A Nomogram Based on Glycolysis-related Gene Expression Profilings Serve as a Novel Prognosis Risk Classifiers for Human Hepatocellular Carcinoma

Lingyu Zhang
The Second Affiliated Hospital of Zhejiang University School of Medicine
Yu Li
Bengbu Medical College
Yibei Dai
The Second Affiliated Hospital of Zhejiang University School of Medicine
Danhua Wang
The Second Affiliated Hospital of Zhejiang University School of Medicine
Xuchu Wang
The Second Affiliated Hospital of Zhejiang University School of Medicine
Ying Cao
The Second Affiliated Hospital of Zhejiang University School of Medicine
Weiwei Liu
The Second Affiliated Hospital of Zhejiang University School of Medicine
Zhihua Tao (zrtzh@zju.edu.cn)
The Second Affiliated Hospital of Zhejiang University School of Medicine

Research Article

Keywords: Hepatocellular Carcinoma, Glycolysis, Gene Set Enrichment Analysis, gene signatures, immune infiltration landscape

Posted Date: March 24th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-324874/v1

Abstract

Metabolic pattern reconstruction is an important element in tumor progression. The metabolism of tumor cells is characterized by the abnormal increase of anaerobic glycolysis, regardless of the higher oxygen concentration, resulting in a large accumulation of energy from glucose sources, and contributes to rapid cell proliferation and tumor growth which is further referenced as the Warburg effect. We tried to reconstruct the metabolic pattern in the progression of cancer to screen which genetic changes are specific in cancer cells.

A total of 12 common types of solid tumors were enrolled in the prospective study. Gene set enrichment analysis (GSEA) was implemented to analyze 9 glycolysis-related gene sets, which are closely related to the glycolysis process. Univariate and multivariate analyses were used to identify independent prognostic variables for the construction of a nomogram based on clinicopathological characteristics and a glycolysis-related gene prognostic index (GRGPI).

The prognostic model based on glycolysis genes has the highest area under the curve (AUC) in LIHC (Liver hepatocellular carcinoma). 8-gene signatures (AURKA, CDK1, CENPA, DEPDC1, HMMR, KIF20A, PFKFB4, STMN1) were related to overall survival (OS) and recurrence-free survival (RFS). Further analysis demonstrates that the prediction model can accurately distinguish between high- and low-risk cancer patients among patients in different clusters in LIHC. A nomogram with a well-fitted calibration curve based on gene expression profiles and clinical characteristics improves discrimination in internal and external cohorts. Furthermore, the altering expression of metabolic genes related to glycolysis may contribute to the reconstruction of the tumor-related microenvironment.

Introduction

Cells are authorized to choose energy metabolism patterns for biosynthesis, depending on cell function and availability of metabolites. In addition to oxidative phosphorylation of glucose, other metabolic pathways, including lipid, nucleotide, and amino acid metabolism can also provide energy to meet its biosynthetic requirements for cell growth and proliferation[1,2]. The energy metabolism pattern of tumor cells has changed dramatically, compared with oxidative phosphorylation (OXPHOS) of normal cells. To maintain survival and meet the synthesis of biological macromolecules, energy metabolism tends to another embodiment, which is referred to as glycolysis or Warburg effect[3–5]. The Warburg effect represents the transformation of glucose utilization by tumor cells from oxidative phosphorylation to glycolysis, which is now acknowledged as a major feature of tumors[6,7]. This change in energy metabolism is determined by complex factors, including pressure on the tumor microenvironment and genetic changes[8–11]. The enhanced glycolysis of tumor cells is mainly due to the increased expression or activity of key glycolysis enzymes[12]. In recent years, people are making a concerted effort to target tumors by inhibiting the activity of key enzymes in the tumor glycolysis pathway. Some studies have shown that specific inhibition of glycolysis is coupled with significant tumor suppression, and induces cell death. Glycolysis key enzymes such as glycokinase 2 (HK2), phosphofructosidase (PFK), and M2-

type acetone kinase (PKM2) have become tumor markers, and their expression and activity can affect tumor glycolysis, which in turn affects tumors of proliferation[13–17]. However, the ability of glycolysis in tumor cells remains a long-standing ill-defined puzzle for refining stratification and management of cancer patients. Early diagnosis and personalized treatment were considered effective methods to improve the survival time of patients. Histopathology is believed to be able to predict the prognosis and outcome of cancer patients to a certain extent. Due to its limitations, patients with the same pathology have different prognoses due to different molecular subtypes[18,19]. With the advent of high-throughput sequencing nucleotide technology of recent years, we are allowed to better understand the dynamic changes at the molecular level. A single gene failed to predict the outcome accurately of cancer patients. In contrast, the utility of these biomarkers combinations may be altered by optimizing the sensitivity and specificity of patient outcomes. Multiple biomarkers that are increasingly related to survival and prognosis can identify high-risk patients, ameliorate the prognosis of cancer patients, and assist with appropriate intervention therapy.

Gene set enrichment analysis (GSEA) was generally used in genomic research to identify potential biological mechanisms. In this study, we tried to develop some potential gene signatures through GSEA analysis, which is closely associated with glycolysis metabolic pathways. Credit to TCGA dataset, we evaluated the tumor glycolysis metabolic patterns of 12 cancer types (Bladder Urothelial Carcinoma, BLCA; Breast invasive carcinoma, BRCA; Colon adenocarcinoma, COAD; Head and Neck squamous cell carcinoma, HNSC; Kidney renal clear cell carcinoma, KIRC; Kidney renal papillary cell carcinoma; KIRP; Liver hepatocellular carcinoma; LIHC; Lung adenocarcinoma, LUAD; Lung squamous cell carcinoma, LUSC; Ovarian serous cystadenocarcinoma, OV; Prostate adenocarcinoma, PRAD; Thyroid carcinoma, THCA), through a comprehensive analysis of genome and transcriptome profiles. Unexpectedly, we developed important GRGPI signatures in LIHC and established multiple risk characteristics that can effectively forecast the prognosis of patients. Surprisingly, glycolysis-related risk characteristics can be applied to recognize patients with dismal outcomes in the high-risk group. Besides, acting following the Cox multivariate hazard ratio analysis, the risk score outperformed other clinical traits in evaluating patient prognosis.

Materials And Methods

Gene expression profiles and patient clinical information. We obtained transcriptome expression profiles from multiple data repositories, including The Cancer Genome Atlas Program (TCGA, https://portal.gdc.cancer.gov/), the International Cancer Genome Consortium (ICGC, http://www.icgc.org), and the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). Datasets without insufficient sample size (< 200) or available clinical information were excluded. The raw counts were transformed into transcripts per kilobase million (TPM) values for subsequent analysis.

Gene Set Enrichment Analysis. The Molecular Signatures Database (MSigDB) was explored to identify gene sets and specific biological processes that are significantly differentially expressed in different groups. This method produced a statistically significant improvement in the connectivity between the

data expression pattern and the biological process, ignoring the clear differential gene threshold[20]. 9 gene sets associated with glycolysis processes including go glycolytic fermentation, go glycolytic process, hallmark glycolysis. kegg glycolysis gluconeogenesis, module 306, reactome glycolysis, and reactome regulation of glycolysis by fructose − 2–6 bisphosphate metabolism, which was downloaded from MsigDB. For each gene set permutations were performed 1000 times, normalized enrichment scores (NES) and FDR values were used to define the enrichment pathways in each phenotype. GSEA was performed to explore whether there is a significant difference in glycolysis-related gene sets between tumor tissues and matched normal tissues. P value and FDR value < 0.05 were set as the threshold.

Construction of risk prediction model and statistical analysis. Univariate Cox regression models were constructed to assess the statistical relationships between mRNA expression levels and RFS or OS. We used a linear regression model with a stepwise forward method to predict significant variations between variables, with the beta value (β) from univariate Cox regression analysis as weighting factors[21]. We performed a multivariable logistic regression analysis after the LASSO (least absolute shrinkage and selection operator) algorithm, which simultaneously selects the variables and penalizes the model coefficients for overoptimism[22]. Multivariate analysis was applied using the Cox proportional hazards (Cox-PH) model to identify independent predictors of survival that involved the above-mentioned variables. After a series of univariate and multivariate analyses, covariates with a P value < 0.05 were used for subsequent risk prediction model construction. The standardized risk score was calculated using a formula described as follows:

$$Risk\ \ Score = \sum_{i=1}^{n} \beta_i \ x\ Expression(G_i)$$

Patients with complete clinicopathological characteristics were segmented into a high- and a low-risk group, given the median value of the risk score. Kaplan-Meier curves served to compare the survival probability variation in low- and high-risk groups. The log-rank test P < 0.05 reveals the significance of survival time differences. All these analyses are done using R version 3.6.1, along with the corresponding R packages[23]. The P value less than 0.05 was regarded as the threshold of statistical significance.

Immunohistochemistry (IHC) analysis. Immunohistochemical slides and relative clinical pathology information were approved from the Human Protein Atlas (HPA, https://www.proteinatlas.org/)[24]. The immunohistochemical staining results were evaluated by two independent pathologists, according to the integrated index by multiplying the intensity by the proportion of immunopositive cells of interest.

Weighted gene co-expression network analysis. To uncover the transcriptomic differences between HCC subgroups, weighted gene co-expression analysis was performed under the unique characteristics of the subgroups to identify potential functional modules that can characterize the biological functions of each subgroup. The optimal soft threshold parameter β (β = 7) was used to construct a scale-free co-expression network. Subsequently, based on Pearson's coefficient, genes with the same expression pattern were concentrated into specific gene modules. The top 2 modules that had the strongest

association with subgroups were selected for further analysis. GO and KEGG pathway enrichment analyses were utilized to examine whether genes from various terms are found more frequently than expected in subgroups.

Results

Screening for glycolysis gene sets with significant differences between tumor tissues and adjacent normal tissues. As demonstrated in Fig. 1, the panoramic flow chart illustrated our comprehensive analysis process. 12 solid tumors with complete clinical information and gene expression profiles were reflected in the present study, consisted BLCA, BRCA, COAD, HNSC, KIRC, KIRP, LIHC, LUAD, LUSC, OV, PRAD, and THCA. All the above data were derived from TCGA and have undergone normalization before the implementation of GSEA. GSEA was performed to carry out analysis with 9 gene sets associated with the glycolysis process. We attempt to phase whether these gene set variants produced significant differences between tumors and their adjacent noncancerous tissues. At least one gene set with an FDR value less than 0.05 was selected for subsequent studies. As presented in Fig. 2, we have identified at least one significant gene set in BLCA, BRCA, HNSC, LIHC, LUAD, and LUSC. In COAD, KIRC, KIRP, OV, PRAD, and THCA, the FDR values of 9 gene sets were revealed to be greater than 0.05. The distribution of the NES value and FDR q-value value of each gene set in the GSEA analysis was shown in the Fig. S1a. A more detailed description of the GSEA results was provided in Fig. S2. Finally, 6 solid tumors (BLCA, BRAC, HNSC, LIHC, LUAD, and LUSC) and corresponding core genes (CORE ENRICHMENT: YES) were used for further analysis (Table S1).

To further investigate whether these core genes participated in the glycolysis process, we performed GO and KEGG pathway analysis by using the R package ClusterProfiler. Further dissection of these genes and pathways indicated enrichment for glucose metabolisms, such as pyruvate metabolic process, pyruvate biosynthetic process, a glycolytic process in GO (Figure S1b-d), and Glycolysis/Gluconeogenesis in KEGG (Figure S1e). These results indicate that these core genes have a profound effect on glucose metabolism, especially glycolysis.

Construction and validation of the prognostic glycolysis associated-gene signature. Core genes were assessed for correlation with OS through univariate regression analysis and future applied in multivariate Cox-PH regression model with a stepwise procedure to identify those important variables. We identified statistically significant gene signatures (GRGPI) in BLCA, BRAC, HNSC, LIHC, LUAD, and LUSC, the detailed results were shown in Table S2. These results highlight the power of GRGPIs to identify patients with adverse outcomes who would be classified as high risk according to these glycolysis gene-related classifiers. We further investigated the area under the time-dependent ROC curves (AUC) values for each cancer type. The highest AUC was observed in LIHC compared with BLCA, BRAC, HNSC, LUAD, and LUSC at 0.5 (0.852), 1(0.840), 2(0.871), 3(0.830), and 5-year (0.756) in Fig. S3. More specifically, in a univariate Cox regression analysis, 92 glycolysis-related genes with significant correlations with overall survival were regarded as significant (Fig. 3a) in LIHC. To avoid overfitting and unnecessary complexity, the independent prognostic factors were restricted to those variables that contributed most toward the final

model coefficients based on the AIC and the model χ2 score. The selected features were incorporated into a least absolute shrinkage and selection operator (LASSO) regression model to penalize for model complexity overfitting. 8 genes (AURKA, DEPDC1, CDK1, CENPA, HMMR, KIF20A, PFKFB4, and STMN1) remained with their individually nonzero LASSO coefficients (Fig. 3b and 3c). Multivariate analysis of these variables contributed to their virtual statistical weighting, determining their impact on prognostic risk, using Cox proportional hazard regression. Finally, the risk score of 8 gene signatures was established as follows: Risk score = (0. 1224 * expression of AURKA) + (0.0534 * expression of CDK1) + (0.0920 * expression of CENPA) + (0.1323 * expression of DEPDC1) + (0.1140 * expression of HMMR) + (0.2425 * expression of KIF20A) + (0.1562 * expression of PFKFB4) + (0.0911 * expression of STMN1). Patients with high or low risk were clustered based on the median risk score of the TCGA discovery cohort. The distribution of risk scores, survival status, and gene expression landscapes of patients varied significantly within the two subgroups as shown in Fig. 4a. Kaplan-Meier survival analysis disclosed that the survival of the low-risk group was significantly longer than the high-risk group (Fig. 4b, $P < 0.001$). The cumulative event probability curve shows that HCC patients in the high-risk group have a significantly higher probability of cumulative events during the entire follow-up period than in low-risk patients (Fig. 4c, $P < 0.001$). We applied the classifier to assess whether the 8-mRNA panel can predict an individual or a specific HCC recurrence. The TCGA dataset containing recurrence events and recurrence time was used as an internal training cohort (TCGA training cohort). Our prognostic evaluations of survival analysis for 8 gene signatures were based on TCGA recurrence-free survival (RFS) outcomes. The distribution of risk score, survival status, and gene expression patterns of patients was demonstrated in Fig. 4d. Patients with low-risk scores also had longer RFS time than patients with high-risk scores (Fig. 4e, $P < 0.001$). The cumulative event occurrence curve revealed a significant cumulative risk (HR) of HCC patients in the high-risk group. (Fig. 4f)The analysis results show that this 8-gene signature can be used as a prognostic indicator for the outcome and recurrence of HCC patients. Subsequently, we performed another two independent analyses on the datasets from GEO and ICGC datasets. Consistently, as described earlier, in two independent validation sets, the 8-gene model sharply divided two risk subgroups (Fig. 5a and 5d). Not surprisingly, the survival analysis and cumulative risk curve indicated that the high-risk group had a shorter OS and higher cumulative risk (Fig. 5b-c and 5e-f). We have observed the robust prognostic value of the classifier in 3 independent cohorts.

Independent predictive value of the 8-mRNA signature. We constructed risk scores and developed predictive models to predict OS and RFS. To verify the assignments of sub-categories, we also performed t-SNE to constraint the dimensionality of the features. The T-SEN analysis revealed that the two risk subgroups are scattered in two discrete directions (Fig. 6a-d). In the TCGA discovery cohort, TCGA training cohort, GEO validation cohort, and ICGC validation cohort, we performed the time-dependent ROC curves analysis to estimate the prognostic accuracy of the 8 glycolysis-related signatures at 0.5-, 1-, 2-, 3- and 5-years, for the prediction of OS and RFS, respectively. These results confirmed the high prognostic accuracy of the 8-signature model (Fig. 6e-h).

The explanation for many clinical situations one can identify some standard variables that have previously been demonstrated to have prognostic value and are generally measured for most patients

having the particular diagnosis. As mentioned earlier, we then explored tumor-related clinicopathological variables associated with our classifier based on TCGA (Fig. 7a, Table S3), GEO (Fig. 7b, Table S4), and ICGC (Fig. 7c, Table S5) cohorts. Patient clinicopathologic characteristics were listed in Table 1. The results of Pearson's Chi-Square test showed that there were several significant correlations between clinicopathological characteristics and HCC risk subtypes in three independent cohorts. More specifically, for the TCGA discovery cohort, T classification ($P = 0.0032$), Stage ($P < 0.001$), Grade ($P = 0.0175$), Family Cancer History ($P = 0.0359$), AFP level ($P = 0.0147$), Cancer Status ($P < 0.001$), Recurrence Event ($P < 0.001$) and patient Status ($P < 0.001$). In the TCGA training set, there was a significant correlation between T classification ($P = 0.0095$) and Stage ($P = 0.0033$) between HCC subgroups (Table S3). Similarly, in the GEO and ICGC cohorts, some important clinicopathological characteristics also have significant correlations between subgroups. More detailed results are shown in Table S4 and Table S5. To prove the independence of GRGPI, a Cox proportional hazard regression analysis was performed in the TCGA, GEO, and ICGC cohorts. The adjustment results of clinical variables suggested that risk score remained an independent prognostic factor, confirming its robust predictive ability for OS (HR = 1.267, $P < 0.001$) and RFS (HR = 1.027, $P < 0.001$) of HCC patients (Fig. 8a and b). In the Cox regression model, some clinical-pathological factors (Cancer Status ($P = 0.017$), Hepatitis Virus Infection ($P = 0.041$), and Child-Pugh Score ($P = 0.011$) for OS; Cancer Status ($P < 0.001$), Hepatitis Virus Infection ($P = 0.017$), and BMI ($P = 0.009$) for RFS) were also considered as independent poor prognostic factors, and could prove valuable for risk stratification in pathological subgroups in Fig. 8a and Fig. 8b. In addition, KM survival analysis revealed that disease-specific survival rates are significantly different in some pathological subgroups, such as T classification (T1-2 vs T3-4, $P < 0.001$), Stage (Stage I-II vs III-IV, $P < 0.001$), Cancer Status (Tumor Free vs With Tumor, $P = 0.011$), Hepatitis Virus Infection (HVI Negative vs Positive, $P = 0.008$), Child-Pugh Score (A/B vs C, $P < 0.001$) and AFP (< = 200 vs > 200, $P = 0.002$) for TCGA OS (Figure S4). These results were consistent with those obtained using a univariate Cox regression for OS with adjustments for the prognostic factors (Fig. 8a, Table S3). In addition, the index can independently predict the OS of GEO (HR (95% CI) = 2.430 (2.054–2.874), $P < 0.001$) (Fig. 8c) and ICGC cohorts (HR (95% CI) = 1.108 (1.069–1.149), $P < 0.001$) (Fig. 8d). All in all, these results strongly suggested that GRGPI was an independent prognostic factor for HCC patients.

Subsequently, further stratified analysis was also performed to investigate the independence of the model within the same subgroups of clinicopathological features. Taking advantage of the clinicopathological parameters, we divided the TCGA discovery cohort into subgroups based on their clinical-pathological features, such as Gender (Male/Female), Age (≤ 65/>65), Grade (G1-2/G3-4), Stage (Stage I-II/III-IV), T classification (T1-2/3–4), Tumor Status (Tumor Free/With Tumor), etc. As revealed in Fig. S5 and Fig. S6, regardless of stratification, 8-mRNA signatures still can make a distinction from high-risk patients. Finally, similar results in the GEO (Fig. S7-S9) and ICGC (Fig. S10-11) cohorts were also statistically significant.

To assess the sensitivity and specificity of the OS/RFS prognostic model and other clinical pathology variables, we performed a multiple ROC curve analysis. To estimate its performance, we applied the 8-mRNA model to the TCGA discovery cohort, TCGA training cohort, GEO, and ICGC cohort, and compared its prediction quality by evaluating the area under the ROC curve. Following the multivariate Cox

regression and AUC analysis, the prognostic model remained a moderate and independent prognostic indicator (TCGA discovery cohort: AUC = 0.860; TCGA training cohort: AUC = 0.801; GEO validation cohort: AUC = 0.834; ICGC validation cohort: AUC = 0.843; Fig. 8e-h). It has demonstrated that the risk score model outperformed the other clinical pathology variables for the prognostic evaluation of HCC patients. These results indicated that GRGPI signatures have a predominately higher favorable value than other parameters in predicting the OS and RFS of HCC patients.

Development and verification of a personalized Nomogram. To provide clinicians with a portable quantitative table to predict the prognosis of liver cancer patients, a nomogram integrating GRGPI and clinicopathological characteristics was constructed in TCGA, GEO, and ICGC cohorts. In the TCGA cohort, the risk score contributed the largest risk point, compared with other clinicopathological characteristics, then followed by T classification, Hepatitis virus infection, Child-Pugh Score and Stage, etc (Fig. 9a). As shown in Fig. 9b and 9c, a total of 371 patients were reclassified in the new Nomogram model for OS NRI (net reclassification index) = 0.415. ROC analysis revealed the accuracy of the Nomo model, which is a good predictor of patient survival, with an AUC value of 0.873 (Fig. 9d). In the decision curve analysis, the novel nomogram showed more net benefit across the range of decision threshold probabilities than the Risk score model and integrated clinicpathology model (Fig. 9e). The calibration curves showed a stable agreement between the prediction by the nomogram and the actual observation for 1-, 2-, and 3-year OS in Fig. 9f. This novel Nomogram model that integrated GRGPI and clinicopathological features maintained good agreement between the predicted and observed survival probabilities in the GEO (AUC = 0.854) and ICGC (AUC = 0.863) cohorts (Fig S12 and Figure S13).

The landscape of immune infiltration in HCC risk subgroups. Due to the significant differences between subtypes, immune infiltration was studied to characterize their immunological characteristics. The CIBERSORT algorithm was used to calculate the abundance of 22 immune-related cell types and displayed in the heatmap and box plot in the TCGA (Fig. 10a-b), GEO (Fig. 10c-d), and ICGC (Fig. 10e-f) cohorts, respectively. Consistently, the frequency of CD8 + T cells in the low-risk group was significantly higher, whereas the proportion of M2 macrophages were rather higher in the high-risk group, in 3 independent cohorts. After calculation of tumor immune infiltration levels of each patient, we found high CD8 + T cell levels could predict better survival, while high levels of M2 cells often indicated worse OS and RFS in HCC tissues (Fig. 10g-k).

WGCNA and GSEA analysis.

We performed WGCNA and GSEA analysis to identify gene expression patterns between different subgroups. Here, based on average clustering, we did not detect outlier samples (Fig. 11a). The soft threshold β was set at 7 to determine a scale-free network (Fig. 11b). Subsequently, the genes were assigned to 16 modules, among which, the gray modules included genes that cannot be clustered (Fig. 11c). The two gene modules most correlated with high- (pink, yellow) and low-risk (greenyellow, turquoise) groups were presented in Fig. 11d. We performed GO and KEGG analysis to identify the potential biological significance of related TOP2 modules in different subgroups (Fig. 11e-h). Besides, we

also performed a GSEA analysis based on the overall TCGA-LIHC expression profiles. The terms existing in both WGCNA and GSEA analysis results were shown in Fig. 11i-l. Specifically, some terms related to cell cycle transition, chromatin separation, DNA replication, and DNA helicase activity were significantly enriched in a high-risk group.

Further verification of the 8-gene signature. To further determine the reliability of the observed gene signatures, we conducted additional verification at the transcriptome level and protein level. We evaluated the expression of 8 gene signatures, based on the TCGA, GEO, and ICGC databases. Inspection of the results revealed a general trend that these 8 gene signatures are dysregulated in HCC tumor tissues, as is demonstrated in Fig. 12a-c. Furthermore, we also examined the expression levels of 8 gene signatures at the protein level using immunohistochemistry (IHC), based on the Human Protein Altas. IHC confirmed the upregulation of protein in HCC tissue. Moderate or high staining intensity of these 8 proteins in HCC tissues contrasted sharply with the low intensity or lack of staining in normal tissues (Fig. 12d).

Table 1. The clinicopathological characteristics of the HCC patients enrolled in the TCGA, GEO, and ICGC cohorts.

Variables	TCGA Cohort (n=371)	ICGC Cohort (n=232)	GEO Cohort (n=221)
Age			
<=65	233	90	200
>65	138	142	21
Gender			
Female	121	61	30
Male	250	171	191
T Classification			
T1	183	N/A	N/A
T2	95	N/A	N/A
T3	80	N/A	N/A
T4	13	N/A	N/A
Stage Classification			
Stage I	179	36	93
Stage II	93	106	78
Stage III	92	71	50
Stage IV	7	19	0
Grade			
G1	57	22	N/A
G2	178	142	N/A
G3	124	59	N/A
G4	12	9	N/A
BCLC Stage			
Zero	N/A	N/A	21
A	N/A	N/A	149
B	N/A	N/A	22
C	N/A	N/A	29

Status			
Alive	241	189	131
Dead	130	43	90
Recurrence			
No	191	N/A	100
Yes	180	N/A	121
Cancer Status			
Tumor Free	250	N/A	N/A
With Tumor	121	N/A	N/A
Family Cancer History			
No	251	152	N/A
Yes	120	80	N/A
Prior Malignancy			
No	N/A	202	N/A
Yes	N/A	30	N/A
Multi Nodular			
No	N/A	N/A	176
Yes	N/A	N/A	45
Cirrhosis			
No	N/A	N/A	18
Yes	N/A	N/A	203
Fibrosis Grade			
No Fibrosis	127	N/A	N/A
Incomplete Cirrhosis	15	N/A	N/A
Established Cirrhosis	111	N/A	N/A
Fibrous Speta	42	N/A	N/A
Portal Fibrosis	76	N/A	N/A
Hepatitis Virus Infection			

None Risk	195	N/A	0
HBV	61	N/A	221
HCV	18	N/A	0
HCV&HBV	97	N/A	0
Child–Pugh Score			
A	198	N/A	97
B	99	N/A	75
C	74	N/A	49
BMI			
<=24	174	N/A	N/A
>24	197	N/A	N/A
AFP level			
Low	190	N/A	121
High	181	N/A	100
ALT level			
<=50	N/A	N/A	130
>50	N/A	N/A	91
Tumor Size			
<=5	N/A	N/A	140
>5	N/A	N/A	81

Abbreviation: BCLC Stage: Barcelona Clinic Liver Cancer Stage; AFP: alpha fetoprotein; ALT: Alanine aminotransferase.

Discussion

Hepatocellular carcinoma (HCC) is classified as a highly malignant tumor that accounts for approximately 90% of total primary liver cancer[25,26]. It has been recognized as the most common malignancy and the most common cause of cancer mortality by absolute cases globally[27]. China represents an area with a high incidence of HCC. According to the World Cancer Report released by the World Health Organization in 2019, new cases of liver cancer in China account for half of the global new

cases, and the total number of death accounts for more than half of the world[28]. Therefore, the treatment of hepatocellular carcinoma has received increasing attention worldwide. Surgical treatment has always been regarded as the main treatment for HCC[29–31]. However, most patients cannot be treated surgically because of tumor anatomical location, tumor size, tumor number, insufficient liver residual volume, or extrahepatic metastasis[32]. Nonsurgical treatment is currently available for most liver cancer patients. The recent development of medical technology and equipment has raised the urgent requirements for the management strategy for HCC[33]. Under the regulation of various ontogenetic modifications, across inflammation, immune suppression, and through direct modulation of host cell behavior. Cancer cells undergo adaptive metabolic programming to maintain their distinctive metabolic state of infinite proliferation. Metabolic rewiring in cancer cells may render them highly dependent on specific metabolic enzymes or processes, may be a viable strategy to design cancer-specific therapeutics[34,35]. Glycogen metabolism is a crucial metabolic process in the liver. Compared to non-cancerous cells of origin, reprogramming of glucose metabolism specifically contributes to facilitates aberrant proliferation and survival in HCC cells[36–38]. A variety of enzymes and proteins engaged in the process of HCC can undergo structural, functional, and expression changes to achieve metabolic reprogramming, which in turn controls the entire glycogen metabolism network to make it suitable for HCC growth[39–42]. Here, we focused in more detail on the abnormalities in their expression patterns of glycolysis-related genes to reflect the metabolic activity of tumors in hypoxic mode. Based on comprehensive bioinformatics analysis, we applied a series of gene sets containing genes responsible for encoding key glycolysis enzymes. Based on the GSEA consequences, we selected those core genes that are significantly enriched in tumor tissues for subsequent analysis. Prospectively, we implemented a combination of transcriptomic analyses followed by a K-M analysis to evaluate the correlation of the expression of glycolysis-associated gene signatures with patient prognosis in 12 solid tumors. After multivariate Cox-PH regression analysis, we identified 8 independent prognostic signatures in HCC, including AURKA, CDK1, CENPA, DEPDC1, HMMR, KIF20A, PFKFB4, and STMN1. We developed and optimized a robust prognostic model based on the 8-gene signatures and verified the powerful performance in 3 independent verification cohorts. Our results suggest that gene signatures related to the glycolysis pathway have the ability to accurately predict the poor prognosis and recurrence of HCC patients. Surprisingly, the prognostic model shows more accurate predictive ability and is superior to other pathological features. In an attempt to evaluate the GRGPI-based risk model in the clinical setting, a nomogram integrating multiple important clinicopathological characteristics has been established, which can be used as a powerful and easy-to-use tool for evaluating the survival probability of HCC patients.

Several studies have shown that some gene signatures are derived from glucose metabolism, including several that are critical for glycolysis and are typically overexpressed in glycolytic cancer cells. Aurora kinase A (AURKA) is an important regulatory protein involved in the regulation of chromosome congression/alignment, regulation of chromosome segregation, and regulation of spindle dynamics[43,44]. Beyond all those effects in the cancer environment, AURKA actively promotes DNA repair and acts as a transcription factor to promote cell migration and invasion[44,45]. It is equally located in the mitochondrial

membrane to regulate mitochondrial dynamics and ATP production[46,47]. It has been recognized as a powerful prognostic indicator that probably integrates multiple oncogenic events in the progression of tumors[43,48,49]. CDK1 (Cyclin-dependentkinase 1) is a serine/threonine-like protein kinase that plays an essential role in controlling cell proliferation at the G2/M point of the cell cycle. Some reports have confirmed that high CDK1 expression is an independent predictor for tumor recurrence in one and five years, and it has been noted that compounds targeting CDK1 could be novel antitumor reagents[50–52]. Growing evidence suggests that CENPA was frequently overexpressed in a variety of cancers, playing an auxiliary but important role in cancer pathogenesis, progression, distant metastases, invasion angiogenesis, etc[53,54]. Previous publications showed that CENPA is abnormally overexpressed in hepatocellular carcinoma (HCC) tumor tissue. It has been declared to be associated with overall survival (OS), disease-free survival (DFS), relapse-free survival (RFS), and progression-free survival (PFS)[55–57]. As a newly discovered gene related to cancer and cell cycle, the key functions and potential regulatory pathways of DEP domain-containing protein 1 (DEPDC1) have been further elucidated in blade cancer, and other human cancers, such as Breast cancer and Prostate cancer. Previous studies revealed the overexpression of DEPDC1 in a variety of tumors and suggested it could drive tumor pathogenesis through multiple potential mechanisms[58–60]. HMMR and KIF20A are important regulators of mitosis and meiosis that revealed carcinogenic properties in various cancers through multiple mechanisms[61–64]. The Warburg pathway enzyme 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 4 (PFKFB4) has been recently reported as a new regulator in the regulation of diverse biological processes to plays an important role in regulating glucose metabolism and guiding macromolecule biosynthesis to maintain the proliferation of cancer cells. Multiple independent research groups have screened and identified PFKFB4 as a poor prognostic factor for multiple tumors through high-throughput[65–67]. STMN1 is considered an oncogene and its aberrant upregulation was closely related to different kinds of tumors[68,69]. As an independent poor predictive factor of the outcome, it is upregulated in hepatocellular carcinoma and promotes migration, invasion, and EMT by activating the PI3K/AKT pathway[69].

The acquisition of a glycolytic strategy is considered a hallmark of oncogenic potential in the field of tumor biology. Reprogrammed glycolytic and mitochondrial pathways are hallmarks of the altered energy generation system of malignant cells, which specifically contributes to the abnormal survival and proliferation energy consumption of tumor cells[70]. Metabolic mode switching from aerobic oxidation to anaerobic glycolysis is an important feature of hepatocellular carcinoma. The Warburg effect results in the accumulation of the lactic acid, as the final product of glycolysis. Mechanisms of immune resolution are often coopted by an acidic microenvironment to mediate their immune escape. This particular acidic environment is formed by the continuous accumulation of highly acidic substances such as lactic acid and ketone bodies[71,72]. Studies in the past decade have found that aerobic glycolysis and the resulting acidification of tumor microenvironment (TME) should exert specific inhibitory effects on the antitumor immune response mediated by T cells and the activity of tumor-infiltrating myeloid cells. Therefore, intervening for sugar metabolism and/or lactic acid production and secretion is an attractive anticancer treatment strategy[73,74]. Also, we evaluated the inflammatory infiltration landscape in HCC tissues based

on 22 immune cells using the cibersort[75]. T cells CD8 have emerged as important immunomodulatory cytokines, which play a critical role at the interface between innate and adaptive immunity, especially in antitumor immune response[76–82]. Macrophages M2 have been proven to promote tumor progression and poor prognosis, especially the formation of metastases in target organs[82–84]. We hypothesize that the enhanced glycolytic activity contributes to the highly acidic environment in the TME. It is widely understood that tumor-reactive T cells are suppressed and resulting in loss-of-function in the acidic TME induced by glycolysis activity, representing a critical barrier for the success of cancer immunotherapy[85]. Most immunotherapies target the immune system but not cancer and, therefore, immunotherapies are believed to be a promising foundation to build treatment regimens for a variety of tumor types. However, the complexity of the metabolic regulation of immune cell subsets and the influence of the TME may have significant implications for the effectiveness of these therapies. Our research revealed 8 gene signatures related to the prognosis of HCC, which are important regulators involved in glucose metabolism and energy production, especially the glycolysis process. Altogether, these observations suggest that the glycolysis pathway is required for the proliferation of most cancer cells and energy production in recreating the microenvironmental characteristics, which enables future efforts for therapeutic optimization to block the glycolysis pathway, thereby controlling tumor progression and improving patients prognosis. These findings may have remarkable prognostic and therapeutic implications for HCC patients.

Inevitably, we acknowledge that some limitations concerning the current study should be considered. First, we use univariate Cox regression analysis and the LASSO method to filtrate glycolytic genes associated with clinical outcomes of HCC and build a prognostic model through multivariate Cox-PH regression analysis. In the linear regression model, adjustments were made stepwise in major groups, which did reveal which variables contributed the most to confounding, some important components with similar contributions may be ignored. Second, we developed and validated the prognostic risk model based on public databases, which was not verified by prospective clinical trials. Subsequent research should consider some traditionally recognized clinical factors, which have a profound impact on tumor progression and prognosis of HCC patients. There are factors related to a clinical interaction that may be missed, such as tumor volume, TP53 mutation, CTNNB1 mutation, lifestyle, and patient follow-up time, and relevant therapeutic information. The factors mentioned above will have more or less impact on the scientific prediction of the model. Therefore, the predictive performance of predictive models based on glycolysis-related gene signatures should be more thoroughly explored in subsequent studies. Importantly, whether clinical decision-making dictated by such approaches leads to improved clinical risk stratification remains an unanswered question that will require well-designed, prospective, multicenter collaborative trials. In order to overcome these limitations, more in vivo and in vitro studies are urgent to verify our model and explore more complex and in-depth biological mechanisms. These will be the focus of our future research.

Conclusion

We have developed and optimized a novel 8-gene signature for identifying outcomes and recurring events in HCC patients. This predictive model improves the accuracy of predicting patient prognosis. Also, the 8-gene signature serves as an independent studies and superior to other clinicopathological features. Based on the GRGPI signatures and clinicopathological characteristics, we established a nomogram, which significantly improved the prognosis in terms of discrimination and effectiveness of clinical decision-making. This study provides the basis of prognostic stratification for designing prospective trials of risk-adapted therapies and surveillance strategies. Our research may have certain clinical value, contributing and benefiting more patients from additional systemic treatment.

Abbreviations

GEO: Gene expression omnibus database; TCGA: The Cancer Genome Atlas; ICGC: International Cancer Genome Consortium; DEGs: Differentially expressed genes; AUC: Area Under roc Curve; CI: Confidence Interval; WGCNA: Weighted Gene Co-expression Network Analysis; PFS: progression-free survival; MSigDB: Molecular signatures database; GSEA: Gene set enrichment analysis; t-SNE: t-distributed stochastic neighbor embedding; HBV: viral hepatitis type B; HCV: viral hepatitis type C; AFP: α-fetoprotein; BCLC: Barcelona Clinic Liver Cancer; DCA: Decision Curve Analysis.

Declarations

Authors contributions

Lingyu Zhang, Yu Li, and Zhihua Tao conceived and designed the study. Yu Li searched a large number of databases and incorporated a series of datasets available for the study. Lingyu Zhang, Weiwei Liu analyzed the datasets and were responsible for the writing of this manuscript. Lingyu Zhang, Yu Li, and Xuchu Wang were under the responsibility of the production of Figures. Lingyu Zhang, Danhua Wang, Ying Cao, and Yibei Dai have searched a large number of literatures and were responsible for reference compilation. Zhihua Tao gave a lot of guidance on manuscript writing. All authors reviewed and considered the final manuscript.

Acknowledgments

Not applicable.

Consent for publication

Not applicable.

Conflict of interest disclosures

The author declares no competing financial interests.

Data availability statement

Underlying research materials can be available upon request by contacting the corresponding author.

Funding

The present study was supported by grants from the National Natural Science Foundation of China Youth Science Foundation Project (Grant nos. 81802571); Zhejiang Medical and Health Science and Technology Project (2019RC039); the National Natural Science Foundation of China (Grant nos. 81902156) and the Natural Science Key Project of Bengbu Medical College (No.BYKY2019012ZD).

Ethics approval and consent to participate

Since the identities of patients in the TCGA and GEO databases cannot be identified, no approval and informed consent from the institutional review board is required.

References

1. Li, X. *et al.* Navigating metabolic pathways to enhance antitumour immunity and immunotherapy. *Nat. Rev. Clin. Oncol.* **16**, 425–441 (2019).
2. Luo, M. *et al.* Targeting Breast Cancer Stem Cell State Equilibrium through Modulation of Redox Signaling. *Cell Metab.* **28**, 69–86 (2018).
3. Chen, P. *et al.* Metabolic Diversity in Human Non-Small Cell Lung Cancer Cells. *Mol. Cell.* **76**, 838–851 (2019).
4. Boroughs, L. K. & DeBerardinis, R. J. Metabolic pathways promoting cancer cell survival and growth. *Nat. Cell Biol.* **17**, 351–359 (2015).
5. Shukla, S. K. *et al.* MUC1 and HIF-1alpha Signaling Crosstalk Induces Anabolic Glucose Metabolism to Impart Gemcitabine Resistance to Pancreatic Cancer. *Cancer Cell.* **32**, 71–87 (2017).
6. Douglas, H. & Robert, A. W. Hallmarks of Cancer: The Next Generation. Cell. 2011;(144):646 – 74 (2011).
7. Li, L. *et al.* Transcriptional Regulation of the Warburg Effect in Cancer by SIX1. *Cancer Cell.* **33**, 368–385 (2018).
8. Sun, L., Suo, C., Li, S., Zhang, H. & Gao, P. Metabolic reprogramming for cancer cells and their microenvironment.Beyond the Warburg Effect. Biochim. Biophys. Acta. Rev. Cancer.1870,51–66(2018).
9. Ebron, J. S. *et al.* MiR-644a Disrupts Oncogenic Transformation and Warburg Effect by Direct Modulation of Multiple Genes of Tumor-Promoting Pathways. *Cancer Res.* **79**, 1844–1856 (2019).
10. Ji, K., Mayernik, L., Moin, K. & Sloane, B. F. Acidosis and proteolysis in the tumor microenvironment. *Cancer Metast. Rev.* **38**, 103–112 (2019).
11. Palsson-McDermott, E. M. *et al.* Pyruvate Kinase M2 Regulates Hif-1a Activity and IL-1b Induction and Is a Critical Determinant of the Warburg Effect in LPS-Activated Macrophages. *Cell Metab.* **21**, 65–80 (2015).

12. Tanner, L. B. *et al.* Four Key Steps Control Glycolytic Flux in Mammalian Cells. *Cell Syst.* **7**, 49–62 (2018).
13. Wang, C. Q. *et al.* Interactome analysis reveals that lncRNA HULC promotes aerobic glycolysis through LDHA and PKM2. *Nat Commun.* **11**, 3162–3176 (2020).
14. Dai, W. X. *et al.* FOXE1 represses cell proliferation and Warburg effect by inhibiting HK2 in colorectal cancer. *Cell Commun. Signal.* **18**, 7–19 (2020).
15. Houles, T. *et al.* RSK Regulates PFK-2 Activity to Promote Metabolic Rewiring in Melanoma. *Cancer Res.* **78**, 2191–2204 (2018).
16. Zhou, Y. Y. *et al.* Benserazide is a novel inhibitor targeting PKM2 for melanoma treatment. *Int J Cancer.* **147**, 139–151 (2020).
17. Feng, J. *et al.* PKM2 is the target of proanthocyanidin B2 during the inhibition of hepatocellular carcinoma. *J. Exp. Clin. Canc. Res.* **38**, 204–228 (2019).
18. Yu, K. *et al.* Association of Omics Features with Histopathology Patterns in Lung Adenocarcinoma. *Cell Syst.* **5**, 620–627 (2017).
19. Constantine, A., Ursula, M., Alex, F., Thomas, P. & Mark, L. Molecular and histopathology directed therapy for advanced bladder cancer. *Nat. Rev. Urol.* **8**, 465–483 (2019).
20. Thomas, M. A., Yang, L., Carter, B. J. & Klaper, R. D. Gene set enrichment analysis of microarray data from Pimephales promelas (Rafinesque), a non-mammalian model organism.BMC. Genomics.12,66–74(2011).
21. Zhang, J. X. *et al.* Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. *Lancet Oncol.* **14**, 1295–1306 (2013).
22. Ron Van, D. *et al.* Tools and techniques–statistics: how many variables are allowed in the logistic and Cox regression models? EuroIntervention. 9,1472–1473(2014).
23. Yu, G. C., Wang, L., Han, Y. & He, Q. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology.* **16**, 284–287 (2012).
24. Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science.* **357**, eaan2507 (2017).
25. Bertuccio, P. *et al.* Global trends and predictions in hepatocellular carcinoma mortality. *J Hepatol.* **67**, 302–329 (2017).
26. Yang, J. D. *et al.* A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat Rev Gastroenterol Hepatol.* **16**, 589–604 (2019).
27. Kulik, L. & El-Serag, H. B. Epidemiology and Management of Hepatocellular Carcinoma. *Gastroenterology.* **156**, 477–491 (2019).
28. Zhou, J. *et al.* Guidelines for Diagnosis and Treatment of Primary Liver Cancer in China (2017 Edition). *Liver Cancer.* **7**, 235–260 (2018).
29. Liu, A. M. *et al.* A novel strategy for the diagnosis, prognosis, treatment, and chemoresistance of hepatocellular carcinoma: DNA methylation. *Med. Res. Rev.* **40**, 1973–2018 (2020).

30. Xu, X. F. *et al.* Risk Factors, Patterns, and Outcomes of Late Recurrence After Liver Resection for Hepatocellular Carcinoma. *JAMA. Surg.* **154**, 209–217 (2019).
31. Kim, H. *et al.* Survival benefit of liver resection for Barcelona Clinic Liver Cancer stage B hepatocellular carcinoma. *Brit. J. Surg.* **104**, 1045–1052 (2017).
32. Chen, Z. H. *et al.* Actual long-term survival in HCC patients with portal vein tumor thrombus after liver resection: a nationwide study. *Hepatol. Int.* **14**, 754–764 (2020).
33. Xu, W. Q. *et al.* Immunotherapy for hepatocellular carcinoma: recent advances and future perspectives. *Ther. Adv. Med. Oncol.* **11**, 1–15 (2019).
34. Ding, Z. B. *et al.* Metabolic pathway analyses identify proline biosynthesis pathway as a promoter of liver tumorigenesis. *J. Hepatol.* **72**, 725–735 (2020).
35. Ahmed, N., Escalona, R., Leung, D., Chan, E. & Kannourakis, G. Tumour microenvironment and metabolic plasticity in cancer and cancer stem cells: Perspectives on metabolic and immune regulatory signatures in chemoresistant ovarian cancer stem cells. *Semin. Cancer Biol.* **53**, 265–281 (2018).
36. Fang, G. X. *et al.* Inhibition of GSK-3β activity suppresses HCC malignant phenotype by inhibiting glycolysis via activating AMPK/mTOR signaling. *Cancer Lett.* **463**, 11–26 (2019).
37. Gjorgjieva, M. *et al.* Dietary exacerbation of metabolic stress leads to accelerated hepatic carcinogenesis in glycogen storage disease type Ia. *J. Hepatol.* **69**, 1074–1087 (2018).
38. Iwagami, Y. *et al.* Aspartate β-hydroxylase modulates cellular senescence through glycogen synthase kinase 3β in hepatocellular carcinoma. *Hepatology.* **63**, 1213–1226 (2016).
39. Zhang, N. *et al.* Glycogen synthase kinase-3β inhibition promotes lysosome-dependent degradation of c-FLIPL in hepatocellular carcinoma. *Cell Death. Dis.* **9**, 230–242 (2018).
40. Chua, H. H. *et al.* RBMY, A Novel Inhibitor of Glycogen Synthase Kinase 3β, Increases Tumor Stemness and Predicts Poor Prognosis of Hepatocellular Carcinoma. *Hepatology (Baltimore, Md.).* **62**, 1480–1496 (2015).
41. Xu, Q. R. *et al.* HSP90 promotes cell glycolysis, proliferation and inhibits apoptosis by regulating PKM2 abundance via Thr-328 phosphorylation in hepatocellular carcinoma. *Mol Cancer.* **16**, 178–191 (2017).
42. Wu, H. *et al.* A negative reciprocal regulatory axis between cyclin D1 and HNF4 α modulates cell cycle progression and metabolism in the liver. *PNAS.* **29**, 17177–17186 (2020).
43. Goos, J. A. C. M. *et al.* Aurora kinase A (AURKA) expression in colorectal cancer liver metastasis is associated with poor prognosis. *Brit. J. Cancer.* **109**, 2445–2452 (2013).
44. Bertolin, G. & Tramier, M. Insights into the non-mitotic functions of Aurora kinase A: more than just cell division. *Cell Mol. Life Sci.* **77**, 1031–1047 (2020).
45. Wang-Bishop, L. *et al.* Gastrointestinal Cancer Cells With Activated KRAS by Preventing Activation of RPS6KB1. *Gastroenterology.* **156**, 662–675 (2019).

46. Yang, C. *et al.* Effects of AURKA-mediated degradation of SOD2 on mitochondrial dysfunction and cartilage homeostasis in osteoarthritis. *J. Cell Physiol.* **234**, 17727–17738 (2019).
47. Bertolin, G. *et al.* Aurora kinase A localises to mitochondria to control organelle dynamics and energy production. *Elife.* **7**, e38111 (2018).
48. Yang, N. *et al.* FOXM1 recruits nuclear Aurora kinase A to participate in a positive feedback loop essential for the self-renewal of breast cancer stem cells. *Oncogene.* **24**, 3428–3240 (2017).
49. Zheng, F. M. *et al.* Nuclear AURKA acquires kinase-independent transactivating function to enhance breast cancer stem cell phenotype. *Nat. Commun.* **7**, 10180–10196 (2016).
50. Wu, C. X. *et al.* Blocking CDK1/PDK1/β-Catenin signaling by CDK1 inhibitor RO3306 increased the efficacy of sorafenib treatment by targeting cancer stem cells in a preclinical model of hepatocellular carcinoma. *Theranostics.* **8**, 3737–3750 (2018).
51. Cai, J. L. *et al.* Prognostic Biomarker Identification Through Integrating the Gene Signatures of Hepatocellular Carcinoma Properties.Ebiomedicine.19,18–30(2017).
52. Moiseeva, T. N., Qian, C., Sugitani, N., Osmanbeyoglu, H. U. & Bakkenist, C. J. WEE1 kinase inhibitor AZD1775 induces CDK1 kinase-dependent origin firing in unperturbed G1- and S-phase cells. *PNAS.* **116**, 23891–23893 (2019).
53. Liu, W. T. *et al.* A novel strategy of integrated microarray analysis identifies CENPA, CDK1 and CDC20 as a cluster of diagnostic biomarkers in lung adenocarcinoma. *Cancer Lett.* **425**, 43–53 (2018).
54. Roulland, Y. *et al.* The Flexible Ends of CENP-A Nucleosome Are Required for Mitotic Fidelity. *Mol. Cell.* **63**, 674–685 (2016).
55. Bayo, J. *et al.* A comprehensive study of epigenetic alterations in hepatocellular carcinoma identifies potential therapeutic targets. *J. Hepatol.* **71**, 78–90 (2019).
56. Li, B. X., Pu, K. & Wu, X. A. Identifying novel biomarkers in hepatocellular carcinoma by weighted gene co-expression network analysis. *J. Cell Biochem.* **120**, 11418–11431 (2019).
57. Long, J. Y. *et al.* A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma. *J. Cell Mol. Med.* **22**, 5928–5938 (2018).
58. Tosi, A. *et al.* Identification of a HLA-A*0201-restricted immunogenic epitope from the universal tumor antigen DEPDC1. *Oncoimmunology.* **6**, e1313371 (2017).
59. Ramalho-Carvalho, J. *et al.* Epigenetic disruption of miR-130a promotes prostate cancer by targeting SEC23B and DEPDC1. *Cancer Lett.* **385**, 150–159 (2017).
60. Yosuke, H. *et al.* Cell-Permeable Peptide DEPDC1-ZNF224 Interferes with Transcriptional Repression and Oncogenicity in Bladder Cancer Cells. *Cancer Res.* **70**, 5829–5839 (2010).
61. Li, W., Pan, T. H., Jiang, W. & Zhao, H. Y. HCG18/miR-34a-5p/HMMR axis accelerates the progression of lung adenocarcinoma.Biomed. Pharmacother.129,110217(2020).
62. Stevens, L. E. *et al.* Extracellular Matrix Receptor Expression in Subtypes of Lung Adenocarcinoma Potentiates Outgrowth of Micrometastases. *Cancer Res.* **77**, 1905–1917 (2017).

63. Khongkow, P. *et al.* Paclitaxel targets FOXM1 to regulate KIF20A in mitotic catastrophe and breast cancer paclitaxel resistance. *Oncogene.* **35**, 990–1002 (2016).
64. Lu, M. Q. *et al.* Aberrant KIF20A expression might independently predict poor overall survival and recurrence-free survival of hepatocellular carcinoma. *Iubmb. Life.* **70**, 328–335 (2018).
65. Yao, L., Wang, L., Cao, Z. G., Hu, X. & Shao, Z. M. High expression of metabolic enzyme PFKFB4 is associated with poor prognosis of operable breast cancer. *Cancer Cell. Int.* **19**, 165–172 (2019).
66. Gao, R. F. *et al.* CD44ICD promotes breast cancer stemness via PFKFB4-mediated glucose metabolism. *Theranostics.* **8**, 6248–6462 (2018).
67. Wang, Q. Y. *et al.* Etk Interaction with PFKFB4 Modulates Chemoresistance of Small-cell Lung Cancer by Regulating Autophagy. *Clin. Cancer Res.* **24**, 950–962 (2018).
68. Li, M. *et al.* Activation of an AKT/FOXM1/STMN1 pathway drives resistance to tyrosine kinase inhibitors in lung cancer. *Brit. J. Cancer.* **117**, 974–983 (2017).
69. Zhang, R. *et al.* STMN1 upregulation mediates hepatocellular carcinoma and hepatic stellate cell crosstalk to aggravate cancer by triggering the MET pathway. *Cancer Sci.* **111**, 406–417 (2020).
70. Xia, Y. *et al.* RNF8 mediates histone H3 ubiquitylation and promotes glycolysis and tumorigenesis. *J. Exp. Med.* **214**, 1843–1855 (2017).
71. Nayak, A. P., Kapur, A., Barroilhet, L. & Patankar, M. S. Oxidative Phosphorylation: A Target for Novel Therapeutic Strategies Against Ovarian Cancer. *Cancers.* **10**, 337–351 (2018).
72. Dias, A. S., Almeida, C. R. & Helguero, L. A. & Duarte I.F. Metabolic crosstalk in the breast cancer microenvironment. *Eur. J. Cancer.* **121**, 154–171 (2019).
73. Ocaña, M. C., Martínez-Poveda, B., Quesada, A. R. & Medina, M. Ã. Metabolism within the tumor microenvironment and its implication on cancer progression: An ongoing therapeutic target. *Med. Res. Rev.* **39**, 70–113 (2019).
74. Locatelli, S. L. *et al.* Targeting Cancer Cells and Tumor Microenvironment in Preclinical and Clinical Models of Hodgkin Lymphoma Using the Dual PI3Kδ/γ Inhibitor RP6530. *Clin. Cancer. Res.* **25**, 1098–1112 (2019).
75. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods.* **12**, 453–457 (2015).
76. Petitprez, F. *et al.* B cells are associated with survival and immunotherapy response in sarcoma. *Nature.* **577**, 556–560 (2020).
77. Helmink, B. A. *et al.* B cells and tertiary lymphoid structures promote immunotherapy response. *Nature.* **577**, 549–555 (2020).
78. Hollern, D. P. *et al.* B Cells and T Follicular Helper Cells Mediate Response to Checkpoint Inhibitors in High Mutation Burden Mouse Models of Breast Cancer. *Cell.* **179**, 1191–1206 (2019).
79. Gu, J. Y. *et al.* Activating miRNA-mRNA network in gemcitabine-resistant pancreatic cancer cell associates with alteration of memory CD4 + T cells. *Ann. Transl. Med.* **8**, 279–293 (2020).

80. Kondo, T. *et al.* Notch-mediated conversion of activated T cells into stem cell memory-like T cells for adoptive immunotherapy. *Nat. Commun.* **8**, 15338 (2017).
81. Ortega, L. S., Yadileiny, P., Sonia, P. Y. & Domingo, F. B. Magnetic targeting of adoptively transferred tumour – specific nanoparticle – loaded CD8 + T cells does not improve their tumour infiltration in a mouse model of cancer but promotes the retention of these cells in tumour – draining lymph nodes. *J. Nanobiotechnol.* **17**, 87–110 (2019).
82. Yang, Y. *et al.* Crosstalk between hepatic tumor cells and macrophages via Wnt/β-catenin signaling promotes M2-like macrophage polarization and reinforces tumor malignant behaviors. *Cell Death. Dis.* **9**, 793–806 (2018).
83. Wen, Z. F. *et al.* Tumor cell-released autophagosomes (TRAPs) promote immunosuppression through induction of M2-like macrophages with increased expression of PD-L1. *J. Immunother. Cancer.* **6**, 116–151 (2018).
84. Zhang, C. R. *et al.* Noninvasive Imaging of CD206-Positive M2 Macrophages as an Early Biomarker for Post-Chemotherapy Tumor Relapse and Lymph Node Metastasis. *Theranostics.* **7**, 4276–4288 (2017).
85. Beckermann, K. E., Dudzinski, S. O. & Rathmell, J. C. Dysfunctional T cell metabolism in the tumor microenvironment. *Cytokine Growth F. R.* **35**, 7–14 (2017).

Figures

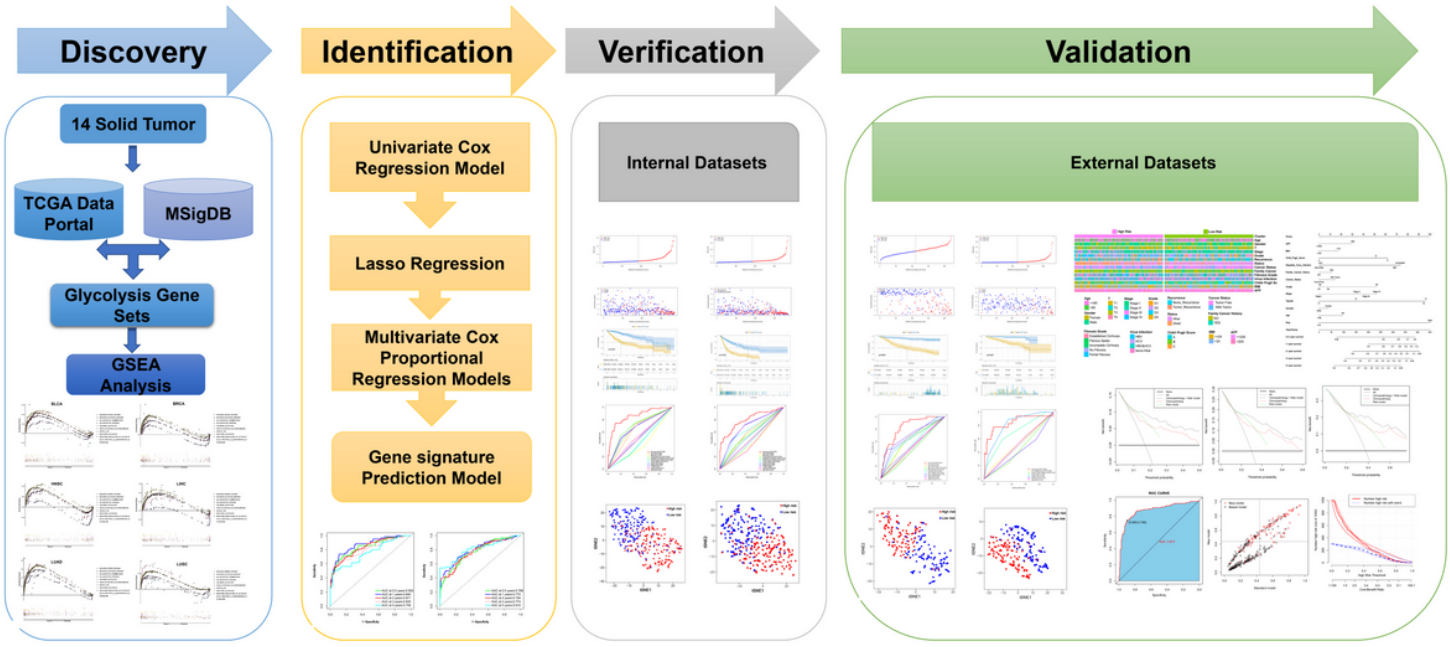

Figure 1

Panoramic flowchart for the development and verification of glycolysis gene signatures.

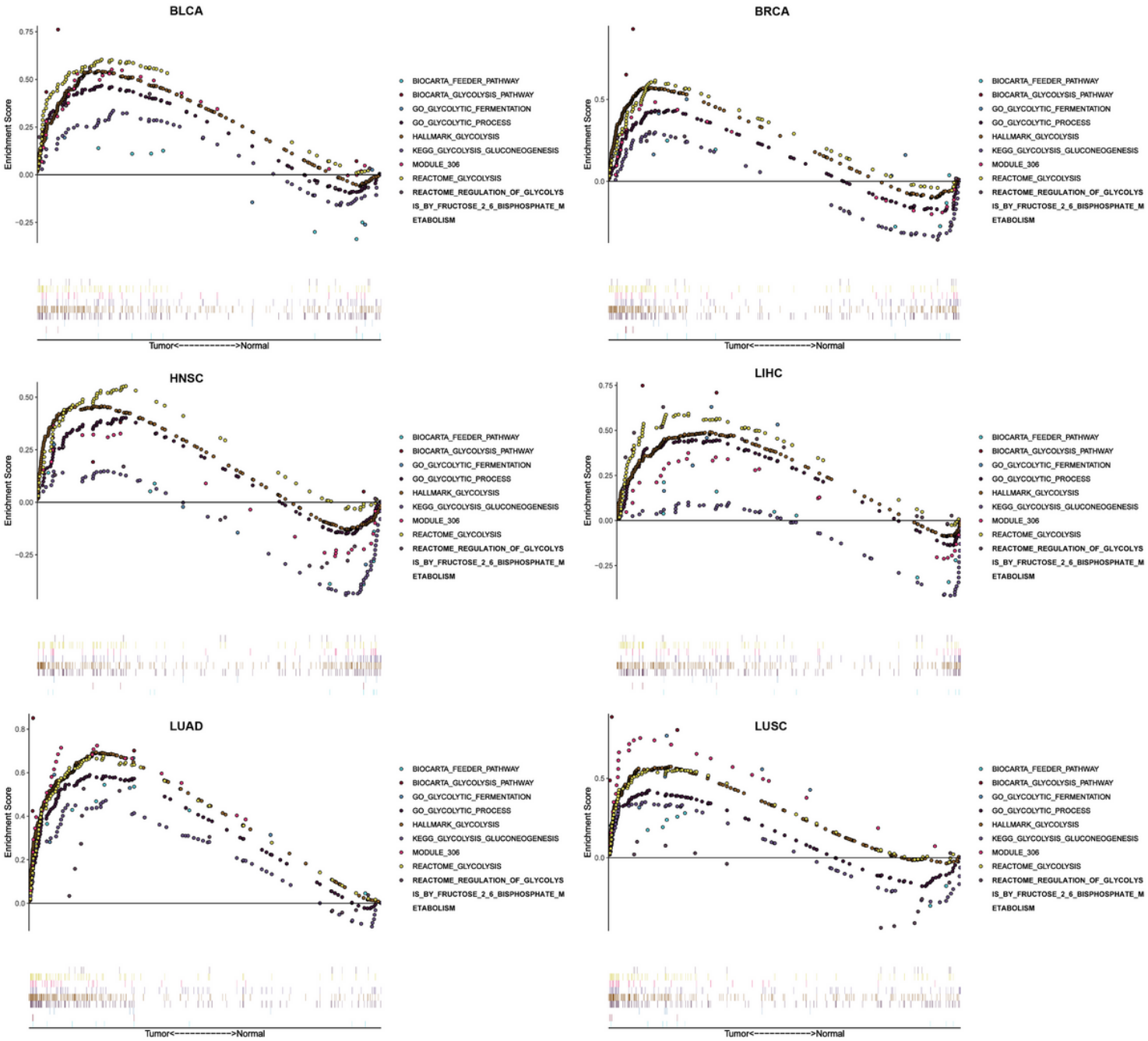

Figure 2

The performance of the enrichment curve of 9 glycolysis-related gene sets in 6 tumors (BLCA, BRCA, HNSC, LIHC, LUAD, LUSC), with FDR less than 0.05 as the statistical threshold.

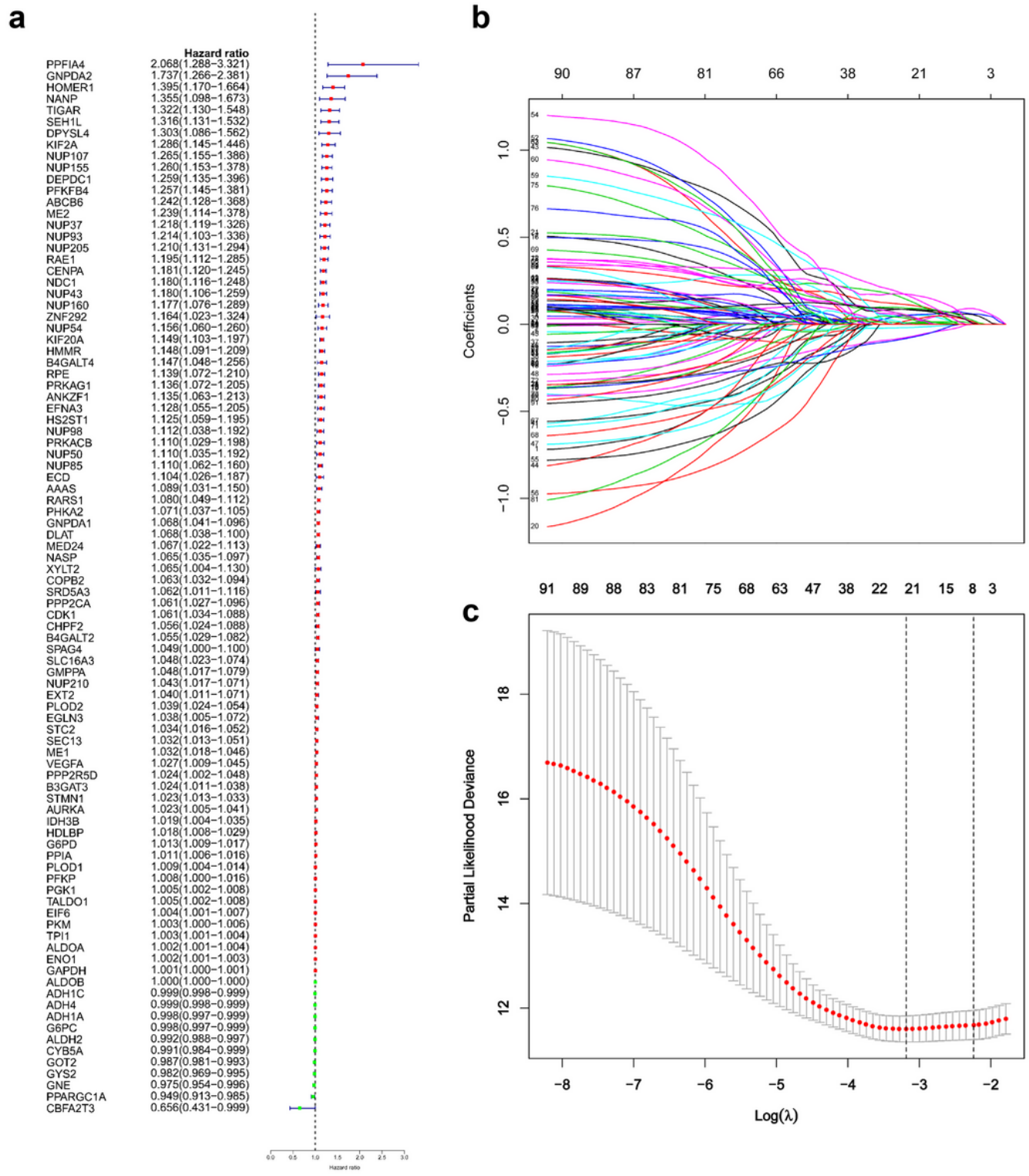

Figure 3

Identification of genetic signatures for HCC prognostic models. a Univariate cox regression analysis candidates 92 glycolytic genes related to OS in HCC patients. b LASSO coefficient profiles of the 92 OS-associated genes. c The adjustment parameter (λ) is selected through a 20-fold cross-validation procedure and plotted as a function of log (λ) in the LASSO model.

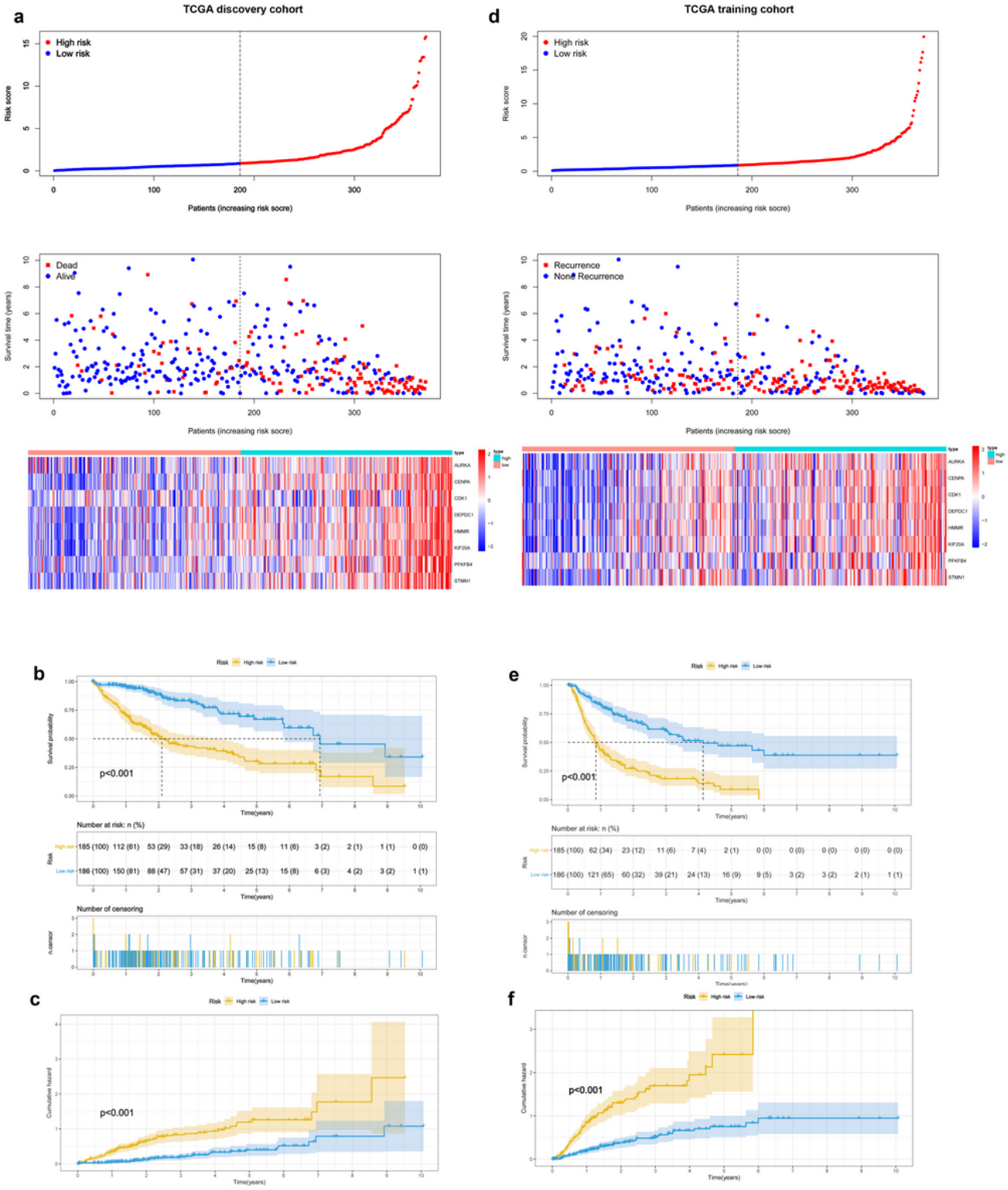

Figure 4

GRGPI signatures serve as a risk factor and promising predictor for overall survival (OS) and recurrence-free survival (RFS) in the TCGA cohort. (a, b) The distribution of risk score, survival status, and gene expression patterns of HCC patients in the high- and low-risk groups for OS and RFS. (c, d) Kaplan–Meier plots for OS and RFS of two risk groups in the TCGA cohort. (e-f) Performance of the cumulative event probability in the two risk subgroups.

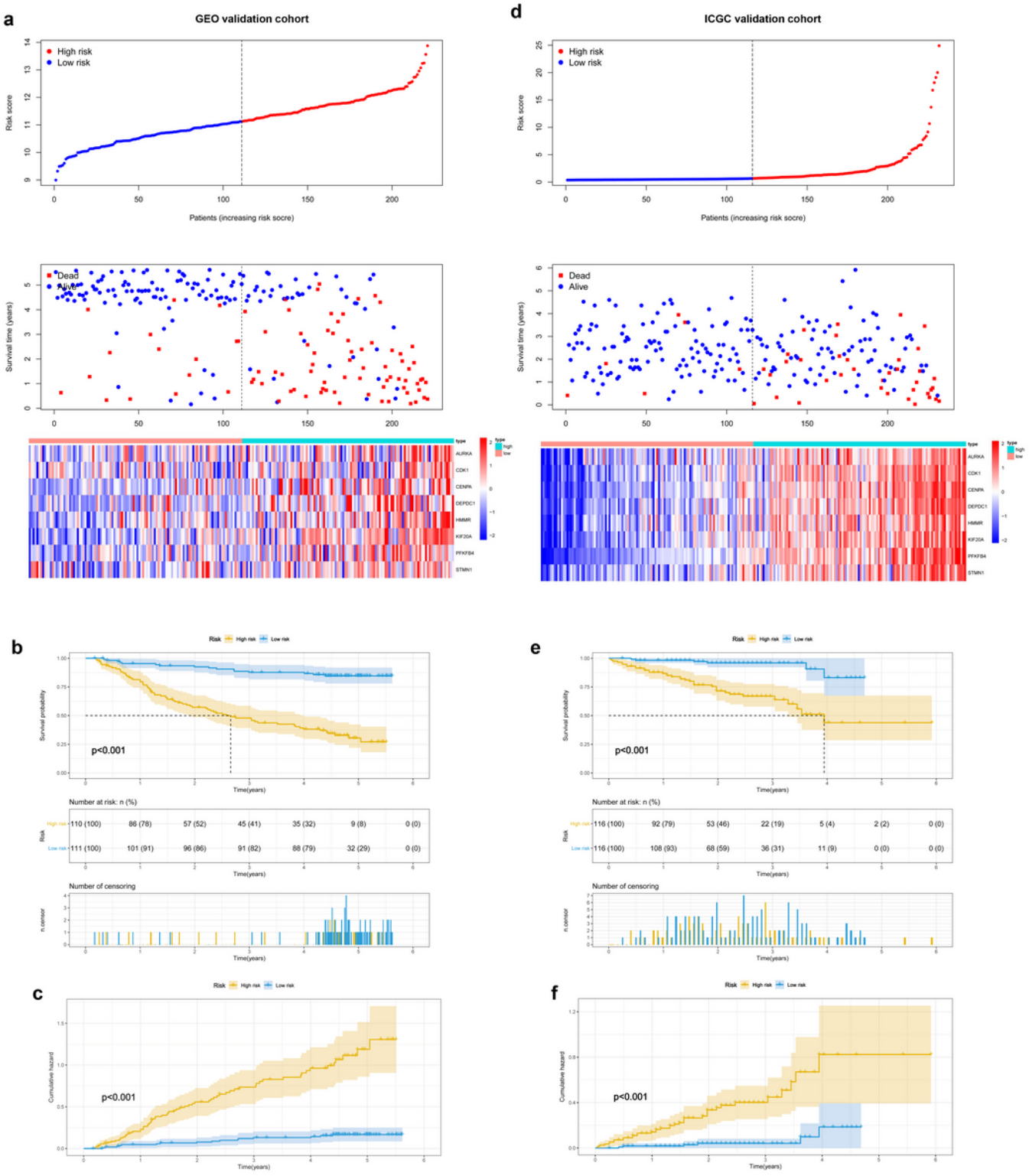

Figure 5

The performance of GRGPI signature in GEO and ICGC validation cohorts. (a, b) The distribution of risk score, survival status, and gene expression patterns of HCC patients in 2 validation cohorts. (c, d) Kaplan–Meier plots for OS in two risk subgroups. (e-f) Performance of the cumulative event probability in the two risk subgroups.

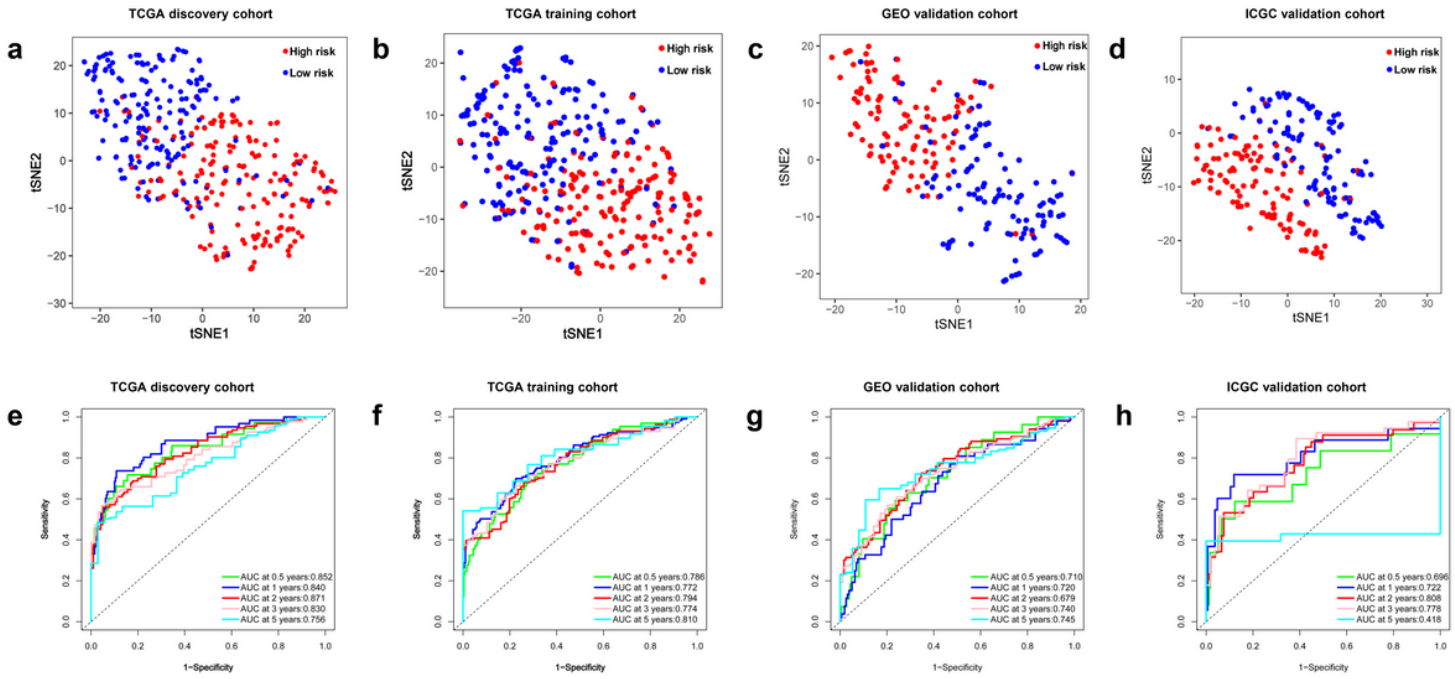

Figure 6

(a-d) T-SNE analysis supported the distribution of high-risk and low-risk patients in discrete directions in TCGA discovery cohort, TCGA training cohort, GEO validation cohort, and ICGC validation cohort. Time-dependent ROC curve analysis of the GRGPI model in the TCGA discovery cohort e, TCGA training cohort (f), GEO cohort (g), and ICGC cohort (h) at 0.5, 1, 2, 3, and 5 years.

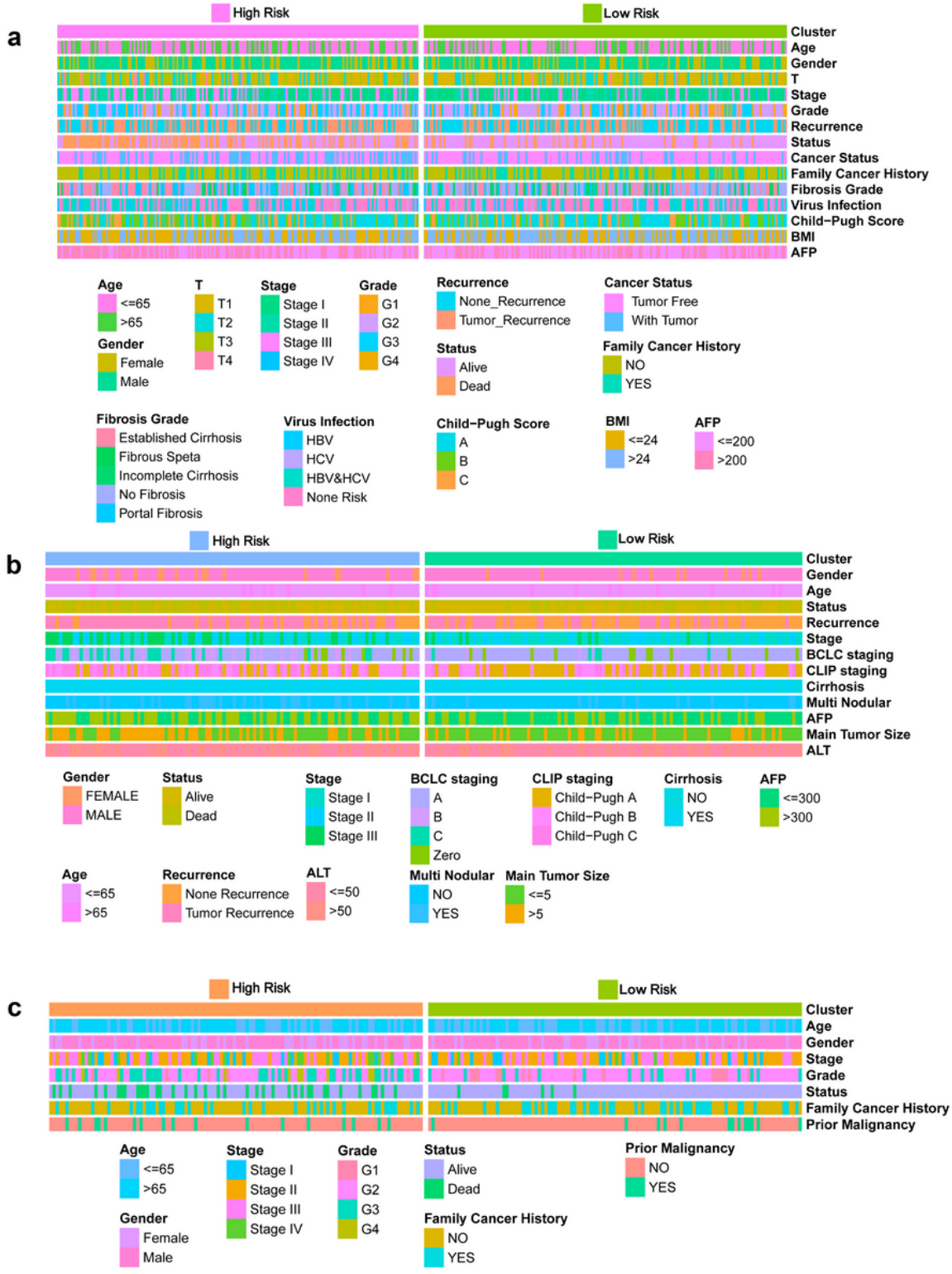

Figure 7

Clinical characteristics of HCC subclasses in the (a) TCGA, (b) GEO, and (c) ICGC cohorts.

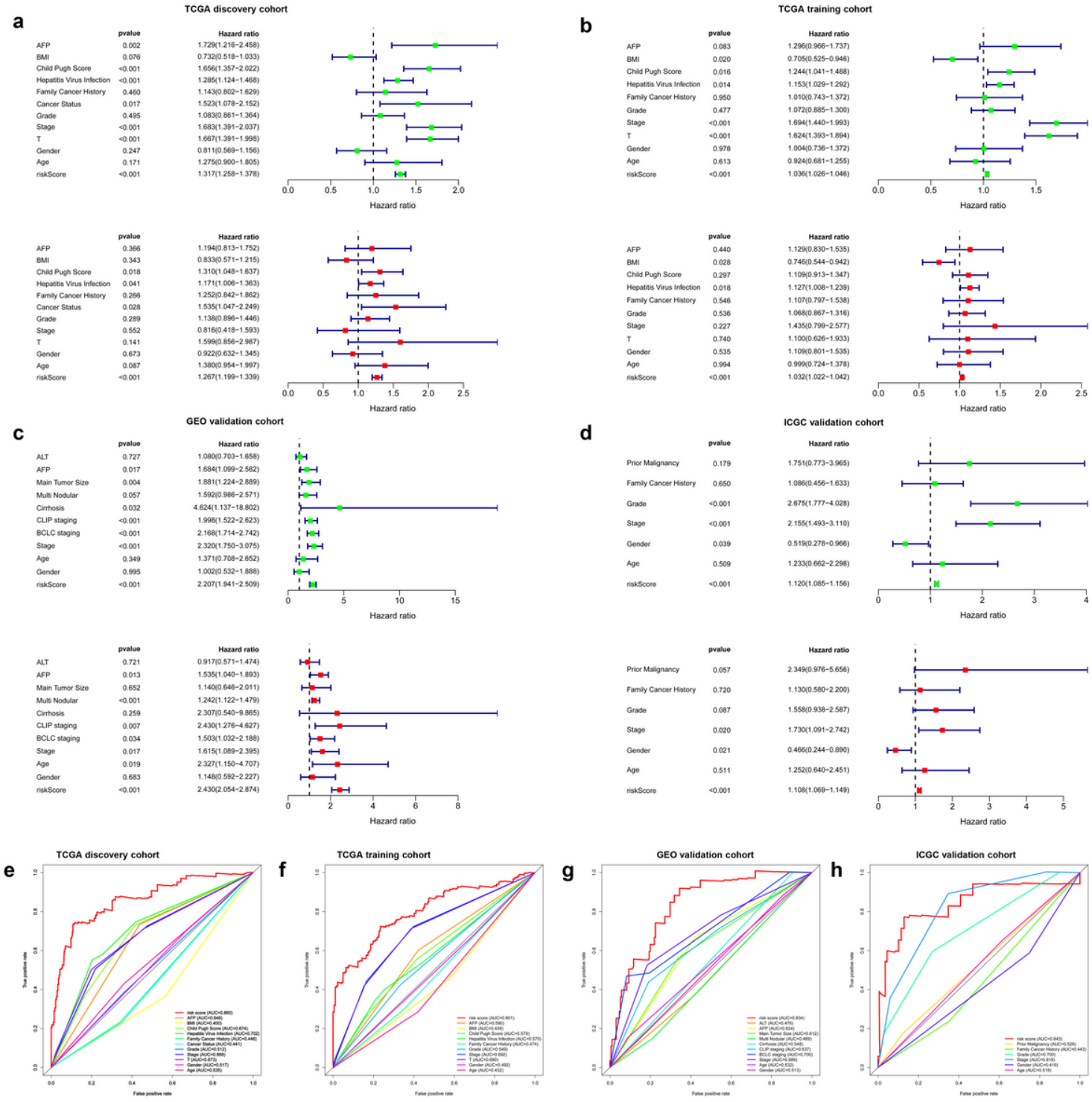

Figure 8

(a-d) Univariate and multivariate Cox regression analysis based on all variables for the prediction of OS and RFS in 4 cohorts. The green dots represent the HR value of univariate cox, and the red dots represent the HR value of multivariate cox, the blue lines indicate the standard error (SE) of HR. Multiple ROC curves compare the predictive power of the GRGPI model and other clinicopathological features in TCGA discovery cohort (e), TCGA training cohort (f), GEO cohort (g), and ICGC cohort (h).

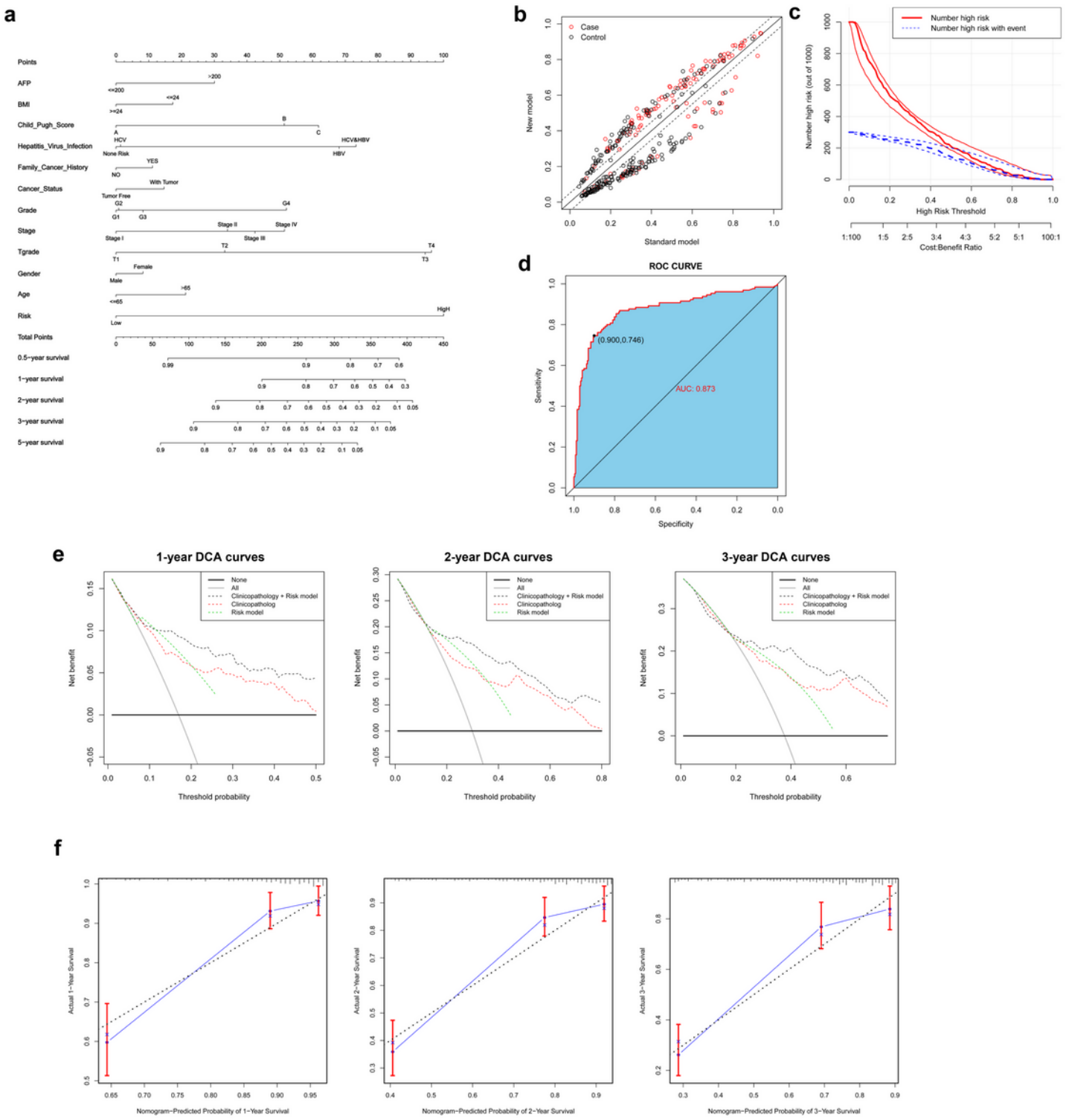

Figure 9

The performance of a personalized Nomogram based on GRGPI and clinicopathological features. A Nomograph for predicting overall survival probability of HCC patients. (b-c) The Nomogram model improves the identification of high-risk patients, and 371 HCC patients are reclassified between the standard model and the Nomogram model. d ROC curve for Nomogram model. e Decision curve analyses

of Nomogram for 1-, 2-, and 3-year. f Calibration curves of 1-, 2-, and 3-year OS for HCC patients in the TCGA discovery cohort.

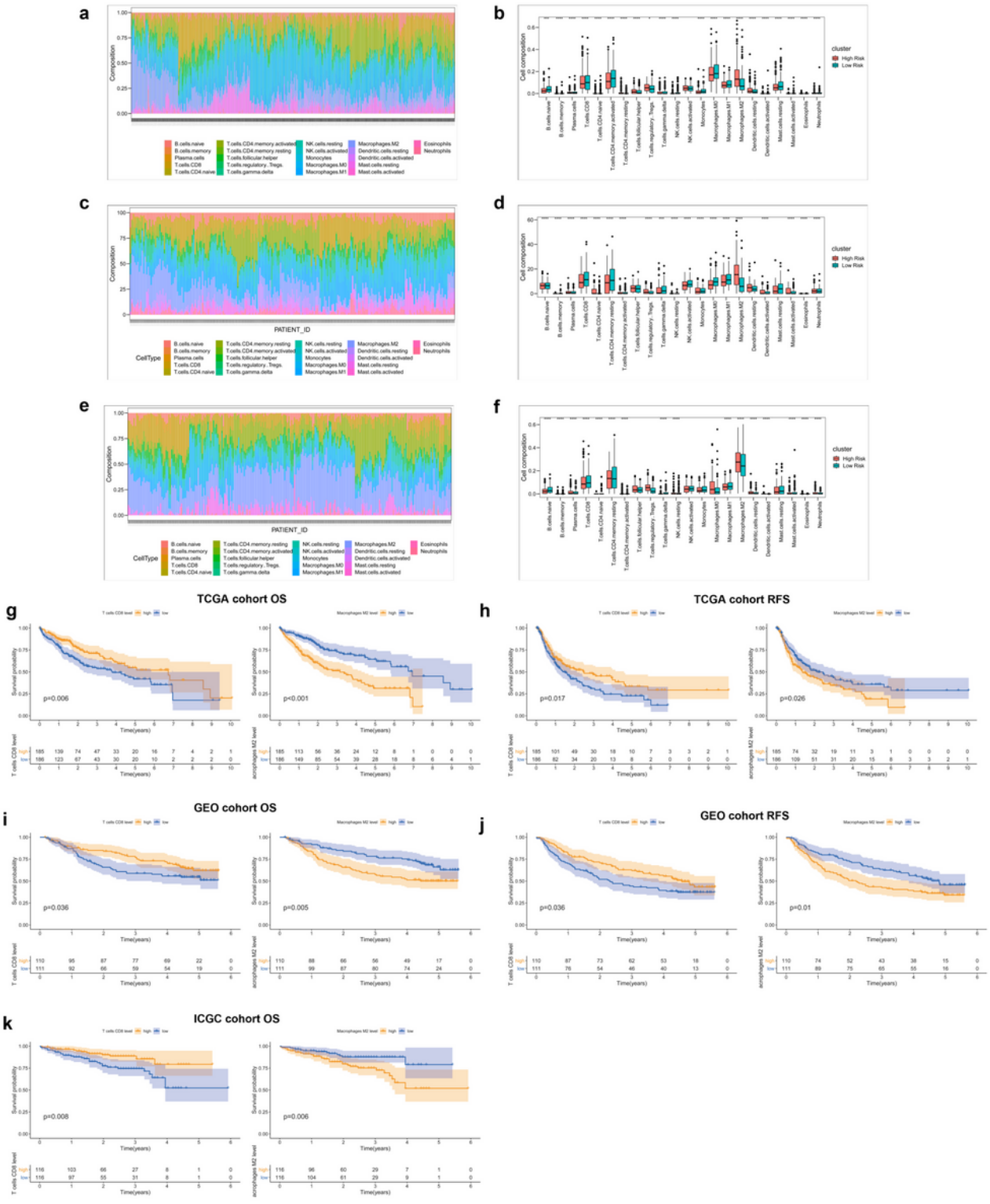

Figure 10

Global inflammatory landscape in HCC patients with high and low-risk groups in TCGA (a), GEO (c), and ICGC (e) cohorts. b The boxplots depict the panoramic distribution of immune cells between the 2 risk subgroups (B, TCGA; D, GEO; F, ICGC), and significance was determined by the Wilcoxon test. d KM-plot

curves revealed that the imbalance of 2 immune cells associated with the OS and RFS status of HCC patients in 3 cohorts, patients were divided based on the median of CD8+ T cells or Macrophage M2.

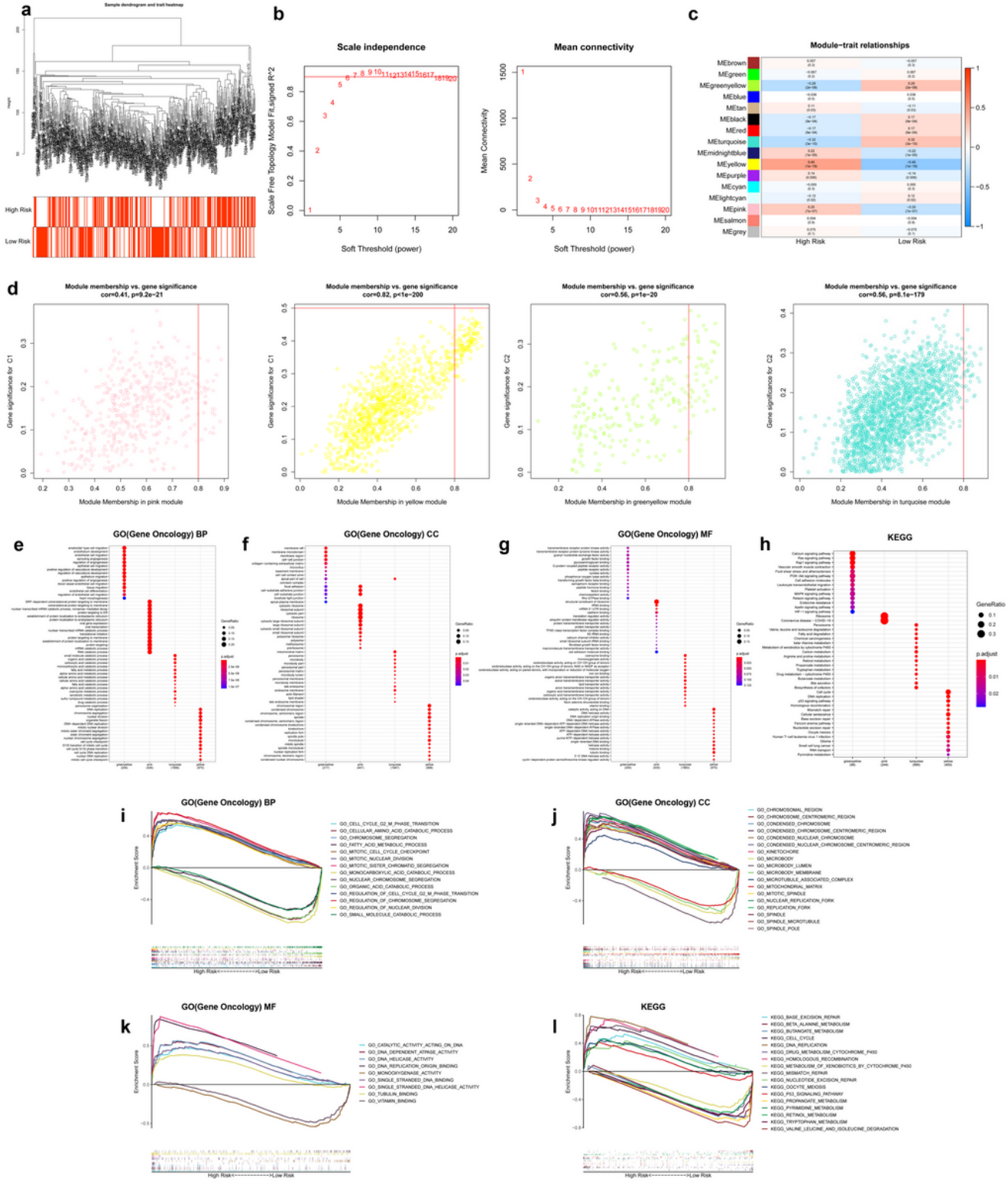

Figure 11

Identification of subtype-specific gene profilings and biological function by WGCNA and GSEA in TCGA cohort. a Clustering dendrograms of HCC samples. b Analysis of the scale-free fit index (left) and average connectivity (right) under various soft threshold powers (β=7). c Correlations between the eigenvector

values of 16 modules and the subtype characteristics. d The scatter diagram of the eigengenes in the top2 modules for each subtype. The heatmap shows the top 15 terms significantly enriched (e) GO(BP), (f) GO(CC), (g) GO(MF), and (h) KEGG for each WGCNA module. (i-l) The multiple GSEA curves describe some important biological terms, which are consistent with the results of WGCNA analysis.

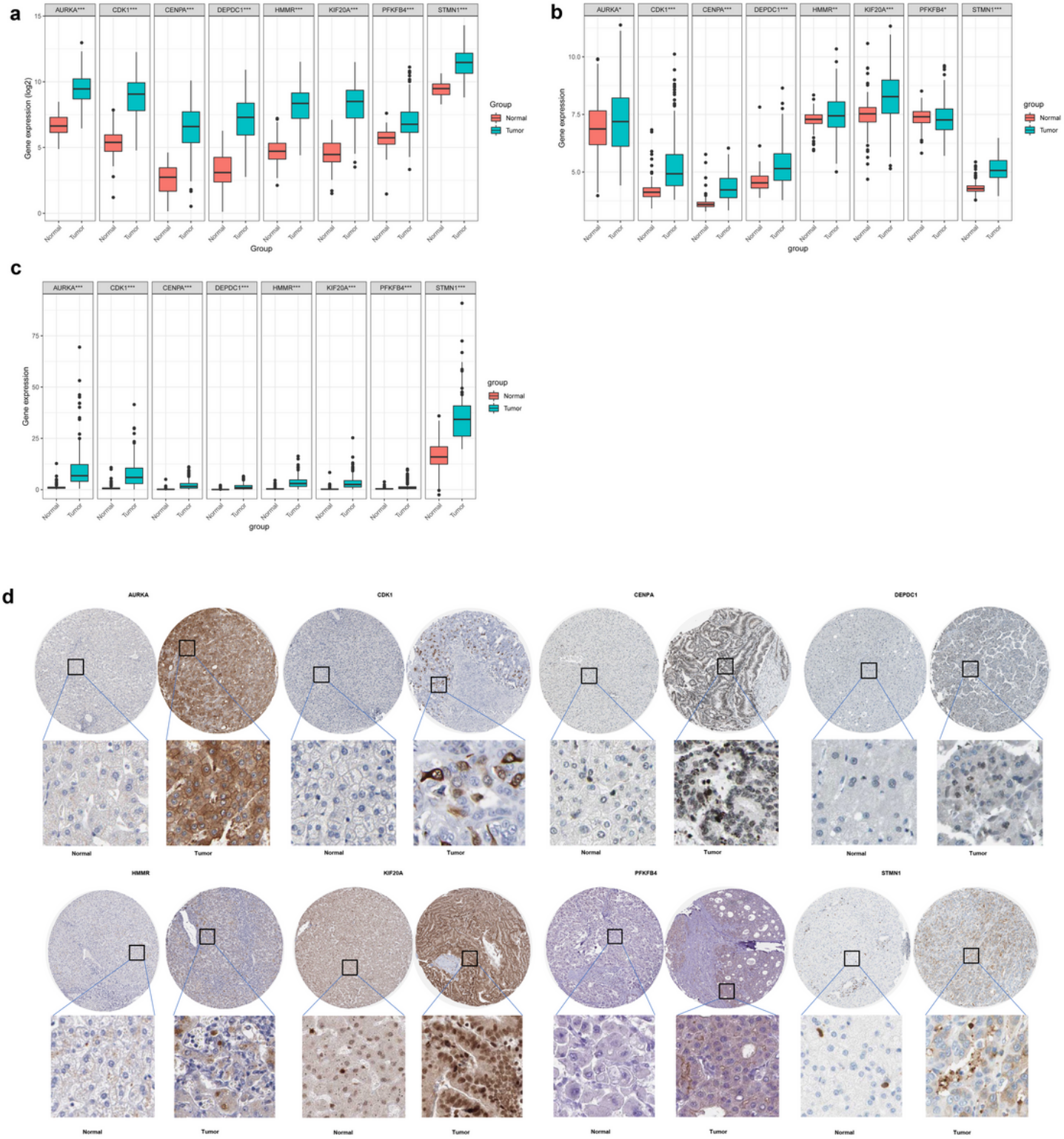

Figure 12

Imbalance of prognostic signature at the gene level and protein level. The expression of three 8 gene signatures in normal and tumor tissues at the transcriptome level in TCGA (a), GEO (b), and ICGC (c) cohorts. (d) Staining intensity of 8 gene signatures in HCC pathological tissue and corresponding normal liver tissue.

Supplementary Files

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

- SupplementalTableS1.docx
- SupplementalTableS2.docx
- SupplementalTableS3.docx
- SupplementalTableS4.docx
- SupplementalTableS5.docx
- FigureS1.tif
- FigureS2.tif
- FigureS3.tif
- FigureS4.tif
- FigureS5.tif
- FigureS6.tif
- FigureS7.tif
- FigureS8.tif
- FigureS9.tif
- FigureS10.tif
- FigureS11.tif
- FigureS12.tif
- FigureS13.tif
- Tableandfigurelegendsforsupplementalmaterials.docx