



Identification of Five Glycolysis-Related Gene Signature and Risk Score Model for Colorectal Cancer

Jun Zhu^{1†}, Shuai Wang^{1†}, Han Bai^{2†}, Ke Wang¹, Jun Hao³, Jian Zhang^{4*} and Jipeng Li^{1*}

OPEN ACCESS

Edited by:

Jiang Chen,
Zhejiang University, China

Reviewed by:

Panpan Yu,
Hangzhou First People's Hospital,
China
Ajaz Bhat,
Sidra Medicine, Qatar
Haruhiko Sugimura,
Hamamatsu University School of
Medicine, Japan

*Correspondence:

Jian Zhang
biozhangj@fmmu.edu.cn
Jipeng Li
jipengli1974@allyun.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Gastrointestinal Cancers,
a section of the journal
Frontiers in Oncology

Received: 17 September 2020

Accepted: 18 January 2021

Published: 04 March 2021

Citation:

Zhu J, Wang S, Bai H, Wang K,
Hao J, Zhang J and Li J (2021)
Identification of Five Glycolysis-Related
Gene Signature and Risk Score
Model for Colorectal Cancer.
Front. Oncol. 11:588811.
doi: 10.3389/fonc.2021.588811

¹ State Key Laboratory of Cancer Biology, Institute of Digestive Diseases, Xijing Hospital, The Fourth Military Medical University, Xi'an, China, ² Department of Radiation Oncology, Xijing Hospital, The Fourth Military Medical University, Xi'an, China, ³ Department of Experiment Surgery, Xijing Hospital, Fourth Military Medical University, Xi'an, China, ⁴ State Key Laboratory of Cancer Biology, Department of Biochemistry and Molecular Biology, The Fourth Military Medical University, Xi'an, China

Metabolic changes, especially in glucose metabolism, are widely established during the occurrence and development of tumors and regarded as biological markers of pancreatic cancer. The well-known 'Warburg effect' demonstrates that cancer cells prefer aerobic glycolysis even if there is sufficient ambient oxygen. Accumulating evidence suggests that aerobic glycolysis plays a pivotal role in colorectal cancer (CRC) development. However, few studies have examined the relationship of glycolytic gene clusters with prognosis of CRC patients. Here, our aim is to build a glycolysis-associated gene signature as a biomarker for colorectal cancer. The mRNA sequencing and corresponding clinical data were downloaded from TCGA and GEO databases. Gene set enrichment analysis (GSEA) was performed, indicating that four gene clusters were significantly enriched, which revealed the inextricable relationship of CRC with glycolysis. By comparing gene expression of cancer and adjacent samples, 236 genes were identified. Univariate, multivariate, and LASSO Cox regression analyses screened out five prognostic-related genes (ENO3, GPC1, P4HA1, SPAG4, and STC2). Kaplan–Meier curves and receiver operating characteristic curves (ROC, AUC = 0.766) showed that the risk model could become an effective prognostic indicator ($P < 0.001$). Multivariate Cox analysis also revealed that this risk model is independent of age and TNM stages. We further validated this risk model in external cohorts (GES38832 and GSE39582), showing these five glycolytic genes could emerge as reliable predictors for CRC patients' outcomes. Lastly, based on five genes and risk score, we construct a nomogram model assessed by C-index (0.7905) and calibration plot. In conclusion, we highlighted the clinical significance of glycolysis in CRC and constructed a glycolysis-related prognostic model, providing a promising target for glycolysis regulation in CRC.

Keywords: glycolytic gene, prognosis analysis, colorectal cancer, GPC1, ENO3, P4HA1, SPAG4, STC2

INTRODUCTION

Colorectal cancer (CRC) is the third most common malignant tumor and the second common cause of cancer-related death worldwide (1). Due to the lack of early clinical symptoms and predictive markers, CRC patients often are diagnosed at advanced stages and forfeit the occasion of surgical treatment. Patients in advanced stages especially with metastatic lesions could have a poor prognosis (2), and the 5-year survival rate of CRC patients is only 13% (3, 4). Therefore, it is crucial to discover and identify novel diagnostic indicator for the early detection and treatment of CRC.

Altered energy metabolism, which fuels tumor cell growth and proliferation, has been heralded as an emerging cancer hallmark (5, 6). In 1924, Warburg found that cancer cells behave bizarrely and cancer cells could maintain a high glycolysis rate compared with adjacent normal tissue even in the presence of oxygen; that is the well-known ‘Warburg effect’ (7, 8). Based on this discovery, (18F)-fluorodeoxyglucose-positron emission tomography (FDG-PET) is applied to scan out a tumor tissue and common metastasis sites because of its generally high uptake of the glucose analog FDG. Many glycolysis-related genes identified upregulated in CRC contains hypoxia-inducible factor-1 α (HIF-1 α) (9–12), glucose transporter family (GLUT) (13–15), hexokinase (HK) (16), pyruvate kinase (PKM) (17, 18), pyruvate dehydrogenase kinases (PDK) (11) and lactate dehydrogenase A (LDHA) (12, 15). LDHA converted to lactic acid metabolic enzymes is identified as the oncogene MYC’s first metabolic target (19). Apart from glycolysis-related genes, some long non-coding RNAs (lncRNAs) were regarded as a significant role in the glycolysis-pathway. lncRNA GLCC1 stabilizes c-Myc transcriptional factor and further facilitates the expression of its target genes (such as LDHA), consequently reprogramming glycolytic metabolism for CRC proliferation (20). Another reporter found that lncRNA FEZF1-AS1 could bind and increase the stability of PKM2 protein, resulting in increased cytoplasmic and nuclear PKM2 level, which further activate STAT3 signaling (21). Therefore, glycolysis-related genes have been a potential target for cancer therapy, and many associated molecules participate in the regulation of glucose metabolism in CRC cells.

CRC is widespread and continuing progress in developing and developed countries. Another worrying phenomenon is the rise in patients diagnosed with CRC younger than 50 years old (22). Thus, there is an urgent need for improved CRC screening and therapeutics. Clinically, superior or comprehensive models to predict the overall survival (OS) of CRC patients are also needed (23). The rapid development of genomic profiling and analytical ability of big data provide great convenience for constructing risk models for cancer patients and prognostic evaluation. In this study, we built a five-gene signature based on glycolysis from The Cancer Genome Atlas (TCGA) and validated it in Gene Expression Omnibus (GEO) database. By multivariate Cox regression and survival analysis, we demonstrated that the risk model could be considered an

independent CRC indicator. Collectively, our results suggested the glycolysis-related risk model may serve as a clinical outcome indicator for CRC, and we provide potential therapeutic targets for CRC patients.

METHODS

Data Collection

The mRNA expression data and corresponding clinical features of CRC patients were downloaded from TCGA (<https://portal.gdc.cancer.gov/>) and Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>). There are 39 paracancerous tissues and 398 cases of cancerous tissues in the TCGA dataset, among which 379 patients had complete follow-up data. As well, there are 122 and 579 CRC patients in the GSE38832 and GSE39582, respectively (**Table S1**). Ten paired CRC samples (cancerous and adjacent tissue) were collected from the Xijing digestive hospital.

Glycolytic Process-Related Gene Sets Enriched by GSEA

Gene set enrichment analysis (GSEA) (24) was performed with GSEA version 4.0.1 software and used to complete glycolysis-related enrichment analysis with 1,000 permutation number. Six gene sets (Kegg glycolysis pathway, Kegg glycolysis gluconeogenesis, Go glycolytic process, Hallmark glycolysis, Reactome regulation of glycolysis, Reactome glycolysis) were downloaded from the MSigDB of GSEA website (www.broadinstitute.org/gsea/) and chosen as reference gene sets. We set the cut-off criteria as gene size ≥ 15 , |normalized enrichment score (NES)| > 1.5 and nominal P-value (NOM P-value) < 0.05 .

Identification of Differentially Expressed Genes and Prognostic Genes

298 genes were extracted from the four gene sets enriched by GSEA, and the linear models for microarray data (LIMMA) R package was performed for DEGs identified between paracancerous and cancer samples. Cut-off values were set at P value < 0.05 and $|\log_2$ fold change| > 0 . DEGs were associated with the OS of patients by univariable Cox analysis. P < 0.05 was considered as a statistically significant difference.

Construction of Risk Score Model by Multivariate Cox and LASSO Regression

Univariate and multivariate Cox regression analyses were used by ‘survival’ R package, and LASSO was analyzed with ‘glmnet’ package to obtain the most useful predictive genes. The risk score model was constructed based on the corresponding coefficients, and risk score of each patient was generated by multiplying the expression and coefficients. Through the median risk scores, two groups were divided (high- and low-risk subgroups). Receiver operating characteristic (ROC) curves and Kaplan–Meier curves (K–M curves) were applied to evaluate the predictive capability

of the risk model. In addition, univariate and multivariate Cox regression analyses were used to explore the prognostic efficiency of the risk score models and other clinicopathological features.

Construction and Validation of a Predictive Nomogram

A nomogram can be performed to predict the prognosis of various cancer (25). In the TCGA datasets, five glycolysis-related genes identified by multivariate Cox regression were included to build a nomogram using the 'rms' package in R to investigate the 3- and 5-year survival rate of CRC patients. To assess the discrimination and accuracy of the nomogram, we calculated the concordance index (C-index) and plotted a calibration curve.

Alternation of Five Genes in the Model

The cBioPortal dataset (<https://www.cbioportal.org/>) possesses multi-dimensional cancer genomics data. We used it to explore the genetic alterations of five genes (TCGA-Pancancer Atlas) and their relationship with other related genes. The network of glycolytic genes and highly related genes was visualized in Cytoscape software (version 3.61).

Immunohistochemistry and RT-PCR

Immunohistochemistry images of the five model genes were downloaded from the Human Protein Atlas (<https://www.proteinatlas.org/>). Messenger RNA (mRNA) of 10 paired tissues and six kinds of cell lines was extracted using TRIzol's method (Invitrogen, Carlsbad, CA, USA), and quantitative real-time PCR was conducted with SYBR-green PCR MasterMix (TaKaRa). The study was approved by the Xijing Hospital Ethics Committee (No. KY20203269-1), and informed consent was taken from all the patients.

Statistical Analysis

R (version 3.63, <http://www.r-project.org/>) was applied to perform statistical analysis. All statistical tests were bilateral, and $P < 0.05$ was considered as statistically significant.

RESULTS

Overall Design of the Study

The flow chart of our research is shown in **Figure 1**. The clinical data of CRC patients and corresponding gene expression profiles were downloaded from the TCGA database. The glycolysis-related gene sets were enriched by GSEA. The risk model of glycolysis related genes was established by multivariate Cox analysis and LASSO algorithm. The five-gene risk model was validated by survival analysis and ROC curves. Patients were divided into two parts based on their clinical characteristics, and survival analysis was performed to determine which group would be predicted accurately. In addition, two external GEO data sets were downloaded to verify this model, and a nomogram based on TCGA data was built. Calibration and C-index were used to assess the nomogram.

Four Glycolysis-Related Gene Clusters Were Significantly Enriched in CRC Patients

In the TCGA-COAD dataset, 39 cases were normal samples, and 398 cases were tumor samples. Four of the six gene sets were significantly enriched (**Figure 2**). The GSEA results documented that glycolysis pathways play a pivotal role in CRC patients' occurrence and progression. 298 genes associated with glycolysis pathway were extracted in transcriptome expression of TCGA colorectal patients for further analysis.

Identification of DEGs in Glycolysis-Associated Clusters

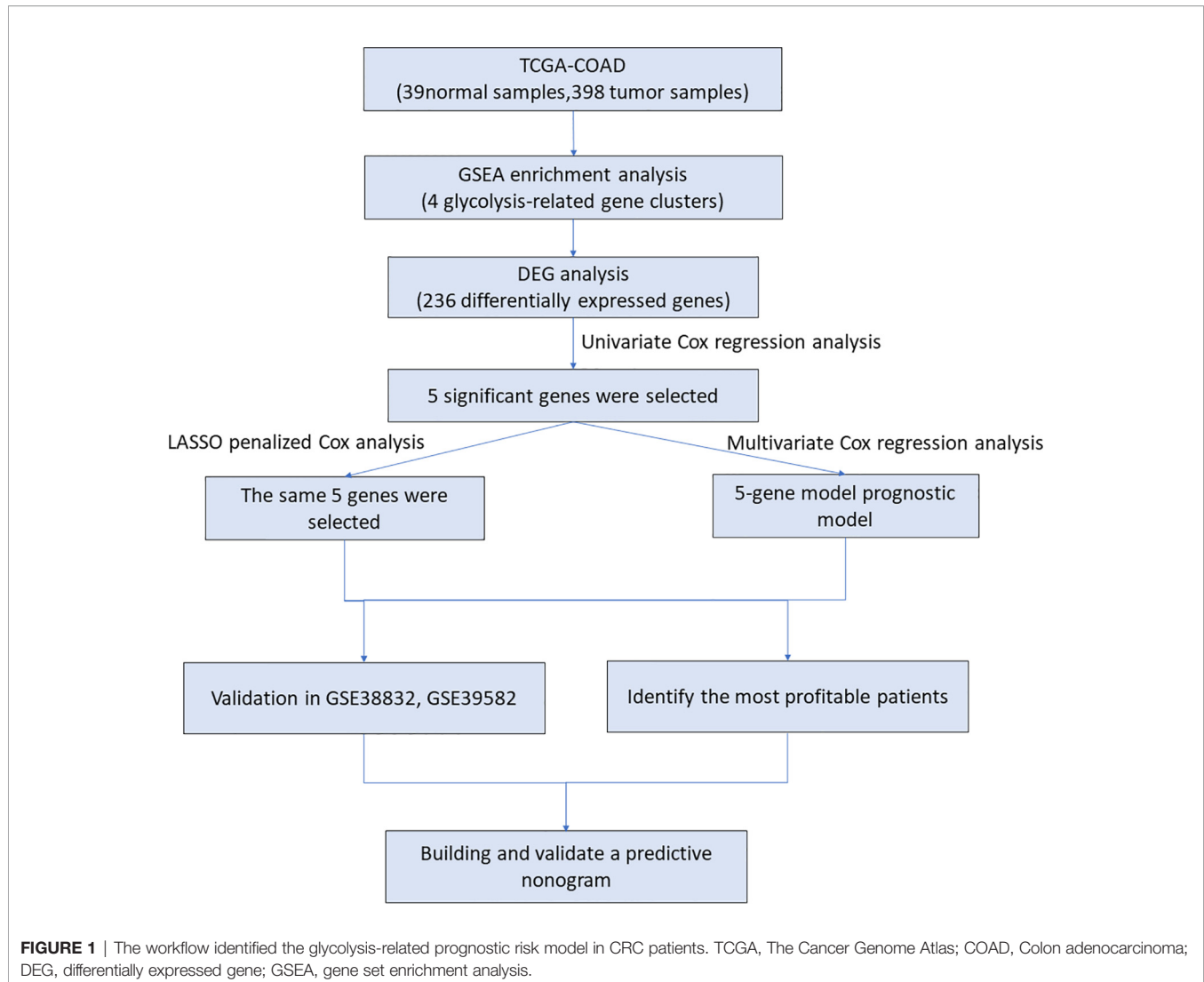
These 298 glycolytic genes were analyzed for differential expression between tumor samples and non-tumor tissues. To construct a better risk model, logFC (log₂ Fold change) was not strictly restricted, and we defined that gene could be retained only if $|\logFC| > 0$ and $P < 0.05$. The results showed that most (236) glycolytic genes were selected, and GSEA analysis was undoubtedly effective and reliable.

Five Glycolysis-Related Genes Were Selected by Univariate Cox Analysis

Univariate Cox regression analysis was performed for preliminary selection of prognostic genes, and five genes were established when the cut-off value was $P < 0.05$ (**Table 1**). The five genes are Enolase 3 (ENO3), Glypican-1 (GPC1), Prolyl 4-hydroxylase subunit alpha 1 (P4HA1), Sperm associated antigen 4 (SPAG4), and Stanniocalcin 2 (STC2). As is expected, these five genes are highly expressed in tumor tissues than in normal samples (**Figures 3A–E**), and patients with low gene expression had a better prognosis than those with high expression (**Figures 3F–J**). The Survminer R package was conducted to find the optimal cut-off value for distinguishing between high and low expression of these genes. These results showed these five glycolytic genes could emerge as reliable predictors for CRC patients' outcomes and were chosen for further building risk score model.

Construction of Risk Score Model by Multivariate and LASSO Regression Analyses

Multivariate Cox and LASSO regression analyses were designed to filter the hub genes in the risk model to predict CRC patients' outcomes. Five genes were all accepted in the risk model, and coefficients were validated by multivariate Cox analysis (strengthened by forward and backward method) (**Table 2**). The minimum Akaike information criterion (AIC) of the risk model is 709.03. Moreover, to verify the role of these genes in the risk model, LASSO algorithm was analyzed, and the results showed that any of the five genes could not be omitted (optimal lambda value 0.0049; **Figure 4**). Based on the five genes and coefficients, we calculated each CRC patient's risk score and divided the samples into two groups (high and low expression) according to the median risk score. After excluding patients whose survival time or status is not available (NA), only



379 CRC patients remained and K–M curves showed that the differences between the high- and low-risk groups were significant ($P = 0.002$; **Figure 5A**). The survival time, survival status, and gene expression of five genes in every CRC patient were vividly shown in **Figures 5E, F**.

Risk Score Model

Five genes were identified and subsequently used to construct a prognostic gene signature. The risk score = $0.317856 \times \text{ENO3} + 0.027138 \times \text{P4HA1} + 0.039965 \times \text{GPC1} + 0.070012 \times \text{SPAG4} + 0.044451 \times \text{STC2}$, while ROC and Kaplan–Meier curves were used to assess the prognostic values of the risk scores (**Figures 5C**). The area under the ROC curves (AUCs) of the risk model were as follows: 1-year AUC: 0.606, 3-year AUC: 0.64, 5-year AUC: 0.766. Time-dependent ROC analysis showed that the risk model exhibits strong predictive ability, especially forecasting 5-year and longer OS (**Figure 5B**).

To verify whether the gene signature could be an independent prognostic indicator, univariate and multivariate Cox analyses were performed. The results of the univariate analysis demonstrated that the risk score, AJCC stage, M stage, N stage, T stage, and age were significantly correlated with OS ($P < 0.05$; **Figure 6A**). Likewise, the results of the multivariate Cox regression analysis documented that the risk score and age were still significantly associated with the OS ($P < 0.05$; **Figure 6B**). Our results indicated that the glycolysis-related risk score model could be a clinically independent prognostic factor for CRC patients.

To find which groups of patients could benefit more from the risk model, we divided CRC patients into these parts: female and male, age ≤ 65 and age > 65 , T1–T2 and T3–T4, N0 and N1–2, and M0 and M1. We found that this model is more suitable for the following patients through K–M curves and log-rank test: female patients, age > 65 , T3–T4, N0, and M0 (**Figure 7**). We explored the different expression of these five genes in the various stages (TNM stage) and found that GPC1 and STC2 had a significant increasing

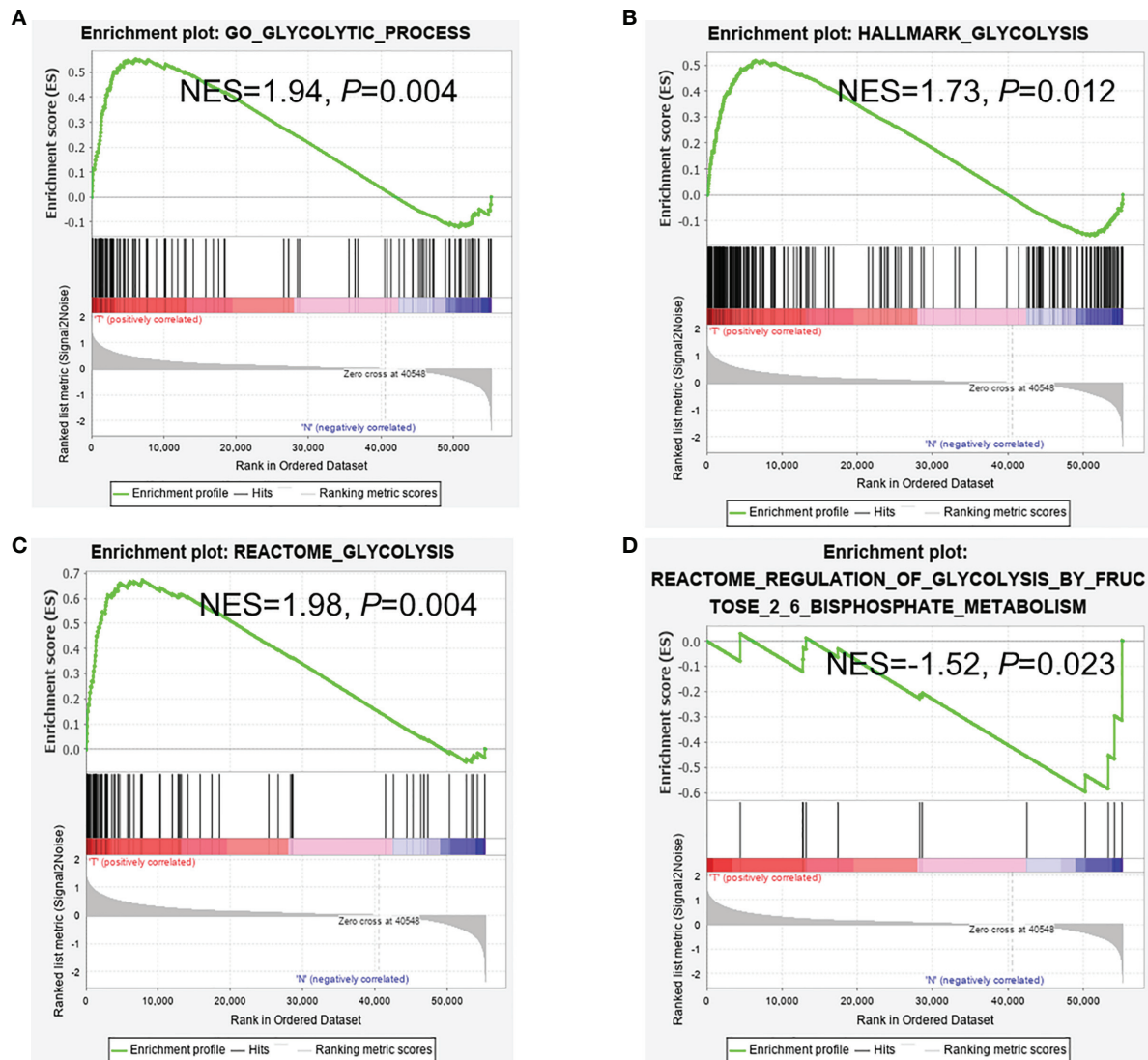


FIGURE 2 | GSEA analysis between cancer and normal samples. **(A–C)** GSEA reveals that glycolysis pathways were enriched in CRC tissues and **(D)** in adjacent normal tissues. **(A)** Go glycolytic process; **(B)** Hallmark glycolysis; **(C)** Reactome glycolysis; **(D)** Reactome regulation of glycolysis.

trend with the advanced stage and N stage (**Figure S1**). The results suggested that GPC1 and STC2 had a close relationship with clinical metastasis in CRC patients.

Validation of Risk Score Model in GEO Data Set

In an external experimental group (GSE38832), there are 122 CRC patients, and the follow-up data was disease-free survival (DFS). Based on TCGA multivariable Cox model coefficients, every patient's score was calculated, and the cut-off value of the risk score is 5.69 determined by Survminer R package. Hence, two subgroups (high- and low-risk groups) were separated, and the K–M curves confirmed that patients with low-risk scores had a more favorable DFS than those with high scores ($P = 0.006$;

Figure 8A). Similarly, another external group (GSE39582) also comes from GEO with 579 CRC patients. The results also demonstrated that there were significant differences in OS between the two groups ($P = 0.011$, **Figure 8B**).

Genetically Alteration of Five Glycolytic Genes

The mutation status of five genes was analyzed by the cBioPortal database. The five genes altered in 114 (35.54%) of 332 colon cancer patients and 37 (27.21%) of 136 rectal cancer patients (**Figure 8C**). The alteration of each gene was shown in **Figure 8D**. SPAG4 and GPC1 altered at 14 and 10%, respectively. Amplification and mRNA high were the primary mutated type. The co-expression networks of these five genes were shown in

TABLE 1 | The results of univariate Cox analysis in CRC.

Gene	β (Cox)	HR	HR.95L	HR.95H	P value
ENO3	0.3989	1.490143	1.141657	1.945002	0.003339
P4HA1	0.0239	1.024144	1.005205	1.043441	0.012244
GPC1	0.02693	1.037619	1.004159	1.072195	0.027237
SPAG4	0.1174	1.12462	1.047549	1.20736	0.001185
STC2	0.0442	1.045169	1.002151	1.090033	0.039384

HR, hazard ratio; β is calculated by univariate Cox analysis.

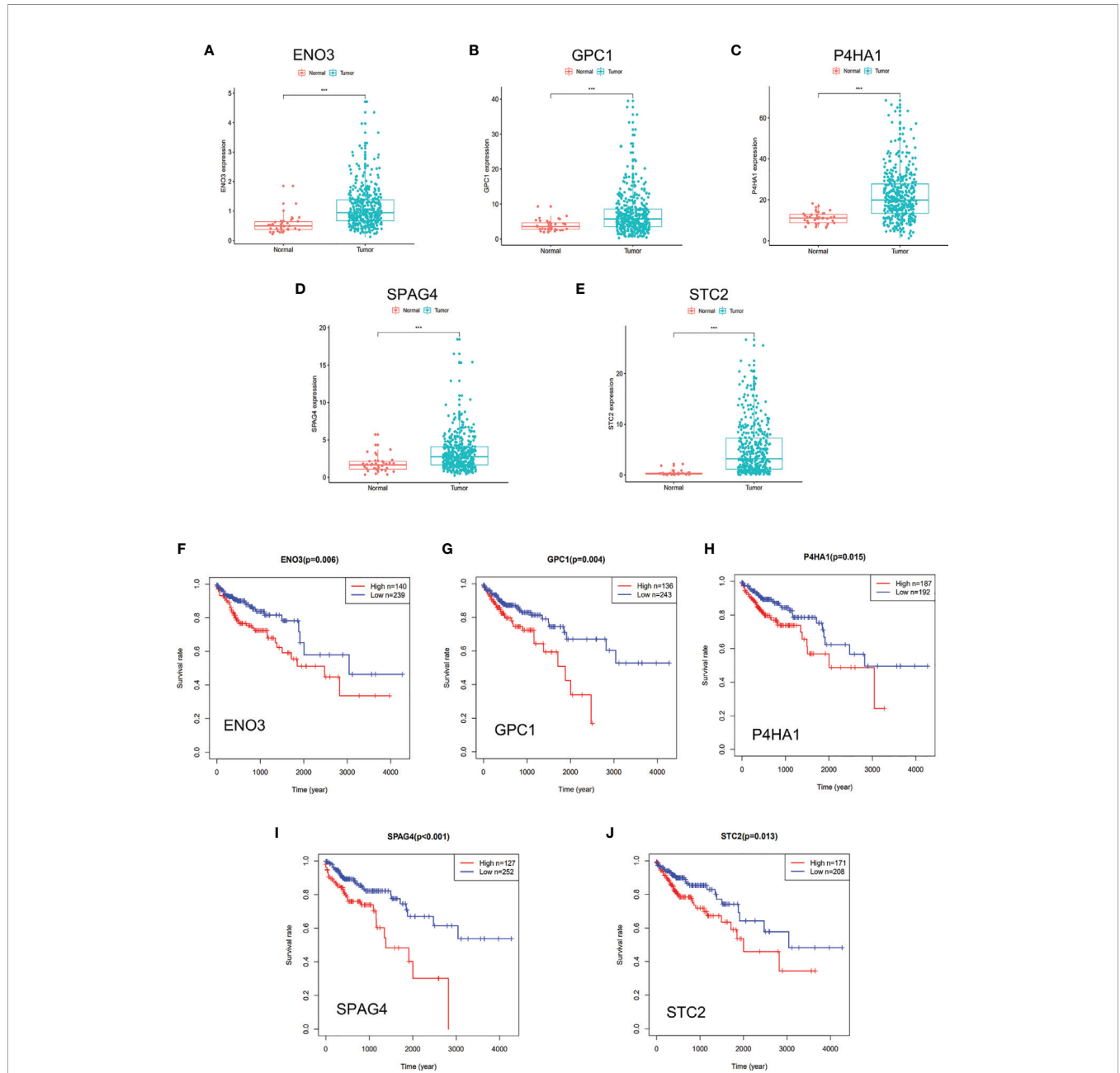


FIGURE 3 | DGE and survival analysis of five glycolysis-related genes. **(A–E)** Five glycolysis-related genes were upregulated in tumor tissue than adjacent in normal tissue. **(F–J)** Kaplan–Meier curves revealed that patients with low gene expression had better outcomes than those with high expression. **(A, F)** ENO3, **(B, G)** GPC1, **(C, H)** P4HA1, **(D, I)** SPAG4, **(E, J)** STC2. *** represents $P < 0.001$.

TABLE 2 | The results of univariate Cox analysis in CRC.

Gene	coefficients	HR
ENO3	0.317857	1.37418
P4HA1	0.027138	1.02751
GPC1	0.039966	1.040775
SPAG4	0.070012	1.072521
STC2	0.044451	1.045454

HR, hazard ratio.

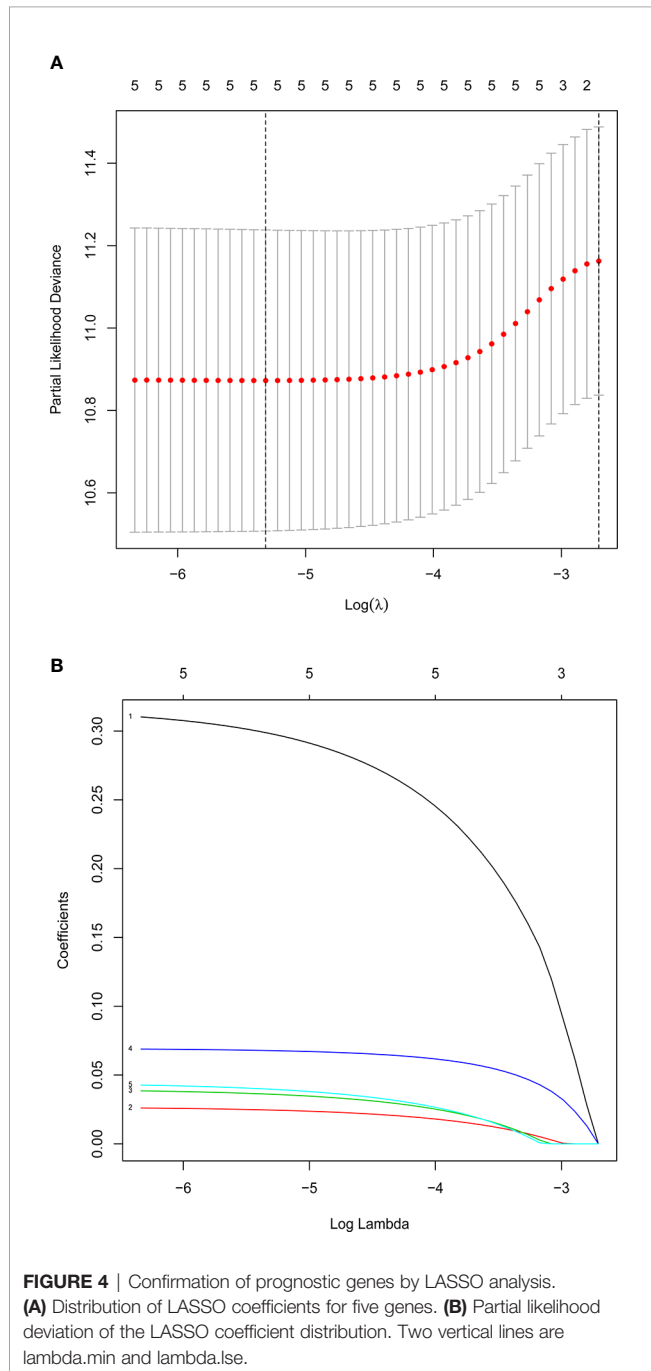


Figure 8E. ENO3, GPC1, and P4HA1 have more robust co-expression networks than SPAG4 and STC2.

Build a Nomogram Based on the Risk Score Models

For predicting 3- and 5-year OS, a nomogram was built (**Figure 9A**), and calibration and C-index were used to evaluate the discrimination and accuracy of the nomogram. In the TCGA dataset, the C-index was 0.7905. The 3- and 5-year survival probability calibration curves for the TCGA datasets demonstrated fair agreements between prediction and observation (**Figures 9B, C**).

Validation of Different Expression in Protein and mRNA Levels

HPA database is the famous database to detect protein expression in various cancers (26). We downloaded the pictures of IHC, and the results showed the protein expression of five genes was highly upregulated in cancer tissues than in normal intestinal tissue (**Figures 10A–E**). Moreover, mRNA expression from my collected 10 CRC patients revealed the same results as the HPA database (**Figures 10 F–J**). In the *in vitro* experiments, we used a normal colorectal cell (NCM460) and five CRC cells (CACO2, HCT116, SW620, SW480 and HT29). The results of RT-PCR indicated that GPC1, STC2, P4HA4, and ENO3 were upregulated in cancer cells while SPAG4 was downregulated in the five cancer cells (**Figure S2**).

DISCUSSION

Recent developments in energy metabolism have led to a renewed interest in a deeper understanding of malignant tumors. Moreover, the interpretation of cancer as a genetic disease has gradually been displaced by a metabolic disease (27). The unique metabolic phenotype of cancer cells was the “Warburg effect” which revealed that even in the presence of oxygen, cancer cells preferentially shift their metabolism toward glycolysis followed by lactic acid fermentation rather than oxidative phosphorylation (OXPHOS). Accumulating evidence suggests that most CRC cells demonstrate the “Warburg” metabolic phenotype, which produces ATP more rapidly than OXPHOS. Besides, numerous glycolysis-related genes and proteins have been demonstrated to be upregulated in CRC (11–18) and correlated with tumor aggressiveness and poor prognosis (21, 28). Here, we analyzed glycolysis-related genes using a series of bioinformatics analysis and constructed a risk model based on glycolysis by multivariate and LASSO Cox regression to identify a novel biomarker for CRC patients. The model could well predict CRC patients’ prognosis and be of vital importance for the diagnosis and treatment of patients clinically.

GESA indicated that four gene clusters were enriched in cancer tissue compared with adjacent normal samples. Then, DEGs between cancer and normal samples were conducted using LIMMA method, and 236 genes were verified. The risk model by

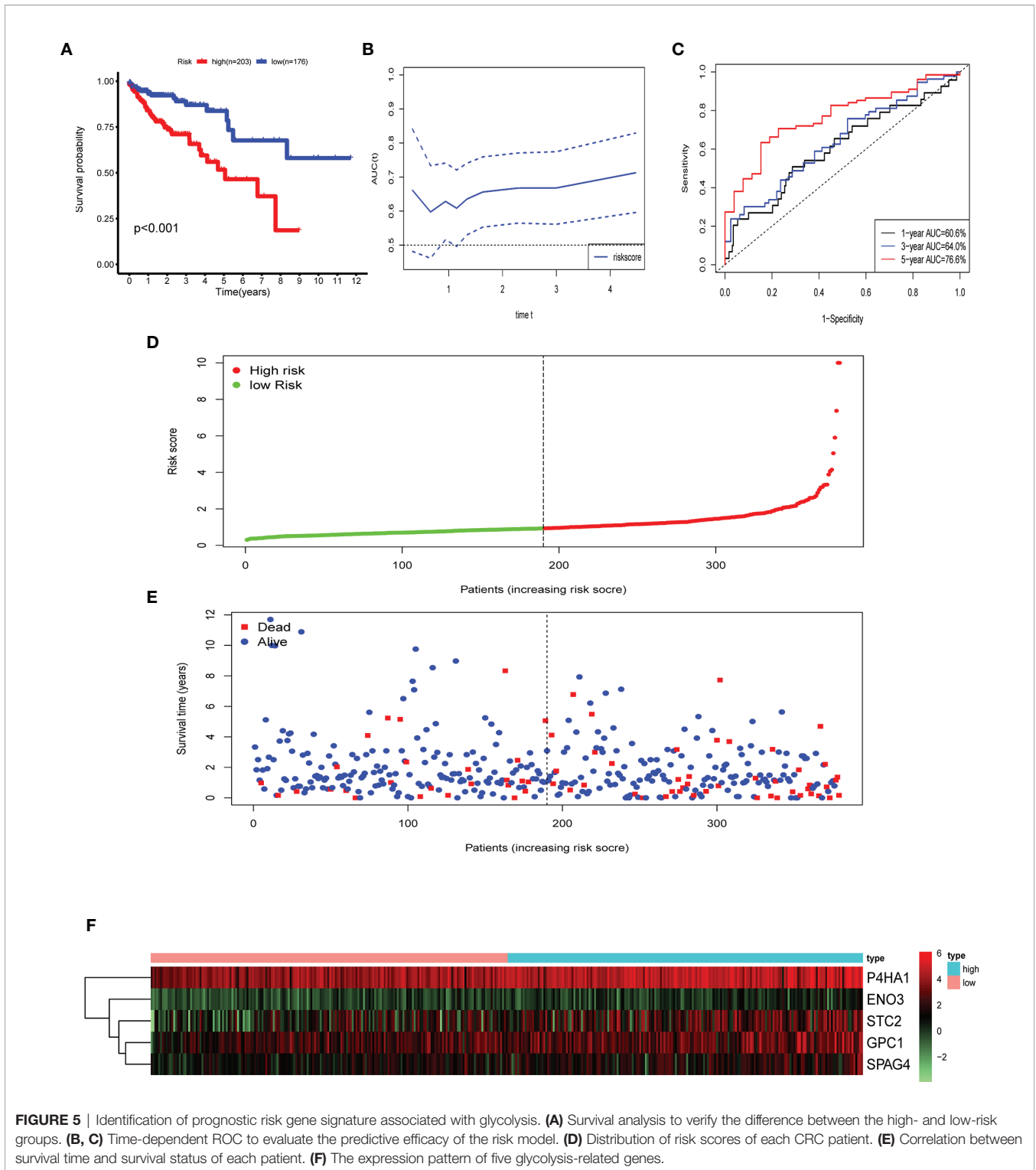


FIGURE 5 | Identification of prognostic risk gene signature associated with glycolysis. **(A)** Survival analysis to verify the difference between the high- and low-risk groups. **(B, C)** Time-dependent ROC to evaluate the predictive efficacy of the risk model. **(D)** Distribution of risk scores of each CRC patient. **(E)** Correlation between survival time and survival status of each patient. **(F)** The expression pattern of five glycolysis-related genes.

multivariate and LASSO analyses contains five genes (ENO3, GPC1, P4HA1, SPAG4, and STC2). K-M and ROC curves showed that survival time in the high-risk group is significantly shorter than in the low-risk group, and the

5-year OS predicted by this model is more reliable than 1- or 3-year OS. Enolase (ENO), also known as phosphohydrolase dehydratase, is a metalloenzyme that catalyzes the transformation of 2-phosphoglycerate to phosphoenolpyruvate

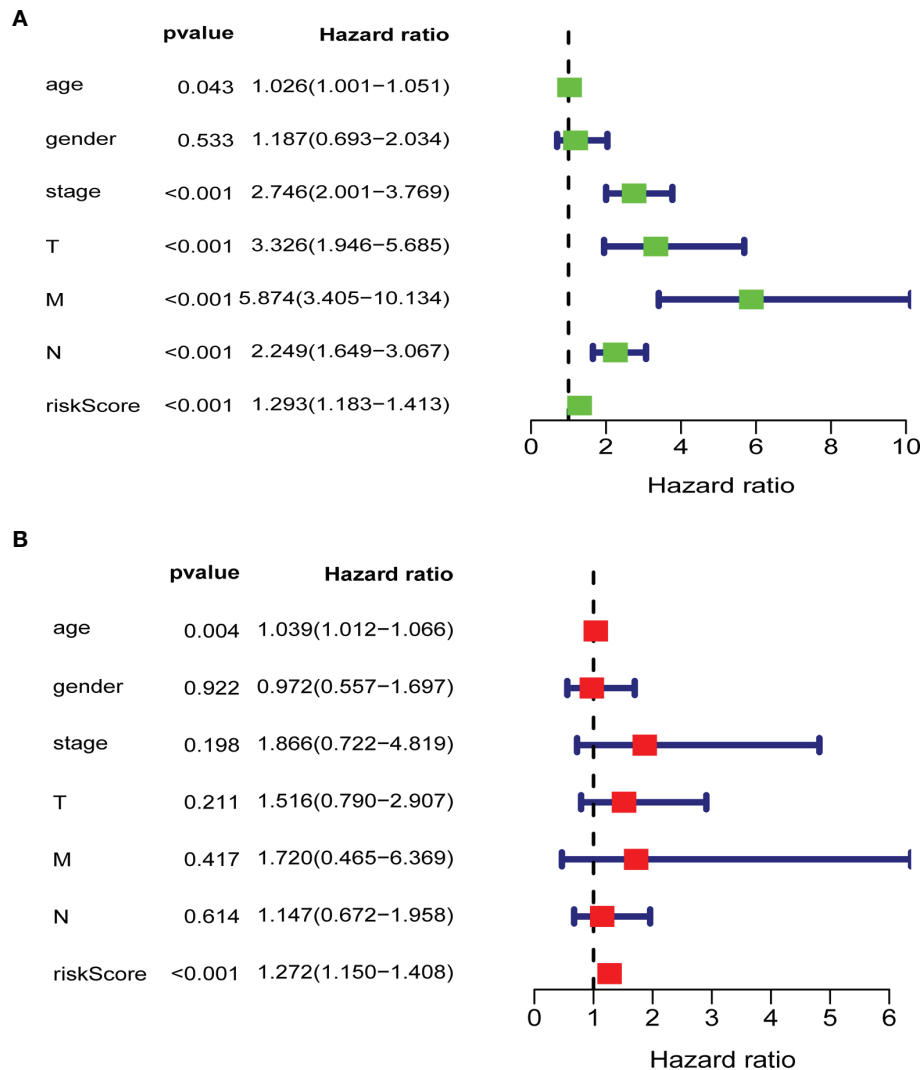


FIGURE 6 | Validation of clinical independence of risk score model. **(A)** Univariable analysis for each clinical feature (age, gender, TNM stage) and risk score model. **(B)** Multivariable analysis for risk score model and clinical characteristics (age, gender, TNM stage). The green and red boxes represent the hazard ratio, and the blue bars mean 95%CI. CI, confidence interval; T, T stage; N, N stage; M, M stage; riskScore, risk score model.

during glycolysis (29). Recent researchers found the knockdown of ENO3 exhibits anticancer effect in STK11 mutant cells and suggest that ENO3-based targeted therapy might be promising for lung cancer patients harboring STK11 mutations (30). GPC1 is a membrane-anchored protein overexpressed in multiple tumors and involved in the tumorigenesis of certain cancers, including breast cancer and pancreas cancer (31, 32). With the development of exosomes in cancer study, GPC1 has been a resurgence of interest because it is a cancer exosomes specific protein (33). GPC1+ circulating exosomes were regarded as a therapeutic target (34) and could facilitate the early detection of cancer (33). Another study demonstrated that combined detection of exosomal GPC1, exosomal CD82, and serum CA19-9 shows excellent promise as a standard method for

pancreatic cancer detection (35). Likewise, P4HA1 is the active catalytic component of prolyl 4-hydroxylase and has been reported to promote tumor progression and metastasis in several cancers (36–38). Gaofeng Xiong et al. suggested that P4HA1 promotes chemoresistance by modulating HIF-1-dependent cancer cell stemness, and targeting collagen P4H is a promising strategy to inhibit tumor progression and sensitize TNBC to chemotherapeutic agents (39). STC2 is a glycoprotein hormone and regulates malignant tumor progression (40–42), which could be a useful biomarker for survival prediction (43). The last validated gene (SPAG4) plays a vital role in spermatogenesis and sperm motility and is a mediating protein between the nucleoskeleton and cytoskeleton (44). SPAG4 promotes survival of cancer cells under hypoxic conditions and

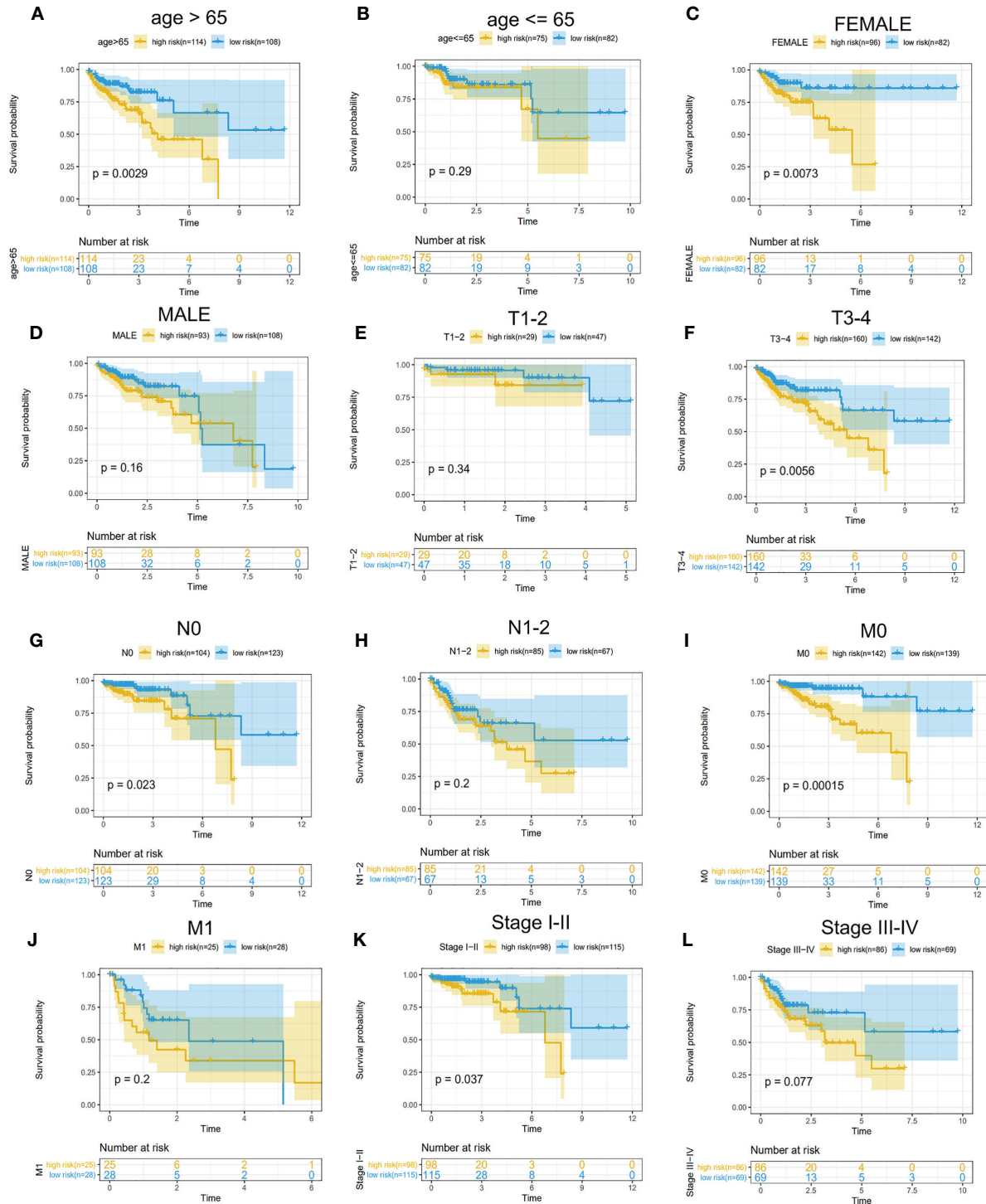


FIGURE 7 | Determination of CRC patients suitable for the model. (A–L) Survival analysis for high- and low-risk groups in different patients. (A, B) Subgroups divided by age. (C, D) Subgroups divided by gender. (E, F) T1–T2 patients were divided into a common group and T3–T4 as another. (G, H) no lymph node metastasis (NO) as a group and lymph node-positive(N1–2) as another. (I, J) subgroup divided by the status of distant metastasis. (K, L) Stages I–II as a group and III–IV as another.

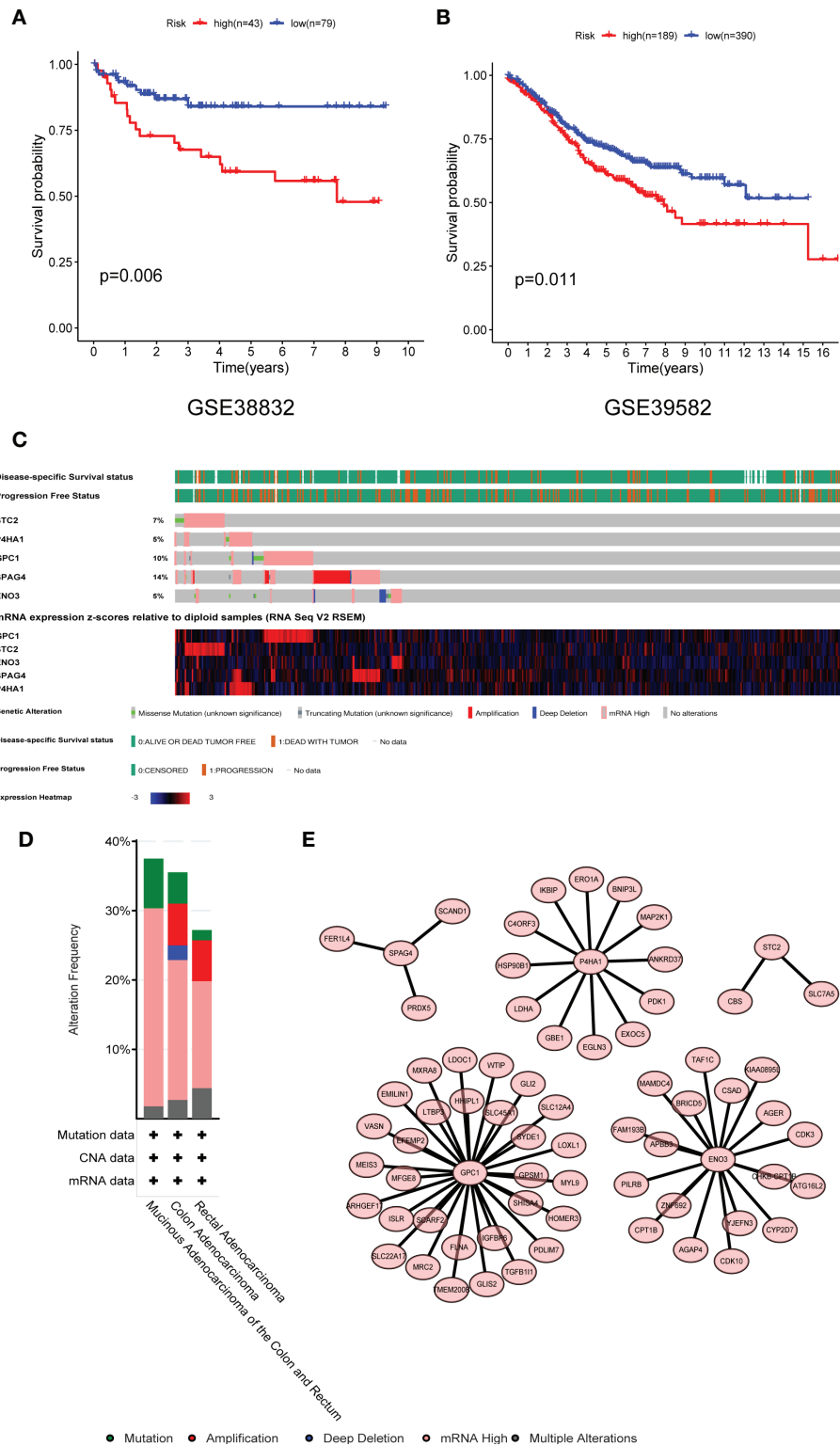
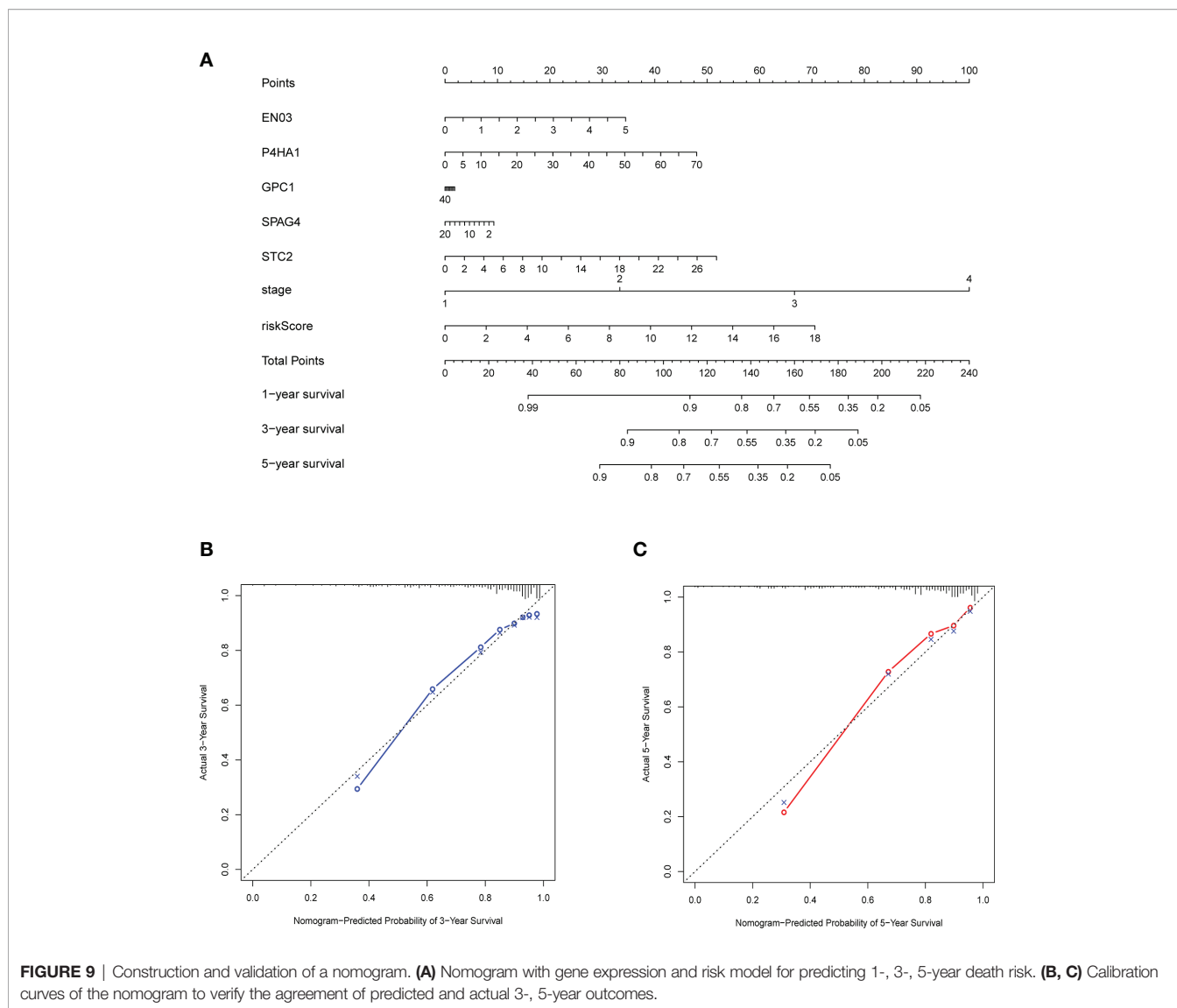


FIGURE 8 | Validation of the risk model in the GEO dataset and mutational profiling of five genes. **(A)** GSE38832 and **(B)** GSE39582 dataset. **(C)** A visual summary of gene alteration from CRC. **(D)** The total alteration of five key genes. **(E)** The network of five glycolytic genes and high related expressed genes. GEO, Gene Expression Omnibus database.



leads to poor prognosis of several cancers, including renal cell carcinoma (45), glioblastoma (46), Pancreatic Ductal Adenocarcinoma (47). The above literature fully demonstrates that all of the selected genes are upregulated in cancer tissues. We used a statistical algorithm to construct a gene model that comprehensively integrated each gene's prediction effect to enhance prediction efficiency.

Considering that the mRNA level of five genes was upregulated in cancerous tissues, we adopted immunohistochemical images of CRC tissues in the HPA database. The results were consistent with the variability of mRNA, which confirmed that the expression level of these proteins in cancer tissues was significantly higher than that in adjacent tissues. Moreover, based on basic clinical features, different subgroups were divided to verify the risk model's prediction. Old female patients with early TNM stage could be predicted more precisely than young male patients with advanced stage. Another research showed that CRC did not have a strong relationship with glycolysis (48), which might be explained

by the fact that they selected only three glycolysis-related gene sets and had no results of multivariate Cox regression.

Nevertheless, the present study has some limitations. Firstly, these five genes are still not reported to be key genes in the glycolysis pathway, and the underlying mechanism needs to be explored. Second, many patients had zero survival days and were regarded as one day survival in our study, which could have biased the results. In the future, we need large cohorts and basic experiments for further exploring the potential mechanism of these glycolysis-related genes.

In the current study, we provide novel insights into the relationship between glycolysis and CRC and established a glycolysis-related gene signature that could be applied to analyze patients' prognosis with CRC. Furthermore, the nomogram models provided an insightful and applicative tool to evaluate CRC prognosis. This signature could be a promising therapeutic target in CRC patients with poor prognoses.

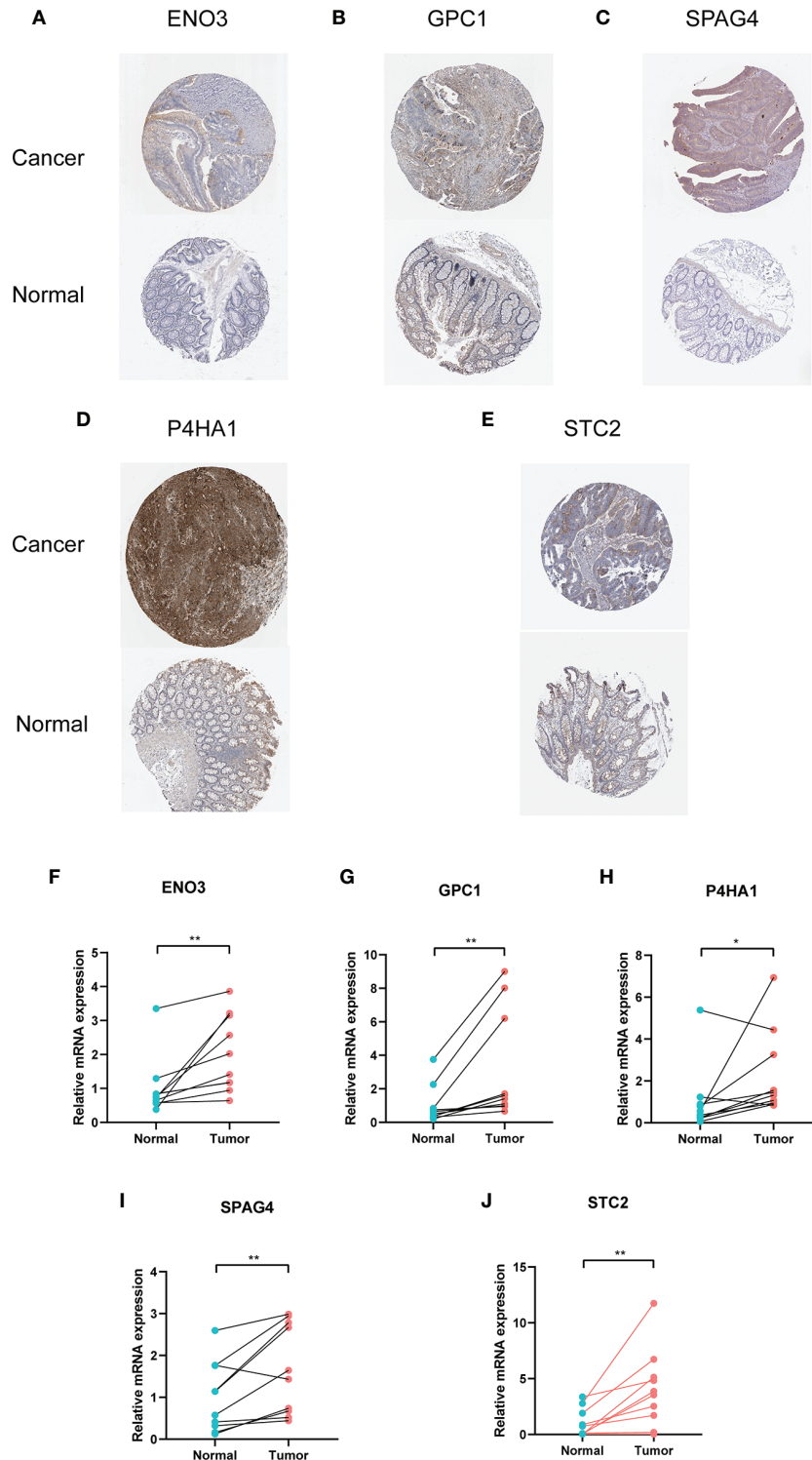


FIGURE 10 | Immunohistochemistry and mRNA of the five genes by HPA database and RT-PCR. **(A)** ENO3, **(B)** GPC1, **(C)** SPAG4, **(D)** P4HA1, **(E)** STC2. **(F–J)** Relative mRNA expression of five genes in cancerous and paracancerous tissue. * means $P < 0.05$ and ** represents $P < 0.01$.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of the First Affiliated Hospital of the Air Force Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JL designed the study. JuZ and SW contributed to the conception of the study. HB, KW, and JH contributed significantly to analysis and manuscript preparation. JuZ and SW performed the data analyses and wrote the manuscript. JL and JiZ helped

perform the analysis with constructive discussions. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by grants from the National Natural Science Foundation of China (81672751) and the Key Research and Development Program of Shaanxi (2019SF-010).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.588811/full#supplementary-material>

Supplementary Figure 1 | The relationship between five genes and clinical stage. (A) Expression of five glycolytic genes in the stage. (B) Expression of five glycolytic genes in the N stage.

Supplementary Figure 2 | The differentially expressed of five genes in cell lines by RT-PCR. (A) GPC1; (B) STC2; (C) P4HA4; (D) SPAG4; (E) ENO3. NCM460: a kind of normal colorectal cells. CACO2, HCT116, SW620, SW480, HT29: five common colorectal cancer cells.

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
- Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* (2015) 136(5):E359–86. doi: 10.1002/ijc.29210
- Huibregtse JM, Scheffner M, Beaudenon S, Howley PM. A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proc Natl Acad Sci USA* (1995) 92(11):5249. doi: 10.1073/pnas.92.11.5249-b
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin* (2018) 68(1):7–30. doi: 10.3322/caac.21442
- Hanahan D, Weinberg R. Hallmarks of cancer: the next generation. *Cell* (2011) 144(5):646–74. doi: 10.1016/j.cell.2011.02.013
- Brown R, Short S, Williams C. Colorectal Cancer and Metabolism. *Curr colorectal Cancer Rep* (2018) 14(6):226–41. doi: 10.1007/s11888-018-0420-y
- Gatenby R, Gillies R. Why do cancers have high aerobic glycolysis? *Nat Rev Cancer* (2004) 4(11):891–9. doi: 10.1038/nrc1478
- WARBURG O. On the origin of cancer cells. *Sci (New York N.Y.)* (1956) 123(3191):309–14. doi: 10.1126/science.123.3191.309
- Furlan D, Sahnane N, Carnevali I, Cerutti R, Uccella S, Bertolini V, et al. Up-regulation and stabilization of HIF-1 α in colorectal carcinomas. *Surg Oncol* (2007) 16 Suppl 1:S25–7. doi: 10.1016/j.suronc.2007.10.014
- Jin F, Yang R, Wei Y, Wang D, Zhu Y, Wang X, et al. HIF-1 α -induced miR-23a~27a~24 cluster promotes colorectal cancer progression via reprogramming metabolism. *Cancer Lett* (2019) 440–1:211–22. doi: 10.1016/j.canlet.2018.10.025
- Leclerc D, Pham DN, Lévesque N, Truongcao M, Foulkes WD, Sapienza C, et al. Oncogenic role of PDK4 in human colon cancer cells. *Br J Cancer* (2017) 116(7):930–6. doi: 10.1038/bjc.2017.38
- Azuma M, Shi M, Danenberg KD, Gardner H, Barrett C, Jacques CJ, et al. Serum lactate dehydrogenase levels and glycolysis significantly correlate with tumor VEGFA and VEGFR expression in metastatic CRC patients. *Pharmacogenomics* (2007) 8(12):1705–13. doi: 10.2217/14622416.8.12.1705
- Waldhart AN, Dykstra H, Peck AS, Boguslawski EA, Madaj ZB, Wen J, et al. Phosphorylation of TXNIP by AKT Mediates Acute Influx of Glucose in Response to Insulin. *Cell Rep* (2017) 19(10):2005–13. doi: 10.1016/j.celrep.2017.05.041
- Rathmell JC, Fox CJ, Plas DR, Hammerman PS, Cinalli RM, Thompson CB. Akt-directed glucose metabolism can prevent Bax conformation change and promote growth factor-independent survival. *Mol Cell Biol* (2003) 23(20):7315–28. doi: 10.1128/mcb.23.20.7315-7328.2003
- Graziano F, Ruzzo A, Giacomini E, Ricciardi T, Aprile G, Loupakis F, et al. Glycolysis gene expression analysis and selective metabolic advantage in the clinical progression of colorectal cancer. *Pharmacogenomics J* (2017) 17(3):258–64. doi: 10.1038/tpj.2016.13
- Wang NN, Zhang PZ, Zhang J, Wang HN, Li L, Ren F, et al. Penfluridol triggers mitochondrial-mediated apoptosis and suppresses glycolysis in colorectal cancer cells through down-regulating hexokinase-2. *Anat Rec (Hoboken)* (2020) 304(3):520–30. doi: 10.1002/ar.24464
- Yang P, Li Z, Fu R, Wu H, Li Z. Pyruvate kinase M2 facilitates colon cancer cell migration via the modulation of STAT3 signalling. *Cell Signal* (2014) 26(9):1853–62. doi: 10.1016/j.cellsig.2014.03.020
- Yan XL, Zhang XB, Ao R, Guan L. Effects of shRNA-Mediated Silencing of PKM2 Gene on Aerobic Glycolysis, Cell Migration, Cell Invasion, and Apoptosis in Colorectal Cancer Cells. *J Cell Biochem* (2017) 118(12):4792–803. doi: 10.1002/jcb.26148
- Shim H, Dolde C, Lewis B, Wu C, Dang G, Jungmann R, et al. c-Myc transactivation of LDH-A: implications for tumor metabolism and growth. *Proc Natl Acad Sci USA* (1997) 94(13):6658–63. doi: 10.1073/pnas.94.13.6658
- Tang J, Yan T, Bao Y, Shen C, Yu C, Zhu X, et al. LncRNA GLCC1 promotes colorectal carcinogenesis and glucose metabolism by stabilizing c-Myc. *Nat Commun* (2019) 10(1):3499. doi: 10.1038/s41467-019-11447-8
- Bian Z, Zhang J, Li M, Feng Y, Wang X, Zhang J, et al. LncRNA-FEZF1-AS1 Promotes Tumor Proliferation and Metastasis in Colorectal Cancer by Regulating PKM2 Signaling. *Clin Cancer Res an Off J Am Assoc Cancer Res* (2018) 24(19):4808–19. doi: 10.1158/1078-0432.ccr-17-2967
- Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet* (2019) 394(10207):1467–80. doi: 10.1016/s0140-6736(19)32319-0
- Martinez-Ledesma E, Verhaak RG, Treviño V. Identification of a multi-cancer gene expression biomarker for cancer clinical outcomes using a network-based algorithm. *Sci Rep* (2015) 5:11966. doi: 10.1038/srep11966

24. Bean DM, Al-Chalabi A, Dobson RJB, Iacoangeli A. A Knowledge-Based Machine Learning Approach to Gene Prioritisation in Amyotrophic Lateral Sclerosis. *Genes* (2020) 11(6):668. doi: 10.3390/genes11060668
25. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol Off J Am Soc Clin Oncol* (2008) 26(8):1364–70. doi: 10.1200/jco.2007.12.9791
26. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science* (2015) 347(6220):1260419. doi: 10.1126/science.1260419
27. Upadhyay M, Samal J, Kandpal M, Singh OV, Vivekanandan P. The Warburg effect: insights from the past decade. *Pharmacol Ther* (2013) 137(3):318–30. doi: 10.1016/j.pharmthera.2012.11.003
28. Xu H, Zeng Y, Liu L, Gao Q, Jin S, Lan Q, et al. PRL-3 improves colorectal cancer cell proliferation and invasion through IL-8 mediated glycolysis metabolism. *Int J Oncol* (2017) 51(4):1271–9. doi: 10.3892/ijo.2017.4090
29. Ho JA, Chang HC, Shih NY, Wu LC, Chang YF, Chen CC, et al. Diagnostic detection of human lung cancer-associated antigen using a gold nanoparticle-based electrochemical immunosensor. *Anal Chem* (2010) 82(14):5944–50. doi: 10.1021/ac1001959
30. Park C, Lee Y, Je S, Chang S, Kim N, Jeong E, et al. Overexpression and Selective Anticancer Efficacy of ENO3 in STK11 Mutant Lung Cancers. *Mol Cells* (2019) 42(11):804–9. doi: 10.14348/molcells.2019.0099
31. Matsuda K, Maruyama H, Guo F, Kleeff J, Itakura J, Matsumoto Y, et al. Glypican-1 is overexpressed in human breast cancer and modulates the mitogenic effects of multiple heparin-binding growth factors in breast cancer cells. *Cancer Res* (2001) 61(14):5562–9.
32. Kleeff J, Ishiwata T, Kumbasar A, Friess H, Büchler MW, Lander AD, et al. The cell-surface heparan sulfate proteoglycan glypican-1 regulates growth factor action in pancreatic carcinoma cells and is overexpressed in human pancreatic cancer. *J Clin Invest* (1998) 102(9):1662–73. doi: 10.1172/jci4105
33. Melo SA, Luecke LB, Kahlert C, Fernandez AF, Gammon ST, Kaye J, et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature* (2015) 523(7559):177–82. doi: 10.1038/nature14581
34. Li J, Chen Y, Guo X, Zhou L, Jia Z, Peng Z, et al. GPC1 exosome and its regulatory miRNAs are specific markers for the detection and target therapy of colorectal cancer. *J Cell Mol Med* (2017) 21(5):838–47. doi: 10.1111/jcmm.12941
35. Xiao D, Dong Z, Zhen L, Xia G, Huang X, Wang T, et al. Combined Exosomal GPC1, CD82, and Serum CA19-9 as Multiplex Targets: A Specific, Sensitive, and Reproducible Detection Panel for the Diagnosis of Pancreatic Cancer. *Mol Cancer Res* (2020) 18(2):300–10. doi: 10.1158/1541-7786.mcr-19-0588
36. Cao XP, Cao Y, Li WJ, Zhang HH, Zhu ZM. P4HA1/HIF1 α feedback loop drives the glycolytic and malignant phenotypes of pancreatic cancer. *Biochem Biophys Res Commun* (2019) 516(3):606–12. doi: 10.1016/j.bbrc.2019.06.096
37. Gilkes DM, Chaturvedi P, Bajpai S, Wong CC, Wei H, Pitcairn S, et al. Collagen prolyl hydroxylases are essential for breast cancer metastasis. *Cancer Res* (2013) 73(11):3285–96. doi: 10.1158/0008-5472.can-12-3963
38. Xiong G, Deng L, Zhu J, Rychahou PG, Xu R. Prolyl-4-hydroxylase α subunit 2 promotes breast cancer progression and metastasis by regulating collagen deposition. *BMC Cancer* (2014) 14:1. doi: 10.1186/1471-2407-14-1
39. Xiong G, Stewart RL, Chen J, Gao T, Scott TL, Samayoa LM, et al. Collagen prolyl 4-hydroxylase 1 is essential for HIF-1 α stabilization and TNBC chemoresistance. *Nat Commun* (2018) 9(1):4456. doi: 10.1038/s41467-018-06893-9
40. Lin C, Sun L, Huang S, Weng X, Wu Z. STC2 Is a Potential Prognostic Biomarker for Pancreatic Cancer and Promotes Migration and Invasion by Inducing Epithelial-Mesenchymal Transition. *BioMed Res Int* (2019) 2019:8042489. doi: 10.1155/2019/8042489
41. Ke J, Zhang BH, Li YY, Zhong M, Ma W, Xue H, et al. MiR-1-3p suppresses cell proliferation and invasion and targets STC2 in gastric cancer. *Eur Rev Med Pharmacol Sci* (2019) 23(20):8870–7. doi: 10.26355/eurrev_201910_19282
42. Liu YN, Tsai MF, Wu SG, Chang TH, Tsai TH, Gow CH, et al. Acquired resistance to EGFR tyrosine kinase inhibitors is mediated by the reactivation of STC2/JUN/AXL signaling in lung cancer. *Int J Cancer* (2019) 145(6):1609–24. doi: 10.1002/ijc.32487
43. Huang R, Meng T, Chen R, Yan P, Zhang J, Hu P, et al. The construction and analysis of tumor-infiltrating immune cell and ceRNA networks in recurrent soft tissue sarcoma. *Aging (Albany NY)* (2019) 11(22):10116–43. doi: 10.18632/aging.102424
44. Ji Y, Jiang J, Huang L, Feng W, Zhang Z, Jin L, et al. Sperm-associated antigen 4 (SPAG4) as a new cancer marker interacts with Nesprin3 to regulate cell migration in lung carcinoma. *Oncol Rep* (2018) 40(2):783–92. doi: 10.3892/or.2018.6473
45. Shoji K, Murayama T, Mimura I, Wada T, Kume H, Goto A, et al. Sperm-associated antigen 4, a novel hypoxia-inducible factor 1 target, regulates cytokinesis, and its expression correlates with the prognosis of renal cell carcinoma. *Am J Pathol* (2013) 182(6):2191–203. doi: 10.1016/j.ajpath.2013.02.024
46. Zhao J, Liu B, Yang JA, Tang D, Wang X, Chen Q. Human sperm-associated antigen 4 as a potential biomarker of glioblastoma progression and prognosis. *Neuroreport* (2019) 30(6):446–51. doi: 10.1097/wnr.0000000000001226
47. Tian G, Li G, Liu P, Wang Z, Li N. Glycolysis-Based Genes Associated with the Clinical Outcome of Pancreatic Ductal Adenocarcinoma Identified by The Cancer Genome Atlas Data Analysis. *DNA Cell Biol* (2020) 39(3):417–27. doi: 10.1089/dna.2019.5089
48. Jiang L, Zhao L, Bi J, Guan Q, Qi A, Wei Q, et al. Glycolysis gene expression profilings screen for prognostic risk signature of hepatocellular carcinoma. *Aging (Albany NY)* (2019) 11(23):10861–82. doi: 10.18632/aging.102489

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Zhu, Wang, Bai, Wang, Hao, Zhang and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.