


# Development and Validation of a Scoring System Based on 9 Glycolysis-Related Genes for Prognosis Prediction in Gastric Cancer

Technology in Cancer Research & Treatment  
 Volume 19: 1-14  
 © The Author(s) 2020  
 Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
 DOI: 10.1177/1533033820971670  
[journals.sagepub.com/home/tct](https://journals.sagepub.com/home/tct)  


Tianqi Luo, MD<sup>1,2</sup>, Yufei Du, MD<sup>1,3</sup> , Jinling Duan, PhD<sup>1,4</sup>,  
 Chengcai Liang, MD<sup>1,2</sup>, Guoming Chen, MD, PhD<sup>1,2</sup>,  
 Kaiming Jiang, MD<sup>1,2</sup>, Yongming Chen, MD<sup>1,2</sup>, and Yingbo Chen, MD<sup>1,2</sup>

## Abstract

Gastric cancer is a malignant tumor with high morbidity and mortality worldwide. However, increasing evidences have revealed the correlation between the glycolysis process and tumorigenesis. This study is aim to develop a list of glycolysis-related genes for risk stratification in gastric cancer patients. We included 500 patients' sample data from GSE62254 and GSE26942 datasets, and classified patients into training ( $n = 350$ ) and testing sets ( $n = 150$ ) at a ratio of 7: 3. Univariate and multivariate Cox regression analysis were performed to screen genes having prognostic value. Based on HALLMARK gene sets, we identified 9 glycolysis-related genes (BPNT1, DCN, FUT8, GMPPA, GPC3, LDHC, ME2, PLOD2, and UGP2). On the basis of risk score developed by the 9 genes, patients were classified into high- and low-risk groups. The survival analysis showed that the high-risk patients had a worse prognosis ( $p < 0.001$ ). Similar finding was observed in the testing cohort and 2 independent cohorts (GSE13861 and TCGA-STAD, all  $p < 0.001$ ). The multivariate Cox regression analysis showed that the risk score was an independent prognostic factor for overall survival ( $p < 0.001$ ). Furthermore, we constructed a nomogram that integrated the risk score and tumor stage, age, and adjuvant chemotherapy. Through comparing the results of the receiver operating characteristic curves and decision curve analysis, we found that the nomogram had a superior predictive accuracy than conventional TNM staging system, suggesting that the risk score combined with other clinical factors (age, tumor stage, and adjuvant chemotherapy) can develop a robust prediction for survival and improve the individualized clinical decision making of the patient.

In conclusion, we identified 9 glycolysis-related genes from hallmark glycolysis pathway. Based on the 9 genes, gastric cancer patients were separated into different risk groups related to survival.

## Keywords

metabolism, nomogram, TCGA, GEO, prediction

## Abbreviations

AUC, Area Under Receiver Operating Characteristic Curve; AIC, Akaike Information Criterion; ACT, Adjuvant Chemotherapy; AJCC, American Joint Committee on Cancer; DCA, Decision Curve Analysis; FPKKM, Fragments Per Kilobase Million; GEO, Gene Expression Omnibus; GC, Gastric Cancer; GSEA, Gene Set Enrichment Analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; ROC, Receiver Operating Characteristic; TCGA, The Cancer Genome Atlas.

Received: June 29, 2020; Revised: October 8, 2020; Accepted: October 13, 2020.

<sup>1</sup> Department of Gastric Surgery, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

<sup>2</sup> State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, China

<sup>3</sup> Department of Biotherapy, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China; Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

<sup>4</sup> Department of Experimental Research (Cancer Institute), Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

## Corresponding Author:

Yingbo Chen, and Yongming Chen, Dongfeng East Road 651, Guangzhou, Guangdong 510060, China.

Emails: [chenyb@sysucc.org.cn](mailto:chenyb@sysucc.org.cn); [chenyongm@sysucc.org.cn](mailto:chenyongm@sysucc.org.cn)



## Introduction

Gastric cancer (GC) is the fifth most frequently diagnosed cancer and the third leading cause of cancer death in 2018.<sup>1</sup> Although surgical resection, chemotherapy, and immunotherapy have been applied to treat patients, the overall survival of patients with advanced GC is still poor, especially when acquiring resistance to chemotherapy agents.<sup>2,3</sup> Currently, the American Joint Committee on Cancer (AJCC) staging system has been widely conducted to predict the prognosis of GC.<sup>4</sup> Nevertheless, it fails to accurately predict prognosis in a considerable number of the patients with same tumor stage, because of high heterogeneity of GC. Therefore, it is urgent to explore an effective way to guide personalized treatment decisions.

Cellular Metabolic changes play an important role in tumor initiation and progression, and one of the most notable alterations is glycolysis.<sup>5</sup> The aerobic glycolysis of tumor cells (the Warburg effect) was discovered in 1927, which showed that tumor tissue could absorb glucose and secrete lactic acid even when oxygen is present.<sup>6</sup> This phenomenon was also discovered in GC,<sup>7</sup> and it may reveal the potential that aerobic glycolysis is responsible for tumor progression and poor prognosis.<sup>8</sup> Presently, numerous bioinformatics methods have been used to explore the link between glycolysis and cancer prognosis.<sup>9,10</sup> which help us to better understand the mechanism of tumorigenesis.

Therefore, this study is aim to develop a signature based on glycolysis-related gene for GC. On the basis of this signature, patients with survival difference were classified into different groups, which might help oncologists to make individualized clinical decision making.

## Materials and Methods

### Data Collection and Process

First, we downloaded the microarray data of GC from the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) database. The detailed inclusion criteria for candidate datasets are following: Human gene expression profile; GC specimen; samples' total count  $\geq 60$ ; available information of follow-up time (overall survival), pathological stage, and adjuvant chemotherapy. Then, 3 datasets (GSE62254, GSE26942, and GSE13861) were included in this study. Furthermore, The Cancer Genome Atlas (TCGA) sequencing data, as the external validation cohort, was obtained from UCSC Xena website (<https://gdc.xenahubs.net>). Similarly, microarray data, GSE13861 data series, was used to validate the model based on glycolysis gene. The patients' clinicopathological characteristics were shown in the Table 1.

Data series from GEO database had been normalized by uploader. For TCGA cohort, sequencing data was normalized as Fragments Per Kilobase Million (FPKM) value. According to the annotation file, the probe identifications were transformed into gene symbols. When a gene symbol was corresponded to multiple probes, the average was taken as the final value. In addition, we merged 2 GEO datasets (GSE62254

and GSE26942) to develop model. Because of the different value ranges between independent data sets, Empirical Bayes method ("sva" package) was performed to diminish the batch effect of merged dataset. Finally, the merged dataset was grouped randomly into training and testing cohorts at the ratio of 7:3.

### Bioinformatics Analysis

Based on the HALLMARK gene set (h.all.v6.1.symbols.gmt), we utilized the pathway named "glycolysis" to perform subsequent analysis, and it included 200 genes encoding proteins involved in glycolysis and gluconeogenesis. Gene Set Enrichment Analysis (GSEA) software 4.0.2 was applied to identify the enriched pathways between the tumor and normal tissues. The univariate Cox regression was carried out to filter glycolysis-related genes having prognostic value, and the genes with  $p < 0.01$  were performed for further multivariate Cox regression analysis to construct risk score ("survival" package).

In order to explore the different tumor-infiltrating patterns of immune cell between the different risk-groups, we used CIBERSORT algorithm (<https://cibersort.stanford.edu/index.php>) to identify 22 types of immune infiltration cells for each sample, and the output result of  $p$ -value  $< 0.05$  can be considered acceptable. Gene MANIA (<http://www.genemania.org>) website is a tool for construction of interaction network and annotation of gene function. We used it to construct a network of identified genes and help to identify hub genes.

### Visualization and Estimation of Model

In order to further perfect the predictive model, available clinical information was combined with risk score to develop a more robust model, and nomogram was conducted to visualize it ("rms" package). Calibration curve was used to display the performance of nomogram. Receiver Operating Characteristic (ROC) analysis ("time ROC" package) was used to estimate the accuracy of nomogram. However, a model that had a much higher specificity but slightly lower sensitivity than another would have a greater the area under ROC curve (AUC) whereas would be a worse decision for patients. Accordingly, Decision curve analysis (DCA) ("stdca" source code) was used to evaluate whether the model would be of benefit in the clinical practice.<sup>11</sup>

### Statistical Analysis

R version 3.6.0 (<http://www.r-project.org>) was performed to analyze data. Wilcoxon and Kruskal-Wallis tests were conducted to deal with continuous variables with non-normal distribution. Categorical variables were analyzed by chi-square analysis. Univariate and multivariate Cox regression analysis to investigate whether the variable was an independent prognostic factor. Overall survival is defined as the period from the diagnosis to death, even the precise cause of death is not specified. And the survival curves and its significant differences

**Table 1.** Clinicopathological Characteristics.

	GSE62254	GSE26942	GSE13861	TCGA-STAD
N	300	200	64	337
Age (median)	64.0 (55.0, 70.0)	59.0 (50.0, 67.0)	62.50 (54.5, 69.0)	67.0 (58.0, 72.0)
Gender				
Female	101 (33.7)	59 (29.5)	19 (29.7)	119 (35.3)
Male	197 (65.7)	140 (70.0)	45 (70.3)	218 (64.7)
Unknown	2 (0.6)	1 (0.5)	0 (0.0)	0 (0.0)
Tumor stage				
I	30 (10.0)	51 (25.5)	12 (18.8)	45 (13.4)
II	96 (32.0)	36 (18.0)	11 (17.2)	107 (31.8)
III	145 (48.3)	99 (49.5)	33 (51.5)	150 (44.4)
IV	27 (9.0)	13 (6.5)	8 (12.5)	21 (6.2)
Unknown	2 (0.7)	1 (0.5)	0 (0.0)	14 (4.2)
Lauren classification				
Diffuse	134 (44.7)	40 (20.0)	29 (45.3)	58 (17.2)
Intestinal	146 (48.6)	141 (70.5)	19 (29.7)	71 (21.1)
Mixed	17 (5.7)	7 (3.5)	12 (18.8)	1 (0.3)
Unknown	3 (1.0)	12 (6.0)	4 (6.2)	207 (61.4)
Adjuvant chemotherapy				
No	156 (52.0)	94 (47.0)	15 (23.4)	35 (10.4)
Yes	144 (48.0)	106 (53.0)	49 (76.6)	16 (4.7)
Unknown	0 (0.0)	0 (0.0)	0 (0.0)	286 (84.9)
Tumor location				
Antrum	150 (50.0)	109 (54.5)	26 (40.6)	120 (35.6)
Body	117 (39.0)	84 (42.0)	32 (50.0)	121 (35.9)
Cardia	32 (10.7)	7 (3.5)	4 (6.2)	84 (24.9)
Unknown	1 (0.3)	0 (0.0)	2 (3.2)	12 (3.6)
Histological grade				
G1/G2	123 (41.0)	0 (0.0)	0 (0.0)	129 (38.3)
G3	177 (59.0)	0 (0.0)	0 (0.0)	199 (59.0)
Unknown	0 (0.0)	200 (100.0)	64 (100.0)	9 (2.7)
Overall survival				
Dead	152 (50.7)	88 (44.0)	29 (45.3)	142 (42.1)
Alive	148 (49.3)	112 (56.0)	35 (54.7)	195 (54.9)
Survival time (median)	57.9 (17.8, 78.9)	38.0 (14.3, 66.5)	87.50 (32.8, 91.0)	16.2 (9.7, 27.3)

were estimated by the Kaplan–Meier method and log-rank test. Based on 2-sided statistical tests, *p*-value is of statistically significant when  $p < 0.05$ .

## Results

### Development of the Glycolysis Gene Signature

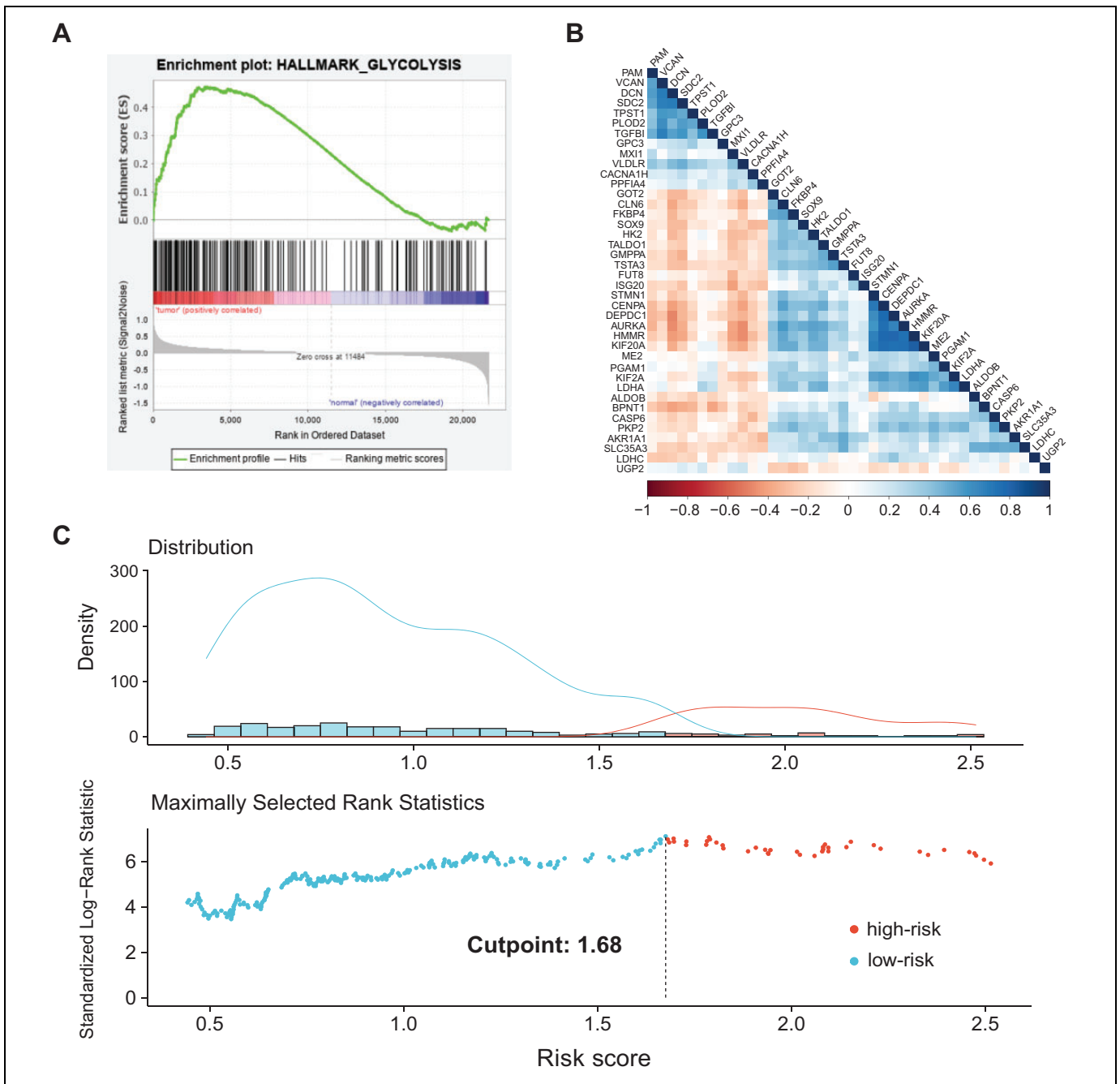
To determine whether the hallmark glycolysis pathway showed significant difference between GC and normal samples, we included GSE66229 dataset to perform GSEA analysis based on HALLMARK gene set. GSE66229 was composed of GSE62254 subseries and 100 case normal tissues, and the result revealed that hallmark glycolysis pathway was significantly enriched in the GC sample group ( $p = 0.031$ , FDR  $q < 0.25$ ,  $|NES| \geq 1$ ; Figure 1A). Before merging the 2 geo datasets (GSE62254, and GSE26942), we excluded 12 cases of normal tissue, 3 cases of gastric stromal tumor, and 2 cases that follow-up time was less than 30 days from GSE26942 dataset. Finally, a total of 500 cases were randomly divided into training set (350) and testing (150) at the ratio of 7:3. In the Figure S1, we

used box plot to demonstrate alteration of gene expression level of merged dataset after eliminating batch effect.

By conducting the univariate Cox regression analysis in the training set, 40 genes related to survival were screened ( $p < 0.01$ ). However, there was a strong correlation observed in these genes (Figure 1B). Therefore, we performed the multivariate Cox regression to further screen these genes. Based on the smallest value of Akaike Information Criterion (AIC) statistics, we obtained 9 glycolysis-related genes lastly (BPNT1, DCN, FUT8, GMPPA, GPC3, LDHC, ME2, PLOD2, and UGP2). The formula consisting of the 9 genes and the corresponding coefficients is the following:

$$\begin{aligned} \text{Risk score} = & \text{BPNT1} \times (-0.361) + \text{DCN} \times (-0.204) \\ & + \text{FUT8} \times (-0.381) + \text{GMPPA} \times (-0.440) \\ & + \text{GPC3} \times 0.185 + \text{LDHC} \times (-1.028) \\ & + \text{ME2} \times (-0.446) + \text{PLOD2} \times 0.300 \\ & + \text{UGP2} \times 0.9666 \end{aligned}$$

Based on the formula developed by the 9 genes, we calculated the risk score for each patient and figure out optimal

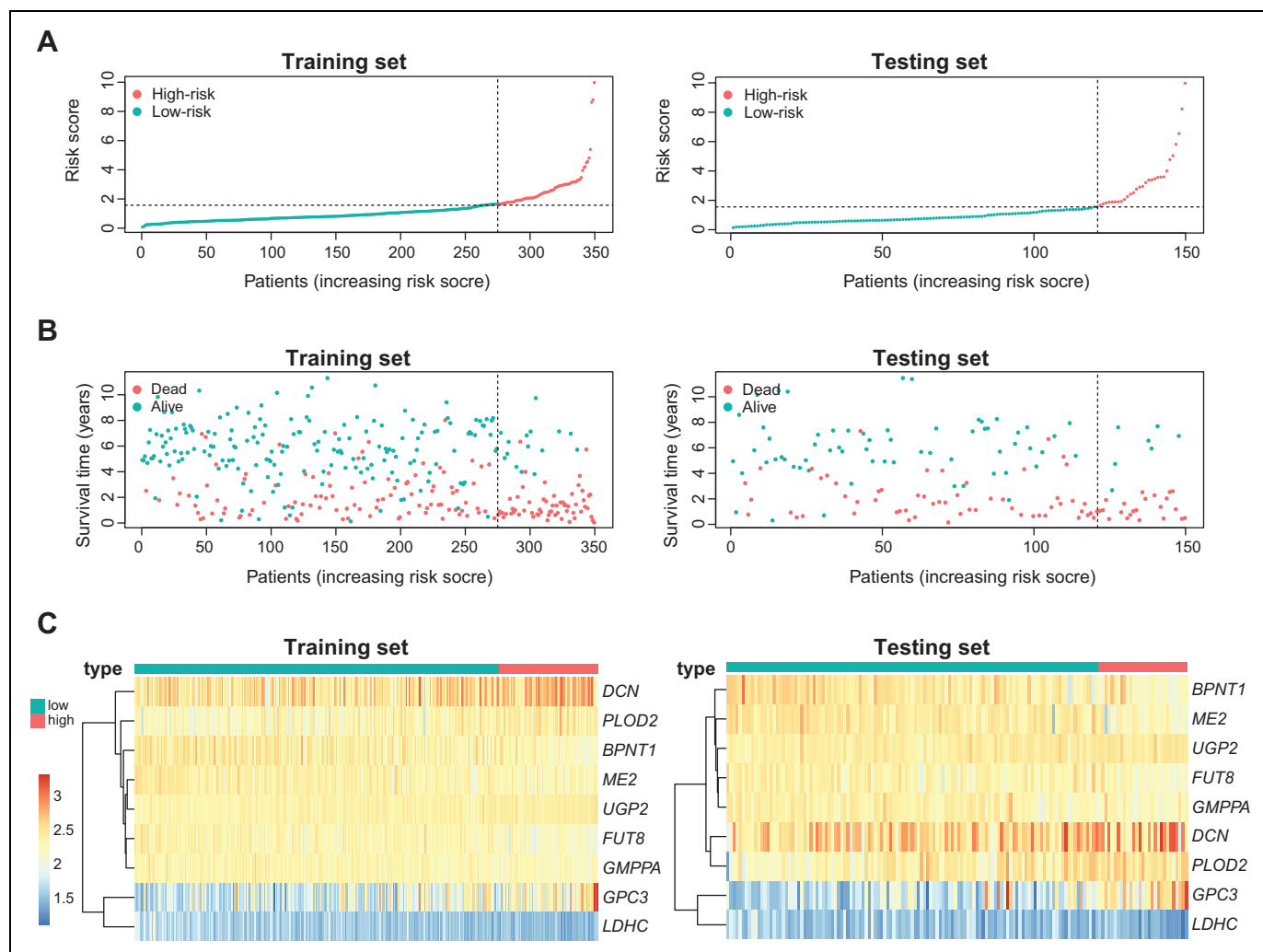


**Figure 1.** Construction of the Risk Score in the Training Cohort. (A). GSEA Plot of HALLMARK Glycolysis Pathway between Gastric cancer and paired normal Tissues. (B). The Correlation analysis of the 40 glycolysis-related genes (C). The Distribution Plot of Optimal cut-off value in the training cohort.

cutoff value (1.68) that can generate largest survival difference between the 2 groups (Figure 1C). Based on the cutoff value, these patients were classified into high- and low-risk groups. The distribution of risk scores, the patients' survival status, the heatmap of the 9 glycolysis genes expression are displayed in Figure 2. It is observed that survival time of patients gradually decreased with an increasing risk score, and high-risk score patients had more end-point events (Figure 2B).

### Validation and Evaluation of the Signature

In the training cohort, the survival analysis demonstrated that patients in the high-risk group had a worse prognosis than their low-risk counterparts (5-years overall survival: 22.5% vs. 64.4%;  $p < 0.001$ ; Figure 3A). After adjusting for age, gender, tumor stage, and adjuvant chemotherapy, the result of multi-variate Cox regression showed that the risk score was



**Figure 2.** Risk Score Based on 9 Glycolysis-Related Genes in the Training and Testing Cohorts. (A). Distribution of the Risk Score in the Training and Testing Cohorts. (B). Survival State of Patients in the Training and Testing Cohorts. (C). The Heatmap of the 9 Glycolysis-related genes in the Training and testing Cohorts.

independently associated with overall survival (HR = 1.60; 95%CI: 1.42 -1.81;  $p < 0.001$ ; Table 2).

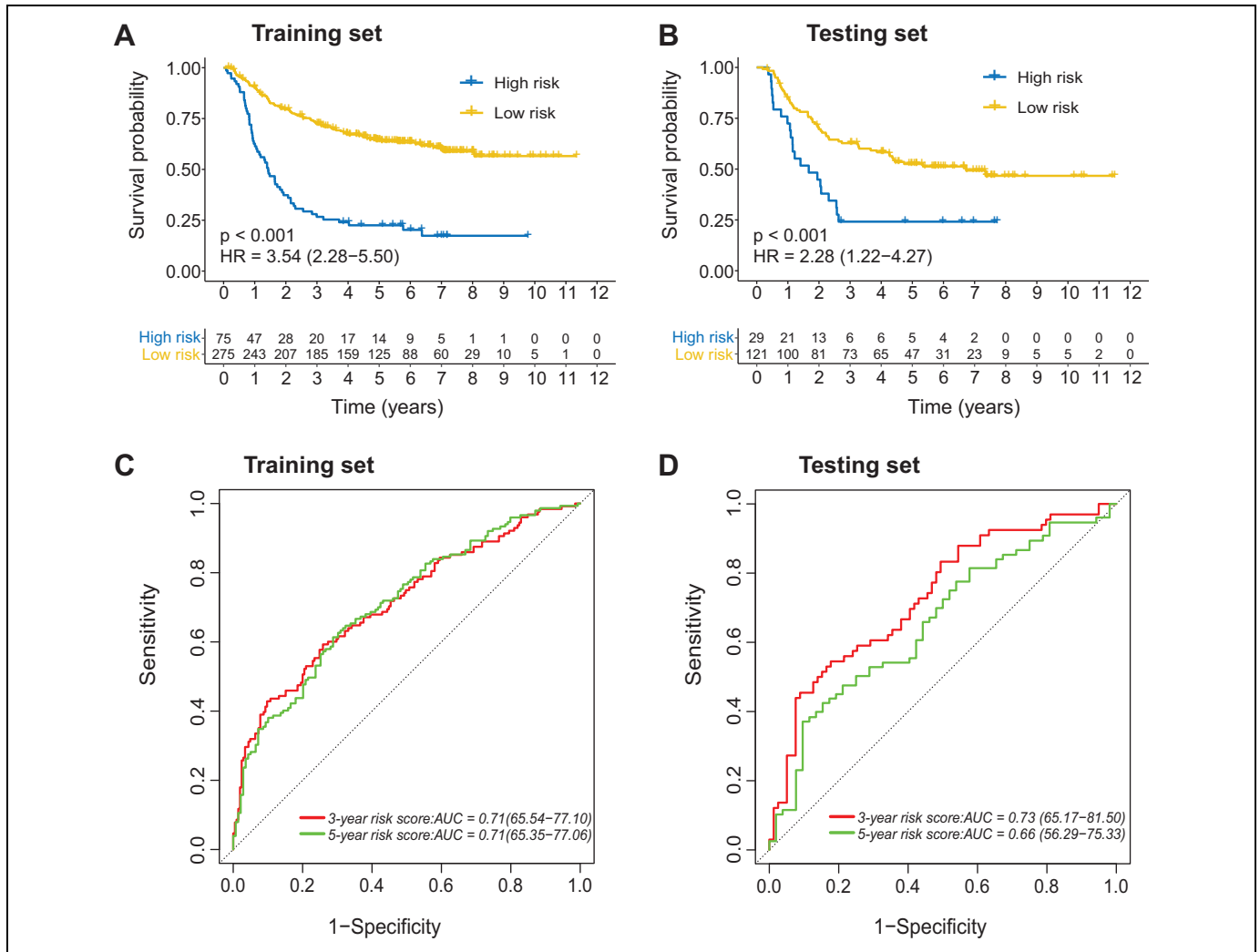
Next, we used the testing set to validate the prognostic value of the signature. Consistently, the result showed that 5-year overall survival differed significantly between high-risk and low-risk groups. (5-year overall survival: 24.1% vs. 52.5%;  $p < 0.001$ ; Figure 3B). The risk score in the testing set was proved to be an independent prognostic factor by multivariate Cox regression (HR = 1.16; 95%CI: 1.03 -1.30;  $p = 0.018$ ; Table S1). Besides, ROC curve was to estimate the predictive accuracy of the signature, and the ROC curves for 3-, and 5-year overall survival in the training and testing sets were respectively depicted in Figure 3C, and D. The result of AUCs basically reached 0.7 in the both of 2 cohorts, which suggested that the risk score had a favorable predictive performance.

Finally, in order to determine whether the risk score can be applied to different groups of patients, we compared the survival curves between the 2 groups in different subgroups

(tumor stage I/II or III/IV, receiving ACT or not). However, the finding indicated that patients with high-risk score were remarkably correlated with poor survival in all subgroups (all  $p < 0.001$ ; Figure 4), suggesting that the risk score has a broad utility in GC patients.

### External Validation in the Independent Dataset

In order to further validate the 9 gene signature, we included 2 independent data series deriving respectively from microarray and sequencing data (GSE13861, and TCGA-STAD). In the TCGA-STAD data, we excluded 32 normal tissues and 38 cases that lacked survival information, and finally included 337 GC samples. The risk scores were calculated for each patient of TCGA data using same formula. Also, the same cutoff value (1.68) was utilized to separated patients into high- and low-risk groups. The finding showed that the low-risk group were significantly associated with lower mortality



**Figure 3.** The Survival Analysis and ROC curves. (A-B). Kaplan-Meier Curves for Overall Survival Between the High-Risk and low-risk Groups in the training, and testing cohorts. (C-D). The Time-dependent ROC curves for the risk score in the training, and testing cohorts.

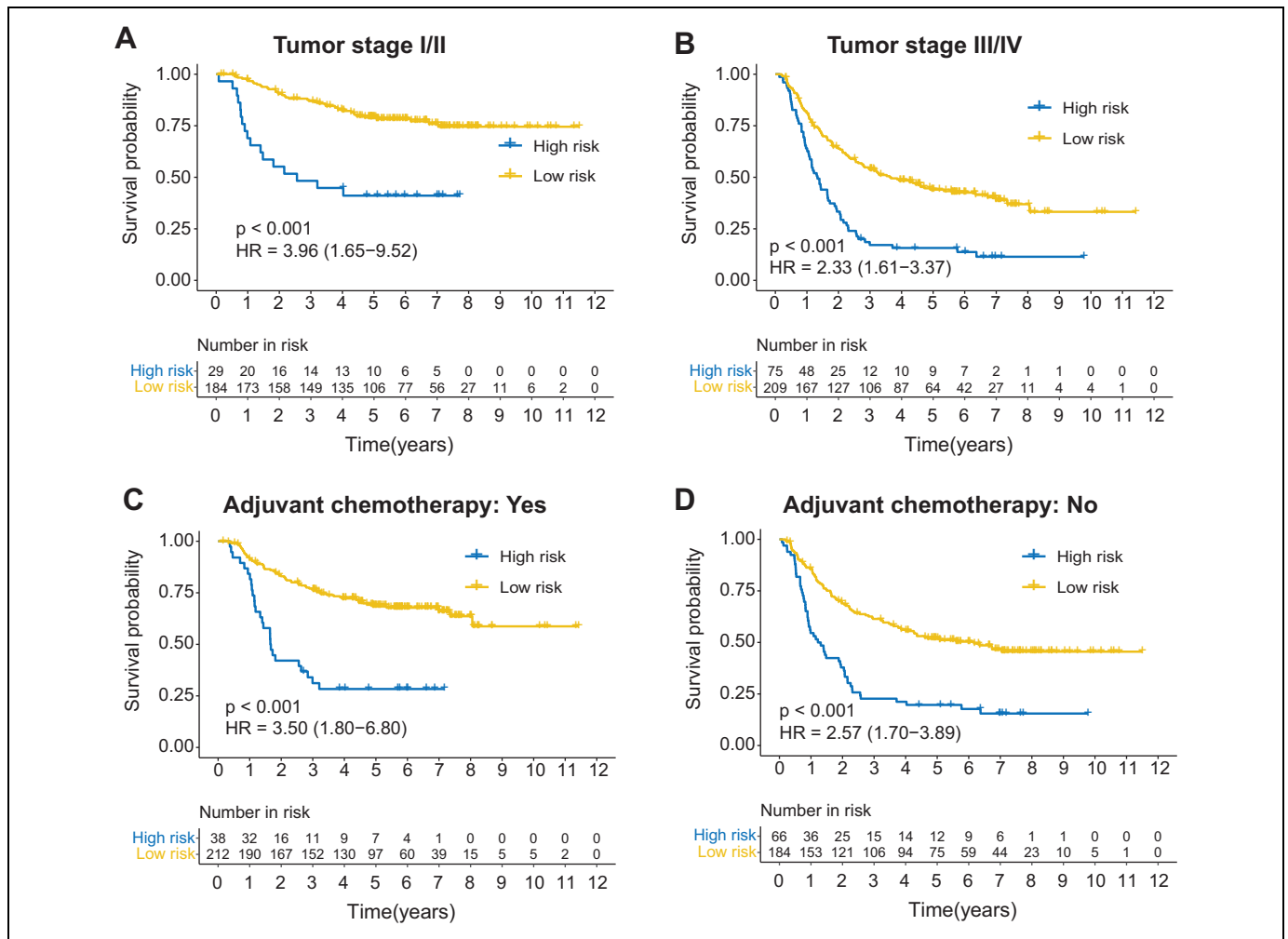
**Table 2.** Univariate and Multivariate Cox Regression Analyses of the Risk Score and Clinical Characteristics With the Overall Survival in the Training Set.

Variable	Univariate analysis			Multivariate analysis		
	HR	95%CI	<i>p</i> -value	HR	95%CI	<i>p</i> -value
Age	1.02	1.00-1.03	0.017	1.03	1.01-1.04	< 0.001
Tumor stage (III/IV vs. I/II)	2.52	2.01-3.15	< 0.001	2.33	1.87-2.90	< 0.001
Adjuvant chemotherapy (Yes vs. No)	0.55	0.40-0.76	< 0.001	0.62	0.44-0.86	0.004
Risk score	1.69	1.52-1.87	< 0.001	1.60	1.42-1.81	< 0.001

(high-risk vs low-risk patients; 5 years overall survival: 13.3% vs 43.1%;  $p < 0.001$ ; Figure 5A).

Consistent with the above methods, we calculated risk scores for each patient of GSE13861 series. Nevertheless, the same cutoff value failed to separate patients into different groups because of small size of samples. Therefore, the median risk score was used to classify the patients into high- and low-

risk groups, and result showed that high-risk group had a worse prognosis in contrast to the low-risk group (high-risk vs low-risk patients; 5 years overall survival: 40.6% vs 75.0%;  $p < 0.001$ ; Figure 5B). These findings were consistent with observations from the training cohort. In the TCGA-STAD cohort, the AUC for 3- and 5-year overall survival were 0.65, and 0.69 (Figure 5C). In the GSE13861 cohort, the AUC for 3- and



**Figure 4.** Stratified Survival Analysis. (A). Kaplan-Meier curves for overall survival between high- and low-risk Group in the Patients with I/II stage; (B). Kaplan-Meier curves for Overall Survival Between High- and Low-Risk group in the patients with III/IV stage; (C). Kaplan-Meier curves for Overall Survival Between the high- and low-Risk Group in the Patients who received Adjuvant Chemotherapy. (D). Kaplan-Meier curves for Overall Survival between the high- and low-Risk Group in the patients who did not received Adjuvant Chemotherapy.

5-year overall survival were 0.70, and 0.68 (Figure 5D), which showed a stable performance for predicting survival.

### Correlation Between the Risk Score and Clinical Features

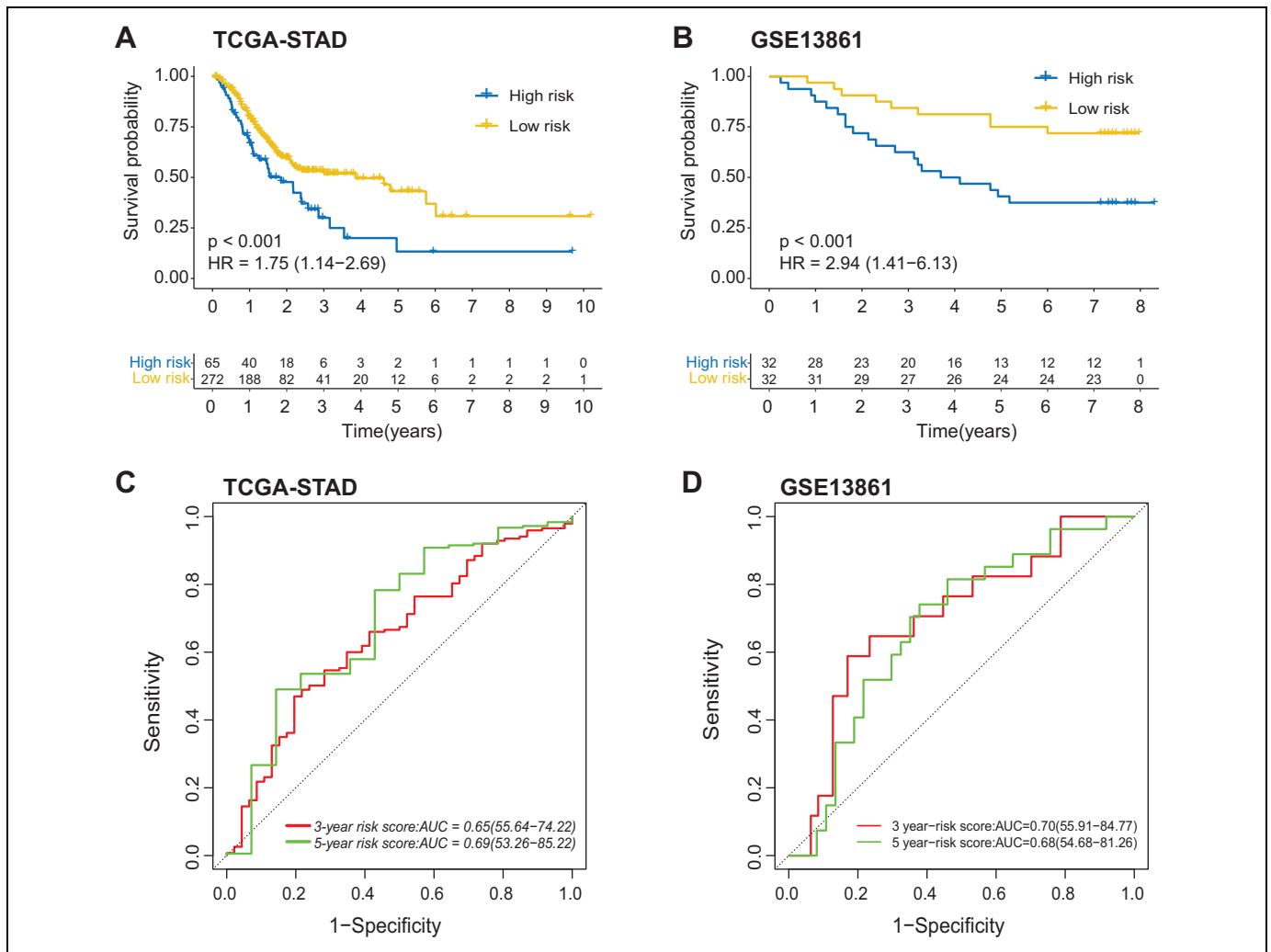
Furthermore, we used the violin plots to compare the risk score between the different clinicopathological characteristics. The results showed that no statistical difference was observed in the risk score between the different age groups, sexes, and tumor locations. However, higher risk score was significantly correlated with poor histological grades ( $p = 0.015$ ) and advanced tumor stages ( $p < 0.001$ ). Moreover, diffuse type was associated with a higher risk score compared to intestinal type ( $p = 0.008$ ) (Figure 6A).

On the one hand, based on CIBERSORT's result, 412 cases with  $p < 0.05$  were included to explore the correlation between the risk groups and tumor-infiltrating immune cells. On the other hand, we did not display the result of naïve CD4+ T cell,

owing to the significantly low percentage in each sample. The subsequent analysis showed that the high-risk group was associated with higher level of macrophages M2 ( $p < 0.001$ ), monocytes ( $p < 0.001$ ), T cells gamma delta ( $p = 0.017$ ), and mast cells resting ( $p = 0.041$ ). In contrast, T cells CD4+ activated memory ( $p = 0.005$ ), T cells CD8+ ( $p = 0.038$ ), dendritic cells resting ( $p = 0.008$ ), T cells regulatory ( $p < 0.001$ ), T cells follicular helper ( $p < 0.001$ ), macrophages M0 ( $p = 0.001$ ), plasma cells ( $p < 0.001$ ) were more enriched in samples of low-risk patients (Figure 6B).

### Construction and Evaluation of Nomogram

Additionally, we constructed a nomogram that integrated age, history of ACT, tumor stage, and risk score. All of included variables in nomogram have been identified as independent prognostic factor according to multivariate Cox regression analysis. As shown in the Figure 7A, the risk score had a



**Figure 5.** External Validation in 2 Independent Datasets. (A-B). Kaplan-Meier curves for overall survival between the high-risk and low-risk groups in the TCGA-STAD, and GSE13861 cohorts. (C-D). The time-dependent ROC curves for the risk score in the TCGA-STAD, and GSE13861 cohorts.

relatively wider range of value compared to that of other variables when calculating nomogram score, indicating that it played a prominent role in determining overall survival. Based on calibration curves for 3-, and 5-year overall survival, the predicted value of the nomogram basically consisted with the actual observation (Figure 7B). The AUC value showed that nomogram score was 0.83, 0.81, and 0.91 in the training, testing, and validation tests, respectively. Based on the comparison of ROC curves, the nomogram score had a better predictive ability than other variables (Figure 8A). Furthermore, the result of DCA showed that nomogram had a better net benefit than other models (Figure 8B).

### Exploration of Biological Function

Based on Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets, we displayed the top 10 of pathways in the high- and low-risk groups (Figure 9A, B), respectively. In

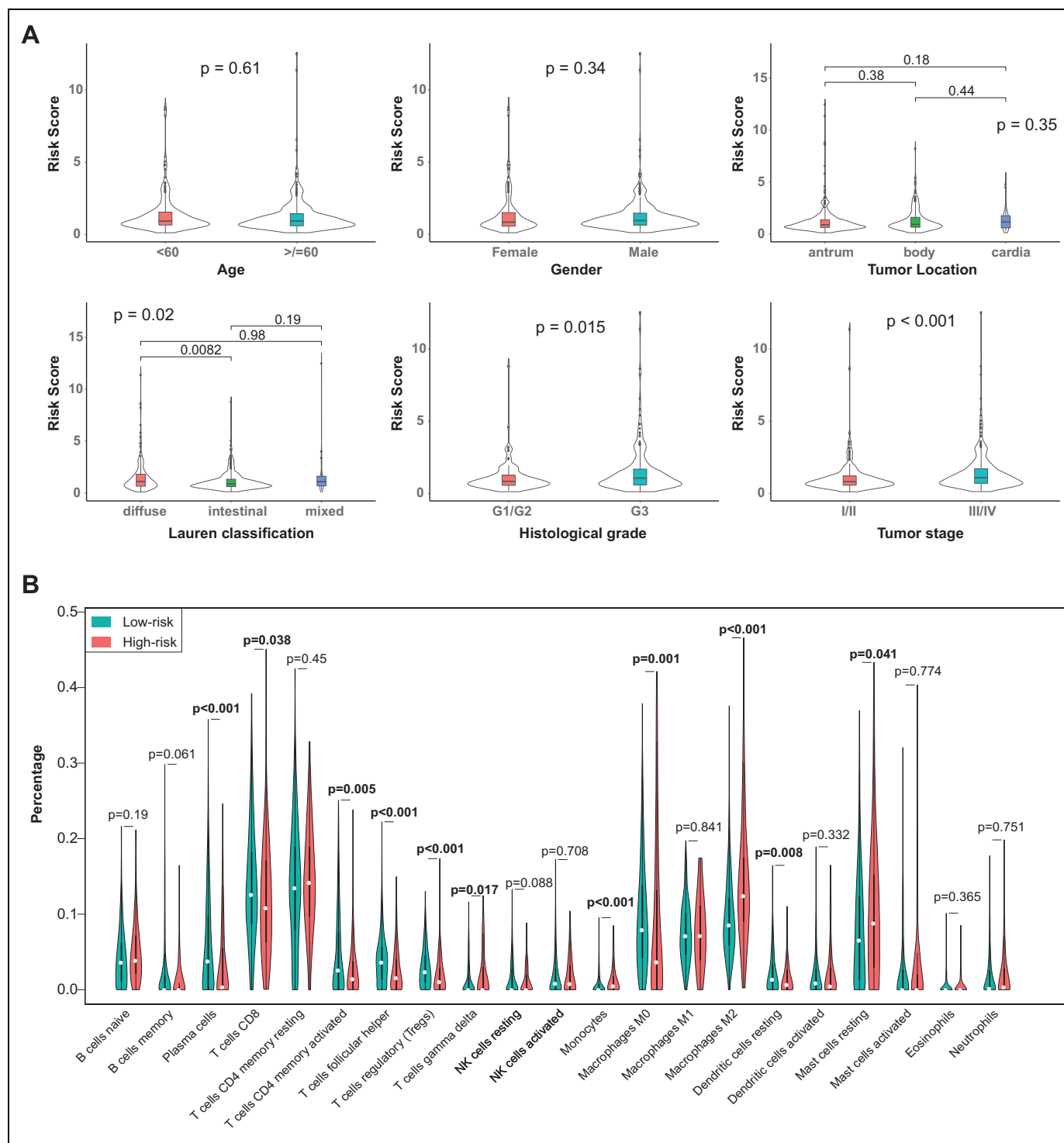
the low-risk group, the finding revealed that the enriched pathways mainly were associated with metabolism-related processes. Next, 10 metabolism-related pathways from the low-risk group were demonstrated through using multiple GSEA plot (Figure 9C).

In addition, we performed GeneMANIA tool to construct gene regulatory network (Figure 9D), which helped to understand the relationship between the 9 glycolysis genes. The network suggested that these genes had a strong independence for each other.

### Discussion

Although numerous prognostic models using molecular signature have been developed by researchers, only conventional Her-2, CEA, CA19-9, and CA72-4 were applied to assisting prediction for GC patients' outcomes in clinical practice.<sup>12</sup> However, these biomarkers were mainly used in diagnosis and

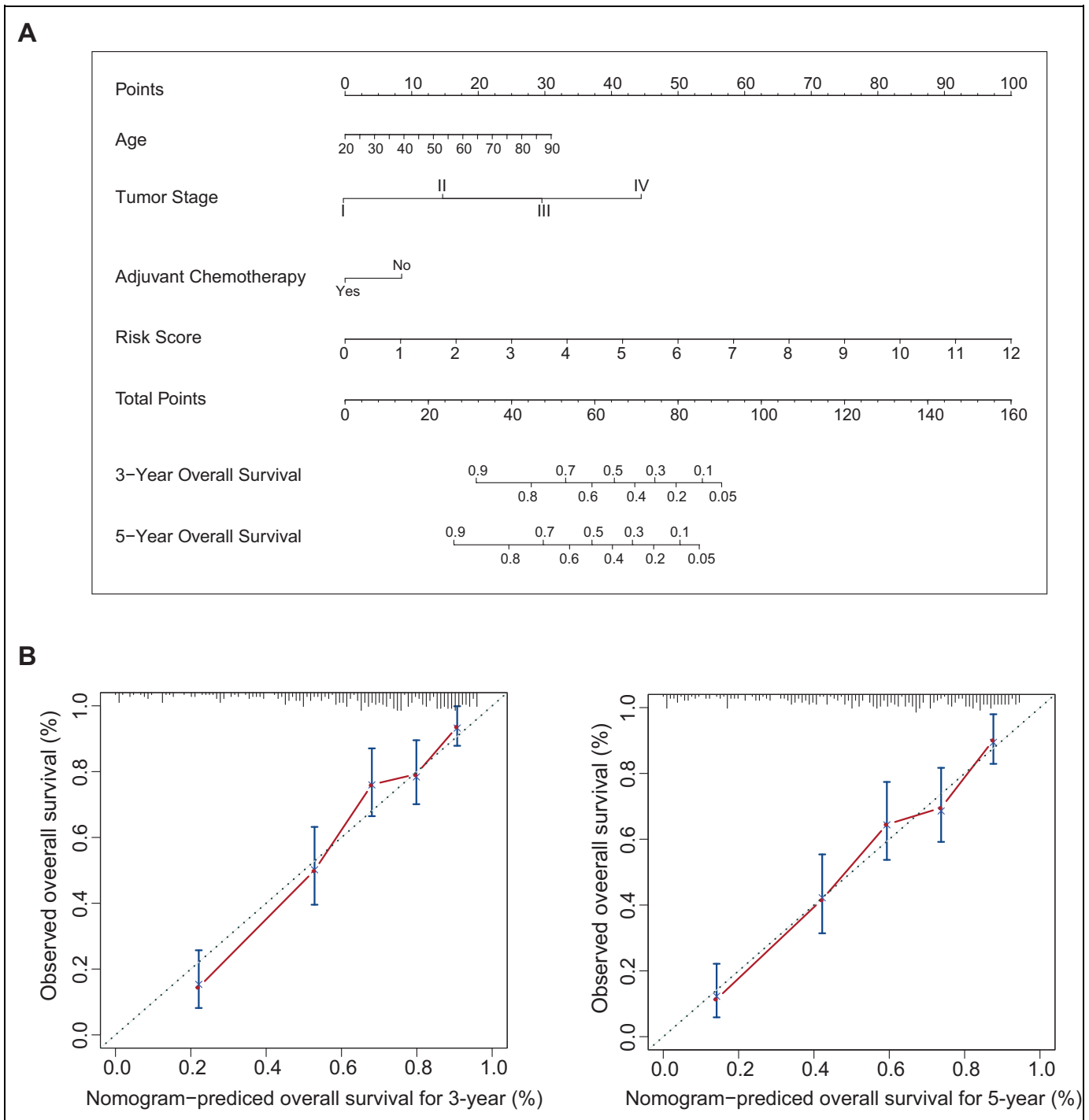




**Figure 6.** (A). The Association Between the Risk Score and Clinicopathological Characteristics. (B). The different tumor-infiltrating immune cells between the high- and low-risk groups.

monitoring of recurrence in GC patients. The biomarker that can play a key role in prognostic prediction for GC was rare. With the rapid development of gene sequencing technology, a lot of solid biomarkers that have prognostic value for GC patients has been identified, which includes immune-related gene, hypoxia-related gene, and immune scores. Meanwhile,

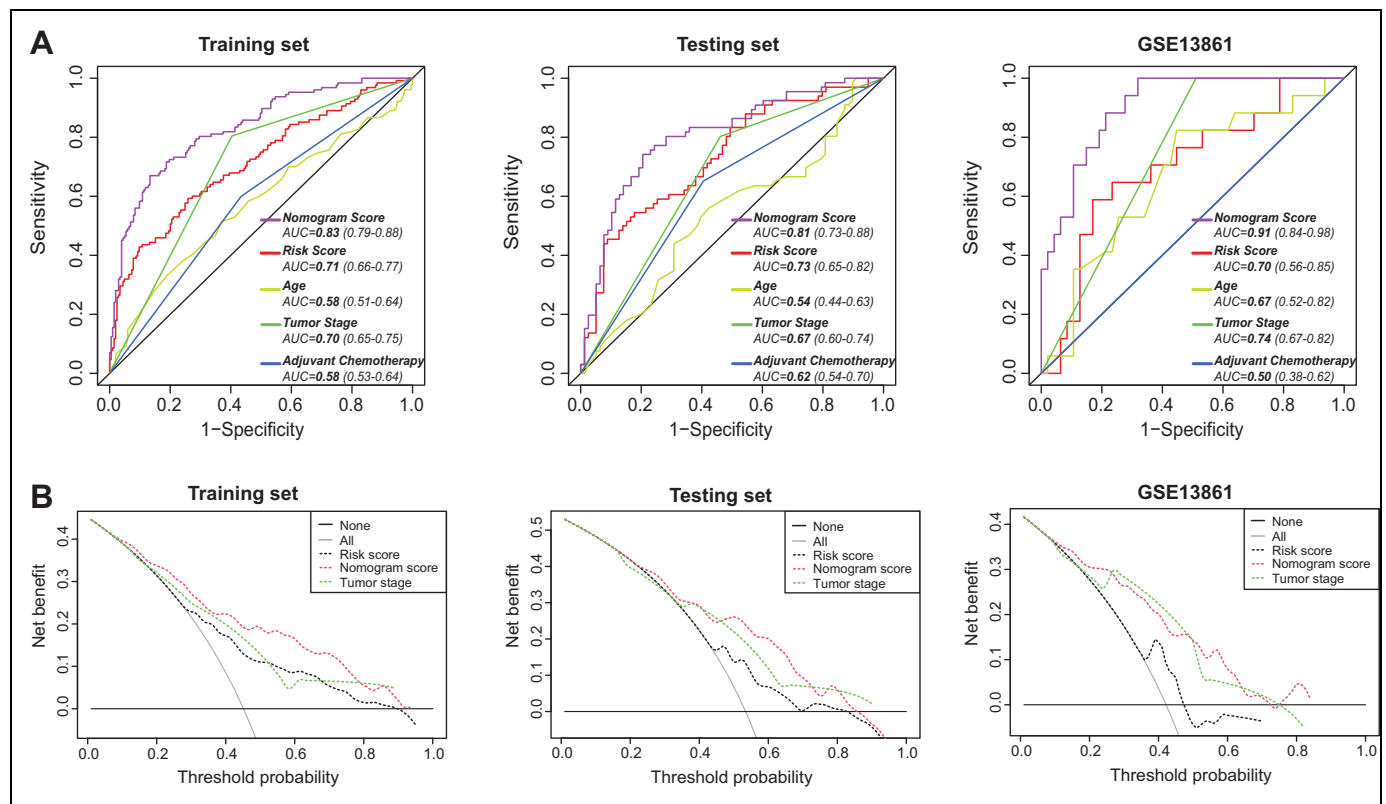
a variety of signatures based on these genes have demonstrated an excellent ability in predicting prognosis for GC, and been validated in multiple independent datasets.<sup>13-15</sup> Owing to the importance of glycolysis pathway in regulating tumorigenesis, this work is aim to develop a list of glycolysis-related genes for risk stratification in GC patients.



**Figure 7.** Construction of the nomogram. (A). Nomogram that integrated age, tumor stage, adjuvant chemotherapy and the risk score. (B). The calibration curves for 3, and 5-year overall survival.

In this study, 9 glycolysis-related genes were identified to construct a novel scoring system. Among these genes, the significantly overexpressing bisphosphate nucleotidase-1 (BPNT1) was found to be associated with the abnormality of 3'-phosphoadenylate metabolism in glioma.<sup>16</sup> Decorin (DCN), as a potential biomarker, was reported to be correlated with the prognosis of colon cancer and lung adenocarcinoma.<sup>17,18</sup> In

lung cancer, mutation of GDP-mannose pyrophosphorylase A (GMPPA) showed significantly negative association with prognosis of patients.<sup>19</sup> Moreover, up-regulation of 6-fucosyltransferase (FUT8) can inhibit proliferation of GC cells,<sup>20</sup> and Glypican-3 (GPC3) may be correlated with nodal and distant metastasis in GC patients.<sup>21</sup> In renal cell carcinoma, patients with overexpressing lactate dehydrogenase C (LDHC)



**Figure 8.** Evaluation of the nomogram. (A). The ROC curves of the nomogram score and other clinical variables in the training, testing, and validation cohorts, respectively. (B). The decision curve analysis in the training, testing, and validation cohorts, respectively; Horizontal solid line indicates an event will occur in no patients; Gray line indicates event will occur in all patients.

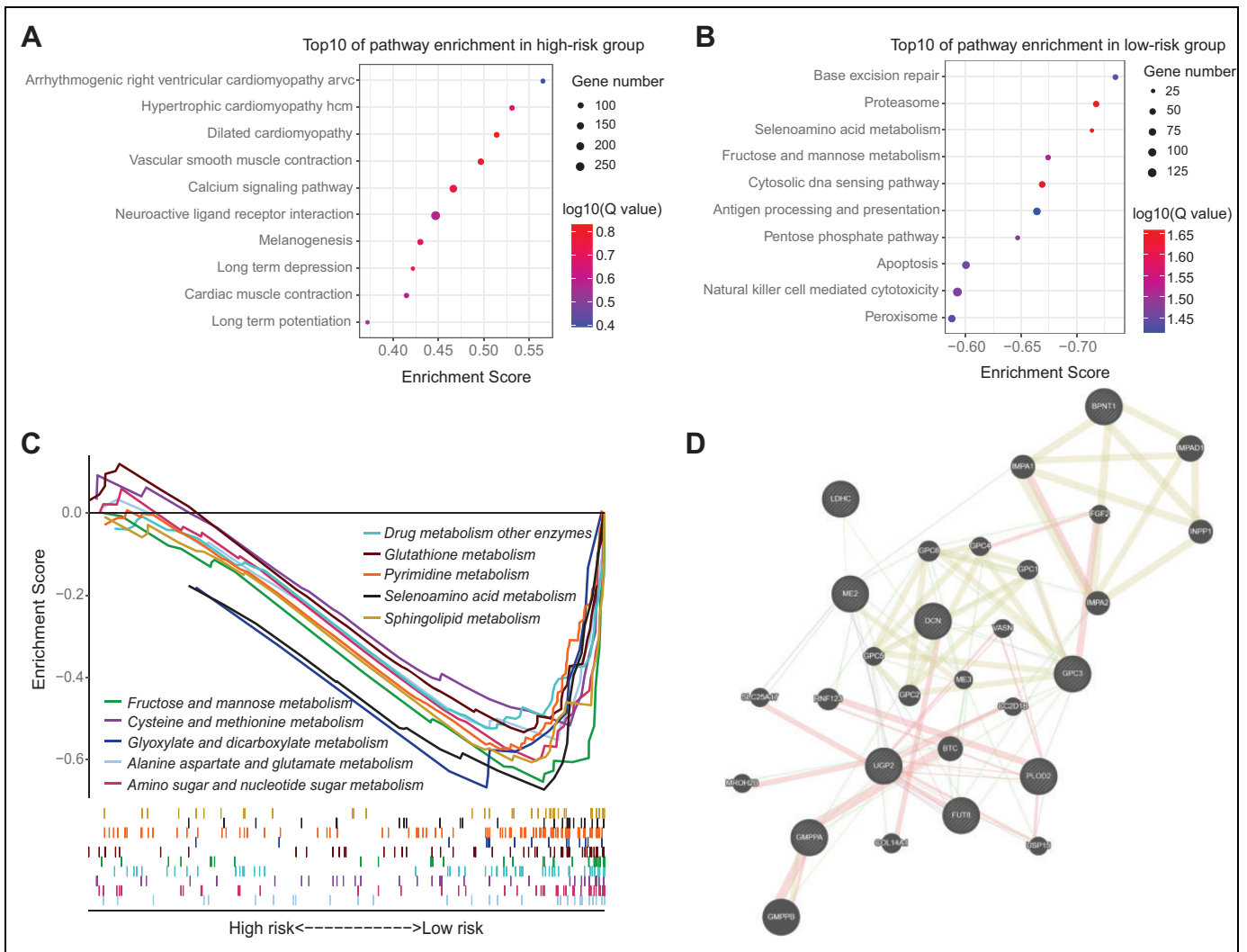
had a worse progression-free survival.<sup>22</sup> A variety of evidences have indicated that malic enzyme-2 (ME2) is a potential therapeutic target and novel biomarker which plays critical role during tumorigenesis.<sup>23</sup> Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2 (PLOD2) was found to be negative correlated with GC patient's prognosis.<sup>24</sup> UTP-glucose-1-phosphate uridylyltransferase (UGP2), as long noncoding RNA, can regulate metabolic pathways, and plays an important role in diagnosis of GC.<sup>25</sup> Interestingly, the prognostic roles of FUT8, GPC3, and PLOD2 were consistent with the observations from our study. However, these genes (ME2, BPNT1, DCN, GMPPA, LDHC, and UGP2) have not been reported to be related to the prognosis for GC. In future studies, the expression levels of these genes will be investigated in GC tissues, since the prognostic value has been observed in this study.

The main finding revealed that patients with high-risk scores were negatively associated with survival and this observation was confirmed in the testing and 2 independent cohorts. To combine molecular information with the clinical feature to reinforce the predictive model, we developed a nomogram that integrated risk score, age, history of ACT, and TNM stage. Through comparing the result of ROC and DCA, we found that the nomogram score had superior predictive ability than conventional factors, indicating that the risk score combined with clinical features can develop a robust prediction for survival and improve the individualized clinical decision making of GC

patients. Compared with previously reported gene signatures for GC, our nomogram also achieved a higher value in AUC.<sup>26-28</sup> Additionally, our nomogram was comparable to the published immunoscore signature<sup>15</sup> in predictive accuracy because both of the 2 models reached above 0.8 in AUC value. Therefore, there is a strong possibility that a more robust model will be developed when including immunoscore into our nomogram, which is worth exploring in the future work.

In GC, the crosstalk between immune cells and tumor cells plays a vital role in predicting prognosis and immunotherapeutic outcomes.<sup>29</sup> After comparing the difference of tumor-infiltrating immune cells between high- and low-risk groups, we observed a high level of T cells gamma delta, monocytes, macrophages M2, and mast cells resting in the high-risk group. Thereinto, tumor-infiltrating monocytes can promote tumor migration and invasion.<sup>30</sup> Increased intratumor mast cells contribute to angiogenesis, and nodal metastasis of GC.<sup>31</sup> Moreover, tumor-associated macrophages were classified into M1 and M2 based on their polarized states,<sup>32</sup> and M2 macrophage was reported to promote GC metastasis via CHI3L1 protein secretion.<sup>33</sup>

On the other hand, 7 types of immune cells (T cells CD8+, T cells CD4+ activated memory, macrophages M0, T cells follicular helper, T cell regulatory, dendritic cells resting, and plasma cells) were more enriched in low-risk patients' samples. According to previous literature, CD4+T cells influence



**Figure 9.** Biological process associated with the signature. (A-B). The top 10 pathways enriched in the high- and low-risk groups based on the KEGG gene set. (C). The GSEA of metabolism-related pathways enriched in low-risk group. (D). The 9 glycolysis-related gene in the interaction network; Red edge indicates physical interactions. Violet edge indicates co-expression. Yellow edge indicates shared protein domain. Green edge indicates genetic interactions.

cancer immunology and immunotherapy by helping cytotoxic T lymphocytes (CTLs) acquire multiple antitumor functions.<sup>34</sup> Liu et al reported that high densities of T cells CD8+ and T cells CD4+ were associated with better clinical outcomes in GC.<sup>35</sup> Additionally, a high level of T cell follicular helper infiltration was related to a favorable prognosis in breast cancer.<sup>36</sup> Wang et al found that high numbers of intra-tumoral T cell regulatory was correlated with improved survival in GC.<sup>37</sup> Extensive literature has shown that tumor-infiltrating plasma cells have a positive prognostic effect for cancer.<sup>38</sup> Based on the above evidences, these results might help to explain why the low-risk group had a better survival outcome to some extent.

However, our study also has several limitations. First, all of data were obtained from public database, and the sample size was relatively small. Second, this study did not conduct an experiment to further validate these genes. Last, this was a

retrospectively designed work, and the potential bias correlated with unbalanced clinicopathological features cannot be ignored. Therefore, well-designed experiments and clinical trials are urgently needed to validate our findings.

## Conclusions

In this study, we identified 9 glycolysis-related genes (BPNT1, DCN, FUT8, GMPPA, GPC3, LDHC, ME2, PLOD2, and UGP2) from hallmark glycolysis pathway. Based on the 9 genes, a nomogram was established, which had a superior predictive ability than conventional prognostic factors.

## Authors' Note

Tianqi Luo, Yufei Du and Jinling Duan are authors contributed equally to this study.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Yufei Du  <https://orcid.org/0000-0001-7933-7439>

## Supplemental Material

Supplemental material for this article is available online.

## References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424. doi:10.3322/caac.21492
- Venerito M, Link A, Rokkas T, Malfertheiner P. Review: gastric cancer—clinical aspects. *Helicobacter*. 2019;24(suppl 1):e12643. doi:10.1111/hel.12643
- Biagioni A, Skalamera I, Peri S, et al. Update on gastric cancer treatments and gene therapies. *Cancer Metastasis Rev*. 2019; 38(3):537-548. doi:10.1007/s10555-019-09803-7
- Fang C, Wang W, Deng JY, et al. Proposal and validation of a modified staging system to improve the prognosis predictive performance of the 8th AJCC/UICC pTNM staging system for gastric adenocarcinoma: a multicenter study with external validation. *Cancer Commun (Lond)*. 2018;38(1):67. doi:10.1186/s40880-018-0337-5
- Vander Heiden MG, DeBerardinis RJ. Understanding the intersections between metabolism and cancer biology. *Cell*. 2017; 168(4):657-669. doi:10.1016/j.cell.2016.12.039
- Warburg O, Wind F, Negelein E. The metabolism of tumors in the body. *J Gen Physiol*. 1927;8(6):519-530.
- Liu Y, Zhang Z, Wang J, et al. Metabolic reprogramming results in abnormal glycolysis in gastric cancer: a review. *Onco Targets Ther*. 2019;12:1195-1204. doi:10.2147/OTT.S189687
- Gatenby RA, Gillies RJ. Why do cancers have high aerobic glycolysis? *Nat Rev Cancer*. 2004;4(11):891-899. doi:10.1038/nrc1478
- Liu J, Li S, Feng G, et al. Nine glycolysis-related gene signature predicting the survival of patients with endometrial adenocarcinoma. *Cancer Cell Int*. 2020;20:183. doi:10.1186/s12935-020-01264-1
- Zhang C, Gou X, He W, Yang H, Yin H. A glycolysis-based 4-mRNA signature correlates with the prognosis and cell cycle process in patients with bladder cancer. *Cancer Cell Int*. 2020; 20:177. doi:10.1186/s12935-020-01255-2
- Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res*. 2019;3:18. doi:10.1186/s41512-019-0064-7
- Matsuoka T, Yashiro M. Biomarkers of gastric cancer: current topics and future perspective. *World J Gastroenterol*. 2018; 24(26):2818-2832. doi:10.3748/wjg.v24.i26.2818
- Liu Y, Wu J, Huang W, et al. Development and validation of a hypoxia-immune-based microenvironment gene signature for risk stratification in gastric cancer. *J Transl Med*. 2020;18(1):201. doi: 10.1186/s12967-020-02366-0
- Wang H, Wu X, Chen Y. Stromal-immune score-based gene signature: a prognosis stratification tool in gastric cancer. *Front Oncol*. 2019;9:1212. doi:10.3389/fonc.2019.01212
- Jiang Y, Zhang Q, Hu Y, et al. ImmunoScore signature: a prognostic and predictive tool in gastric cancer. *Ann Surg*. 2018; 267(3):504-513. doi:10.1097/SLA.0000000000002116
- Li W, Jia H, Li Q, et al. Glycerophosphatidylcholine PC(36:1) absence and 3'-phosphoadenylylate (pAp) accumulation are hallmarks of the human glioma metabolome. *Sci Rep*. 2018;8(1): 14783. doi:10.1038/s41598-018-32847-8
- Li G, Li M, Liang X, et al. Identifying DCN and HSPD1 as potential biomarkers in colon cancer using 2D-LC-MS/MS combined with iTRAQ technology. *J Cancer*. 2017;8(3):479-489. doi: 10.7150/jca.17192
- Yan Y, Xu Z, Qian L, et al. Identification of CAV1 and DCN as potential predictive biomarkers for lung adenocarcinoma. *Am J Physiol Lung Cell Mol Physiol*. 2019;316(4): L630-L643. doi:10. 1152/ajplung.00364.2018
- Cho H-J, Lee S, Ji YG, Lee DH. Association of specific gene mutations derived from machine learning with survival in lung adenocarcinoma. *PLoS One*. 2018;13(11):e0207204. doi:10.1371/ journal.pone.0207204
- Zhao Y-P, Xu X-Y, Fang M, et al. Decreased core-fucosylation contributes to malignancy in gastric cancer. *PLoS One*. 2014;9(4): e94536. doi:10.1371/journal.pone.0094536
- Ushiku T, Uozaki H, Shinozaki A, et al. Glypican 3-expressing gastric carcinoma: distinct subgroup unifying hepatoid, clear-cell, and alpha-fetoprotein-producing gastric carcinomas. *Cancer Sci*. 2009;100(4):626-632. doi:10.1111/j.1349-7006.2009.01108.x
- Hua Y, Liang C, Zhu J, et al. Expression of lactate dehydrogenase C correlates with poor prognosis in renal cell carcinoma. *Tumour Biol*. 2017;39(3):1010428317695968. doi:10.1177/10104283176 95968
- Sarfraz I, Rasul A, Hussain G, et al. Malic enzyme 2 as a potential therapeutic drug target for cancer. *IUBMB Life*. 2018;70(11): 1076-1083. doi:10.1002/iub.1930
- Li S-S, Lian Y-F, Huang Y-L, Huang Y-H, Xiao J. Over-expressing PLOD family genes predict poor prognosis in gastric cancer. *J Cancer*. 2020;11(1):121-131. doi:10.7150/jca. 35763
- Mo X, Wu Y, Chen L, et al. Global expression profiling of metabolic pathway-related lncRNAs in human gastric cancer and the identification of RP11-555H23.1 as a new diagnostic biomarker. *J Clin Lab Anal*. 2019;33(2):e22692. doi:10.1002/ jcla.22692
- Cheng P. A prognostic 3-long noncoding RNA signature for patients with gastric cancer. *J Cell Biochem*. 2018;119(11): 9261-9269. doi:10.1002/jcb.27195
- Deng X, Xiao Q, Liu F, Zheng C. A gene expression-based risk model reveals prognosis of gastric cancer. *PeerJ*. 2018;6:e4204. doi:10.7717/peerj.4204

28. Cai C, Yang L, Tang Y, et al. Prediction of overall survival in gastric cancer using a nine-lncRNA. *DNA Cell Biol.* 2019;38(9): 1005-1012. doi:10.1089/dna.2019.4832
29. Zeng D, Li M, Zhou R, et al. Tumor microenvironment characterization in gastric cancer identifies prognostic and immunotherapeutically relevant gene signatures. *Cancer Immunol Res.* 2019; 7(5):737-750. doi:10.1158/2326-6066.CIR-18-0436
30. Lim SY, Yuzhalin AE, Gordon-Weeks AN, Muschel RJ. Tumor-infiltrating monocytes/macrophages promote tumor invasion and migration by upregulating S100A8 and S100A9 expression in cancer cells. *Oncogene.* 2016;35(44):5735-5745. doi:10.1038/onc.2016.107
31. Sammarco G, Varricchi G, Ferraro V, et al. Mast cells, angiogenesis and lymphangiogenesis in human gastric cancer. *Int J Mol Sci.* 2019;20(9). doi:10.3390/ijms20092106
32. Mantovani A, Sozzani S, Locati M, Allavena P, Sica A. Macrophage polarization: tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. *Trends Immunol.* 2002;23(11):549-555.
33. Chen Y, Zhang S, Wang Q, Zhang X. Tumor-recruited M2 macrophages promote gastric and breast cancer metastasis via M2 macrophage-secreted CHI3L1 protein. *J Hematol Oncol.* 2017; 10(1):1-13. doi:10.1186/s13045-017-0408-0
34. Borst J, Ahrends T, Babala N, Melief CJM, Kastenmuller W. CD4(+) T cell help in cancer immunology and immunotherapy. *Nat Rev Immunol.* 2018;18(10):635-647. doi:10.1038/s41577-018-0044-0
35. Liu K, Yang K, Wu B, et al. Tumor-infiltrating immune cells are associated with prognosis of gastric cancer. *Medicine.* 2015; 94(39): e1631. doi:10.1097/MD.0000000000001631
36. Gu-Trantien C, Loi S, Garaud S, et al. CD4(+) follicular helper T cell infiltration predicts breast cancer survival. *J Clin Invest.* 2013;123(7):2873-2892. doi:10.1172/JCI67428
37. Wang B, Xu D, Yu X, et al. Association of intra-tumoral infiltrating macrophages and regulatory T cells is an independent prognostic factor in gastric cancer after radical resection. *Ann Surg Oncol.* 2011;18(9):2585-2593. doi:10.1245/s10434-011-1609-3
38. Wouters MCA, Nelson BH. Prognostic significance of tumor-infiltrating B cells and plasma cells in human cancer. *Clin Cancer Res.* 2018;24(24):6125-6135. doi:10.1158/1078-0432.CCR-18-1481