

# A glycolysis-related gene expression signature in predicting recurrence of breast cancer

Jianing Tang

Wuhan University Zhongnan Hospital

Gaosong Wu (✉ [wugaoson@163.com](mailto:wugaoson@163.com))

Wuhan University Zhongnan Hospital

---

## Research

**Keywords:** Breast cancer, glycolysis, prognosis

**Posted Date:** May 8th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-27021/v1>

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

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Aging on November 16th, 2020. See the published version at <https://doi.org/10.18632/aging.103806>.

# Abstract

**Background** Metabolic change is the hallmark of cancer. Even in the presence of oxygen, cancer cells reprogram their glucose metabolism to enhance glycolysis and reduce oxidative phosphorylation. In the present study, we aimed to develop a glycolysis-related gene signature to predict the prognosis of breast cancer patients.

**Methods** Gene expression profiles and clinical data of breast cancer patients were obtained from the GEO database. Univariate, Lasso-penalized, and multivariate Cox analysis were performed to construct the glycolysis-related gene signature.

**Results** A four-gene based signature (ALDH2, PRKACB, STMN1 and ZNF292) was developed to separate patients into high-risk and low-risk groups. Kaplan-Meier survival analysis demonstrated that patients in low-risk group had significantly better prognosis than those in the high-risk group. Time-dependent ROC analysis demonstrated that the glycolysis-related gene signature had excellent prognostic accuracy. We further confirmed the expression of the four prognostic genes in breast cancer and paracancerous tissues samples using qRT-PCR analysis. Expression level of PRKACB was higher in paracancerous tissues, while STMN1 and ZNF292 were overexpressed in tumor samples. No difference was found in ALDH2 expression. The same results were observed in the IHC data from the human protein atlas. Global proteome data of 105 TCGA breast cancer samples obtained from the Clinical Proteomic Tumor Analysis Consortium were used to evaluate the prognostic value of their protein levels. Consistently, high expression of PRKACB protein level was associated with better prognosis, while high ZNF292 and STMN1 protein expression levels indicated poor prognosis.

**Conclusions** The glycolysis-related gene signature might provide an effective prognostic predictor and a new view for individual treatment of breast cancer patients.

## Background

Breast cancer is the most common malignant tumor and is also the primary cause of cancer death among women around the world. According to the Global Cancer Statistics 2018, breast cancer accounts for 24% of cancer diagnosed and 15% of cancer mortality in female[1]. Although incidence of breast cancer is markedly higher in developed countries, almost 50% of new breast cancer diagnoses and approximately 60% breast cancer deaths occur in the developing countries. Breast cancer survival rate also varies largely worldwide, 5-year survival rate is estimated at 80% in developed countries while 40% in developing countries[2]. Due to the improve of therapeutic strategies, breast cancer-related deaths have been observed to decrease in recent decades. Unfortunately, some breast cancer patients initially diagnosed with advanced stage are still incurable. And nearly 30% patients diagnosed at early stage disease will eventually develop locoregional or distant tumor recurrence[3]. In the majority of breast cancer patients, metastatic disease is the underlying cause of death and current clinical strategies fall short in accurately identifying patients at high risk of recurrence. It is important to understand the

molecular mechanisms underlying the recurrent process of breast cancer and innovative biomarkers for prognosis predication.

The metabolic alterations are hallmarks of cancer cells, which could distinguish them from the normal cells. This altered metabolism is necessary for cancer cells to sustain high proliferative rates in a hostile environment[4]. Under aerobic conditions, glucose is first processed to pyruvate in the cytosol via glycolysis, and thereafter to CO<sub>2</sub> in the mitochondria. Under anaerobic conditions, most cells prefer the glycolysis, little pyruvate is dispatched to the oxygen-consuming mitochondria. In the 1920s, Otto Warburg first observed a unique metabolic phenotype of cancer cells: even in the presence of oxygen, cancer cells can reprogram their glucose metabolism, thus enhance glycolysis and reduce oxidative phosphorylation, which has been termed “aerobic glycolysis”[5, 6]. Upregulated glycolysis is associated with the capabilities of attenuation of apoptosis, avoidance of cytostatic controls, and cell proliferation[7]. Activation of oncogenes and mutations of tumor suppressors are implicated in the metabolism of breast cancer cells. The MYC proto-oncogene is critical regulator of cell proliferation, differentiation, and apoptosis[8]. In addition, c-MYC activation upregulates expression of glycolytic target genes (PFK1, GLUT, ENO, and HK) and LDH, which contribute directly to the aerobic glycolysis[9]. Under hypoxic conditions, HIF1 $\alpha$  upregulates glycolysis and downregulates phosphorylation by inducing expression of glucose transporters, glycolytic enzymes, pyruvate dehydrogenase kinase 1 and lactate dehydrogenase A[10–12]. Wild type TP53 inhibits the expression of glucose transporters and glycolytic enzymes, thereby suppressing glycolysis[13, 14]. Targeting glycolysis is a promising therapeutic strategy for cancer therapy.

In the present study, we identified a glycolysis-related risk signature and constructed a nomogram for patients with breast cancer. These results might provide an effective prognostic predictor and a new view for individual treatment of breast cancer patients.

## Methods

### Patient clinical parameter and the gene expression data

Gene expression profiles and clinical data of breast cancer patients (GSE21653, GSE20685, and GSE25055) were downloaded from the GEO database, cases with unknown survival information were excluded from our study. GSE21653 was assigned as the training set, GSE42568 and GSE25055 were used for validation. We used the Robust Multi-array Average (RMA) method to normalize raw microarray datasets, including background correction, log<sub>2</sub> transformation and normalization. Probes were changed into gene symbols using corresponding annotation files. And we scaled the RNA expression data with a standard deviation of 1 and a mean of 0. A total of 290 glycolysis-related genes were obtained from the Molecular Signature Database v7.0.

### Construction of the prognostic gene signature and nomogram

Univariate, Lasso and multivariate Cox regression analysis were performed to identify the prognostic gene signature. Univariate Cox regression analysis was firstly carried out to investigate the correlation between the expression level of each gene and patient survival. Genes with P value <0.05 in the univariate Cox regression analysis were considered as prognostic genes and selected for further analysis. Then we applied Lasso-penalized Cox analysis with 10-fold cross-validation to narrow the genes for prognostic prediction. 26 glycolysis-related genes were screened out. Finally, we conducted a stepwise multivariate Cox regression analysis to assess the contribution of a gene as an independent prognostic factor for patient survival. The predictive model was established by the expression levels of the glycolysis-related genes and their relative coefficient ( $\beta$ ) derived from the multivariate Cox regression model. The prognostic score = ( $\beta \times$  expression level of ALDH2) + ( $\beta \times$  expression level of PRKACB) + ( $\beta \times$  expression level of ZNF292) + ( $\beta \times$  expression level of STMN1). Subsequently, we constructed a nomogram using “rms” R package and plotted the calibrate curve.

### **Validation using proteomic data**

In order to investigate the prognostic value of their protein levels, global proteome data of 105 TCGA breast cancer samples and their corresponding clinical data were downloaded from the Clinical Proteomic Tumor Analysis Consortium (<https://cptac-data-portal.georgetown.edu/cptacPublic/>). The Human Protein Atlas (<http://www.proteinatlas.org>) was also used to validate the immunohistochemistry of prognostic genes.

### **Validation in human breast cancer samples**

25 paired breast cancer and paracancerous tissues samples were collected from patients undergoing modified radical mastectomy at Zhongnan Hospital of Wuhan University. All the breast tumors samples were confirmed by two pathological specialists independently. The samples were immediately frozen and stored in liquid nitrogen. We isolated total RNA from breast cancer and paracancerous tissue samples and performed qRT-PCR analysis to validate the expression of prognostic genes in human samples. This study was reviewed and approved by the Ethical Board at the Zhongnan Hospital of Wuhan University with written informed consent from all the patients

### **Statistical analysis**

To investigate the prediction accuracy of the glycolysis-related classifier, we performed time-dependent receiver operating characteristic (ROC) analysis using the “survivalROC” package in R software. Patients were separated into low- and high-risk groups based on the optimal cut-off value calculated by ‘survminer’ package. Kaplan-Meier survival analysis was conducted to assess the survival differences

between low-/ high-risk groups. A two-sided log-rank test was performed using 'survival' package in R. Student's t test was used to compare the difference between two groups. We performed all statistical analysis using R software 3.6.1 and a  $P < 0.05$  was set as the probability value of statistical significance.

## Results

### Patient characteristics

A flow chart of this study was shown in figure 1. Here, we established a glycolysis-related gene signature based on 4 prognostic genes. A total of 878 patients from three datasets (GSE21653, GSE20685, and GSE25055) were included in our analysis. 241 patients from GSE21653 were assigned in the training set, 327 patients from GSE20685 and 310 patients from GSE25055 were assigned in the validation sets. In GSE21653, median follow up was 5.73 years in the low-risk group and 2.99 years in the high-risk group. 28 patients in the low-risk group (20.1%) and 48 patients in the high-risk group (47.1%) had locoregional or distant recurrence during the follow-up period. Similar results were found in the validation sets (Table 1).

### Establishment of the prognostic glycolysis-related gene signature

Based on the univariate Cox regression analysis, 129 glycolysis-related genes significantly associated with the prognosis of breast cancer patients were considered as prognostic genes and they were included in the subsequent analysis. After primary filtration, we used Lasso regression analysis to further narrow the genes for the construction of prognostic model. As a result, 26 glycolysis related genes were screened out for the stepwise multivariate Cox regression analysis. And four genes were finally selected to construct a prognostic model. Risk score was calculated for every patient based on the expression levels of the four genes and weighted by the multivariate Cox regression coefficients as follows: risk score =  $(-0.34 \times \text{expression level of ALDH2}) + (-0.28 \times \text{expression level of PRKACB}) + (0.33 \times \text{expression level of ZNF292}) + (0.32 \times \text{expression level of STMN1})$ . Consistent with the univariate Cox regression analysis, ALDH2 and PRKACA showed negative coefficients in the prognostic model, indicating their expression was associated with good prognosis. ZNF292 and STMN1 had positive coefficients, their high expression implying a shorter RFS (Figure 2). Patients in training set were divided into low- and high-risk groups using 'survminer' package in R based on the optimal cut-off of risk score (0.05). As shown in figure 2, time-dependent ROC analysis demonstrated that this signature had good prognostic accuracy. AUCs of the four-gene prognostic model were 0.771, 0.825 and 0.810 at 1, 3, and 5-year survival time. Kaplan-Meier survival analysis indicated patients in the high-risk group had significantly poor survival rate. To confirm the prognostic value of the glycolysis-related signature in different population, we stratified patients by clinicopathological risk factors. We also noted that prognosis of patients in the low-risk group was better than those in the high-risk group (Figure 3). We further developed a nomogram combining the glycolysis-related signature and clinicopathological risk factors to provide a quantitative method for

prediction of 3- and 5-year RFS. Calibration plots indicated that the nomogram had a similar performance compare to the ideal model (Figure 4).

### **Validation of the signature**

We used two external validation sets (GSE20685 and GSE 25055) to further access the prognostic value of glycolysis-related signature. Patients in the two validation sets were separated into low- and high-risk groups according to the glycolysis-related signature identified above. Consistently, in comparison with the high-risk group, significantly higher survival rates were observed in the low-risk group. In addition, the 1-, 3-, and 5-year AUCs of the glycolysis-related signature demonstrated that this signature had excellent prognostic accuracy (Figure 5).

### **Expression validation of the four genes**

To further confirm the expression of the four prognostic genes, we collected breast cancer and paracancerous tissues samples from patients undergoing modified radical mastectomy and performed qRT-PCR analysis. As show in figure 6, expression level of PRKACB was higher in paracancerous tissues, while STMN1 and ZNF292 were overexpressed in tumor samples. No difference was found in ALDH2 expression. We then investigated the prognostic value of their proteins. High expression of PRKACB protein was associated with better RFS, while high ZNF292 and STMN1 protein expression levels indicated poor prognosis, expression of ALDH2 protein was not associated with the prognosis of breast cancer patients (Figure 6E-G). Immunohistochemistry data from the Human Protein Atlas were also used to validate the expression of prognostic genes. Consistent with our qRT-PCR results, ZNF292 and STMN1 staining were higher in tumor samples, while staining of PRKACB was lower in tumor. No difference was found in ALDH2 staining (Figure 7).

## **Discussion**

It is well known that high rate of energy metabolism is the hallmark of cancer. In order to fuel cell growth and division, glucose uptake and utilization is markedly increased in various human cancer types. Interestingly, even in the presence of enough oxygen, malignant tumors prefer glycolysis rather than oxidative phosphorylation in the mitochondria[6]. The high rate of glycolysis not only ensures an adequate amount of energy for cancer cells, but also provide ample intermediate macromolecules to sustain a rapid cell proliferation and tumor mass expansion[15]. Thus, targeting tumor metabolism is a promising therapeutic strategy for cancer treatment. However, limited studies have systematically investigated the metabolic process of tumor and its prognostic values. And most prognostic biomarkers are identified in a wide range. Here, we investigate the associations between tumor and glycolysis to provide further understanding the mechanism of tumorigenesis.

In the present study, we used high-throughput expression data downloaded from GEO database and constructed a glycolysis-related signature for predicting the prognosis of patients with breast cancer. After univariate, Lasso and multivariate Cox analysis, four genes (ALDH2, PRKACB, STMN1, and ZNF292) were screened out as prognostic genes for the construction of prognostic model. We first examined the prognostic value of this signature in the training set. Patients were divided into low- and high-risk groups based on the optimal cut-off of risk score. Kaplan-Meier survival analysis with log-rank test demonstrated that patients in low-risk group had significantly better prognosis than those in the high-risk group. When stratified by clinicopathological risk factor, this glycolysis-related signature was still a significant prognostic model. The prognostic accuracy of this prognostic signature was assessed with time-dependent ROC analysis at various survival times, AUCs of the four-gene prognostic model were 0.771 at 1 year, 0.825 at 3 years and 0.810 at 5 years. A nomogram combining the glycolysis-related signature and clinicopathological risk factors was developed to provide a quantitative method for prediction of 3- and 5-year RFS. The calibration plots demonstrated that actual survival corresponded closely with predicted survival, suggesting good predictive performance of the nomogram. Two external validation sets were used to further estimate the prognostic value of the signature. Our results indicated that this glycolysis-based classifier was still a clinically and statistically significant prognostic model. The expression levels of the four prognostic genes were further validated using breast cancer and paracancerous tissues samples from patients undergoing modified radical mastectomy. Our qRT-PCR analysis demonstrated that STMN1 and ZNF292 were overexpressed in tumor samples, while PRKACB expression was higher in paracancerous tissues. Due to the small size of the patients, no difference was found in ALDH2 expression. The same results were observed in the IHC data from the human protein atlas. Global proteome data of 105 TCGA breast cancer samples obtained from the Clinical Proteomic Tumor Analysis Consortium were used to evaluate the prognostic value of their protein levels. Consistently, high expression of PRKACB protein level was associated with better RFS, while high ZNF292 and STMN1 protein expression levels indicated poor prognosis.

The biological role of genes in our panel has been roughly illustrated in previous studies. ALDH2 encodes a protein belonging to the aldehyde dehydrogenase family which is the second enzyme of the major oxidative pathway of alcohol metabolism. The major role of ALDH2 is to detoxify acetaldehyde to non-toxic acetic acid. Accumulating evidences indicate that dysfunction of ALDH2 may contribute to human cancer[16, 17]. ALDH2 was suppressed in human lung adenocarcinoma, its repression leads to ACE accumulation in lung adenocarcinoma cells and induces DNA damage and metastatic features. Suppression of ALDH2 also promotes proliferation and stemness of lung adenocarcinoma cells[18]. While its role in breast cancer has not been clearly investigated. The protein encoded by PRKACB is a catalytic subunit of cAMP (cyclic AMP)-dependent protein kinase, and its biological function is unclear. The mRNA and protein level of STMN1 were significantly higher in various types of cancer, including colon cancer, lung cancer, pancreatic ductal adenocarcinoma, and hepatocellular carcinoma. In colon cancer patients, STMN1 expression is significantly related to lymph node metastasis and TNM staging. Prognosis of STMN1-positive patients is significantly poorer than STMN1-negative patients[19–22]. ZNF292 is a transcription factor-encoding gene and is considered as a candidate tumor suppressor[23].

However, high expression of ZNF292 is associated with poor prognosis of breast cancer patients in our present study. The underlying mechanism remains unclear.

## Conclusions

In conclusion, we identified a novel glycolysis-related gene signature and constructed a nomogram that can predict the outcome of patients with breast cancer. Our results indicated this signature could be a promising prognostic target in clinical practice and provide new insights into the underlying mechanism of breast cancer.

## Declarations

### Ethics approval and consent to participate

The research was carried out according to the World Medical Association Declaration of Helsinki and was approved by the Ethics Committee at Zhongnan Hospital of Wuhan University.

### Consent for publication

Not applicable

### Availability of data and materials

Not applicable

### Competing interests

The authors have no conflicts of interest

### Funding

None

### Authors' contributions

JT and GW reviewed relevant literature and drafted the manuscript. JT conducted all statistical analyses. All authors read and approved the final manuscript.

## Acknowledgements

None.

## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2018.
2. Coleman MP, Quaresma M, Berrino F, Lutz JM, De Angelis R, Capocaccia R, Baili P, Rachet B, Gatta G, Hakulinen T, Micheli A, Sant M, Weir HK, Elwood JM, Tsukuma H, Koifman S, GA ES, Francisci S, Santaquilani M, Verdecchia A, Storm HH, Young JL. Cancer survival in five continents: a worldwide population-based study (CONCORD). *The Lancet Oncology*. 2008; 9: 730–56.
3. O'Shaughnessy J. Extending survival with chemotherapy in metastatic breast cancer. *Oncologist*. 2005;10(Suppl 3):20–9.
4. Anastasiou D. Tumour microenvironment factors shaping the cancer metabolism landscape. *British journal of cancer*. 2017;116:277–86.
5. Weinhouse S. **WARBURG O, BURK D, SCHADE ALJS**. On Respiratory Impairment in Cancer Cells. 124: 267–72.
6. Person P. Otto warburg: “on the origin of cancer cells”. 10: 412–21.
7. Semenza GL. Tumor metabolism: cancer cells give and take lactate. *J Clin Investig*. 2008;118:3835–7.
8. Adhikary S, Eilers M. Transcriptional regulation and transformation by Myc proteins. *Nature reviews Molecular cell biology*. 2005;6:635–45.
9. Miller DM, Thomas SD, Islam A, Muench D. **Sedoris KJCCRAOJotAAfCR**. c-Myc and Cancer Metabolism. 2012; 18: 5546–53.
10. Iyer NV, Kotch LE, Agani F, Leung SW, Laughner E, Wenger RH, Gassmann M, Gearhart JD, Lawler AM, Yu AY, Semenza GL. Cellular and developmental control of O<sub>2</sub> homeostasis by hypoxia-inducible factor 1 alpha. *Genes Dev*. 1998;12:149–62.
11. Ryan HE, Lo J, Johnson RS. HIF-1 alpha is required for solid tumor formation and embryonic vascularization. *EMBO J*. 1998;17:3005–15.
12. Kim JW, Tchernyshyov I, Semenza GL, Dang CV. HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. *Cell Metabol*. 2006;3:177–85.

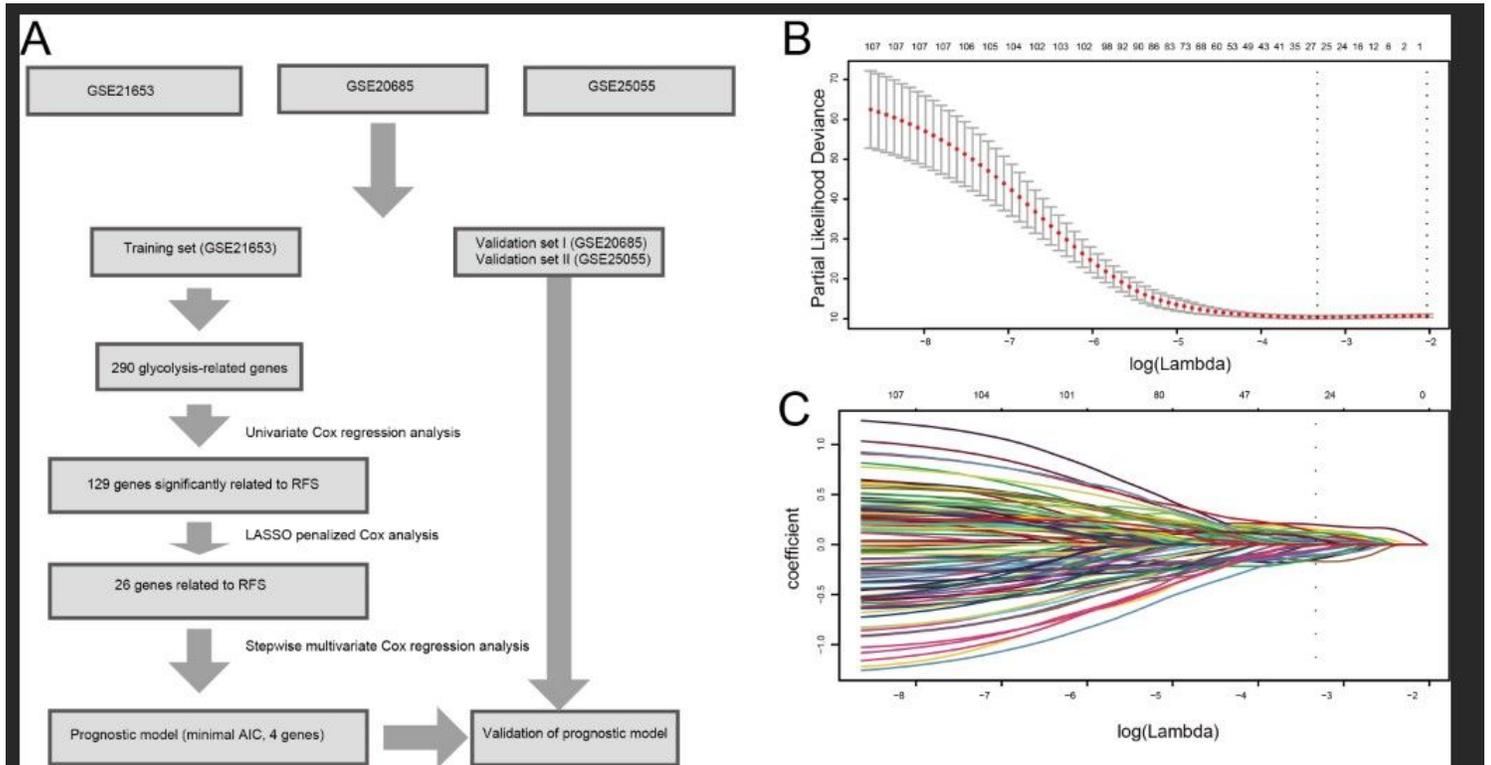
13. DeBerardinis RJ. Is cancer a disease of abnormal cellular metabolism? New angles on an old idea. *Genetics in medicine: official journal of the American College of Medical Genetics*. 2008;10:767–77.
14. Kruiswijk F, Labuschagne CF, Vousden KH. p53 in survival, death and metabolic health: a lifeguard with a licence to kill. *Nature reviews Molecular cell biology*. 2015;16:393–405.
15. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science*. 2009;324:1029–33.
16. Chang JS, Hsiao JR, Chen CH. ALDH2 polymorphism and alcohol-related cancers in Asians: a public health perspective. *Journal of biomedical science*. 2017;24:19.
17. Chen CH, Ferreira JC, Gross ER, Mochly-Rosen D. Targeting aldehyde dehydrogenase 2: new therapeutic opportunities. *Physiological reviews*. 2014;94:1–34.
18. Li K, Guo W, Li Z, Wang Y, Sun B, Xu D, Ling J, Song H, Liao Y, Wang T, Jing B, Hu M, Kuang Y, Wang Q, Yao F, Sun A, Zhu L, Wang L, Deng J. ALDH2 Repression Promotes Lung Tumor Progression via Accumulated Acetaldehyde and DNA Damage. *Neoplasia (New York NY)*. 2019;21:602–14.
19. Zhang HQ, Guo X, Guo SQ, Wang Q, Chen XQ, Li XN, Guo LS. STMN1 in colon cancer: expression and prognosis in Chinese patients. *Eur Rev Med Pharmacol Sci*. 2016;20:2038–44.
20. Suzuki K, Watanabe A, Araki K, Yokobori T, Harimoto N, Gantumur D, Hagiwara K, Yamanaka T, Ishii N, Tsukagoshi M, Igarashi T, Kubo N, Gombodorj N, Nishiyama M, Hosouchi Y, Kuwano H, Shirabe K. High STMN1 Expression Is Associated with Tumor Differentiation and Metastasis in Clinical Patients with Pancreatic Cancer. *Anticancer research*. 2018;38:939–44.
21. Jiang L, Zhao L, Bi J, Guan Q, Qi A, Wei Q, He M, Wei M, Zhao L. Glycolysis gene expression profilings screen for prognostic risk signature of hepatocellular carcinoma. *Aging*. 2019;11:10861–82.
22. Li M, Yang J, Zhou W, Ren Y, Wang X, Chen H, Zhang J, Chen J, Sun Y, Cui L, Liu X, Wang L, Wu C. Activation of an AKT/FOXM1/STMN1 pathway drives resistance to tyrosine kinase inhibitors in lung cancer. *British journal of cancer*. 2017;117:974–83.
23. Lee JH, Song SY, Kim MS, Yoo NJ, Lee SH. Frameshift mutations of a tumor suppressor gene ZNF292 in gastric and colorectal cancers with high microsatellite instability. *APMIS: acta pathologica, microbiologica, et immunologica Scandinavica*. 2016; 124: 556–60.

## Tables

**Table 1. Clinicopathologic characteristics of three sets of breast cancer patients according to the integrated mRNA signature**

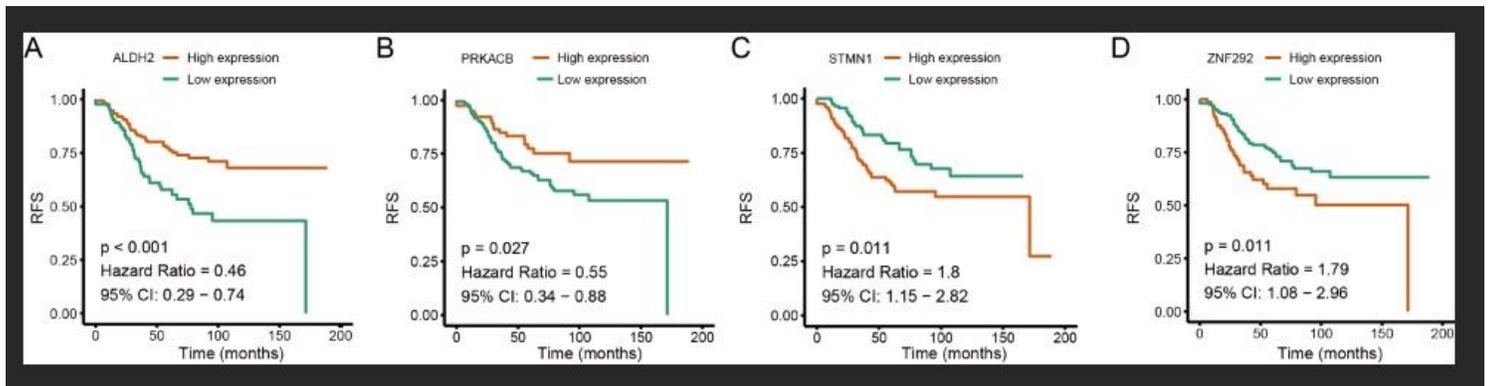
Variables	GSE21653 (n=241)		GSE20685 (n=327)		GSE25055 (n=310)	
	Low risk (%)	High risk (%)	Low risk (%)	High risk (%)	Low risk (%)	High risk (%)
<b>Age at diagnosis (years)</b>						
Median	55	56	47	44	48	50
≤ 50	44(31.7)	38(37.3)	120(63.5)	89(64.5)	100(55.2)	68(52.7)
> 50	95(68.3)	64(62.7)	69(36.5)	49(35.5)	81(44.8)	61(47.3)
<b>T stage</b>						
T0	0	0	0	0	0	2(1.6)
T1	32(23.0)	23(22.5)	68(36.0)	33(23.9)	16(8.8)	4(3.1)
T2	73(52.5)	48(47.1)	108(57.1)	80(58.0)	101(55.8)	64(49.6)
T3	34(24.5)	31(30.4)	9(4.8)	17(12.3)	39(21.5)	35(27.1)
T4	0	0	4(2.1)	8(5.8)	25(13.8)	24(18.6)
<b>Lymph node status</b>						
Negative	62(44.6)	51(50.0)	94(49.7)	43(31.2)	67(37.0)	20(15.5)
Positive	77(55.4)	51(50.0)	95(50.3)	95(68.8)	114(63.0)	109(84.5)
<b>Metastases status</b>						
Negative			188(58.9)	1(12.5)		
Positive			131(41.1)	7(87.5)		
<b>Grade</b>						
I	37(26.6)	6(5.9)			16(8.8)	3(2.3)
II	60(43.2)	23(22.5)			90(49.7)	27(20.9)
III	42(30.2)	73(71.6)			65(35.9)	86(66.7)
Indeterminate					10(5.5)	13(10.1)
<b>Disease-relapse event</b>						
	28(20.1)	48(47.1)	75(30.7)	63(75.9)	14(7.7)	52(40.3)
<b>Median Follow-up (years)</b>						
	5.73	2.99	8.7	5.7	3.0	1.8

# Figures



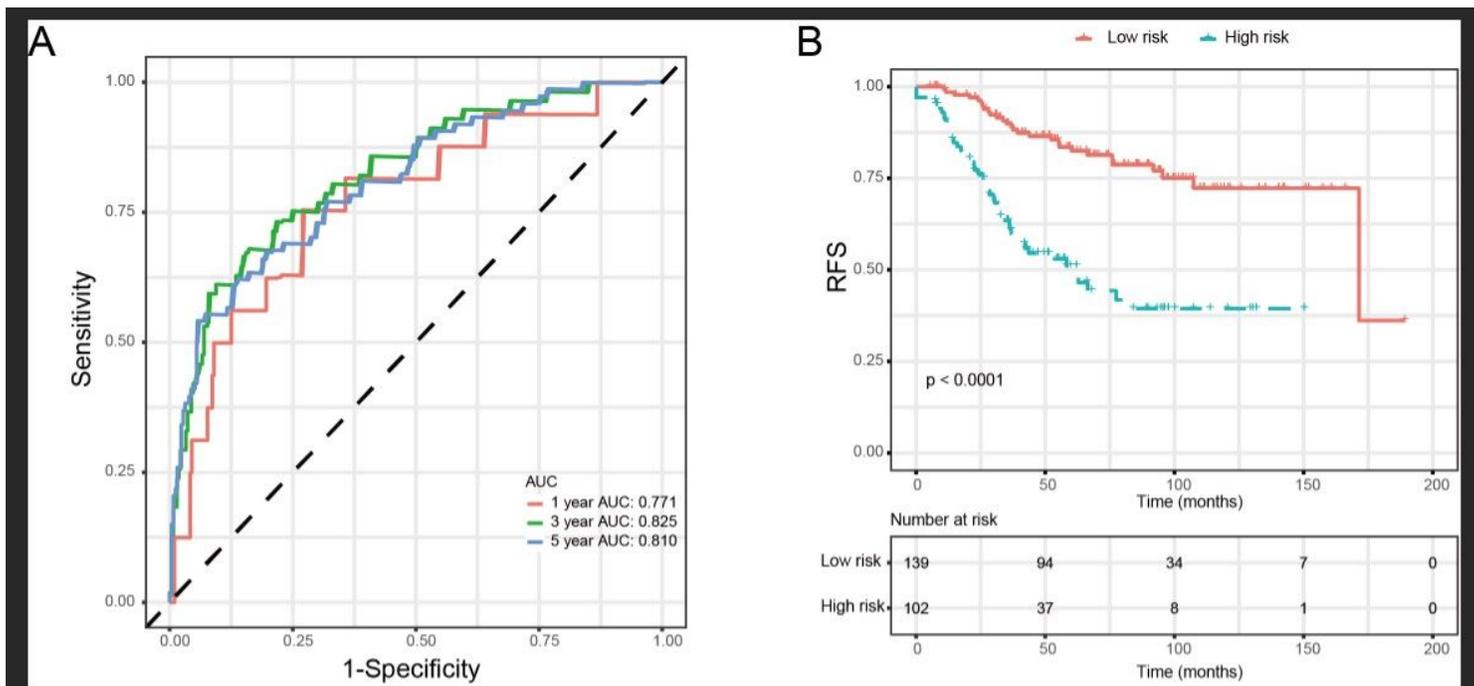
**Figure 1**

Flow chart and 10-time cross-validation for tuning parameter selection. (A). Flow chart indicating the process used to select target genes included in the analysis. (B) Ten-time cross validation for tuning parameter selection in the lasso model. (C). LASSO coefficient profiles of the 129 prognostic genes. A vertical line is drawn at the value chosen by 10-fold cross-validation.



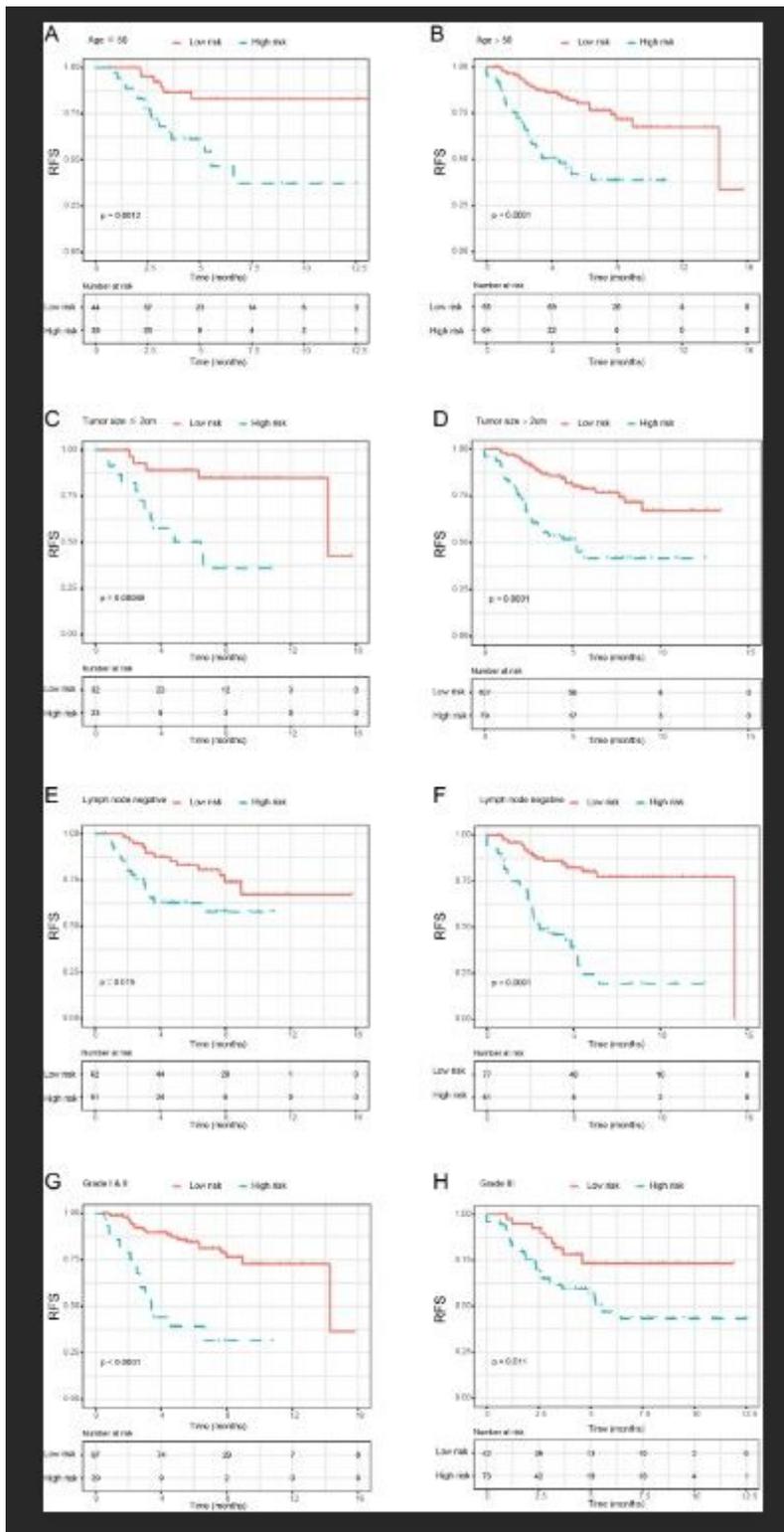
**Figure 2**

Univariate Cox regression analysis of the four prognostic genes in the signature. (A). ALDH2. (B). PRKACB. (C). STMN1. (D). ZNF292.



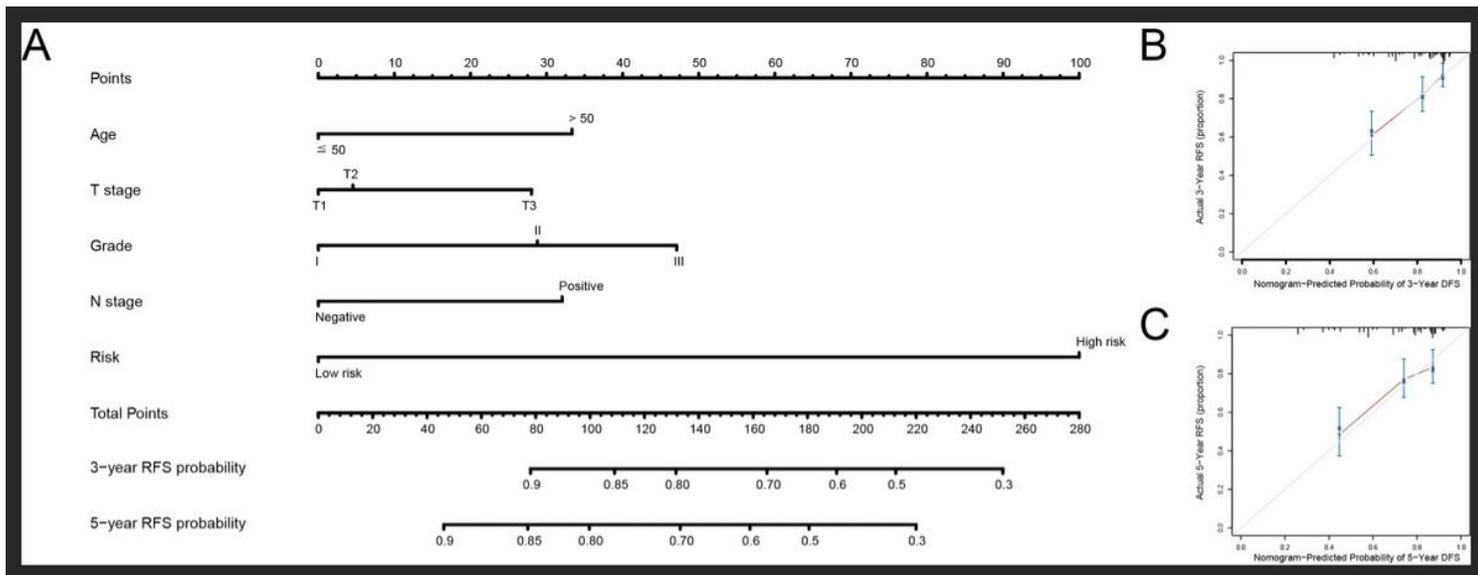
**Figure 3**

Validation of prognostic risk score model in training set. (A). Time-dependent ROC curves of the glycolysis-related signature. (B). Kaplan-Meier survival analysis of the glycolysis-related signature.



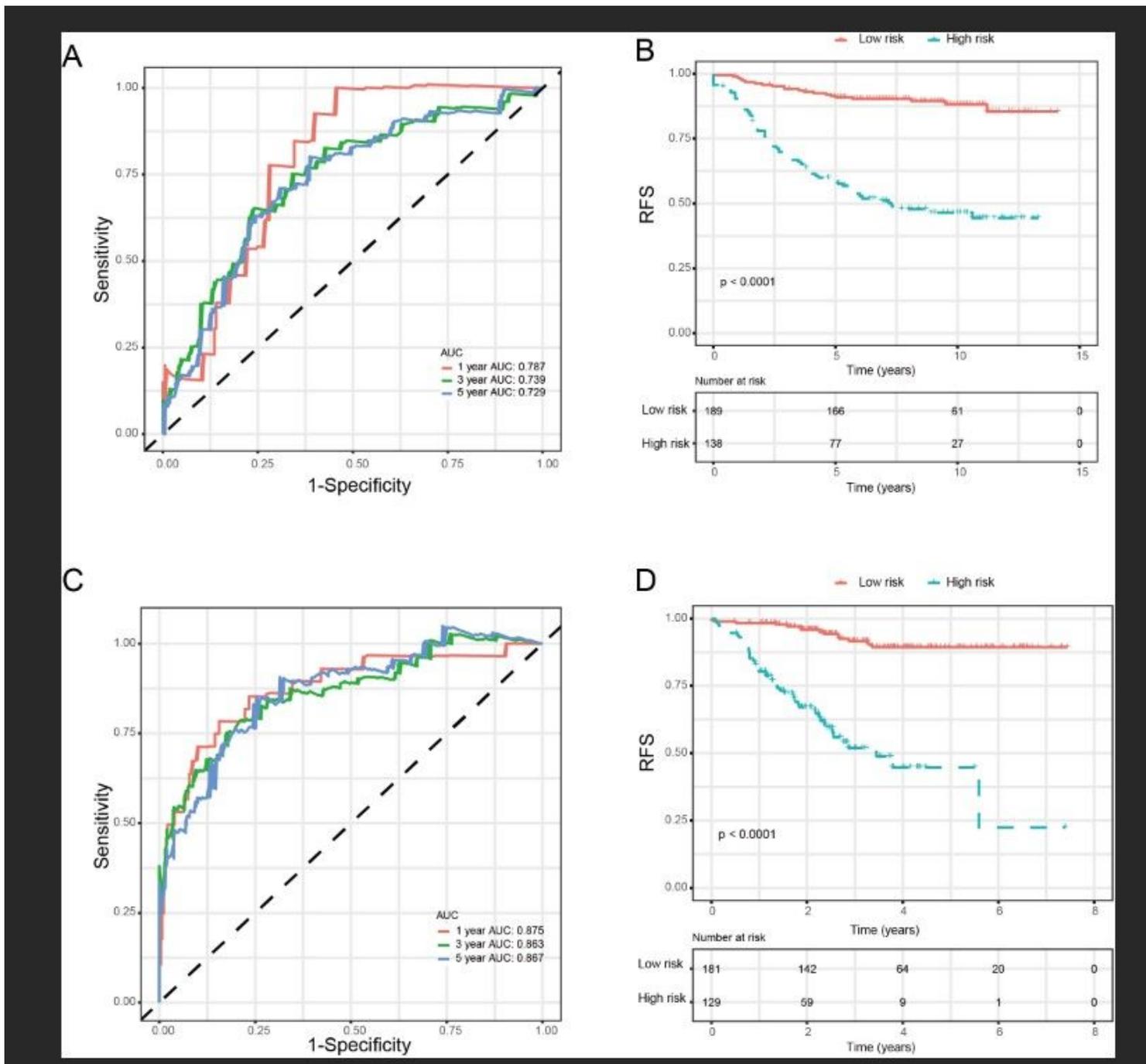
**Figure 4**

Kaplan-Meier survival analysis for patients according to the glycolysis-related signature stratified by clinicopathological risk factors. (A, B). Age. (C, D). Tumor size. (E, F). Lymph node status. (G, H). Tumor grade



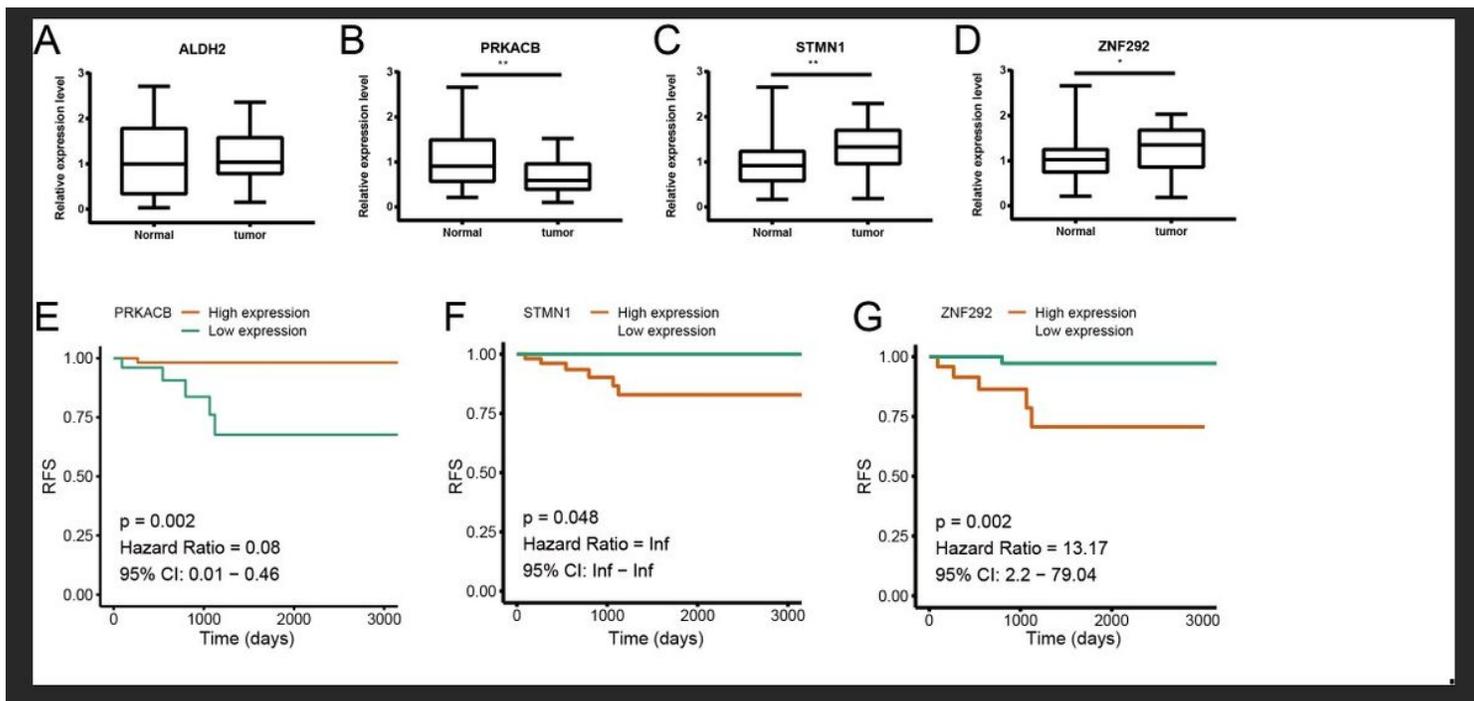
**Figure 5**

Nomogram to predict risk of cancer recurrence. (A). Nomograms to predict risk of cancer recurrence. (B). 3-year nomogram calibration curves of training set. (C). 5-year nomogram calibration curves of training set.



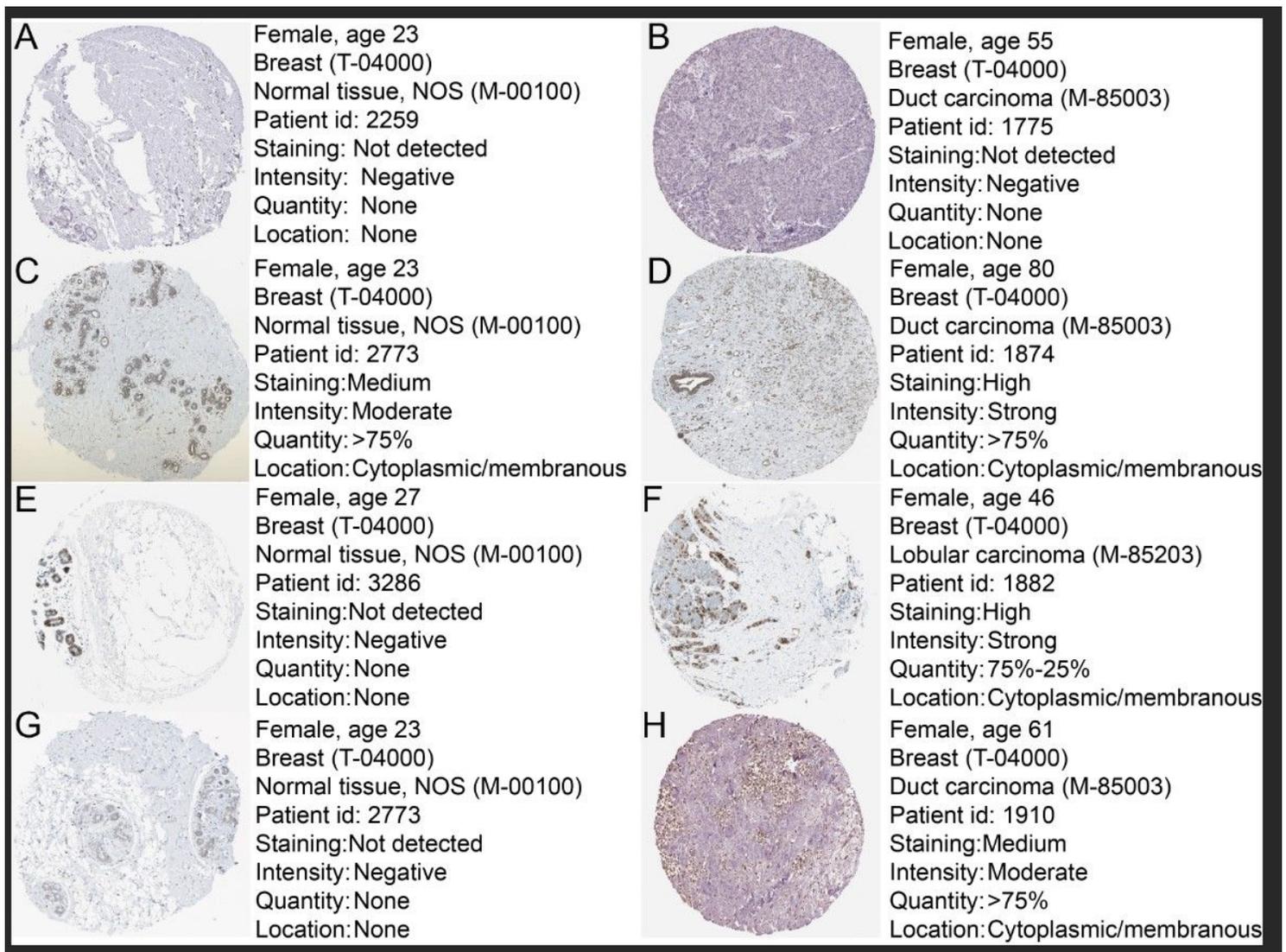
**Figure 6**

Validation of glycolysis-related signature in validation sets. (A) Time-dependent ROC curves of the glycolysis-related signature in validation set GSE20685. (B) Kaplan-Meier survival analysis of the glycolysis-related signature in validation set GSE20685. (C). Time-dependent ROC curves of the glycolysis-related signature in validation set GSE25055. (D). Kaplan-Meier survival analysis of the glycolysis-related signature in validation set GSE25055.



**Figure 7**

Validation of the four prognostic genes. (A-D). Relative mRNA expression of ALDH2, PRKACB, STMN1 and ZNF292 in breast cancer and paracancerous tissues samples. (E-G). Prognostic value of PRKACB, STMN1 and ZNF292 protein levels.



**Figure 8**

Immunohistochemistry of the four prognostic genes based on the Human Protein Atlas. (A). IHC staining of ALDH2 in normal tissue. (B). IHC staining of ALDH2 in tumor tissue. (C). IHC staining of PRKACB in normal tissue. (D). IHC staining of PRKACB in tumor tissue. (E). IHC staining of STMN1 in normal tissue. (F). IHC staining of STMN1 in tumor tissue. (G). IHC staining of ZNF292 in normal tissue. (H). IHC staining of ZNF292 in tumor tissue.