

Nine glycolysis-related gene signature predicting the survival of patients with endometrial adenocarcinoma

liu jinhui

The first affiliated hospital of nanjing medical university

Li siyue

The first affiliated hospital of nanjing medical university

Gao feng

the first affiliated hospital of nanjing medical university

meng huangyang

The first affiliated hospital of nanjing medical university

Nie sipei

The first affiliated hospital of nanjing medical university

Sun rui

The first affiliated hospital of nanjing medical university

yang jing

The first affiliated hospital of nanjing medical university

wenjuncheng (✉ wenjunchengdoc@163.com)

Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

<https://orcid.org/0000-0002-3872-4384>

Primary research

Keywords: Endometrial cancer, Glycolysis, prognostic model, TCGA, GSEA

Posted Date: February 4th, 2020

DOI: <https://doi.org/10.21203/rs.2.22618/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 Cancer Cell International on May 24th, 2020. See the published version at <https://doi.org/10.1186/s12935-020-01264-1>.

Abstract

Background: Endometrial cancer is the fourth most common cancer in women. The death rate for endometrial cancer has increased. Glycolysis of cellular respiration is a complex reaction and is the first step in most carbohydrate catabolism, which was proved to participate in tumors. **Methods:** We analyzed the sample data of over 500 patients from TCGA database. The bioinformatic analysis included GSEA, cox and lasso regression analysis to select prognostic genes, as well as construction of a prognostic model and a nomogram for OS evaluation. The immunohistochemistry staining, survival analysis and expression level validation were also performed. Maftools package was for mutation analysis. GSEA identified Glycolysis was the most related pathway to EC. **Results:** According to the prognostic model using the train set, 9 glycolysis-related genes including B3GALT6, PAM, LCT, GMPPB, GLCE, DCN, CAPN5, GYS2 and FBP2 were identified as prognosis-related genes. Based on nine gene signature, the EC patients could be classified into high and low risk subgroups, and patients with high risk score showed shorter survival time. Time-dependent ROC analysis and Cox regression suggested that the risk score predicted EC prognosis accurately and independently. Analysis of test and train sets yielded consistent results A nomogram which incorporated the 9-mRNA signature and clinical features was also built for prognostic prediction. Immunohistochemistry staining and TCGA validation showed that expression levels of these genes do differ between EC and normal tissue samples. GSEA revealed that the samples of the low-risk group were mainly concentrated on Bile Acid Metabolism. Patients in the low-risk group displayed obvious mutation signatures compared with those in the high-risk group. **Conclusion:** This study found that the Glycolysis pathway is associated with EC and screened for hub genes on the Glycolysis pathway, which may serve as new target for the treatment of EC.

Background

Endometrial cancer is a kind of female malignancy. In female tumors, EC ranked fourth. In 2015, the American Cancer Society (ACS) predicted that the number of new cases of EC was 54,870, of which the number of patients died was 10,170. This means that in the past 20 years, the fatality rate of EC has increased by more than 100%. The average age of patients at diagnosis is 63. Among them, 90% of patients are over 50 years old, and only 20% of patients can get diagnosed before they menopause[1]. Though quite a lot of research have been conducted. EC is not amenable to screening, hence needs to be effectively managed once the diagnosis is made[2–5]. Glycolysis of cellular respiration is a complex reaction and is the first step in most carbohydrate catabolism. Most of the glycolysis occurs in the cytoplasm. It does not utilize any molecular oxygen to react. It is a special metabolic pathway. Increased glycolysis is the main source of energy supply in cancer cells that use this metabolic pathway for ATP generation. Increased glycolysis can produce ATP for cancer cells, becoming the main energy source for cancer cell growth and metabolism. Hence Altered energy metabolism was seen to be “hallmarks of cancer”[6]. Research by Ganapathy-Kanniappan S et al. Showed that Tumor glycolysis working as a target to treat cancer was very promising[7]. They also found that tumor glycolysis is closely related to immune evasion in cancer, which might be a brand new therapeutic opportunities[8]. Akins NS et al.

found that Inhibition of Glycolysis can fight tumors[9]. Li C et al. also found that glycolysis can be used as a new target for tumor therapy[10]. Qin W et al. found that glycolysis can regulate metastasis of gastric cancer cells[11]. Feinberg T et al. also demonstrated that glycolysis was involved in the metabolism of lung cancer[12]. This study focused on the relationship between Glycolysis and endometrial cancer treatment, screening appropriate targets, and opening up new ideas for the treatment of EC.

Material And Method

Source of obtaining data

TCGA provided mRNA data and the corresponding EC clinical information[13], which was proceeded on platform Illumina HiSeq RNA-seq[14], containing 552 EC patient samples and 35 normal tissues. After integrating clinical information, 520 samples were obtained. These samples were classified into the training cohort randomly (n = 260), the testing cohort (n = 260). The training cohort was used for prognostic model construction, while the testing cohort and entire cohort were chosen for validation.

Gene set enrichment analysis (GSEA)

GSEA (<http://software.broadinstitute.org/gsea/index.jsp>)[15] was used to determine the Molecular Signatures Database (MSigDB) provided hallmark gene sets in order to predict biological processes between the healthy samples and EC samples. the P-value < 0.05 and FDR (false discovery rate) < 0.01 were set as the cutoff.

Identification of prognostic genes and their characteristics

Univariate regression analysis, Lasso analysis and multivariate regression analysis were applied to explore the correlation between expression levels of glycolysis-related genes and patients' overall survival (OS). In the univariate Cox regression analysis, gene was seen to be a candidate prognostic gene when P-value was < 0.05. Lasso-penalized and multivariate analysis were next performed for further screening. Hazard ratios (HRs) and regression coefficient were calculated for each gene, and the satisfactory mRNAs were ultimately included. The gene alteration type and frequency, as well as the most frequently altered neighbor genes of satisfactory genes were exhibited by the cBioPortal (cBio Cancer Genomics Portal) tool[16].

Construction of the gene-related prognostic model

The prognostic risk-score model for outcomes prediction of EC patients was the combination of each optimal prognostic mRNA expression level multiplying relative regression coefficient weight calculated from the multivariate model according to the following way:

All patients from the training cohort were divided into high- and low-risk groups on the basis of the median risk score. The Kaplan–Meier survival curves of both groups were plotted and the ROC (receiver

operating characteristic) curve for OS prediction was present to assess the sensitivity and specificity of the model[17]. Cox multivariate analysis regarding several clinicopathological features of EC patients were also performed to exam the independency of the prognostic model without clinical characters.

Validation of the efficacy of the prognostic risk model

By comparing the testing cohort and entire cohort patient's risk score with the cut-off value calculated from the training cohort, each patient was categorized as the high-risk or low-risk group. Time-dependent ROC, Kaplan–Meier curve and cox multivariate analysis were also performed. Meanwhile, the stratification analysis was operated based on clinicopathological features.

Validation of the hub genes

A nomogram and calibrate curve was built by the “rms” package on R. The correctness was examined to check the consistency index between actual observation frequency and predicted probability. Then, we presented the predicted and observed results in the calibration curve to visualize the performance of the nomogram. And the 45° line represents the best prediction. TCGA data was also used so as to validate the glycolysis-related genes expression level between EC and normal samples. And the immunohistochemistry staining of both the normal and EC samples were downloaded from the Human Protein Atlas database (<https://www.proteinatlas.org/>). Survival analysis was also conducted for hub genes using “survminer” R package and “survival” R package. EC samples from TCGA were divided into two groups based on each hub gene's best-separation cut-off value to plot the Kaplan-Meier (K-M) survival curves.

Mutation analysis

The mutation data were processed and visualized by R package maftools.

Result

Functional pathway screening using GSEA

Clinical features information of a total of 587 samples including 552 EC and 35 healthy samples were achieved from the TCGA. Based on the mentioned data, GSEA indicated that whether the identified gene sets showed significant differences between EC and adjacent healthy tissues. And we found that these genes are significantly enriched in glycolysis, cholesterol homeostasis, fatty acid metabolism and xenobiotic metabolism. Glycolysis was shown to be the most relevant pathway (Fig. 1).

Establishment of Glycolysis-related genes and EC prognosis models

Firstly we integrated mRNA expression profiles and clinical information so as to screen out 520 EC samples. We analyzed 520 EC samples and found a total of 179 participating genes on the Glycolysis pathway in order to research the relationship between Glycolysis and the prognosis of EC patients. Then

we randomly selected 260 samples as training cohort and built a prognostic model for 260 samples. univariate Cox regression analysis screened out 11 genes with the cutoff of $p < 0.05$ (TABLE1). These prognosis-related Glycolysis genes were further analyzed with the least absolute shrinkage and selection operator (LASSO) Cox regression algorithm (Fig. 2A-B). Then multivariate Cox proportional hazards regression analysis built the risk signature. We constructed prognostic models and the risk scores were calculated (TABLE2). Nine genes including B3GALT6, PAM, LCT, GMPPB, GLCE, DCN, CAPN5, GYS2 and FBP2 were identified as prognosis-related genes. The risk score come out as the followed: $0.000755345 * B3GALT6 - 5.19E-05 * PAM + 0.029807032 * LCT - 0.000708518 * GMPPB + 0.000784398 * GLCE - 0.00015091 * DCN + 0.000258397 * CAPN5 + 0.031956259 * GYS2 + 0.00431111 * FBP2$. According to the median levels of risk score, EC patients were classified into low- and high-risk groups. In the model, survival analysis indicated that low-risk patients had significantly longer overall survival time than high-risk patients (Fig. 2C). We also performed the receiver operating characteristic curve (ROC) analysis. As shown in Fig. 2D, ROC curve analysis was also completed according to the 1,3,5-year survival of the area under the receiver operating characteristic curve (AUC) value, the specificity and sensitivity were highest when the risk score was 0.794, 0.765, 0.773. The risk score and survival status indicated by the prognostic model was displayed in Fig. 3A-C. To assess whether the model was an independent predictor of EC, univariate and multivariate Cox regression analyses were conducted, including risk scores and clinical factors. And the results showed that this prognostic model showed moderate and independent prognostic power for Glycolysis pathway (Fig. 3D-E).

Table 1
Univariate Cox regression analysis screened out 11 genes with the cutoff of $p < 0.05$.

ID	HR	HR.95L	HR.95H	p-value
CAPN5	1.000211613	1.000078089	1.000345156	0.001894248
LCT	1.026275024	1.009246	1.043591	0.002381
FBP2	1.003031	1.000989	1.005076	0.003603
B3GALT6	1.000592	1.000172	1.001012	0.005692
PAM	0.999942	0.999895	0.99999	0.016807
GYS2	1.025056	1.004418	1.046119	0.017089
GMPPB	0.99935	0.998794	0.999906	0.022055
GLCE	1.000492	1.000069	1.000916	0.022703
B4GALT4	0.999228	0.998555	0.999901	0.02458
ALDOB	1.001479	1.000093	1.002866	0.036525
DCN	0.999835	0.99967	1	0.049346

Table 2

Prognostic models were constructed and the risk scores were calculated.

ID	Coef	HR	HR.95L	HR.95H	p-value
B3GALT6	0.000755	1.000756	1.000327	1.001184	0.000545
PAM	0.0000519	0.999948	0.999895	1.000001	0.056424
LCT	0.029807	1.030256	1.013231	1.047567	0.000455
GMPPB	-0.00071	0.999292	0.998756	0.999828	0.009621
GLCE	0.000784	1.000785	1.000348	1.001222	0.000431
DCN	-0.00015	0.999849	0.999664	1.000035	0.110895
CAPN5	0.000258	1.000258	1.000093	1.000424	0.002159
GYS2	0.031956	1.032472	1.009372	1.056101	0.005641
FBP2	0.004311	1.00432	1.002169	1.006476	8.12E-05

Validation of Glycolysis-related genes and EC prognosis

In order to verify the authenticity of the above prognostic model, we built another prognostic model using the testing cohort(260 samples). Based on the training cohort' cut-off, samples were divided into low- and high-risk group according to the median levels of risk score. Survival analysis indicated that low-risk patients had significantly longer overall survival time than high-risk patients (Fig. 4A). ROC curve analysis showed that the specificity and sensitivity were highest when the risk score was 0.717, 0.613, 0.643 according to the 1,3,5-year survival of the area under the receiver operating characteristic curve (AUC) value (Fig. 4B). The risk score and survival status indicated by the prognostic model was displayed in Fig. 4C-E. To assess whether the model was an independent predictor of EC, univariate and multivariate analyses were completed, including clinical factors and risk scores. The results showed that this prognostic model showed moderate and independent prognostic power for Glycolysis pathway (Fig. 4F-G). These conclusions were all consistent with previous prognostic model trends, validating the reliability of our speculation that Glycolysis is involved in the development of EC and affects the prognosis of EC.

Complete Glycolysis-related prognostic model

We finally built a complete prognostic model based on entire cohort(520 samples). Based on the training cohort' cut-off and median levels of risk score, samples were classified into low- and high-risk group, and survival analysis indicated that low-risk patients had significantly longer overall survival time than high-risk patients (Fig. 5A). ROC curve analysis showed that the specificity and sensitivity were highest when the risk score was 0.763, 0.692, 0.705 according to the 1,3,5-year survival of the area under the receiver operating characteristic curve (AUC) value (Fig. 5B). The risk score and survival status indicated by the prognostic model was displayed in Fig. 5C-E. To assess whether the model was an independent predictor

of EC, univariate and multivariate analyses were done, including clinical factors and risk scores. The results showed that this prognostic model showed moderate and independent prognostic power for Glycolysis pathway (Fig. 5F-G). These further validate the reliability of our previous two prognostic models.

Hierarchical analysis of clinical features and Glycolysis-related hub genes

Univariate and multivariate Cox proportional hazards regression analysis identified nine genes including B3GALT6, PAM, LCT, GMPPB, GLCE, DCN, CAPN5, GYS2 and FBP2 to be prognosis-related. Among the 9 genes, we found significant differences in the expression levels of 7 genes in the high-risk and low-risk groups (Fig. 6A). In addition, the heatmap showed the expression of the nine genes in high- and low-risk patients in the TCGA dataset. We observed significant differences between the high- and low-risk groups associated with tumor status, grade, histological type and stage (Fig. 6B). We further analyzed the relationship between nine genes and various clinical features including risk, tumor status, grade, histological type, stage and age. We found that tumor status, grade, histological type and stage were significantly related with the 9 genes. We analyzed 9 genes for different clinical features respectively. We found that expression level of CAPN5, DCN, GLCE and GMPPB were significantly different in different age groups (FIGS1A-D). For different histological type, the expression level of B3GAL, CAPN5, GLCE, GMPPB and PAM were significantly different (FIGS2A-E). For different grade, the expression level of CAPN5, DCN, GLCE, GMPPB and PAM were significantly different (FIGS3A-E). For different tumor status, DCN, GMPPB and PAM expressed differently (FIGS4A-C). In next, the stratification analysis was done according to histological type, grade, stage, tumor status and age. Patients were stratified into endometrioid subgroups, grade G1&G2 subgroup, grade G3&G4 subgroup, tumor free subgroup, with tumor subgroup, stage I & stage II subgroup, stage III & stage IV subgroup, age > 60 subgroup and age \leq 60 subgroup. For the patients in endometrioid subgroup, the survival time of patients in the low-risk case was significantly longer than that of patients in the high-risk case (Fig. 7A), which was consistent with the results belonging to the grade G1&G2 subgroup, grade G3&G4 subgroup, stage I & stage II subgroup, stage III & stage IV subgroup, tumor free subgroup, with tumor subgroup, age > 60 subgroup and age \leq 60 subgroup. (Fig. 7B-I).

Building predictive nomogram

For the goal of establishing a clinically method to predict the survival probability with EC patients, we created a nomogram based on the TCGA cohort to estimate the probability of the 3- and 5-year OS. The predictors of the nomogram contained 6 independent prognostic factors including stage, age, histological type, grade, tumor status and risk score Fig. 8A). The C-index of the model for evaluation of OS was 0.871. The 45° line represented the best prediction. Calibration plots suggested that the nomogram performed well (Fig. 8B-C). ROC curve analysis also showed that the risk score AUC value of the model was 0.757, the clinical factors AUC value was 0.772, both much significantly higher than the clinical stage (AUC = 0.690), grade (AUC = 0.622), histological type (AUC = 0.608), tumor status (AUC = 0.751) and

patients' age (AUC = 0.578). Interestingly, when combined the risk score with clinical factors, the ROC curve of combination model was much higher than each alone (AUC = 0.805).

Based on 9 glycolysis-related gene expression, principal component analysis of the training cohort, testing cohort, and entire EC cohort displayed a significantly different distribution pattern of high and low risk which indicating their difference in glycolysis phenotype (Fig. 9A-C).

Genetic Information of the Glycolysis-related genes

The genetic alteration in the Glycolysis-related genes was analyzed with cBioPortal software. The network constructed by B3GAL, DCN, GLCE, GYS2 and their most associated neighbor genes were exhibited (only four out of the 9 genes had a joint node, while the remaining 5 genes had no junctions and were not shown) (Fig. 10A). Figure 10B-C illustrated that the 9 genes were altered in 92 (17%) from the 547 patients; LCT and CAPN5 showed most diverse alteration including amplification, missense mutation etc.

Validation of Glycolysis-related hub genes

All of the 9 glycolysis-related hub genes were validated in TCGA data. We found that DCN had the higher expression level of EC tissues than that of healthy tissues, while the B3GALT6, PAM, LCT, GMPPB, GLCE, CAPN5, GYS2 and FBP2 had the lower expression levels of EC tissues than that of healthy tissues (FIGS5A-I). We further validated the 9 glycolysis-related hub genes including B3GALT6, PAM, LCT, GMPPB, GLCE, DCN, CAPN5, GYS2 and FBP2 using immunohistochemistry. PAM, GMPPB, GLCE, CAPN5, GYS2 and FBP2 had the consistent expression trend. B3GALT6 and LCT were not available in the database (Fig. 11A-G). AUC value was used to identify the diagnostic efficacy of distinguishing normal and cancerous tissues, AUC value of 9 genes combined diagnosis was 0.992, which means the 9 genes can well identify cancer tissue and normal tissue (Fig. 11H). Regarding prognosis, Kaplan-Meier curves showed that higher expression of CAPN5, FBP2 and GYS2 correlated significantly with poor overall survival (OS), while the lower expression of DCN, GMPPB and PAM correlated significantly with OS (FIGS7).

Association between Nine hub genes and genetic mutations

We compared the frequency of genetic mutations between high- and low- risk score groups through R package maftools. The high-risk group had somatic mutations in the following order: TP53 > PTEN > PIK3CA > ARID1A > TTN > PIK3R1 > KMT2D > CTNNB1 > CTCF > MUC16 (FIGS6A). The low-risk group had somatic mutations in the following order: PTEN > ARID1A > PIK3CA > TTN > PIK3R1 > CTCF > KMT2D > ZFH3 > MUC16 > MUC5B (FIGS6B). Furthermore, we found that the patients in the low-risk group showed obvious mutation signatures, compared with patients in the high-risk group (FIGS6C-D).

Identification of nine cell cycle-related genes risk score associated biological pathways

GSEA further analyzes high- and low- risk group samples, revealing the main enrichment pathway. The samples of the high-risk group were mainly concentrated on the pathways such as dna repair, g2m checkpoint, myc targets v1 and myc targets v2. The samples of the low-risk group were mainly concentrated on the pathways such as bile acid metabolism, fatty acid metabolism, heme metabolism and xenobiotic metabolism (Fig. 12).

Discussion

Endometrial cancer is a common malignant tumor that threatens women's lives. It often occurs in postmenopausal women and is difficult to diagnose in the early stage. Therefore, it needs to be widely regarded. Glycolysis is a common energy metabolism pathway in human body. Many studies have shown that Glycolysis affected the biological behavior of tumor cells. We reasonably speculated that Glycolysis was related to the development of endometrial cancer, so we analyzed endometrial cancer samples by GSEA. Glycolysis was found to be the most highly enriched pathway, initially confirming our hypothesis.

EC samples were randomly classified into training cohort, testing cohort and entire cohort. We used training cohort to construct Cox regression prognostic model, testing cohort and entire cohort for validation. Nine glycolysis-related prognostic genes including B3GALT6, PAM, LCT, GMPPB, GLCE, DCN, CAPN5, GYS2 and FBP2 were screened out. After a comprehensive analysis of the clinical information, we found that these nine genes are associated with multiple clinical features of EC respectively.

After reviewing the existing literature, we found that these nine genes are more or less related to tumors. For example, Saldise L et al. found that the distribution of peptidyl-glycine alpha-amidating mono-oxygenase (PAM) enzymes in normal human lung and in lung epithelial tumors were different[18]. Lactase-phlorizin hydrolase was researched to be associated with colorectal cancer patients by Piepoli A[19]. D-glucuronyl C5-epimerase (GLCE) was shown as a potential tumor suppressor gene which participated in lung and breast carcinogenesis [20, 21]by inhibiting tumor angiogenesis and invasion/metastasis pathways, which was also proved to affect angiogenesis in prostate cancer cells[22]. Decorin (DCN), as an important component of the extracellular matrix (ECM), is a small leucine-rich proteoglycan and synthesized by fibroblasts, the deficiency of which promoted renal cell carcinoma growth and metastasis[23]. DCN was also seen to be potential biomarker of Colon Cancer[24]. Zhang W et al.'s research proves that DCN affects the microenvironment of tumors[25]. glycogen synthase 2 (GYS2) was demonstrated to participate in a feedback loop which restricted HBV-Related Hepatocellular Carcinoma growth[26]. Far-upstream element (FUSE)-binding protein 2 (FBP2) belongs to single-stranded DNA-binding protein family; it usually acts in regulating transcription and post-transcription and has been widely learned in liver tumors[27, 28] Kajiwarra T et al. demonstrated that in colon cancer tissues c-myc suppressor far-upstream element-binding protein-interacting repressor splicing variants were activated[29], which was also proved to induce invasion and migration of non-small cell lung cancer cells[30]. Wang YY et al. found that FBP2 was correlated with proliferation and doxorubicin resistance in

human breast cancer cell lines[31]. B3GALT6, GMPPB and CAPN5 have not been thoroughly studied in tumors.

The somatic mutations analysis between samples of high- and low- risk group were conducted, the result of which showed that the different mutated genes could contribute to the different sore of genes in the EC patients. The mutation rate of PTEN, ARID1A and PIK3R1 in the low-risk group is higher than that in the high-risk group. Interestingly, these three genes have been proved to have a certain tumor suppressive effect by previous studies[32–34]. PIK3CA has been found to play a role in gynecological tumors such as cervical cancer[35]. The mutations of other genes have not been explored in EC, and it is worth studying in detail in the future.

GSEA displayed that the samples of the high-risk group were mainly concentrated on the pathways such as dna repair, while the samples of the low-risk group were mainly concentrated on the pathways such as bile acid metabolism. Research proved that when DNA repair fails, this damage can lead to carcinogenesis and tumor genomic instability. In this pathway, biological targets involved in immunotherapy can be found[36]. Bile Acid Metabolism has been found to be related to Signaling in Cholestasis, Inflammation, and Cancer[37].

The advantage of this study is that, firstly, it was found that the Glycolysis pathway is related to the mechanism of EC, opening a new perspective for the regulation of metabolic processes and the treatment of EC. Secondly, we found hub genes closely related to EC survival in this pathway. Most of these genes have been confirmed to affect tumor progression and are likely to be used for targeted therapy. B3GALT6, GMPPB and CAPN5 have not been thoroughly studied. We firstly discovered that these three genes were related to EC and might become an innovative research direction in the future.

Conclusion

This study found that the Glycolysis pathway is associated with EC and screened for hub genes on the Glycolysis pathway, which may serve as new target for the treatment of EC.

Declarations

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Written informed consent for publication was obtained from all participants.

DATA AVAILABILITY STATEMENT

The data and materials can be found from the first author and corresponding author.

ACKNOWLEDGEMENTS

Not applicable.

CONFLICTS OF INTEREST

The authors have no conflicts of interests to declare.

FUNDING

This work was supported by the National Nature Science Foundation of China (81872119) and Jiangsu province medical innovation team (CXTDA2017008).

AUTHOR'S CONTRIBUTION

JinHui Liu and Gao Feng had contribution on Data curation. SiYue Li had contribution on Writing – original draft. HuangYang Meng and SiPei Nie had contribution on Formal analysis. Rui Sun had contribution on Conceptualization. Jing Yang and WenJun Cheng had contribution on review & editing. All authors reviewed and approved the final manuscript.

References

1. Braun MM, Overbeek-Wager EA, Grumbo RJ: **Diagnosis and Management of Endometrial Cancer.** *American family physician* 2016, **93**(6):468-474.
2. Suri V, Arora A: **Management of Endometrial Cancer: A Review.** *Reviews on recent clinical trials* 2015, **10**(4):309-316.
3. McAlpine JN, Temkin SM, Mackay HJ: **Endometrial cancer: Not your grandmother's cancer.** *Cancer* 2016, **122**(18):2787-2798.
4. Bendifallah S, Ballester M, Darai E: **[Endometrial cancer: Predictive models and clinical impact].** *Bulletin du cancer* 2017, **104**(12):1022-1031.
5. Lee YC, Lheureux S, Oza AM: **Treatment strategies for endometrial cancer: current practice and perspective.** *Current opinion in obstetrics & gynecology* 2017, **29**(1):47-58.
6. Gill KS, Fernandes P, O'Donovan TR, McKenna SL, Doddakula KK, Power DG, Soden DM, Forde PF: **Glycolysis inhibition as a cancer treatment and its role in an anti-tumour immune response.** *Biochimica et biophysica acta* 2016, **1866**(1):87-105.
7. Ganapathy-Kanniappan S, Geschwind JF: **Tumor glycolysis as a target for cancer therapy: progress and prospects.** *Molecular cancer* 2013, **12**:152.
8. Ganapathy-Kanniappan S: **Linking tumor glycolysis and immune evasion in cancer: Emerging concepts and therapeutic opportunities.** *Biochimica et biophysica acta Reviews on cancer* 2017, **1868**(1):212-220.

9. Akins NS, Nielson TC, Le HV: **Inhibition of Glycolysis and Glutaminolysis: An Emerging Drug Discovery Approach to Combat Cancer.** *Current topics in medicinal chemistry* 2018, **18**(6):494-504.
10. Li C, Zhang G, Zhao L, Ma Z, Chen H: **Metabolic reprogramming in cancer cells: glycolysis, glutaminolysis, and Bcl-2 proteins as novel therapeutic targets for cancer.** *World journal of surgical oncology* 2016, **14**(1):15.
11. Qin W, Li C, Zheng W, Guo Q, Zhang Y, Kang M, Zhang B, Yang B, Li B, Yang H *et al*: **Inhibition of autophagy promotes metastasis and glycolysis by inducing ROS in gastric cancer cells.** *Oncotarget* 2015, **6**(37):39839-39854.
12. Feinberg T, Herbig J, Kohl I, Las G, Cancilla JC, Torrecilla JS, Ilouze M, Haick H, Peled N: **Cancer metabolism: the volatile signature of glycolysis-in vitro model in lung cancer cells.** *Journal of breath research* 2017, **11**(1):016008.
13. Tomczak K, Czerwinska P, Wiznerowicz M: **The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge.** *Contemporary oncology (Poznan, Poland)* 2015, **19**(1a):A68-77.
14. Toh H, Shirane K, Miura F, Kubo N, Ichiyanagi K, Hayashi K, Saitou M, Suyama M, Ito T, Sasaki H: **Software updates in the Illumina HiSeq platform affect whole-genome bisulfite sequencing.** *BMC genomics* 2017, **18**(1):31.
15. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP: **GSEA-P: a desktop application for Gene Set Enrichment Analysis.** *Bioinformatics (Oxford, England)* 2007, **23**(23):3251-3253.
16. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E *et al*: **The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.** *Cancer discovery* 2012, **2**(5):401-404.
17. Heagerty PJ, Zheng Y: **Survival model predictive accuracy and ROC curves.** *Biometrics* 2005, **61**(1):92-105.
18. Saldise L, Martinez A, Montuenga LM, Treston A, Springall DR, Polak JM, Vazquez JJ: **Distribution of peptidyl-glycine alpha-amidating mono-oxygenase (PAM) enzymes in normal human lung and in lung epithelial tumors.** *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 1996, **44**(1):3-12.
19. Piepoli A, Schirru E, Mastroilli A, Gentile A, Cotugno R, Quitadamo M, Merla A, Congia M, Usai Satta P, Perri F: **Genotyping of the lactase-phlorizin hydrolase c/t-13910 polymorphism by means of a new rapid denaturing high-performance liquid chromatography-based assay in healthy subjects and colorectal cancer patients.** *Journal of biomolecular screening* 2007, **12**(5):733-739.
20. Belyavskaya VA, Prudnikova TY, Domanitskaya NV, Litviakov NV, Maksimov VN, Cherdyntseva NV, Grigorieva EV: **GLCE rs3865014 (Val597Ile) polymorphism is associated with breast cancer susceptibility and triple-negative breast cancer in Siberian population.** *Gene* 2017, **628**:224-229.
21. Grigorieva EV, Prudnikova TY, Domanitskaya NV, Mostovich LA, Pavlova TV, Kashuba VI, Zabarovsky ER: **D-glucuronyl C5-epimerase suppresses small-cell lung cancer cell proliferation in vitro and tumour growth in vivo.** *British journal of cancer* 2011, **105**(1):74-82.

22. Rosenberg EE, Prudnikova TY, Zabarovsky ER, Kashuba VI, Grigorieva EV: **D-glucuronyl C5-epimerase cell type specifically affects angiogenesis pathway in different prostate cancer cells.** *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* 2014, **35**(4):3237-3245.
23. Xu Y, Xia Q, Rao Q, Shi S, Shi Q, Ma H, Lu Z, Chen H, Zhou X: **DCN deficiency promotes renal cell carcinoma growth and metastasis through downregulation of P21 and E-cadherin.** *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* 2016, **37**(4):5171-5183.
24. Li G, Li M, Liang X, Xiao Z, Zhang P, Shao M, Peng F, Chen Y, Li Y, Chen Z: **Identifying DCN and HSPD1 as Potential Biomarkers in Colon Cancer Using 2D-LC-MS/MS Combined with iTRAQ Technology.** *Journal of Cancer* 2017, **8**(3):479-489.
25. Zhang W, Ge Y, Cheng Q, Zhang Q, Fang L, Zheng J: **Decorin is a pivotal effector in the extracellular matrix and tumour microenvironment.** *Oncotarget* 2018, **9**(4):5480-5491.
26. Chen SL, Zhang CZ, Liu LL, Lu SX, Pan YH, Wang CH, He YF, Lin CS, Yang X, Xie D *et al.*: **A GYS2/p53 Negative Feedback Loop Restricts Tumor Growth in HBV-Related Hepatocellular Carcinoma.** *Cancer research* 2019, **79**(3):534-545.
27. Samarin J, Laketa V, Malz M, Roessler S, Stein I, Horwitz E, Singer S, Dimou E, Cigliano A, Bissinger M *et al.*: **PI3K/AKT/mTOR-dependent stabilization of oncogenic far-upstream element binding proteins in hepatocellular carcinoma cells.** *Hepatology (Baltimore, Md)* 2016, **63**(3):813-826.
28. Malz M, Bovet M, Samarin J, Rabenhorst U, Sticht C, Bissinger M, Roessler S, Bermejo JL, Renner M, Calvisi DF *et al.*: **Overexpression of far upstream element (FUSE) binding protein (FBP)-interacting repressor (FIR) supports growth of hepatocellular carcinoma.** *Hepatology (Baltimore, Md)* 2014, **60**(4):1241-1250.
29. Kajiwara T, Matsushita K, Itoga S, Tamura M, Tanaka N, Tomonaga T, Matsubara H, Shimada H, Habara Y, Matsuo M *et al.*: **SAP155-mediated c-myc suppressor far-upstream element-binding protein-interacting repressor splicing variants are activated in colon cancer tissues.** *Cancer science* 2013, **104**(2):149-156.
30. Muller B, Bovet M, Yin Y, Stichel D, Malz M, Gonzalez-Vallinas M, Middleton A, Ehemann V, Schmitt J, Muley T *et al.*: **Concomitant expression of far upstream element (FUSE) binding protein (FBP) interacting repressor (FIR) and its splice variants induce migration and invasion of non-small cell lung cancer (NSCLC) cells.** *The Journal of pathology* 2015, **237**(3):390-401.
31. Wang YY, Gu XL, Wang C, Wang H, Ni QC, Zhang CH, Yu XF, Yang LY, He ZX, Mao GX *et al.*: **The far-upstream element-binding protein 2 is correlated with proliferation and doxorubicin resistance in human breast cancer cell lines.** *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* 2016, **37**(7):9755-9769.
32. Wise HM, Hermida MA, Leslie NR: **Prostate cancer, PI3K, PTEN and prognosis.** *Clinical science (London, England : 1979)* 2017, **131**(3):197-210.

33. Zhu YP, Sheng LL, Wu J, Yang M, Cheng XF, Wu NN, Ye XB, Cai J, Wang L, Shen Q *et al*: **Loss of ARID1A expression is associated with poor prognosis in patients with gastric cancer.** *Human pathology* 2018, **78**:28-35.
34. Yan LX, Liu YH, Xiang JW, Wu QN, Xu LB, Luo XL, Zhu XL, Liu C, Xu FP, Luo DL *et al*: **PIK3R1 targeting by miR-21 suppresses tumor cell migration and invasion by reducing PI3K/AKT signaling and reversing EMT, and predicts clinical outcome of breast cancer.** *International journal of oncology* 2016, **48**(2):471-484.
35. Chung TKH, Cheung TH, Yim SF, Yu MY, Chiu RWK, Lo KWK, Lee IPC, Wong RRY, Lau KKM, Wang VW *et al*: **Liquid biopsy of PIK3CA mutations in cervical cancer in Hong Kong Chinese women.** *Gynecologic oncology* 2017, **146**(2):334-339.
36. Bever KM, Le DT: **DNA repair defects and implications for immunotherapy.** *The Journal of clinical investigation* 2018, **128**(10):4236-4242.
37. Li T, Apte U: **Bile Acid Metabolism and Signaling in Cholestasis, Inflammation, and Cancer.** *Advances in pharmacology (San Diego, Calif)* 2015, **74**:263-302.

Supplemental Information

Figure S1 Expression level of CAPN5, DCN, GLCE and GMPPB in different age groups. (A) CAPN5, (B) DCN, (C) GLCE, (D) GMPPB.

Figure S2 Expression level of B3GAL, CAPN5, GLCE, GMPPB and PAM in different histological type. (A) B3GAL, (B) CAPN5, (C) GLCE, (D) GMPPB, (E) PAM.

Figure S3 Expression level of CAPN5, DCN, GLCE, GMPPB and PAM in different grade. (A) CAPN5, (B) DCN, (C) GLCE, (D) GMPPB, (E) PAM.

Figure S4 Expression level of DCN, GMPPB and PAM in different tumor status. (A) DCN, (B) GMPPB, (C) PAM.

Figure S5 TCGA Expression level Validation of 9 glycolysis-related hub genes. (A) B3GALT6, (B) CAPN5, (C) DCN, (D) FBP2, (E) GLCE, (F) GMPPB, (G) GYS2, (H) LCT, (I) PAM.

Figure S6 Somatic mutation analysis. (A) Oncoplot displaying the somatic landscape of EC with high-risk score. (B) Oncoplot displaying the somatic landscape of EC with low-risk score. Stacked bar chart and cohort summary plot displaying distribution of variants according to variant classification, type, and SNV class. Bottompart (from left to right) indicates mutation load for each sample, variant classification type of the high-risk group(C) and low-risk group (D).

Figure S7 Kaplan-Meier curves showed that higher expression of CAPN5, FBP2 and GYS2 correlated significantly with poor OS, while the lower expression of DCN, GMPPB and PAM correlated significantly with OS. The yellow line indicates samples with highly expressed genes (above best-separation value), and the green line designates the samples with lowly expressed genes (below best-separation value)

Figures

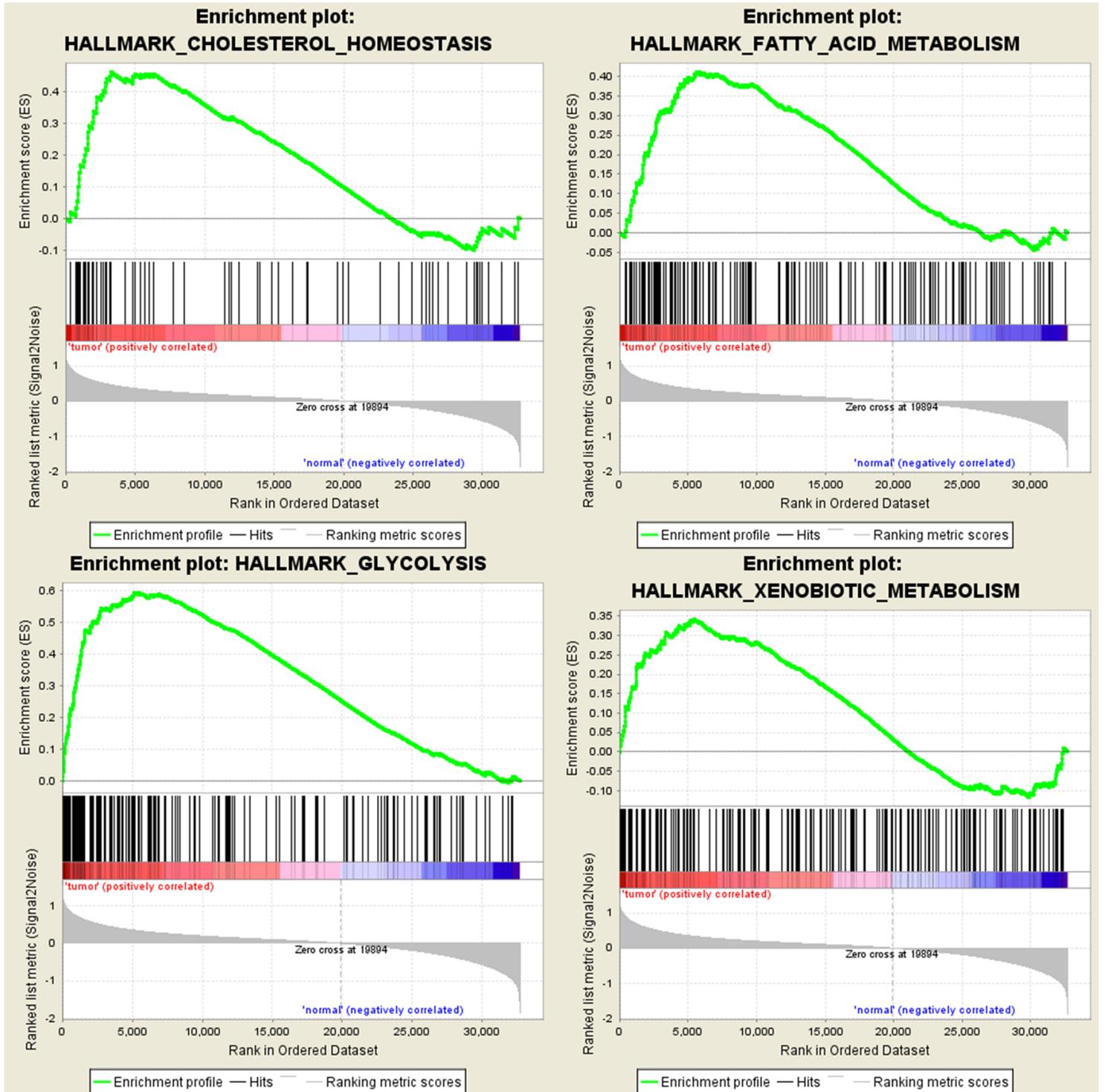


Figure 1

GSEA identified that four gene sets were significantly enriched including glycolysis, cholesterol homeostasis, fatty acid metabolism and xenobiotic metabolism.

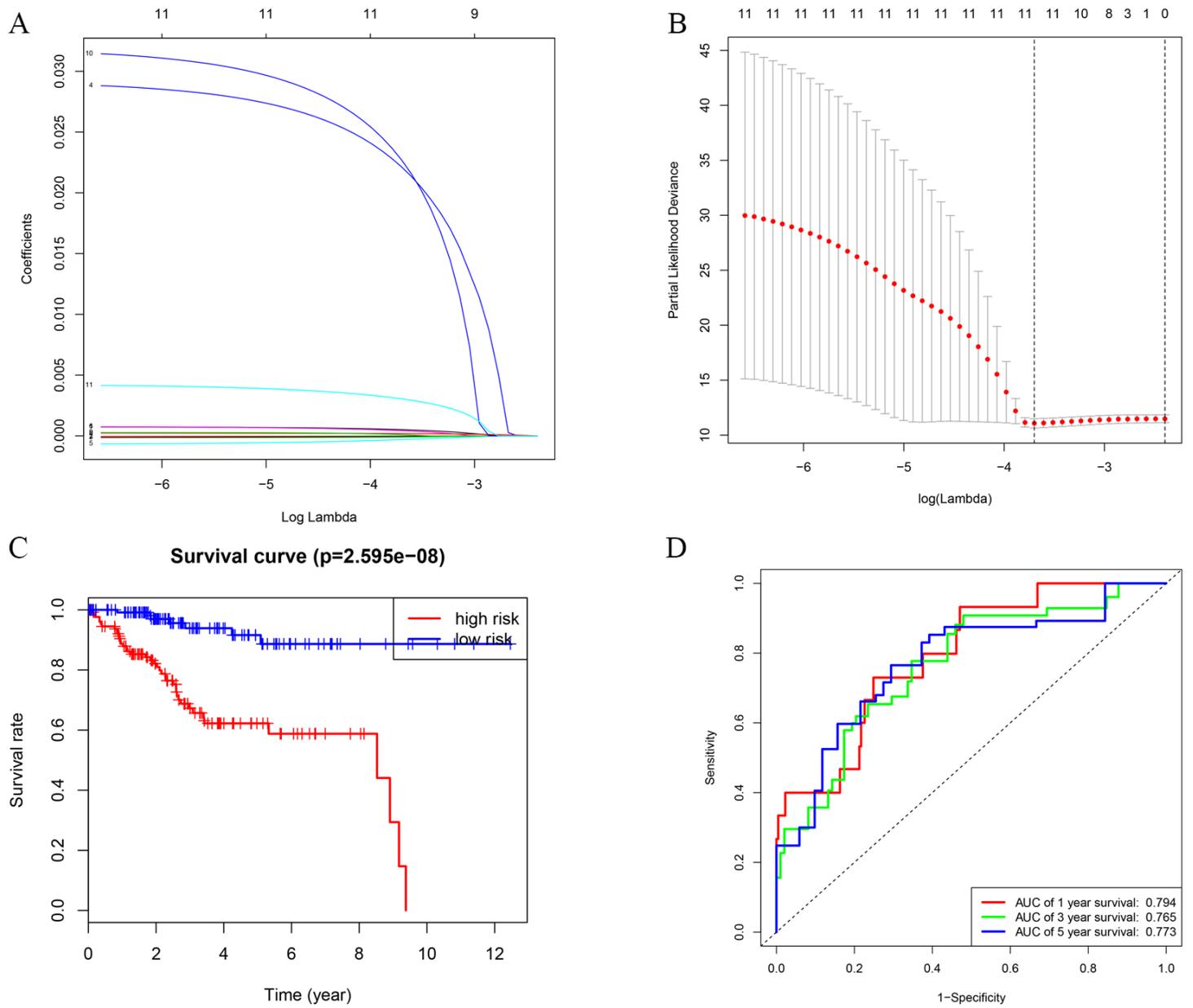


Figure 2

Prognostic model of the training cohort. (A-B) the coefficients calculated by LASSO. (C) Kaplan-Meier survival analysis of the low- and high- risk group patients in the training cohort. (D) ROC curve analysis according to the 1,3,5-year survival of the area under the AUC value.

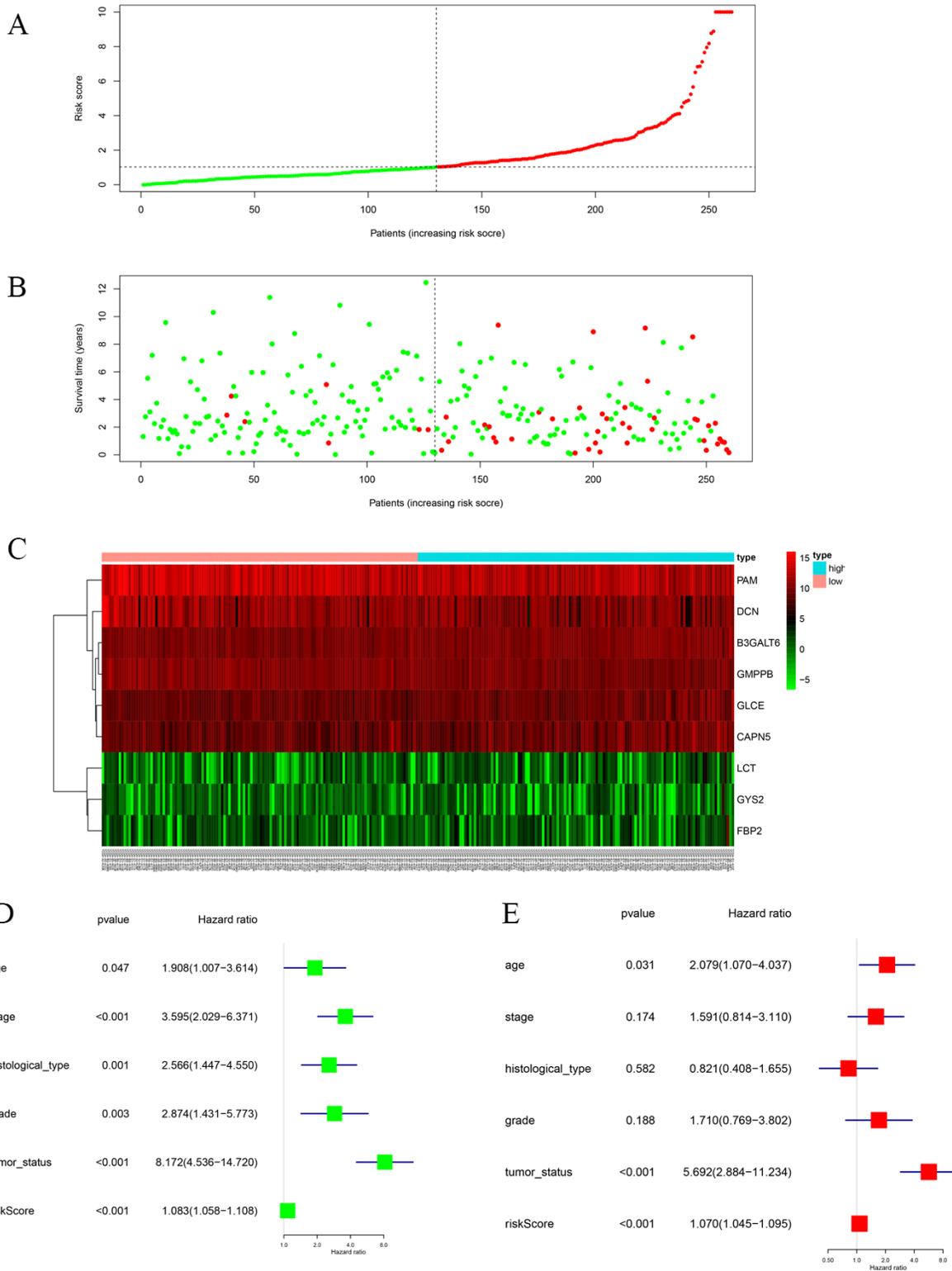


Figure 3

Risk signature with the 9 glycolysis-related hub genes. (A-B) The risk scores for all patients in training cohort are plotted in ascending order and marked as low risk (blue) or high risk (red), as divided by the threshold (vertical black line). (C) The distribution of risk score, survival status, and the expression of 9 genes of each patient in training cohort by z-score, with red indicating higher expression and light blue indicating lower expression. (D) univariate regression model. (E) multivariate regression model.

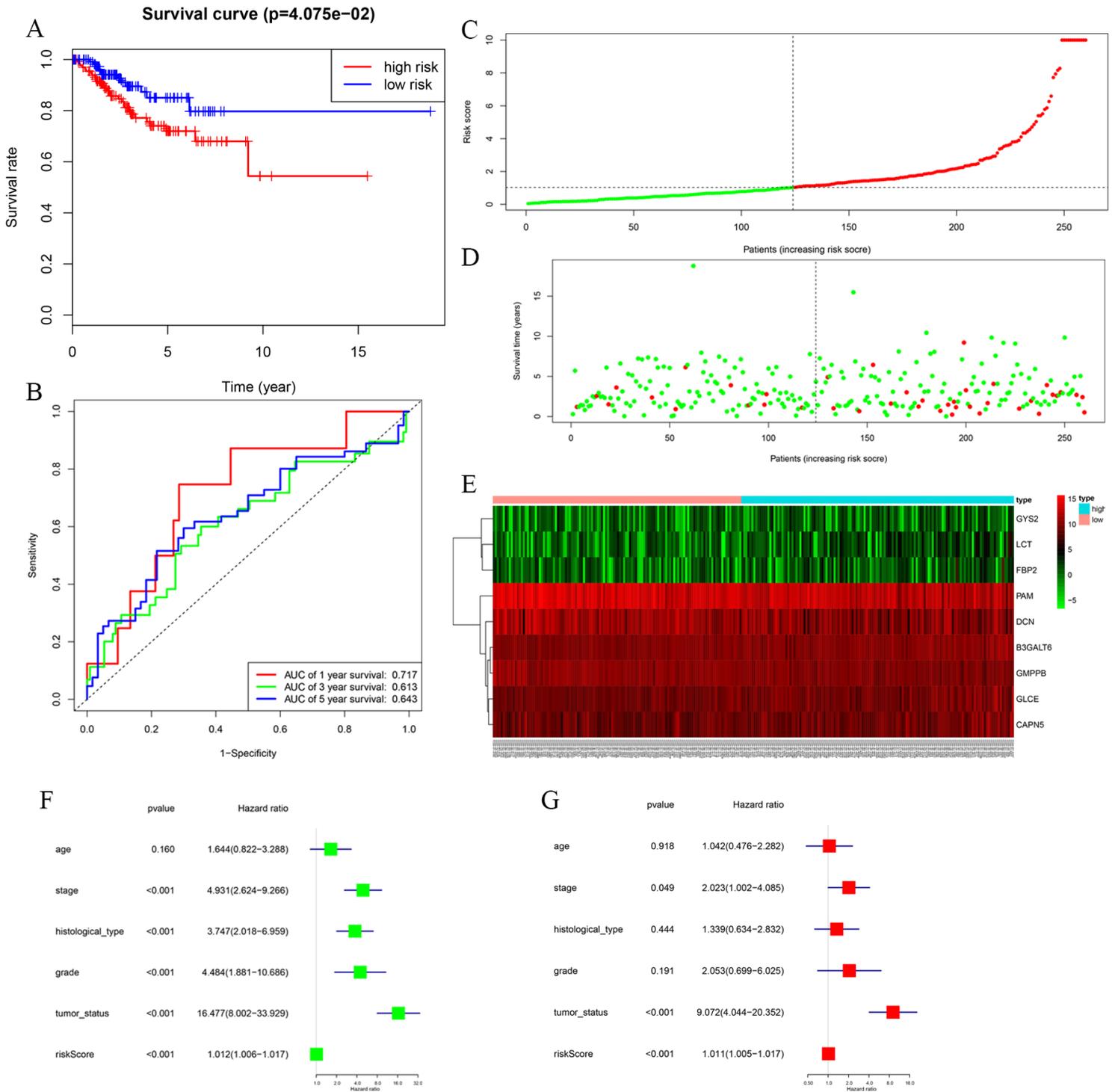


Figure 4

Prognostic model of the testing cohort and Risk signature with the 9 glycolysis-related hub genes. (A) Kaplan-Meier survival analysis of the low- and high-risk group patients in the testing cohort. (B) ROC curve analysis according to the 1,3,5-year survival of the area under the AUC value. (C-D) The risk scores for all patients in testing cohort are plotted in ascending order and marked as low risk (blue) or high risk (red), as divided by the threshold (vertical black line). (E) The distribution of risk score, survival status, and the expression of 9 genes of each patient in testing cohort by z-score, with red indicating higher

expression and light blue indicating lower expression. (F) univariate regression model. (G) multivariate regression model.

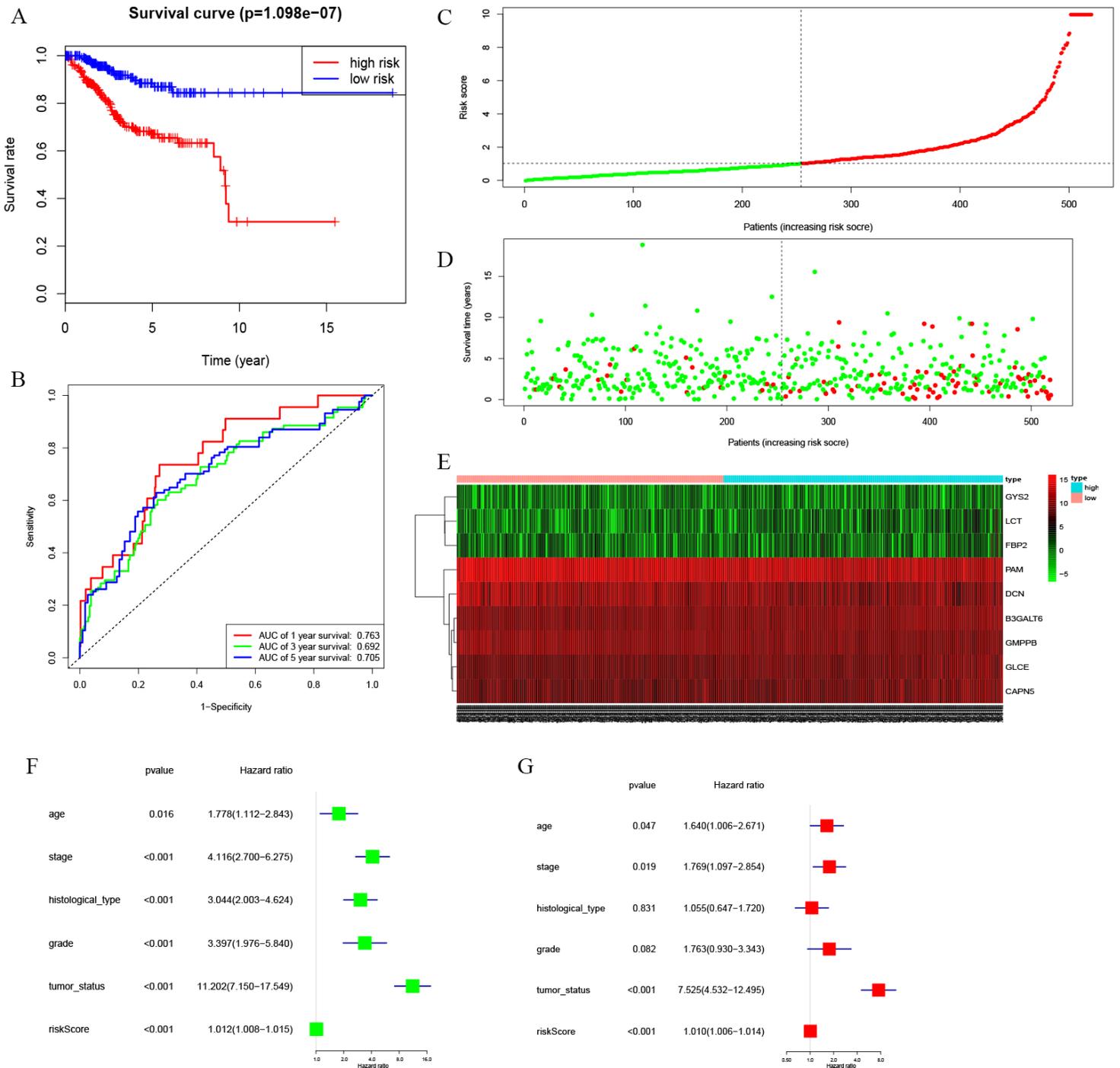
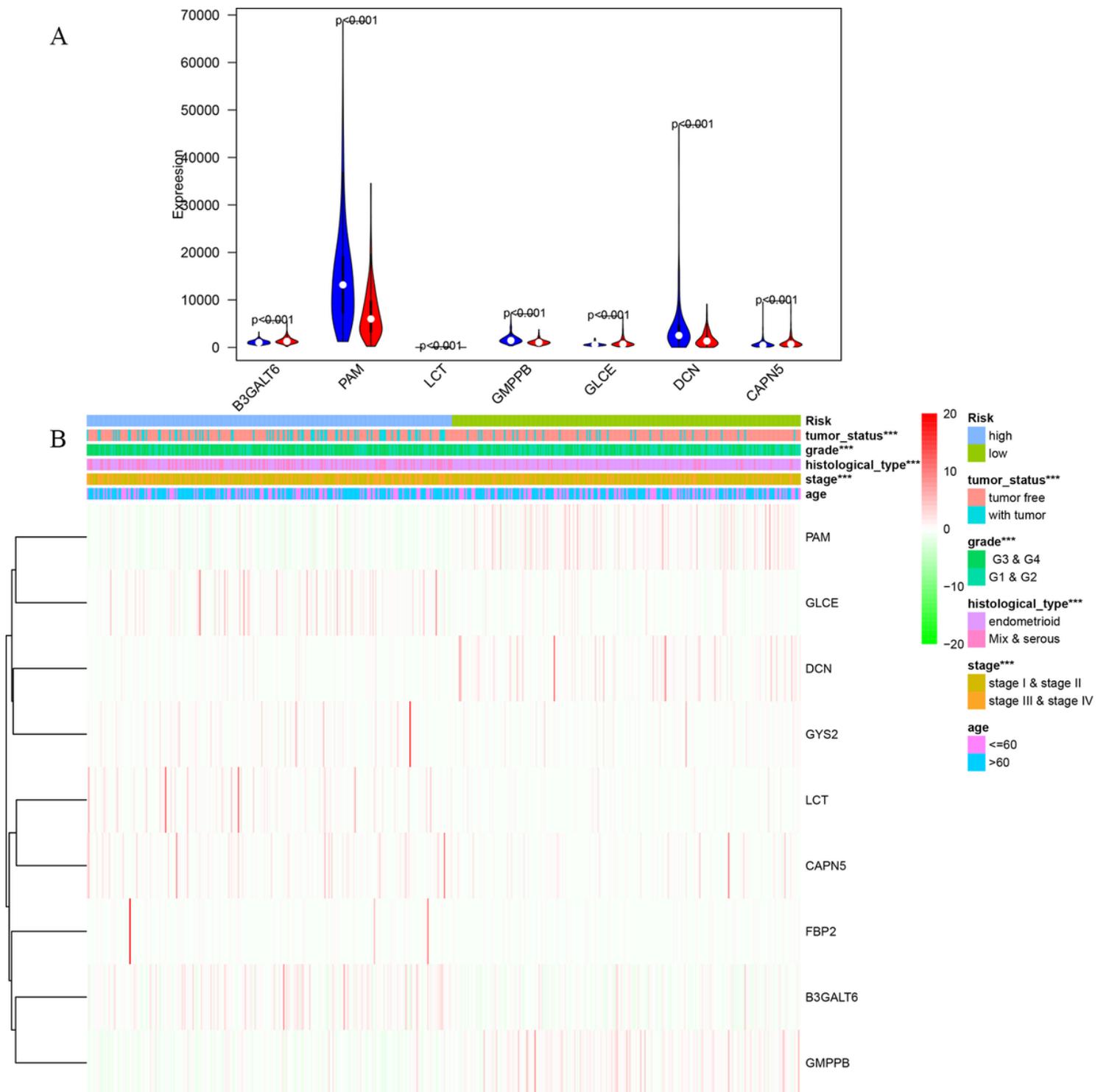


Figure 5

Prognostic model of the entire cohort and Risk signature with the 9 glycolysis-related hub genes. (A) Kaplan-Meier survival analysis of the low- and high-risk group patients in the entire cohort. (B) ROC curve analysis according to the 1,3,5-year survival of the area under the AUC value. (C-D) The risk scores for all patients in entire cohort are plotted in ascending order and marked as low risk (blue) or high risk (red), as divided by the threshold (vertical black line). (E) The distribution of risk score, survival status, and the

expression of 9 genes of each patient in entire cohort by z-score, with red indicating higher expression and light blue indicating lower expression. (F) univariate regression model. (G) multivariate regression model.



genes in high- and low-risk patients in the TCGA dataset associated with tumor status, grade, histological type and stage.

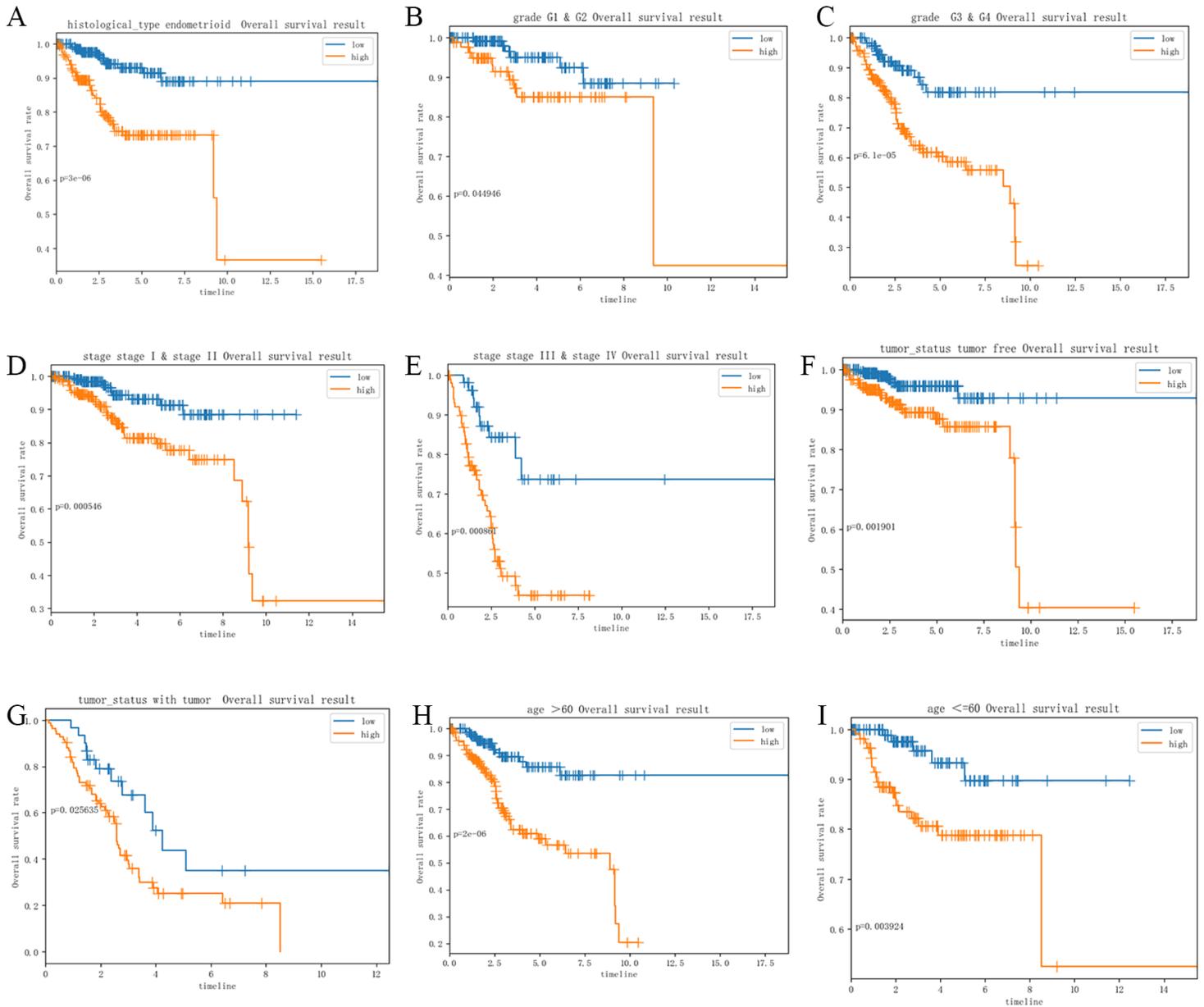


Figure 7

Survival time of patients in high-risk and low-risk group of different subgroups. (A) endometrioid subgroup, (B) grade G1&G2 subgroup, (C) grade G3&G4 subgroup, (D) stage I & stage II subgroup, (E) stage III & stage IV subgroup, (F) tumor free subgroup, (G) with tumor subgroup, (H) age>60 subgroup, (I) age≤60 subgroup.

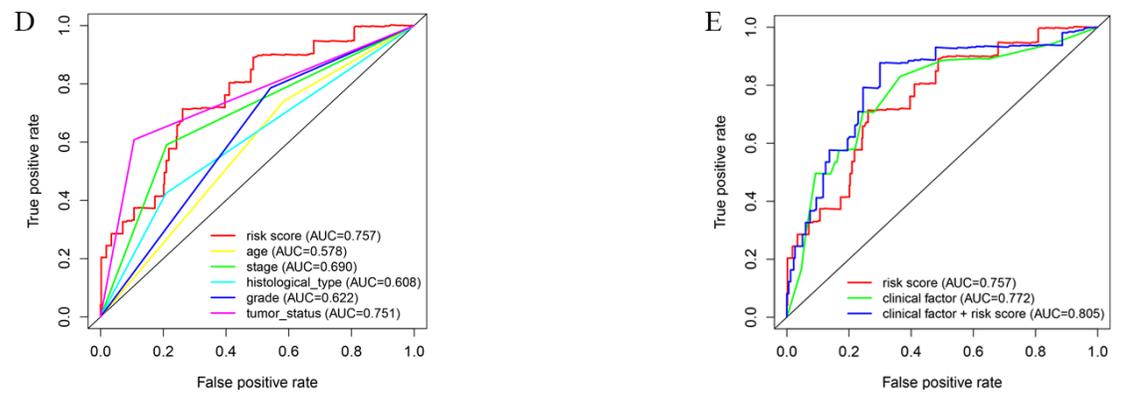
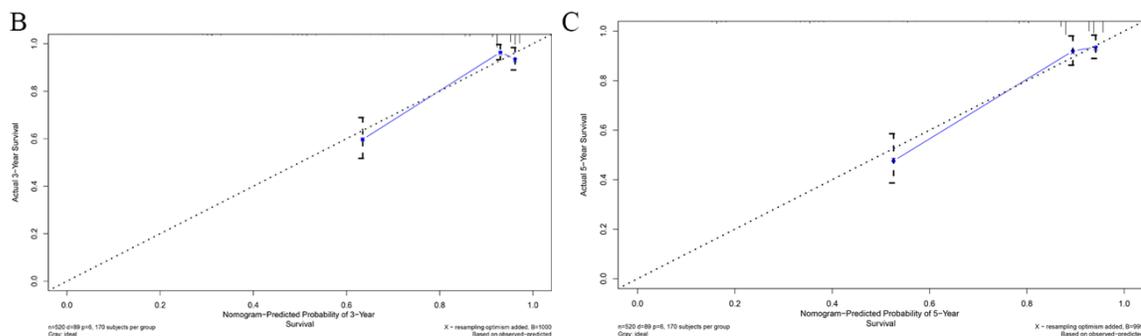
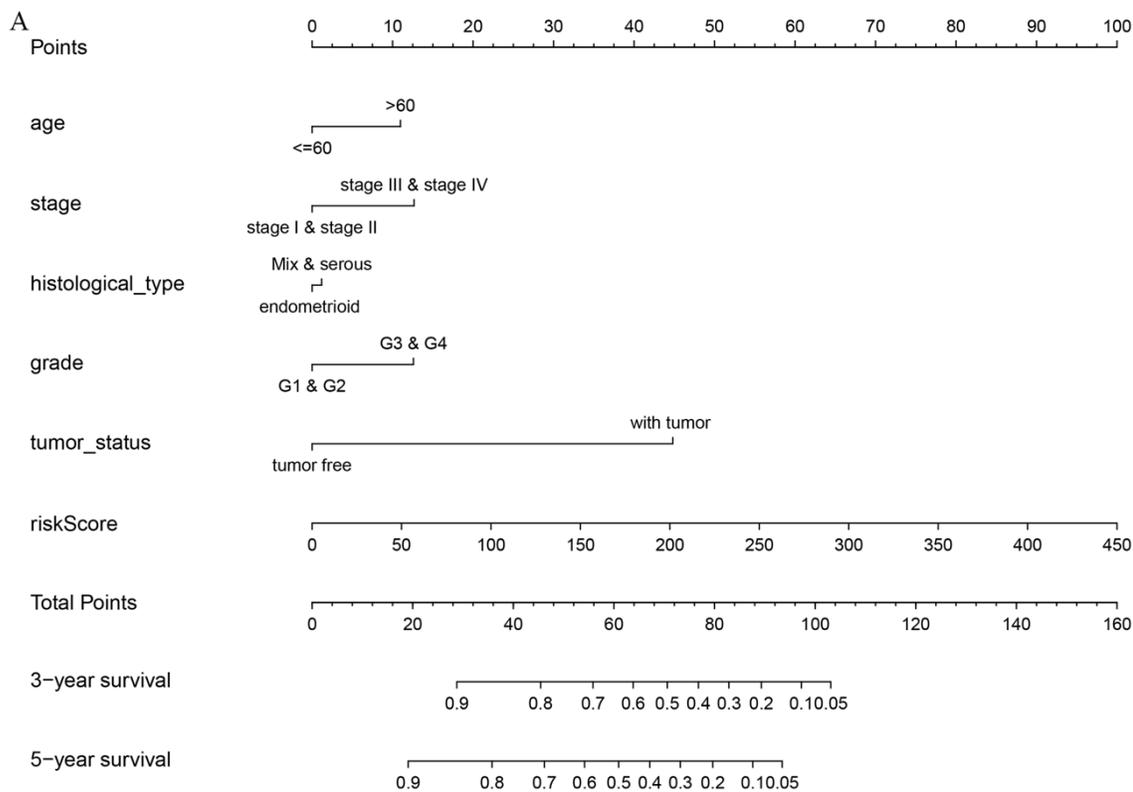


Figure 8

The nomogram to predict 3- or 5-year OS and prognostic value of 9 genes in the entire set. (A) The nomogram for predicting proportion of patients with 3- or 5-year OS. (B-C) The calibration plots for predicting patient 3- or 5- year OS. Nomogram-predicted probability of survival is plotted on the x-axis; actual survival is plotted on the y-axis. (D-E) The relationship between 9-mRNA signature and different clinical features.

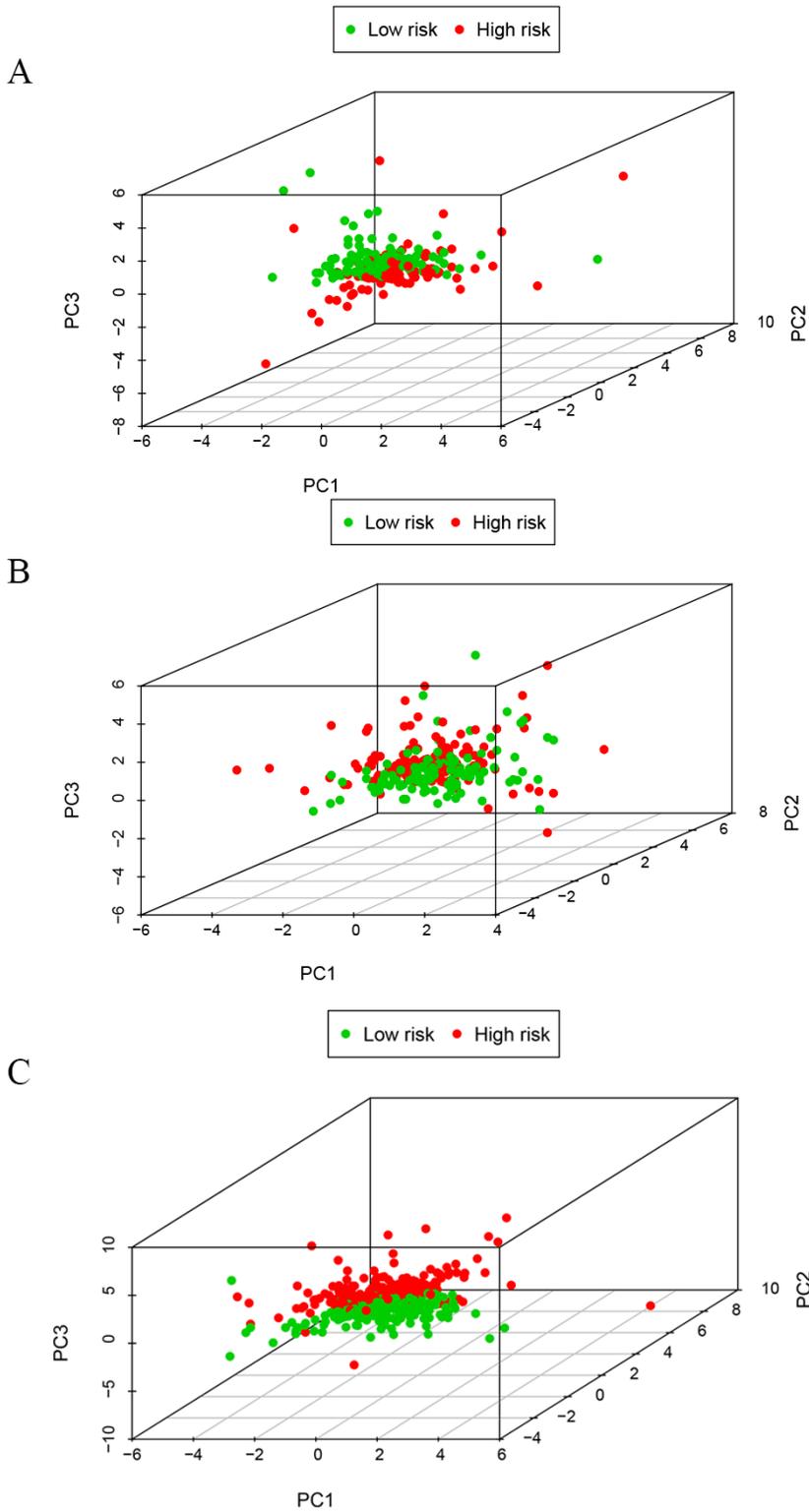


Figure 9

Principal component analysis of the training cohort, testing cohort, and entire EC cohort. (A) training cohort, (B) testing cohort, (C) entire cohort.

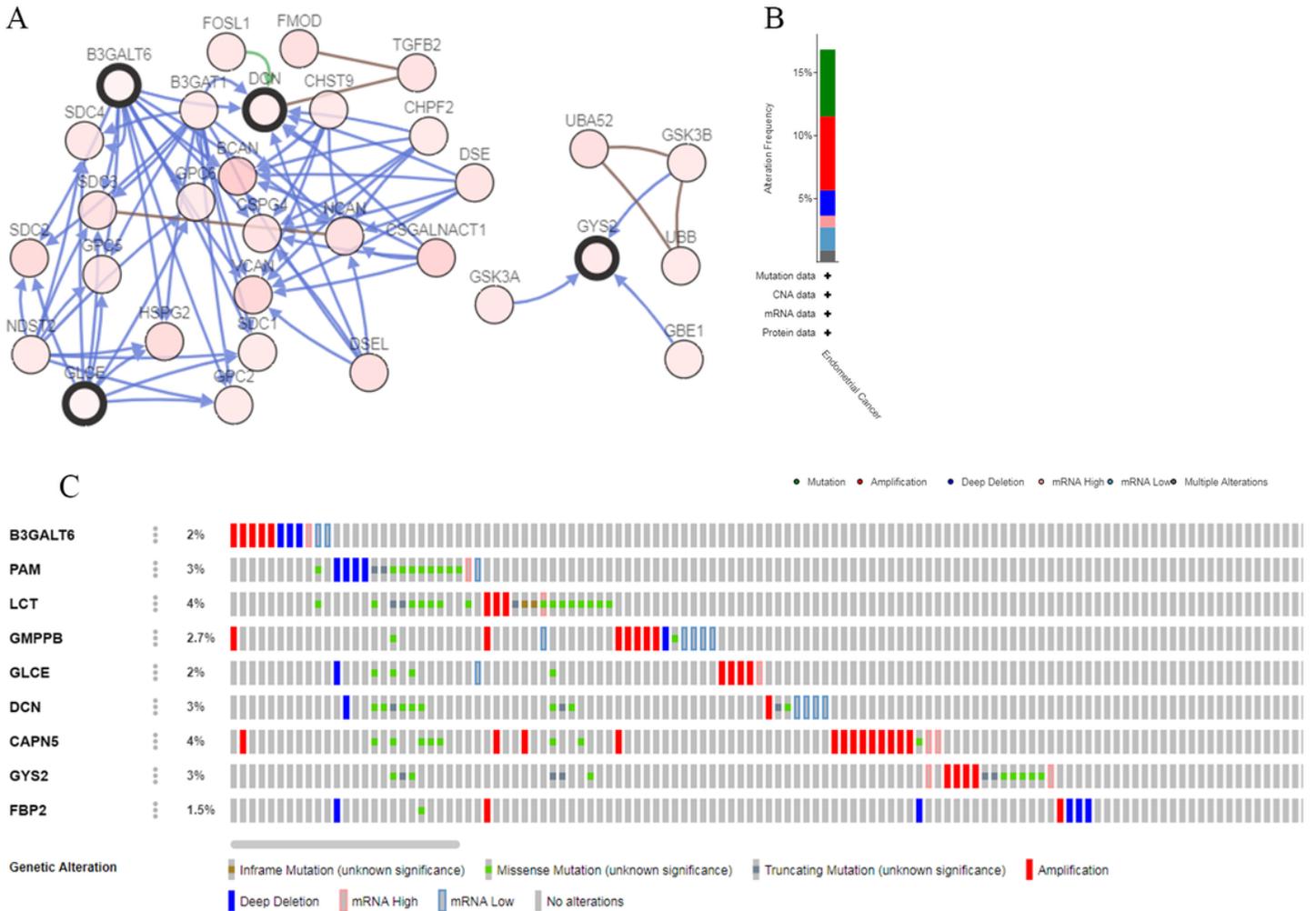


Figure 10

The gene mutation overview of 9 prognostic glycolysis-related genes in TCGA EC patients. (A) The network of 9 genes and the most frequently altered neighbor genes. (B) The summary of mutation type of 9 genes in EC patients. (C) Nine genes were altered in 92 (17%) from the 547 patients.

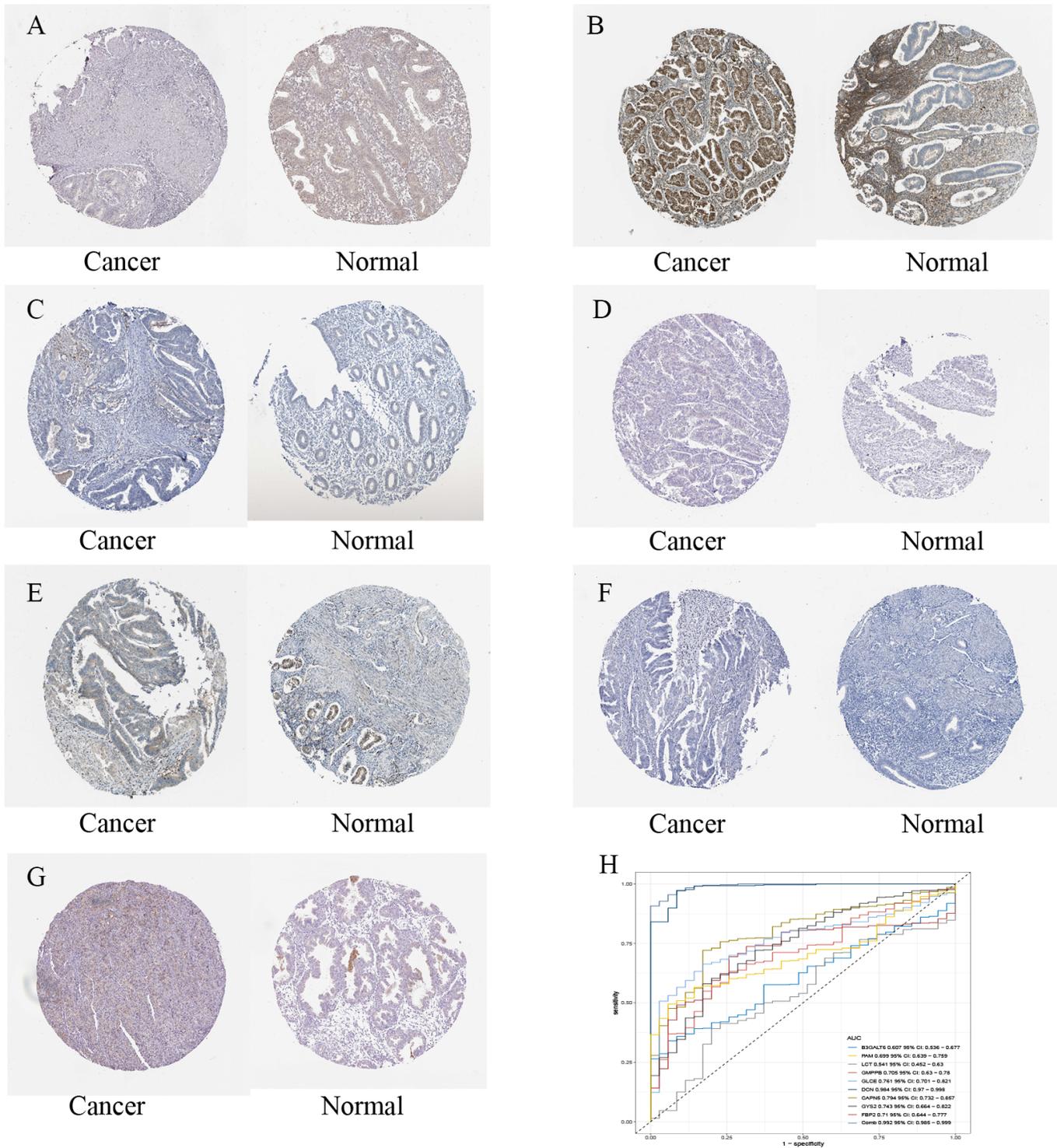


Figure 11

The protein expression difference of (A) CAPN5, (B) DCN, (C) FBP2, (D) GLCE, (E) GMPPB, (F) GYS2, (G) PAM between EC and normal samples from TCGA. (H) AUC value was used to identify the diagnostic efficacy of distinguishing normal and cancerous tissues.

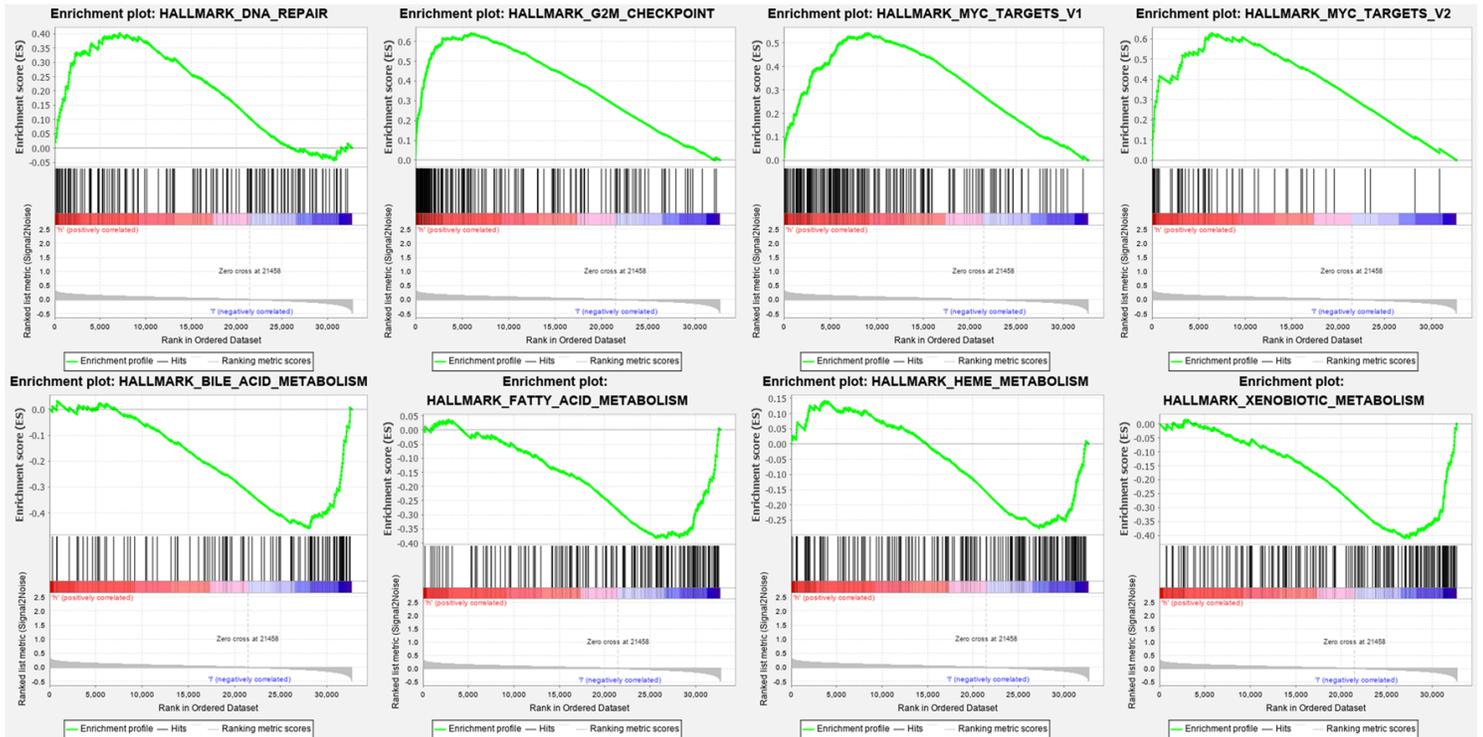


Figure 12

Gene Set Enrichment Analysis in TCGA database. Enrichment Map were used for visualization of the GSEA results.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [S1.tif](#)
- [S4.tif](#)
- [S5.tif](#)
- [S6.tif](#)
- [S3.tif](#)
- [S2.tif](#)
- [S7.tif](#)