on seven metabolism-related genes that could distinguish the risk and immune microenvironment characteristics of breast cancer patients. And this metabolic signature could predict the responsiveness of patients to immunotherapy and chemotherapy. It can also help to screen out patients those tend to benefit from a specific clinical therapy, and improve prognosis. These results may also contribute to the discovery of novel metabolic molecular targets with therapeutic value and provide new insights into the understanding of breast cancer and its microenvironment.

Acknowledgements

The datasets presented and analyzed in this study were obtained from the TCGA database

(<https://portal.gdc.cancer.gov>) and the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). This research was supported by the National Key Research & Development Program of China (2018YFA0507401) and the National Natural Science Foundation of China (318-70910, 81871253 and 82101854).

Disclosure of conflict of interest

None.

Address correspondence to: Shulei Yin, Sheng Xu and Yizhi Yu, National Key Laboratory of Medical Immunology and Institute of Immunology, Naval Medical University, Shanghai 200433, China. E-mail: yinsl@immunol.org (SLY); xusheng@immunol.org (SX); yuyz@immunol.org (YZY)

References

- [1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021; 71: 209-249.
- [2] Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, Watson M, Davies S, Bernard PS, Parker JS, Perou CM, Ellis MJ and Nielsen TO. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J Natl Cancer Inst* 2009; 101: 736-750.
- [3] Zuo S, Yu J, Pan H and Lu L. Novel insights on targeting ferroptosis in cancer therapy. *Biomark Res* 2020; 8: 50.
- [4] Vander Heiden MG and DeBerardinis RJ. Understanding the intersections between metabolism and cancer biology. *Cell* 2017; 168: 657-669.
- [5] Fouad YA and Aanei C. Revisiting the hallmarks of cancer. *Am J Cancer Res* 2017; 7: 1016-1036.
- [6] Liu Y, Zhou Q, Song S and Tang S. Integrating metabolic reprogramming and metabolic imaging to predict breast cancer therapeutic responses. *Trends Endocrinol Metab* 2021; 32: 762-775.
- [7] Pavlova NN and Thompson CB. The emerging hallmarks of cancer metabolism. *Cell Metab* 2016; 23: 27-47.
- [8] Girgis H, Masui O, White NM, Scorilas A, Rotondo F, Seivwright A, Gabriel M, Filter ER, Girgis AH, Bjarnason GA, Jewett MA, Evans A, Al-Haddad S, Siu KM and Yousef GM. Lactate dehydrogenase A is a potential prognostic marker in clear cell renal cell carcinoma. *Mol Cancer* 2014; 13: 101.

A metabolism-related prognostic signature for breast cancer patients

- [9] Klysz D, Tai X, Robert PA, Craveiro M, Cretenet G, Oburoglu L, Mongellaz C, Floess S, Fritz V, Matias MI, Yong C, Surh N, Marie JC, Huehn J, Zimmermann V, Kinet S, Dardalhon V and Taylor N. Glutamine-dependent alpha-ketoglutarate production regulates the balance between T helper 1 cell and regulatory T cell generation. *Sci Signal* 2015; 8: ra97.
- [10] Bader JE, Voss K and Rathmell JC. Targeting metabolism to improve the tumor microenvironment for cancer immunotherapy. *Mol Cell* 2020; 78: 1019-1033.
- [11] Oh MH, Sun IH, Zhao L, Leone RD, Sun IM, Xu W, Collins SL, Tam AJ, Blosser RL, Patel CH, Englert JM, Arwood ML, Wen J, Chan-Li Y, Tenora L, Majer P, Rais R, Slusher BS, Horton MR and Powell JD. Targeting glutamine metabolism enhances tumor-specific immunity by modulating suppressive myeloid cells. *J Clin Invest* 2020; 130: 3865-3884.
- [12] Li W, Tanikawa T, Kryczek I, Xia H, Li G, Wu K, Wei S, Zhao L, Vatan L, Wen B, Shu P, Sun D, Kleer C, Wicha M, Sabel M, Tao K, Wang G and Zou W. Aerobic glycolysis controls myeloid-derived suppressor cells and tumor immunity via a specific CEBPB isoform in triple-negative breast cancer. *Cell Metab* 2018; 28: 87-103. e6.
- [13] Gong Y, Ji P, Yang YS, Xie S, Yu TJ, Xiao Y, Jin ML, Ma D, Guo LW, Pei YC, Chai WJ, Li DQ, Bai F, Bertucci F, Hu X, Jiang YZ and Shao ZM. Metabolic-pathway-based subtyping of triple-negative breast cancer reveals potential therapeutic targets. *Cell Metab* 2021; 33: 51-64. e9
- [14] Balar AV, Galsky MD, Rosenberg JE, Powles T, Petrylak DP, Bellmunt J, Loriot Y, Necchi A, Hoffman-Censits J, Perez-Gracia JL, Dawson NA, van der Heijden MS, Dreicer R, Srinivas S, Retz MM, Joseph RW, Drakaki A, Vaishampayan UN, Sridhar SS, Quinn DI, Duran I, Shaffer DR, Eigl BJ, Grivas PD, Yu EY, Li S, Kadel EE 3rd, Boyd Z, Bourgon R, Hegde PS, Mariathasan S, Thastrom A, Abidoye OO, Fine GD and Bajorin DF; IMvigor210 Study Group. Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. *Lancet* 2017; 389: 67-76.
- [15] Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, Berent-Maoz B, Pang J, Chmielowski B, Cherry G, Seja E, Lomeli S, Kong X, Kelley MC, Sosman JA, Johnson DB, Ribas A and Lo RS. Genomic and transcriptomic features of response to Anti-PD-1 therapy in metastatic melanoma. *Cell* 2016; 165: 35-44.
- [16] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43: e47.
- [17] Yu G, Wang LG, Han Y and He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012; 16: 284-287.
- [18] Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szgyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J and Ponten F. Proteomics. Tissue-based map of the human proteome. *Science* 2015; 347: 1260419.
- [19] Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B and Liu XS. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res* 2017; 77: e108-e110.
- [20] Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, Liu J, Freeman GJ, Brown MA, Wucherpfennig KW and Liu XS. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* 2018; 24: 1550-1558.
- [21] Geeleher P, Cox N and Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* 2014; 9: e107468.
- [22] Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, Sivakumar IKA, Gentles AJ, Malhotra R, Farshidfar F, Colaprico A, Parker JS, Mose LE, Vo NS, Liu J, Liu Y, Rader J, Dhankani V, Reynolds SM, Bowlby R, Califano A, Cherniack AD, Anastassiou D, Bedognetti D, Mokrab Y, Newman AM, Rao A, Chen K, Krasnitz A, Hu H, Malta TM, Noushmehr H, Pedomallu CS, Bullman S, Ojesina AI, Lamb A, Zhou W, Shen H, Choueiri TK, Weinstein JN, Guinney J, Saltz J, Holt RA and Rabkin CS; Cancer Genome Atlas Research Network, Lazar AJ, Serody JS, Demicco EG, Disis ML, Vincent BG and Shmulevich I. The immune landscape of cancer. *Immunity* 2018; 48: 812-830. e814.
- [23] Salemm V, Centonze G, Cavallo F, Defilippi P and Conti L. The crosstalk between tumor cells and the immune microenvironment in breast cancer: implications for immunotherapy. *Front Oncol* 2021; 11: 610303.
- [24] Voss K, Larsen SE and Snow AL. Metabolic reprogramming and apoptosis sensitivity: defining the contours of a T cell response. *Cancer Lett* 2017; 408: 190-196.

A metabolism-related prognostic signature for breast cancer patients

- [25] Guerra B, Recio C, Aranda-Tavio H, Guerra-Rodriguez M, Garcia-Castellano JM and Fernandez-Perez L. The mevalonate pathway, a metabolic target in cancer therapy. *Front Oncol* 2021; 11: 626971.
- [26] Oguro H. The roles of cholesterol and its metabolites in normal and malignant hematopoiesis. *Front Endocrinol (Lausanne)* 2019; 10: 204.
- [27] Miki Y, Kidoguchi Y, Sato M, Taketomi Y, Taya C, Muramatsu K, Gelb MH, Yamamoto K and Murakami M. Dual roles of group IID phospholipase A2 in inflammation and cancer. *J Biol Chem* 2016; 291: 15588-15601.
- [28] Xiong Y, Si Y, Feng Y, Zhuo S, Cui B and Zhang Z. Prognostic value of lipid metabolism-related genes in head and neck squamous cell carcinoma. *Immun Inflamm Dis* 2021; 9: 196-209.
- [29] Ye Z, Zou S, Niu Z, Xu Z and Hu Y. A novel risk model based on lipid metabolism-associated genes predicts prognosis and indicates immune microenvironment in breast cancer. *Front Cell Dev Biol* 2021; 9: 691676.
- [30] Liu H, Xu R, Gao C, Zhu T, Liu L, Yang Y, Zeng H, Huang Y and Wang H. Metabolic molecule PLA2G2D is a potential prognostic biomarker correlating with immune cell infiltration and the expression of immune checkpoint genes in cervical squamous cell carcinoma. *Front Oncol* 2021; 11: 755668.
- [31] Kaushik AK, Shojaie A, Panzitt K, Sonavane R, Venghatakrishnan H, Manikkam M, Zaslavsky A, Putluri V, Vasu VT, Zhang Y, Khan AS, Lloyd S, Szafran AT, Dasgupta S, Bader DA, Stossi F, Li H, Samanta S, Cao X, Tsouko E, Huang S, Frigo DE, Chan L, Edwards DP, Kaiparettu BA, Mitsiades N, Weigel NL, Mancini M, McGuire SE, Mehra R, Ittmann MM, Chinnaiyan AM, Putluri N, Palapattu GS, Michailidis G and Sreekumar A. Inhibition of the hexosamine biosynthetic pathway promotes castration-resistant prostate cancer. *Nat Commun* 2016; 7: 11612.
- [32] Liu W, Jiang K, Wang J, Mei T, Zhao M and Huang D. Upregulation of GNPAT1 predicts poor prognosis and correlates with immune infiltration in lung adenocarcinoma. *Front Mol Biosci* 2021; 8: 605754.
- [33] Chokchaitaweek C, Kobayashi T, Izumikawa T and Itano N. Enhanced hexosamine metabolism drives metabolic and signaling networks involving hyaluronan production and O-GlcNAcylation to exacerbate breast cancer. *Cell Death Dis* 2019; 10: 803.
- [34] Zheng X, Li Y, Ma C, Zhang J, Zhang Y, Fu Z and Luo H. Independent prognostic potential of GNPAT1 in lung adenocarcinoma. *Biomed Res Int* 2020; 2020: 8851437.
- [35] Sahn F, Oezen I, Opitz CA, Radlwimmer B, von Deimling A, Ahrendt T, Adams S, Bode HB, Guillemin GJ, Wick W and Platten M. The endogenous tryptophan metabolite and NAD+ precursor quinolinic acid confers resistance of gliomas to oxidative stress. *Cancer Res* 2013; 73: 3225-3234.
- [36] Jacobs KR, Castellano-Gonzalez G, Guillemin GJ and Lovejoy DB. Major developments in the design of inhibitors along the kynurenine pathway. *Curr Med Chem* 2017; 24: 2471-2495.
- [37] Liu CL, Cheng SP, Chen MJ, Lin CH, Chen SN, Kuo YH and Chang YC. Quinolinic acid promotes invasiveness of breast cancer through myosin light chain phosphorylation. *Front Endocrinol (Lausanne)* 2020; 11: 621944.
- [38] Tang X, Benesch MG and Brindley DN. Lipid phosphate phosphatases and their roles in mammalian physiology and pathology. *J Lipid Res* 2015; 56: 2048-2060.
- [39] Busnelli M, Manzini S, Parolini C, Escalante-Alcalde D and Chiesa G. Lipid phosphate phosphatase 3 in vascular pathophysiology. *Atherosclerosis* 2018; 271: 156-165.
- [40] Flanagan JM, Funes JM, Henderson S, Wild L, Carey N and Boshoff C. Genomics screen in transformed stem cells reveals RNASEH2A, PPAP2C, and ADARB1 as putative anticancer drug targets. *Mol Cancer Ther* 2009; 8: 249-260.
- [41] Batai K, Imler E, Pangilinan J, Bell R, Lwin A, Price E, Milinic T, Arora A, Ellis NA, Bracamonte E, Seligmann B and Lee BR. Whole-transcriptome sequencing identified gene expression signatures associated with aggressive clear cell renal cell carcinoma. *Genes Cancer* 2018; 9: 247-256.
- [42] Alcedo KP, Guerrero A, Basrur V, Fu D, Richardson ML, McLane JS, Tsou CC, Nesvizhskii AI, Welling TH, Lebrilla CB, Otey CA, Kim HJ, Omary MB and Snider NT. Tumor-selective altered glycosylation and functional attenuation of CD73 in human hepatocellular carcinoma. *Hepatol Commun* 2019; 3: 1400-1414.
- [43] Kim M, Min YK, Jang J, Park H, Lee S and Lee CH. Single-cell RNA sequencing reveals distinct cellular factors for response to immunotherapy targeting CD73 and PD-1 in colorectal cancer. *J Immunother Cancer* 2021; 9: e002503.
- [44] Lu JC, Zhang PF, Huang XY, Guo XJ, Gao C, Zeng HY, Zheng YM, Wang SW, Cai JB, Sun QM, Shi YH, Zhou J, Ke AW, Shi GM and Fan J. Amplification of spatially isolated adenosine pathway by tumor-macrophage interaction induces anti-PD1 resistance in hepatocellular carcinoma. *J Hematol Oncol* 2021; 14: 200.

A metabolism-related prognostic signature for breast cancer patients

- [45] Ma XL, Shen MN, Hu B, Wang BL, Yang WJ, Lv LH, Wang H, Zhou Y, Jin AL, Sun YF, Zhang CY, Qiu SJ, Pan BS, Zhou J, Fan J, Yang XR and Guo W. CD73 promotes hepatocellular carcinoma progression and metastasis via activating PI3K/AKT signaling by inducing Rap1-mediated membrane localization of P110beta and predicts poor prognosis. *J Hematol Oncol* 2019; 12: 37.
- [46] Magagna I, Gourdin N, Kieffer Y, Licaj M, Mhaidly R, Andre P, Morel A, Vincent-Salomon A, Paturel C and Mechta-Grigoriou F. CD73-mediated immunosuppression is linked to a specific fibroblast population that paves the way for new therapy in breast cancer. *Cancers (Basel)* 2021; 13: 5878.
- [47] Petruk N, Tuominen S, Akerfelt M, Mattsson J, Sandholm J, Nees M, Yegutkin GG, Jukkola A, Tuomela J and Selander KS. CD73 facilitates EMT progression and promotes lung metastases in triple-negative breast cancer. *Sci Rep* 2021; 11: 6035.
- [48] Ochoa de Olza M, Navarro Rodrigo B, Zimmermann S and Coukos G. Turning up the heat on non-immunoreactive tumours: opportunities for clinical development. *Lancet Oncol* 2020; 21: e419-e430.
- [49] Wang J, Wang Y, Chu Y, Li Z, Yu X, Huang Z, Xu J and Zheng L. Tumor-derived adenosine promotes macrophage proliferation in human hepatocellular carcinoma. *J Hepatol* 2021; 74: 627-637.
- [50] Bernhardt S, Bayerlova M, Vetter M, Wachter A, Mitra D, Hanf V, Lantzsch T, Uleer C, Peschel S, John J, Buchmann J, Weigert E, Burring KF, Thomssen C, Korf U, Beissbarth T, Wiemann S and Kantelhardt EJ. Proteomic profiling of breast cancer metabolism identifies SHMT2 and ASCT2 as prognostic factors. *Breast Cancer Res* 2017; 19: 112.
- [51] Yin K. Positive correlation between expression level of mitochondrial serine hydroxymethyltransferase and breast cancer grade. *Oncotargets Ther* 2015; 8: 1069-1074.
- [52] Chen L, He J, Zhou J, Xiao Z, Ding N, Duan Y, Li W and Sun LQ. EIF2A promotes cell survival during paclitaxel treatment in vitro and in vivo. *J Cell Mol Med* 2019; 23: 6060-6071.
- [53] Li X, Zhang K, Hu Y and Luo N. ERRalpha activates SHMT2 transcription to enhance the resistance of breast cancer to lapatinib via modulating the mitochondrial metabolic adaptation. *Biosci Rep* 2020; 40: BSR20192465.
- [54] Agarwal S, Chakravarthi B, Behring M, Kim HG, Chandrashekar DS, Gupta N, Bajpai P, Elkholy A, Balasubramanya SAH, Hardy C, Duffalha SA, Varambally S and Manne U. PAICS, a purine nucleotide metabolic enzyme, is involved in tumor growth and the metastasis of colorectal cancer. *Cancers (Basel)* 2020; 12: 772.
- [55] Du B, Zhang Z, Di W, Xu W, Yang L, Zhang S, He G, Yang R and Wang M. PAICS is related to glioma grade and can promote glioma growth and migration. *J Cell Mol Med* 2021; 25: 7720-7733.
- [56] Meng M, Chen Y, Jia J, Li L and Yang S. Knockdown of PAICS inhibits malignant proliferation of human breast cancer cell lines. *Biol Res* 2018; 51: 24.
- [57] Zhou S, Yan Y, Chen X, Wang X, Zeng S, Qian L, Wei J, Yang X, Zhou Y, Gong Z and Xu Z. Roles of highly expressed PAICS in lung adenocarcinoma. *Gene* 2019; 692: 1-8.
- [58] Zhu Z, Zhang M, Wang W, Zhang P, Wang Y and Wang L. Global characterization of metabolic genes regulating survival and immune infiltration in osteosarcoma. *Front Genet* 2021; 12: 814843.
- [59] Cassetta L and Pollard JW. Targeting macrophages: therapeutic approaches in cancer. *Nat Rev Drug Discov* 2018; 17: 887-904.
- [60] Wang X, Tokheim C, Gu SS, Wang B, Tang Q, Li Y, Traugh N, Zeng Z, Zhang Y, Li Z, Zhang B, Fu J, Xiao T, Li W, Meyer CA, Chu J, Jiang P, Cejas P, Lim K, Long H, Brown M and Liu XS. In vivo CRISPR screens identify the E3 ligase Cop1 as a modulator of macrophage infiltration and cancer immunotherapy target. *Cell* 2021; 184: 5357-5374, e22.
- [61] Duhamel M, Rodet F, Delhem N, Vanden Abeele F, Kobeissy F, Nataf S, Pays L, Desjardins R, Gagnon H, Wisztorski M, Fournier I, Day R and Salzet M. Molecular consequences of proprotein convertase 1/3 (PC1/3) inhibition in macrophages for application to cancer immunotherapy: a proteomic study. *Mol Cell Proteomics* 2015; 14: 2857-2877.
- [62] Fridman WH, Zitvogel L, Sautes-Fridman C and Kroemer G. The immune contexture in cancer prognosis and treatment. *Nat Rev Clin Oncol* 2017; 14: 717-734.
- [63] Zhang SC, Hu ZQ, Long JH, Zhu GM, Wang Y, Jia Y, Zhou J, Ouyang Y and Zeng Z. Clinical implications of tumor-infiltrating immune cells in breast cancer. *J Cancer* 2019; 10: 6175-6184.
- [64] Lai J, Lin X, Cao F, Mok H, Chen B and Liao N. CDKN1C as a prognostic biomarker correlated with immune infiltrates and therapeutic responses in breast cancer patients. *J Cell Mol Med* 2021; 25: 9390-9401.
- [65] Ali HR, Chlon L, Pharoah PD, Markowitz F and Caldas C. Patterns of immune infiltration in breast cancer and their clinical implications: a gene-expression-based retrospective study. *PLoS Med* 2016; 13: e1002194.
- [66] Zahran AM, Shaltout AS, Fakhry H, Khallaf SM, Abdel Fattah ON, Temerik DF and Rayan A.

A metabolism-related prognostic signature for breast cancer patients

- Prognostic significance of circulating CD28 negative suppressor T cells and memory B cells in patients with breast cancer. *Iran J Immunol* 2020; 17: 95-110.
- [67] Tsvetkov P, Coy S, Petrova B, Dreishpoon M, Verma A, Abdusamad M, Rossen J, Joesch-Cohen L, Humeidi R, Spangler RD, Eaton JK, Frenkel E, Kocak M, Corsello SM, Lutsenko S, Kanarek N, Santagata S and Golub TR. Copper induces cell death by targeting lipoylated TCA cycle proteins. *Science* 2022; 375: 1254-1261.
- [68] Hernando-Calvo A, Cescon DW and Bedard PL. Novel classes of immunotherapy for breast cancer. *Breast Cancer Res Treat* 2022; 191: 15-29.

A metabolism-related prognostic signature for breast cancer patients

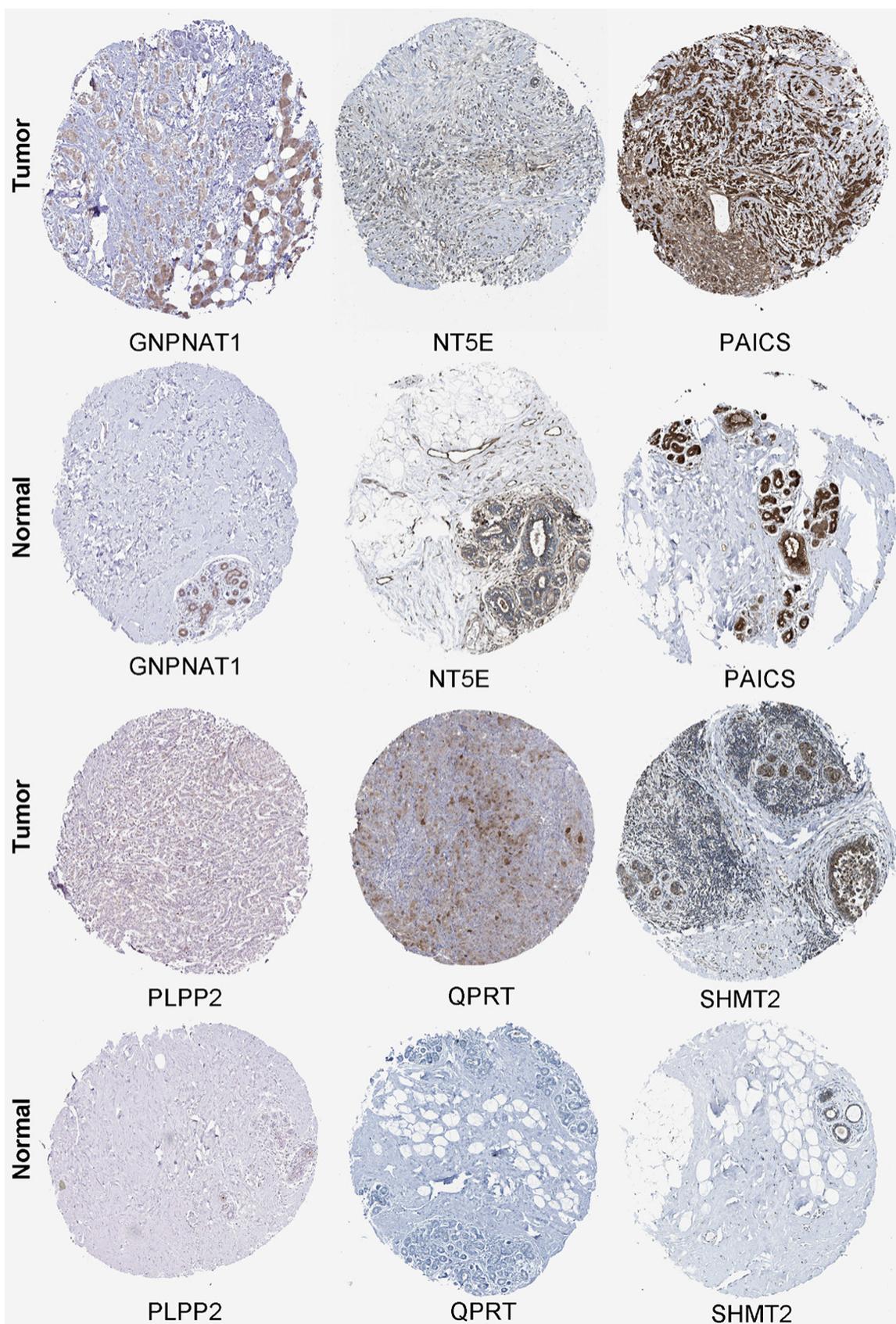
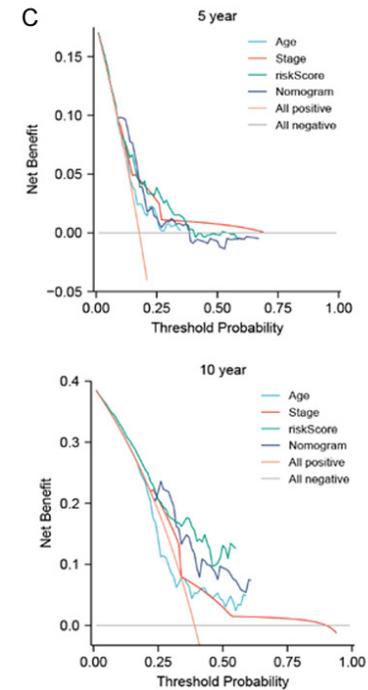
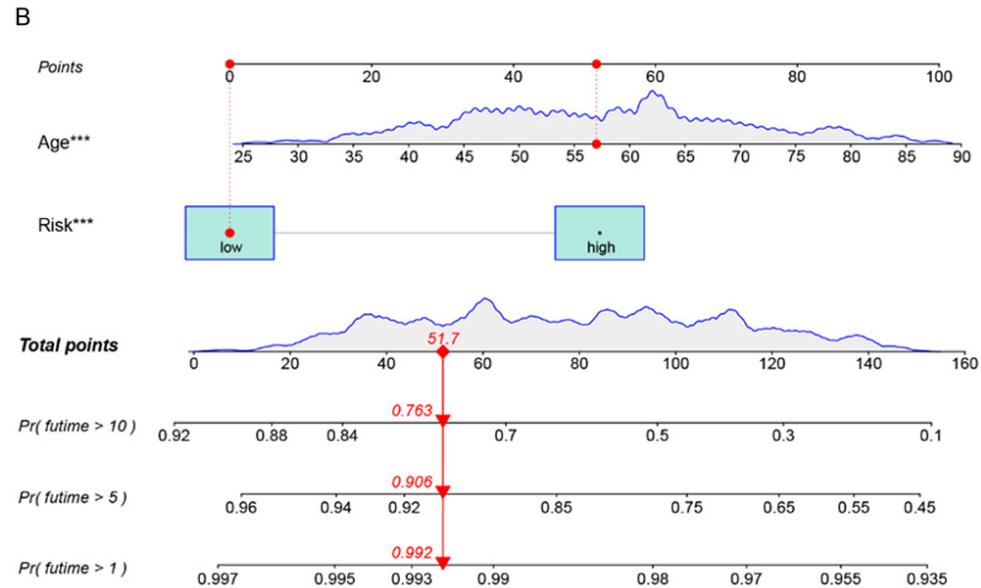
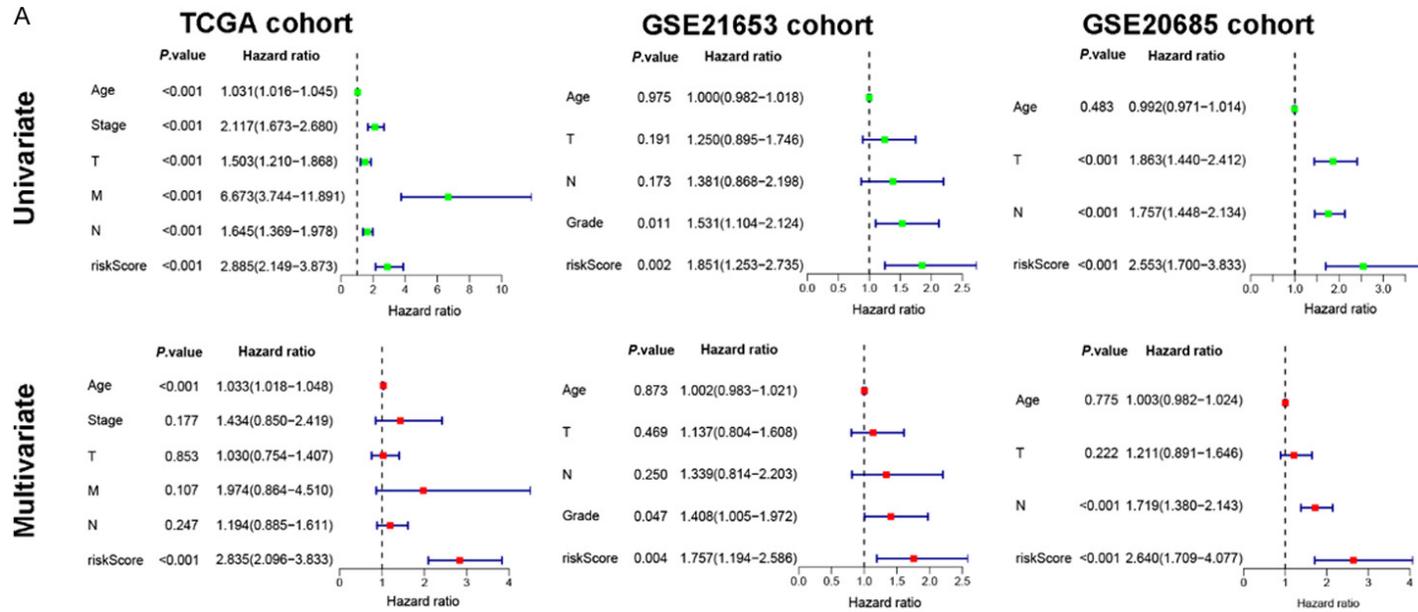


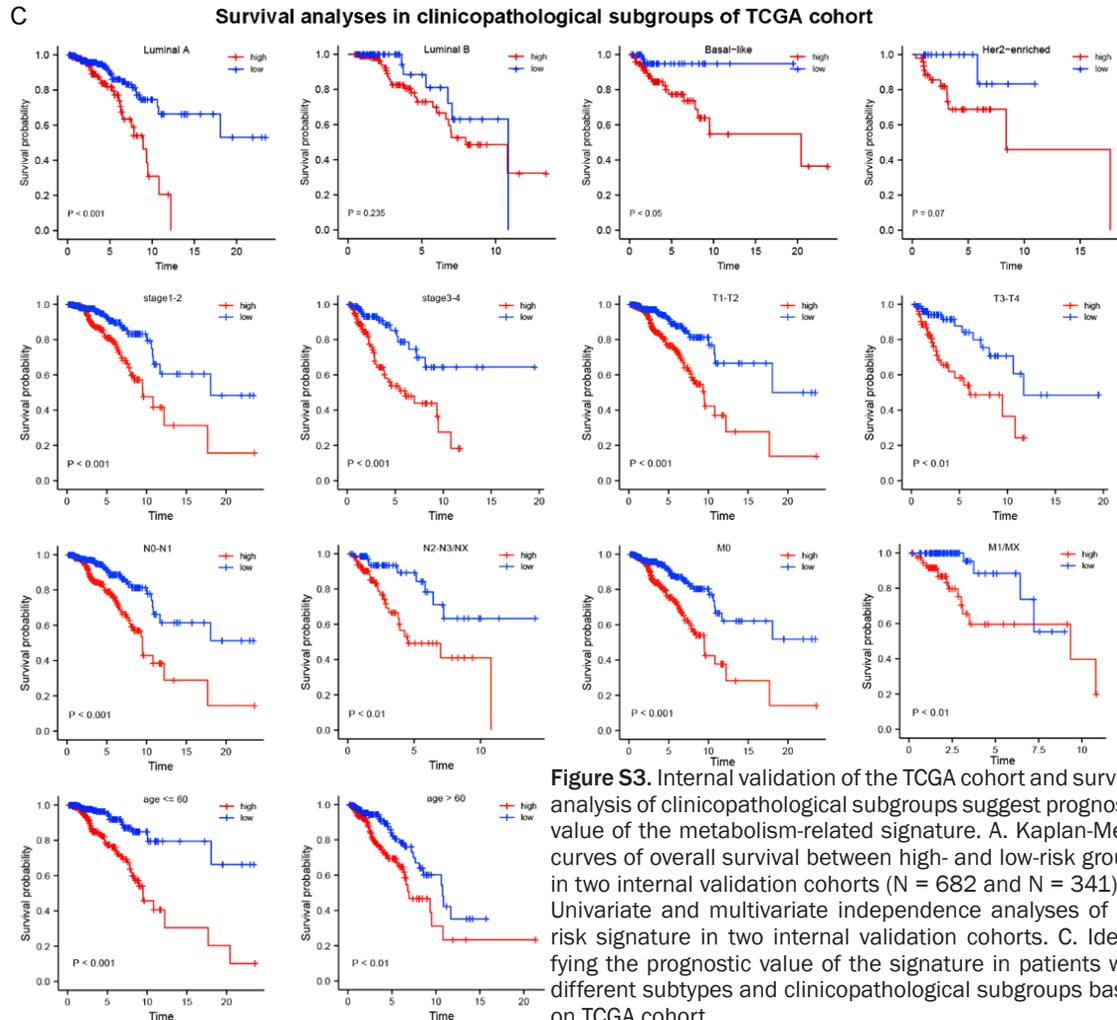
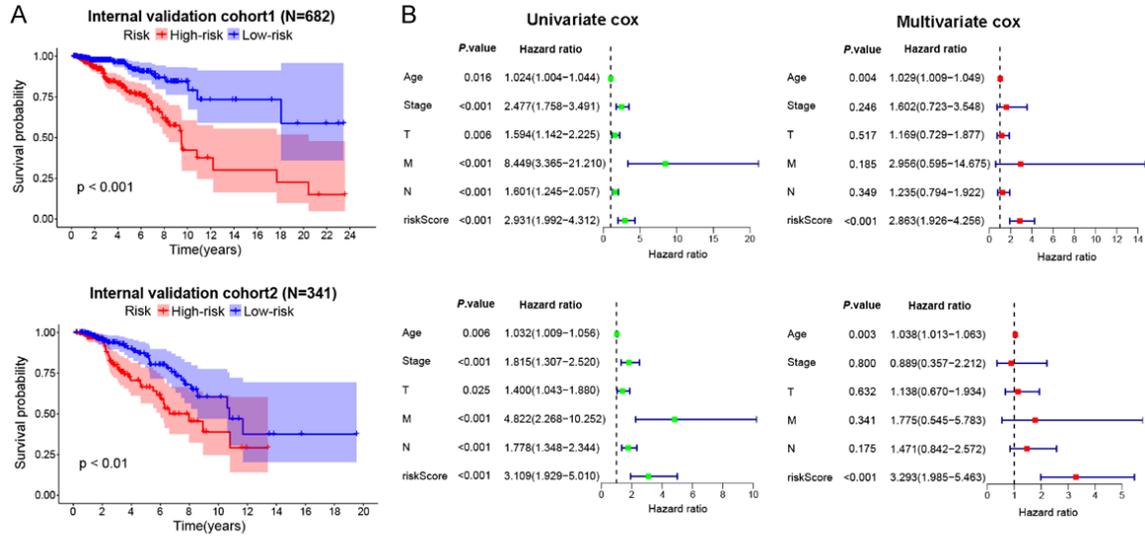
Figure S1. Validation on the protein expression of MRGs using HPA database.

A metabolism-related prognostic signature for breast cancer patients



A metabolism-related prognostic signature for breast cancer patients

Figure S2. Independence analysis and construction of the nomogram. A. Univariate and multivariate independence analysis of the prognostic signature. B. The nomogram integrating the risk score and independent prognostic factors for predicting patient survival. C. DCA curves for evaluating the performance of the risk signature and combined nomogram.



A metabolism-related prognostic signature for breast cancer patients

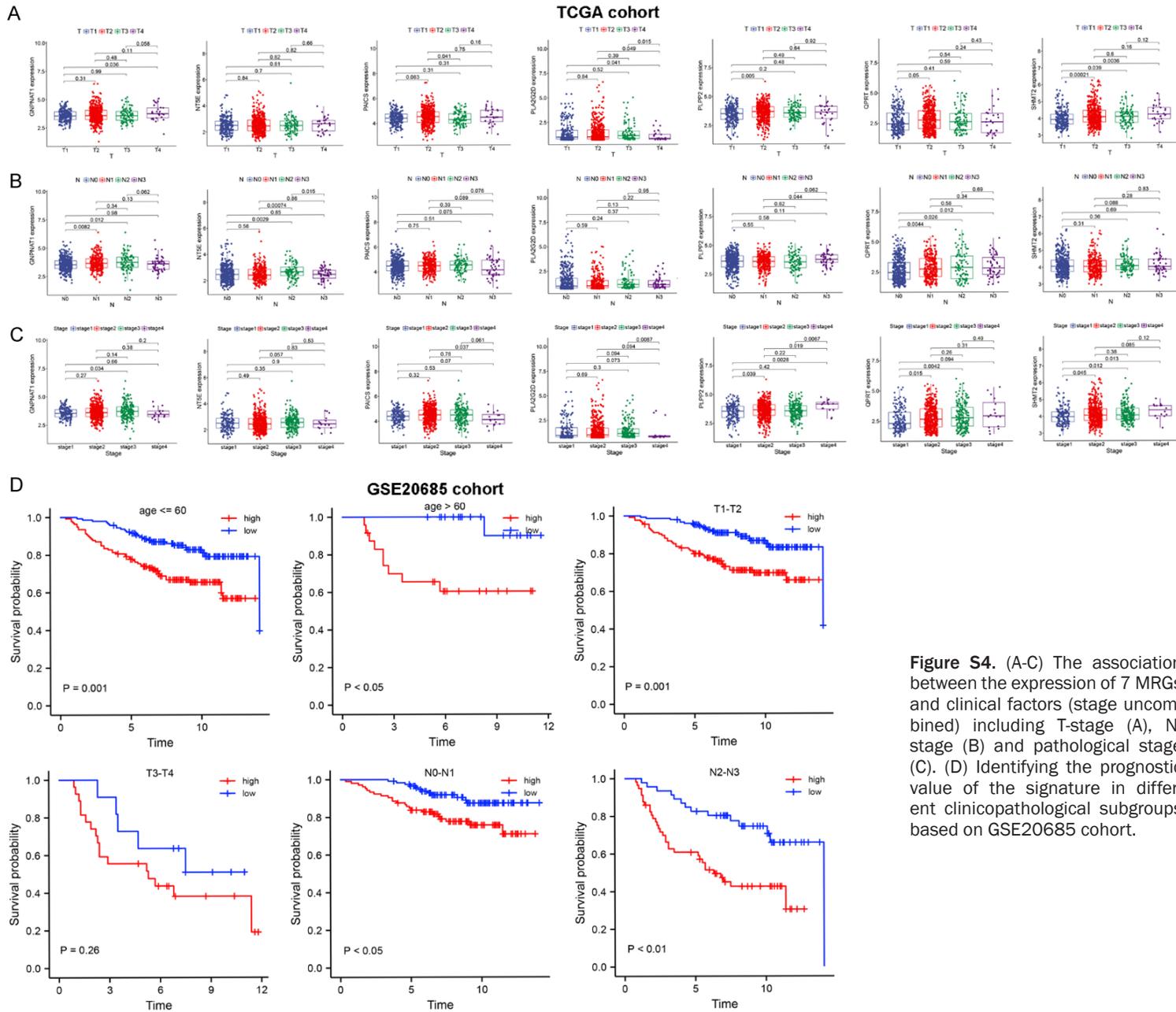


Figure S4. (A-C) The association between the expression of 7 MRGs and clinical factors (stage uncombined) including T-stage (A), N-stage (B) and pathological stage (C). (D) Identifying the prognostic value of the signature in different clinicopathological subgroups based on GSE20685 cohort.

