

Co-amplification of Genes in Chromosome 8q24: a Robust Prognostic Marker in Hepatocellular Carcinoma

Yuan Cheng

Zhujiang Hospital

Yongjian Zheng

Zhujiang Hospital

Cheng Zhang

Zhujiang Hospital

Shunjun Fu

Zhujiang Hospital

Guolin He

Zhujiang Hospital

Lei Cai

Zhujiang Hospital

Ling Qiu

Zhujiang Hospital

Huakun Huang

Zhujiang Hospital

Qunhui Chen

Zhujiang Hospital

Wenzhuan Xie

3D Medicines Inc.

Tingting Chen

3D Medicines Inc.

Mengli Huang

3D Medicines Inc.

Yuezong Bai

3D Medicines Inc.

Mingxin Pan (✉ pmxwxy@sohu.com)

Zhujiang Hospital <https://orcid.org/0000-0002-3979-5571>

Keywords: Hepatocellular carcinoma, 8q24, amplification, TCGA, ICGC, prognostic model

Posted Date: May 24th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-272648/v2>

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

[Read Full License](#)

Co-amplification of genes in chromosome 8q24: a robust prognostic marker in Hepatocellular Carcinoma

Yuan Cheng^{1†}, Yongjian Zheng^{1†}, Cheng Zhang^{1†}, Shunjun Fu¹, Guolin He¹, Lei Cai¹, Ling Qiu², Kunhua Huang¹, Qunhui Chen¹, Wenzhuan Xie³, Tingting Chen³, Mengli Huang³, Yuezong Bai³ and Mingxin Pan^{1*}

*Correspondence: pmxwxy@sohu.com

† Yuan Cheng, Yongjian Zheng and Cheng Zhang contributed equally to this work and share first authorship.

¹ Second Department of Hepatobiliary Surgery, Zhujiang Hospital of Southern Medical University, Guangzhou, China

Full list of author information is available at the end of the article.

Abstract

Background: Hepatocellular carcinoma (HCC) is one of the major cause of cancer related deaths worldwide, due to high 5 year postoperative recurrence rate and individual heterogeneity. Thus, prognostic model has dramatically urgently needed on HCC in recent years. Serval research have reported that copy number amplification of the 8q24 chromosomal region is associated with low survival in many cancers. The objective of this study was to construct a multi-gene model to predict the prognosis of HCC.

Methods: RNAseq and copy number variant (CNV) data of tumor tissue samples of HCC from TCGA (N = 328) was used to identify differentially expressed mRNAs of genes located on chromosomal 8q24 regions by Wilcox test. Univariate Cox and Lasso Cox regression were performed to screen and construct prognostic multi-gene signature in TCGA cohort (N = 119). The multi-gene signature was validated in ICGC cohort (N = 240). A nomogram for prognosis prediction was built and Gene Set Enrichment Analysis (GSEA) was used to further study the underlying molecular mechanisms.

Results: A 7-gene prognosis signature model was established for predicting HCC prognosis. Using the model, we divided individuals into high-risk and low-risk sets with significantly different overall survival in training dataset (HR = 0.17, 95% CI 0.1–0.28; P < 0.001) and in testing dataset (HR = 0.42, 95% CI 0.23–0.74; P = 0.002). Multivariate Cox regression analysis indicated that this signature was an independent prognostic factor of HCC survival. Nomogram including the prognostic signature was constructed and showed better predictive performance in short year (1 and 3 year) than long year (5 year) survival. Furthermore, GSEA revealed several significantly pathways, which may help explain the underlying molecular mechanism.

Conclusions: The 7-gene signature was a reliable prognostic marker in HCC, which may provide meaningful information to therapeutic customization and treatment decision making.

Keywords: Hepatocellular carcinoma, 8q24, amplification, TCGA, ICGC, prognostic model

Background

Hepatocellular carcinoma (HCC) is one of the most shared type in liver malignancies, which is rising in most areas of the world [1]. The HCC prognosis is exceedingly poor, with only 12% HCC

patient survival over 5 years [2-4]. Thus, prognostic model are urged to help predict the outcome HCC patients and to develop an individual management decision. Most previous relevant studies constructed prognostic models using parameters from clinical characteristics and serum tumor markers [5-7]. But with the rapid development of the genome-sequencing technologies, accumulating evidence shown that gene signatures and traditional parameters had immense potential in predicting HCC prognosis. For example, Li et al. established a three-gene predictive signature that could effectively identify HCC patients at a high risk of death [8]. Bai et al. identified a six-miRNA signature as an independent factor to predict recurrence of HCC [9]. Several other reports also constructed messenger RNA (mRNA) expression signatures comprising different numbers of genes using similar methods [10-13].

Accumulated research suggest that multiple gene mutations occurring sequentially in cell lines, at chromosomal and nucleotide levels, are the cause of tumor lesions. Chromosomal DNA amplification is one of the molecular mechanisms capable of changing genes. The over expression of mRNA contributes to occurrence and progression of a variety of tumors [14, 15]. Amplification of the chromosome 8q24 are frequently detected in human cancer, and several studies have been emphasized the significance of this sub-chromosomal region in cancer occurrence and progression, such as prostate [16, 17], gastrointestinal [18, 19] and breast cancer [20]. Besides, the 8q24 chromosome locus is one of the most common amplified genomic region of HCC [18, 21-23].

Although the functional annotation genes in this region are still not fully expounded, several studies have been reported the amplification of some genes located on 8q24 chromosome region separately, such as MYC[24], MAL2[22], BOP1[4], which were associated with the occurrence and progression of HCC in recent year. Through the informatics analysis of the genomic data of 150 HCC tumors, we further confirmed that the increase of 8q24 is the major contributing event in the occurrence and development of HCC [25]. Besides, from our previous study we retrospectively analyzed the genetic testing and clinical follow-up information of 80 HCC patients in clinical. Our results indicated that a gain of 8q24 was significantly associated with poor survival of HCCs ($P < .05$), with the alteration frequency was 22%~30%.

Much attention has been drawn to structure effective and powerful predicting model in recent years for benefits of convenient and predictability, which can provide valuable information to clinical medical decisions. In the present study, we focused on copy number variation (CNV) of genes located on chromosome 8q24 region to conduct a HCC prognostic predicting model. The mRNA expression data of HCC from TCGA cohort were analyzed to identify differential expression genes. Lasso Cox regression analysis were performed to establish a 7-gene signature. The prognostic signature was validated with ICGC cohort, and nomogram combining 7-gene prognostic signature and clinical prognostic factors to predict survival. Gene set enrichment analysis (GSEA) was used to further analyze the potential mechanism of the predictive effect of our model.

Methods

Data collection

The mRNA expression and clinical data were obtained from The Cancer Genome Atlas-Liver Hepatocellular Carcinoma dataset (TCGA-LIHC, <https://portal.gdc.cancer.gov/>), cBioportal for Cancer Genomics and International Cancer Genome Consortium (ICGC, <http://cbioportal.org/>) [26]. All data were obtained from publicly available and open-access databases, thus additional ethical approval were not applicable.

Identification of intersected differentially expressed mRNA in TCGA-LIHC

The raw count of HCC mRNA expression data were standardized using the transcripts per million (TPM) method and underwent a log₂ conversion. 167 protein coding genes located in chromosome 8q24 region were accurately annotated. The differentially expressed mRNA (DEMs) were estimated using the Limma version 3.36.2 R package. The DEMs of datasets with an absolute log₂ fold change (FC) > 1 and an adjusted P value of <0.05 were selected for following analysis (<https://cancerbiomedcentral.com/>).

Construction of the prognostic gene signature

A TCGA cohort of 331 patients was used as a discovery cohort to develop a genetic signature model to predict prognosis. Only patients with initial treatment and an overall survival greater than 1 month were included in the analysis. Patients with two or more genes located on chromosome 8q24 region amplification were divided into amplification group (mutation group), while none or one gene amplification were non-amplification group (wild-type group). The differentially expressed genes (DEGs) were identified by univariate Cox regression analysis. Lasso-penalized Cox regression analysis was designed to further screen prognostic genes for overall survival in HCC patients using the glmnet R package [27]. Finally, a 7-gene signature prognosis model was constructed in TCGA discovery cohort based on a linear combination of the regression coefficient of the lasso Cox regression model coefficients (β) and its mRNA expression level. The prognostic gene signature was presented as risk score = (β_{mRNA1} * expression level of mRNA1) + (β_{mRNA2} * expression level of mRNA2) + ... + (β_{mRNA_n} * expression level of mRNA_n). In order to verify the prediction efficiency of the prediction gene signature, patients from TCGA cohort were 6:4 randomly separated into training dataset and testing dataset by the createDataPartition function in the caret R package. SurvMiner R package was used to investigate the optimal cutoff values. Patients were divided into high-risk group and low-risk group according to the correlation between the signature genes expression levels and patients' OS in the training data set. Kaplan-Meier (KM) survival curves combined with log-rank test were used to analyze the survival difference between high- and low-risk groups by using the survival R package. Time-dependent receiver operating characteristic (ROC) analysis and Concordance index (C-index) were used to assess the predictive performance of prognostic signature on 1-, 3- and 5-year survival. The estimated value of the gene signatures of prognosis was further verified in the testing dataset and TCGA discovery cohort. The mRNA expression level of the 7 genes in prognostic gene signature was analyzed between high-risk group and low-risk group in TCGA discovery cohort using Wilcoxon signed-rank test.

External validation of the prognostic gene signature

A total of 240 HCC patients from the ICGC cohort were used as a separate validation cohort. The risk score of patients was calculated according to the prognostic gene characteristics. The ROC curve and the Kaplan-Meier curve were performed to validate the prognostic gene signature predictive value.

Independence of the prognostic gene signature of other clinical parameters in TCGA

To determine whether the prediction efficiency of the prognostic model could be independent of other clinical parameters [including age, gender, alpha-fetoprotein (AFP), body mass index (BMI),

tumor grade, vascular tumor invasion(MVI), inflammation, residual tumor, tumor mutation burden(TMB), tumor stage], univariate and multivariate analyses were performed by Cox regressive model. P value < 0.05 was considered as statistically significant difference. The hazard ratio (HR) and 95% confidence intervals were further calculated.

Building and validation of a predictive nomogram

Nomograms are commonly applied to predict prognosis of cancer. All independent prognostic factors were selected to evaluate the probability of 1-, 3-, and 5-year OS rate of HCC using multivariable Cox regression method. Validation of the nomogram was assessed by discrimination and calibration. The calibration curve of the nomogram was plotted to test the nomogram prediction probability and observed rates. Decision curve analysis (DCA) was adopted to compare the reliability of the nomogram versus that of tumor stage or risk group.

Gene set enrichment analyses (GSEA)

To reveal the potential molecular mechanism of prognostic prognostic gene signature, GSEA was used between the high-risk and low-risk subgroups of the TCGA discovery cohort using GSEA v.3.0. Molecular Signatures Database v.7.0 was retrieved to identify enrichment pathways related to survival in two risk groups separately. Gene sets with P value < 0.05 and false discovery rate (FDR) < 25% were identified significantly enriched.

Statistical analysis

Statistical analyses were adopted using R software v3.6.0 (R Foundation for Statistical Computing, Vienna, Austria). If not specified as two-tailed tests, p values less than 0.05 were considered statistically significant.

Results

Differentially expressed genes identification

A flow chat of the analysis process was presented in Fig. 1. Totally 109 genes were screened as differentially mRNA expression level between 8q24 amplification group (n=184) and non-amplification group (n=109; Supplementary Table S1).

Construction and validation of the prognostic gene signature

184 primary HCC patients of TCGA discovery cohort were 6:4 randomly separated into a training dataset (n=111) and a testing dataset (n=73). The baseline characteristics of two datasets were summarized in Table 1. None clinical parameters were significantly different in the testing dataset and training dataset. We further narrowed candidate mRNAs using Lasso-penalized Cox analysis. Finally, 7 target genes were identified and then used to construct prognostic gene characteristics. The seven genes included Zinc Finger Protein 7 (ZNF7), Trafficking Protein Particle Complex 9 (TRAPPC9), Transmembrane Protein 65(TMEMP65), POU domain class 5 transcription factor 1B (POU5F1B), DEP Domain Containing MTOR Interacting Protein (DEPTOR), BOP1 Ribosomal Biogenesis Factor (BOP1), Rho GTPase Activating Protein 39 (ARHGAP39), which were identified to be connected with prognosis of HCC patients using the regsubsets function of leaps R package. The risk score= $(-1.465 * Expr_{ZNF7}) + (-1.317 * Expr_{TRAPPC9}) + (0.432 * Expr_{TMEMP65}) + (-0.728 * Expr_{POU5F1B}) + (-0.253 * Expr_{DEPTOR}) + (0.357 * Expr_{BOP1}) + (0.553 * Expr_{ARHGAP39})$. Using the

training set's survival data and the survcutpoint function of the survminer R package, a risk score of 2.074 was defined as the optimal cut-off value to dissect the population into a high-risk group (N = 44) and a low-risk group (N = 169), and this cut-off value was used for all subsequent stratification. Then the training cohort's survival data were used to calculate the optimal cut-off value using the the survcutpoint function of the survminer R package. The risk score of 2.074 was finally defined as the optimal cut-off value, and patients of training dataset were dissected into a high-risk group (N = 44) and a low-risk group (N = 169). All following stratification were based on this optimal cut-off value. Time dependent ROC curve and Kaplan-Meier survival curve were calculated to evaluate the prognostic predicting capacity of this seven-gene signature. The same procedure were employed in the testing dataset and the whole TCGA dataset. The Area under the ROC curves (AUCs) for 1-year, 3-year and 5-year OS were 0.77, 0.741, 0.784 and 0.7, 0.636, 0.633 and 0.739, 0.704, 0.725 for the training dataset (Fig. 2a), the testing dataset (Fig. 2b) and the whole TCGA cohort (Fig. 2c), separately. The OS in high-risk groups of three datasets all significantly poorer than in the low-risk groups (P < 0.001 for the training dataset, P = 0.002 for the testing dataset, and P < 0.001 for the whole TCGA cohort, Fig. 3a-c). This results indicated a moderate sufficient and specificity of the 7-genes signature for prognostic prediction. Among the 368 HCC patients from TCGA cohort, 64 (18%) gained one or more genetic variations of this seven selected genes. Amplification was the most common gene variation type. The mRNA expression of BOP1, TMEM65 and ARHGAP39 were significantly up-regulated, in contrast, TRAPPC9, POU5F1B and DEPTOR were significantly down-regulated compared with wild-type group (Fig. 4a).

External validation of the prognostic gene signature in ICGC

In order to further assess the predictive capacity of seven-gene signature, the prognostic signature were evaluate in the ICGC cohort. A total 240 HCC patients were divided into a high-risk group (n=34) and a low-risk group (n=206) with the same optimal risk cut-off value above. In consistent with the results in ICGC cohort, the OS in the high-risk group was significantly poorer than low-risk group (P < 0.001, Fig. 2d). The AUC for 1 year, 3 year and 5 year OS were 0.836, 0.717, 0.608, respectively (Fig. 3d).

Independent prognostic role of the gene signature

Regression analysis Cox univariate and multivariate were conducted in TCGA cohort to assess 7-gene signature defined methodological risk groups and clinicopathological variables including age, gender, AFP, BMI, inflammation, vascular tumor invasion, residual tumor, tumor grade, tumor stage and tumor mutational burden (TMB) as independent prognosis predictors. The risk group and the tumor stage were independent prognostic factors of the TCGA cohort (risk group, HR: 0.32; 95% CI: 0.2 - 0.52; P < 0.001; tumor stage, HR: 2.13; 95% CI: 1.4 - 3.19; P < 0.001). Details are provided in Table 2.

Building and validating a predictive nomogram

To establish a clinically applicable method for predicting the survival probability of HCC patients, For the purpose of constructing a clinically applicable predicting survival model for HCC patients, a nomogram was built based on stage and risk score (Fig. 5a) [28]. Calibration plots resulted that the nomogram predicting capacity of short-term survival (1- and 3- year) was better than long-term

survival (5- year), which were consistent with the predicted and actual survival rates (Fig. 5b). DCA resulted that the combined predictive model showed the best net benefit for 1 year OS and 3 year OS but not 5-year OS (Fig. 5c).

Gene set enrichment analysis

To reveal the potential molecular mechanism of 7-gene signature, GSEA analysis was performed on the TCGA cohort. As demonstrated in Table 3 and Fig. 6, the tumors of high-risk group defined by 7-gene predicting signature were significantly enriched in the G2M checkpoint ($P < 0.001$, FDR = 0.005), DNA repair ($P = 0.010$, FDR = 0.048), MYC targets ($P = 0.002$, FDR = 0.006) and MTORC1 SIGNALING($P = 0.006$, FDR = 0.028) pathways or terms. And the significantly enriched pathways in the low-risk group were related to several metabolic processes (including xenobiotic metabolism, bile acid, and fatty acid), adipogenesis, coagulation, and peroxisome function.

Discussion

Because of its complex molecular mechanisms and cellular heterogeneity, the morbidity and mortality of HCC are still rising in many countries [11, 25]. The discovery of new prognostic markers and the establishment of more accurate prognostic models are urgent needs to improve the clinical dilemma of HCC. Compared with a single biomarker, the genetic prognostic prediction model constructed with traditional clinical parameters may have better prediction efficiency. In this study, we focused on the amplification variant types of genes locating on 8q24 chromosomal region, which successfully established a novel seven-gene signature based on TCGA cohort and validated in ICGC cohort. The risk of 7-gene signature was an independent prognostic factor for HCC, and the survival rate of patients in the high-risk group was lower than that in the low-risk group. ROC and DCA results showed that the nomogram combining with 7-gene signature combining tumor stage performed the best efficacy for predicting short-term survival (1 and 3 years), but not for long-term survival (5 years) of HCC patients. Altogether, the results demonstrated that this 7-gene signature could be a useful and promising survival predicting indicator for HCC.

Three of genes in the seven-gene signature have been reported to be related to HCC. BOP1 (block of proliferation 1) is a WD40 protein, which was first isolated from the cDNA library screening for growth related sequences in mouse embryonic fibroblasts. BOP1 in promoting HCC cell migratory and invasive activities through the induction of epithelial-mesenchymal transition (EMT), which is a key regulatory program activated during cancer invasion and metastasis [4]. Moreover, BOP1 stimulates actin stress fiber assembly and RhoA activation, and plays a vital role in enhancing cell contractility via actin rearrangement and stress fiber formation [29-31]. DEPTOR (DEP-domain containing mTOR-interacting protein), also known as DEPDC6 (DEP-domain containing protein 6), has become a research hotspot in recent years because of its discovery as a natural negative regulator of mTOR (mammalian Target of Rapamycin). The main component of the protein is mTORC1 and mTORC2 complexes, which inhibit its kinase activity by interacting with mTOR [32, 33]. The deregulation of mTOR is commonly associated with tumor pathogenesis, consistent with inhibitory mTOR activity, the expression levels of DEPTOR was mostly low in most of cancers, and few exceptions such multiple myeloma (MM), thyroid carcinoma or lung cancer [34, 35]. DEPTOR promote EMT and metastatic HCC cells through mTOR inhibition of activated the TGF- β 1-smad3/sm4-smad4-snail pathway [36]. OCT4 (also called as POU5F1) is essential role in the process of embryonic stem cell self-renewal, development and somatic cell reprogramming to become

pluripotent stem cells. OCT4 have six pseudogenes that can be transcribed in many tumors [37, 38]. POU5F1B (POU domain class5 transport factor 1B) is an OCT4 pseudogene that encodes 95% homology to OCT4A. The HCC survival of high expression POU5F1B was shorter than that of low expression, and the AKT phosphorylation level of high POU5F1B expression patients was higher. POU5F1B promoted hepatoma cell proliferation by activating AKT, thus the inhibition of AKT in POU5F1B overexpressed hepatoma cells can inhibit cell proliferation. Moreover, AKT could also be phosphorylated by TSC2 at Ser473 sites, thereby activating Rheb activity stimuli mTORC1, increasing protein synthesis by phosphorylation of eukaryotic translation initiation factor 4E (eIF4E) and ribosomal proteins S6 [39].

The roles of Arhgap39, TRAPPC9, ZNF7 and TMEM65 have not been reported in HCC. As we all known, Rho GTPases participate in cell growth, cell dynamics, apoptosis and intracellular membrane trafficking. Arhgap39 [also called as preoptic regulatory factor-2 (Porf-2) or Vilse] is a multidomain protein containing of WW domain, MyTH4 (myosin tail homolog 4) domain, and RhoGAP domain, which can inactivate Rho GTPases [40]. Studies have shown that Arhgap39 regulate Rac/cdc42 dependent cytoskeleton and neurogenesis in hippocampal neurons of *Drosophila* and mice through WW domains [41]. TRAPPC9 (transport protein particle complex 9), a protein subunit of transport protein particle II (TRAPP II), has been illustrated to play a vital role in the transport of cargo from endoplasmic reticulum (ER) to Golgi body and endosome-to-Golgi in yeast cells [42]. Besides, studies have indicated that TRAPPC9 mutations are associated with an autosomal recessive disorder characterized by mental retardation, autism in mothers, neonatal microcephaly, hearing loss and stroke [43, 44]. A20 (TNF inducible protein 3, TNFAIP3) is a deubiquitinating enzyme with significant functions of anti-inflammatory. Genome-wide studies have illustrated that A20 is a susceptible gene in a variety of human inflammatory and autoimmune diseases, such as systemic lupus erythematosus, psoriasis, rheumatoid arthritis and Crohn's disease [45, 46]. The binding ability of A20's ZnF domains 4 and 7 affects its ability to bind to ubiquitin chains. Notably, ZNF7 in A20 was involved in the regulation of tumor necrosis factor receptor 1 (TNFR1) signaling pathway, which specifically binding to the linear ubiquitin chain [47, 48] and inhibiting TNFR1-mediated apoptosis and death [49]. The ZnF7 can inhibit NF- κ B, and the NF- κ B constitutive activation was associated with tumorigenesis. Therefore, it can be speculated that the dysregulated NF- κ B activation leads to A20 mutation or deletion, which is the main cause of B cell lymphoma [50]. TMEM65 (Transmembrane Protein 65) is an endometrial protein that plays an important role in the function of mitochondrial respiratory chain [19, 51]. The biological function of TMEM65 protein has not yet been fully elucidated. A case reported that a TMEM65 gene mutation in a mitochondrial encephalomyopathy patient, funding the TMEM65 functional changes in mitochondrial function were caused by TMEM65 gene mutations, resulting severe cellular and clinical consequences. Besides, the low expression of TMEM65 in skin fibroblasts affected the content and respiration rate of mitochondrial [52].

In recent years, many studies have constructed effective multi-gene signatures prognostic models of HCC. However, the gene signature predicting prognostic model of HCC based on copy number amplification of genes have not been reported previously. Gains of 8q as the most common change have been reported in 27.7% of pan-cancer arising in 27 different tissues and in 31% to 66% of HCCs [23]. Amplification of DNA in 8q24 chromosome region is thought to be associated with progression of many cancers including HCC, which indicated that the 8q region harbors one or more targeted genes whose amplification promotes them oncogenic.

Herein, our study focused on copy number variant type of genes located in 8q24 chromosome region, constructing a predictive signature based on the expression levels of seven genes. The nomogram combining seven-gene signature and clinical parameters to determine patient's prognosis. These DEMs may be potential molecular targets to fight HCC. Four of the genes in signature have not been reported to be correlated with HCC before, which may be potential research interests or molecular therapeutic target.

However, some certain limitations of our study need to be pointed out. Firstly, the analysis data and clinical data were obtained from TCGA and ICGC databases, which majority races are European or American. Secondly, the seven-gene signature prognostic predicting model showed a better performance in predicting short-term (1 and 3 year survival) but not in long-term (5-year survival). Third, further investigation and experimentation are needed to explore the underlying function mechanisms of predictive signature genes in HCC development and progression.

Conclusions

Overall, our study selected the mRNA expression data of 8q24 genes of HCC patients and focused on the copy number variations of mutation types. A seven-gene signature was successfully constructed to predict the prognosis of HCC. A prognostic nomogram combining the clinical factors to predictive prognosis reliably recognize high-risk HCC patients with poor survival and develop precise individualized therapy.

Abbreviations

HCC: hepatocellular carcinoma; DEMs: differentially expressed mRNA; DEGs: differentially expressed genes; AUC: area under the curve; ROC: receiver operating characteristic; DCA: Decision curve analysis; GSEA: gene set enrichment analysis; OS: overall survival; HR: hazard ratio; CNV: copy number variation; ICGC: international cancer genome consortium; TCGA: the cancer genome atlas.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Publicly available datasets were analyzed in this study. All data can be found here: <https://portal.gdc.cancer.gov/repository>, <https://dcc.icgc.org/projects/LIRI-JP>.

Competing interests

Author Tingting Chen, Wenzhuan Xie, Mengli Huang and Yuezong Bai were employed by the company 3D Medicines Inc. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

None.

Authors' contributions

Conceptualization, Mingxin Pan; Formal analysis, Yuan Cheng, Yongjian Zheng and Cheng Zhang; Investigation, Mengli Huang, Yuezong Bai and Shunjun Fu; Validation, Guolin He, Lei Cai and Ling Qiu; Writing – original draft, Tingting Chen; Writing – review & editing Wenzhuan Xie, Huakun Huang and Qunhui Chen. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

We thank the patients and investigators who participated in TCGA, GEO and ICGC for providing data.

Author details

¹ Second Department of Hepatobiliary Surgery, Zhujiang Hospital of Southern Medical University, Guangzhou, China

² Second Department of Surgery, Dongfeng People's Hospital, Guangzhou, China

³ The Research and Development Center of Precision Medicine, 3D Medicines Inc., Shanghai, China

References

1. Kato H, Shibata T, Kokubu A, Ojima H, Loukopoulos P, Kanai Y, Kosuge T, Fukayama M, Kondo T, Sakamoto M, Hosoda F, Ohki M, Imoto I, Inazawa J, Hirohashi S. Genetic profile of hepatocellular carcinoma revealed by array-based comparative genomic hybridization: identification of genetic indicators to predict patient outcome. *J Hepatol* 2005, 43(5):863-874.
2. Llovet JM, Montal R, Sia D, Finn RS. Molecular therapies and precision medicine for hepatocellular carcinoma. *Nat Rev Clin Oncol* 2018, 15(10):599-616.
3. Villanueva A. Hepatocellular Carcinoma. *N Engl J Med* 2019, 380(15):1450-1462.
4. Chung KY, Cheng IK, Ching AK, Chu JH, Lai PB, Wong N. Block of proliferation 1 (BOP1) plays an oncogenic role in hepatocellular carcinoma by promoting epithelial-to-mesenchymal transition. *Hepatology* 2011, 54(1):307-318.
5. Budhu A, Forgues M, Ye QH, Jia HL, He P, Zanetti KA, Kammula US, Chen Y, Qin LX, Tang ZY, Wang XW. Prediction of venous metastases, recurrence, and prognosis in hepatocellular carcinoma based on a unique immune response signature of the liver microenvironment. *Cancer Cell* 2006, 10(2):99-111.
6. He YZ, He K, Huang RQ, Wang ZL, Ye SW, Liu LW, Luo QJ, Hu ZM. Preoperative evaluation and prediction of clinical scores for hepatocellular carcinoma microvascular invasion: a single-center retrospective analysis. *Ann Hepatol* 2020.
7. Yoo J, Lee MW, Lee DH, Lee JH, Han JK. Evaluation of a serum tumour marker-based recurrence prediction model after radiofrequency ablation for hepatocellular carcinoma. *Liver Int* 2020, 40(5):1189-1200.
8. Li B, Feng W, Luo O, Xu T, Cao Y, Wu H, Yu D, Ding Y. Development and Validation of a Three-gene Prognostic Signature for Patients with Hepatocellular Carcinoma. *Sci Rep* 2017, 7(1):5517.

9. Bai F, Zhou H, Ma M, Guan C, Lyu J, Meng QH. A novel RNA sequencing-based miRNA signature predicts with recurrence and outcome of hepatocellular carcinoma. *Mol Oncol* 2018, 12(7):1125-1137.
10. Chen PF, Li QH, Zeng LR, Yang XY, Peng PL, He JH, Fan B. A 4-gene prognostic signature predicting survival in hepatocellular carcinoma. *J Cell Biochem* 2019, 120(6):9117-9124.
11. Ke K, Chen G, Cai Z, Huang Y, Zhao B, Wang Y, Liao N, Liu X, Li Z, Liu J. Evaluation and prediction of hepatocellular carcinoma prognosis based on molecular classification. *Cancer Manag Res* 2018, 10:5291-5302.
12. Qiao GJ, Chen L, Wu JC, Li ZR. Identification of an eight-gene signature for survival prediction for patients with hepatocellular carcinoma based on integrated bioinformatics analysis. *PeerJ* 2019, 7:e6548.
13. Wang Z, Teng D, Li Y, Hu Z, Liu L, Zheng H. A six-gene-based prognostic signature for hepatocellular carcinoma overall survival prediction. *Life Sci* 2018, 203:83-91.
14. Brisbin AG, Asmann YW, Song H, Tsai YY, Aakre JA, Yang P, Jenkins RB, Pharoah P, Schumacher F, Conti DV, Duggan DJ, Jenkins M, Hopper J, Gallinger S, Newcomb P, Casey G, Sellers TA, Fridley BL. Meta-analysis of 8q24 for seven cancers reveals a locus between NOV and ENPP2 associated with cancer development. *BMC Med Genet* 2011, 12:156.
15. Wang XC, Wu YP, Ye B, Lin DC, Feng YB, Zhang ZQ, Xu X, Han YL, Cai Y, Dong JT, Zhan QM, Wu M, Wang MR. Suppression of anoikis by SKP2 amplification and overexpression promotes metastasis of esophageal squamous cell carcinoma. *Mol Cancer Res* 2009, 7(1):12-22.
16. Fromont G, Godet J, Peyret A, Irani J, Celhay O, Rozet F, Cathelineau X, Cussenot O. 8q24 amplification is associated with Myc expression and prostate cancer progression and is an independent predictor of recurrence after radical prostatectomy. *Hum Pathol* 2013, 44(8):1617-1623.
17. Ciampa J, Yeager M, Amundadottir L, Jacobs K, Kraft P, Chung C, Wacholder S, Yu K, Wheeler W, Thun MJ, Divers WR, Gapstur S, Albanes D, Virtamo J, Weinstein S, Giovannucci E, Willett WC, Cancel-Tassin G, Cussenot O, Valeri A, Hunter D, Hoover R, Thomas G, Chanock S, Chatterjee N. Large-scale exploration of gene-gene interactions in prostate cancer using a multistage genome-wide association study. *Cancer Res* 2011, 71(9):3287-3295.
18. White NM, Maher CA. The potential use of lncRNAs found in the 8q24 region as biomarkers for colon cancer. *Ann Oncol* 2017, 28(8):1688-1689.
19. Sharma P, Abbasi C, Lazic S, Teng ACT, Wang D, Dubois N, Ignatchenko V, Wong V, Liu J, Araki T, Tiburcy M, Ackerley C, Zimmermann WH, Hamilton R, Sun Y, Liu PP, Keller G, Stajlar I, Scott IC, Kislinger T, Gramolini AO. Evolutionarily conserved intercalated disc protein Tmem65 regulates cardiac conduction and connexin 43 function. *Nat Commun* 2015, 6:8391.
20. Meng F, Liu B, Xie G, Song Y, Zheng X, Qian X, Li S, Jia H, Zhang X, Zhang L, Yang YL, Fu L. Amplification and overexpression of PSCA at 8q24 in invasive micropapillary carcinoma of breast. *Breast Cancer Res Treat* 2017, 166(2):383-392.
21. Zhao XM, Xiang ZL, Chen YX, Yang P, Hu Y, Zeng ZC. A sequence polymorphism on 8q24 is associated with survival in hepatocellular carcinoma patients who received radiation therapy. *Sci Rep* 2018, 8(1):2264.
22. Lopez-Coral A, Del Vecchio GJ, Chahine JJ, Kallakury BV, Tuma PL. MAL2-Induced Actin-

- Based Protrusion Formation is Anti-Oncogenic in Hepatocellular Carcinoma. *Cancers (Basel)* 2020, 12(2).
23. Okamoto H, Yasui K, Zhao C, Arai S, Inazawa J. PTK2 and EIF3S3 genes may be amplification targets at 8q23-q24 and are associated with large hepatocellular carcinomas. *Hepatology* 2003, 38(5):1242-1249.
 24. Wu M, Zhou Y, Fei C, Chen T, Yin X, Zhang L, Ren Z. ID1 overexpression promotes HCC progression by amplifying the AURKA/Myc signaling pathway. *Int J Oncol* 2020, 57(3):845-857.
 25. Poon TC, Wong N, Lai PB, Rattray M, Johnson PJ, Sung JJ. A tumor progression model for hepatocellular carcinoma: bioinformatic analysis of genomic data. *Gastroenterology* 2006, 131(4):1262-1270.
 26. Liu GM, Xie WX, Zhang CY, Xu JW. Identification of a four-gene metabolic signature predicting overall survival for hepatocellular carcinoma. *J Cell Physiol* 2020, 235(2):1624-1636.
 27. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010, 33(1):1-22.
 28. Long J, Zhang L, Wan X, Lin J, Bai Y, Xu W, Xiong J, Zhao H. A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma. *J Cell Mol Med* 2018, 22(12):5928-5938.
 29. Narumiya S, Tanji M, Ishizaki T. Rho signaling, ROCK and mDia1, in transformation, metastasis and invasion. *Cancer Metastasis Rev* 2009, 28(1-2):65-76.
 30. Hall A. The cytoskeleton and cancer. *Cancer Metastasis Rev* 2009, 28(1-2):5-14.
 31. Etienne-Manneville S, Hall A. Rho GTPases in cell biology. *Nature* 2002, 420(6916):629-635.
 32. Peterson TR, Laplante M, Thoreen CC, Sancak Y, Kang SA, Kuehl WM, Gray NS, Sabatini DM. DEPTOR is an mTOR inhibitor frequently overexpressed in multiple myeloma cells and required for their survival. *Cell* 2009, 137(5):873-886.
 33. Caron A, Briscoe DM, Richard D, Laplante M. DEPTOR at the Nexus of Cancer, Metabolism, and Immunity. *Physiol Rev* 2018, 98(3):1765-1803.
 34. Catena V, Fanciulli M. Deptor: not only a mTOR inhibitor. *J Exp Clin Cancer Res* 2017, 36(1):12.
 35. Yao H, Tang H, Zhang Y, Zhang QF, Liu XY, Liu YT, Gu WT, Zheng YZ, Shang HB, Wang Y, Huang JY, Wei YX, Zhang X, Zhang J, Wu ZB. DEPTOR inhibits cell proliferation and confers sensitivity to dopamine agonist in pituitary adenoma. *Cancer Lett* 2019, 459:135-144.
 36. Chen J, Zhu H, Liu Q, Ning D, Zhang Z, Zhang L, Mo J, Du P, Liu X, Song S, Fan Y, Liang H, Liu J, Zhang B, Chen X. DEPTOR induces a partial epithelial-to-mesenchymal transition and metastasis via autocrine TGFbeta1 signaling and is associated with poor prognosis in hepatocellular carcinoma. *J Exp Clin Cancer Res* 2019, 38(1):273.
 37. Hayashi H, Arao T, Togashi Y, Kato H, Fujita Y, De Velasco MA, Kimura H, Matsumoto K, Tanaka K, Okamoto I, Ito A, Yamada Y, Nakagawa K, Nishio K. The OCT4 pseudogene POU5F1B is amplified and promotes an aggressive phenotype in gastric cancer. *Oncogene* 2015, 34(2):199-208.
 38. Yi J, Zhou LY, Yi YY, Zhu X, Su XY, Zhao Q, Lin J, Qian J, Deng ZQ. Low Expression of Pseudogene POU5F1B Affects Diagnosis and Prognosis in Acute Myeloid Leukemia (AML). *Med Sci Monit* 2019, 25:4952-4959.
 39. Pan Y, Zhan L, Chen L, Zhang H, Sun C, Xing C. POU5F1B promotes hepatocellular carcinoma

- proliferation by activating AKT. *Biomed Pharmacother* 2018, 100:374-380.
40. Nowak FV. Porf-2 = Arhgap39 = Vilse: A Pivotal Role in Neurodevelopment, Learning and Memory. *eNeuro* 2018, 5(5).
 41. Lee JY, Lee LJ, Fan CC, Chang HC, Shih HA, Min MY, Chang MS. Important roles of Vilse in dendritic architecture and synaptic plasticity. *Sci Rep* 2017, 7:45646.
 42. Mbimba T, Hussein NJ, Najeed A, Safadi FF. TRAPPC9: Novel insights into its trafficking and signaling pathways in health and disease (Review). *Int J Mol Med* 2018, 42(6):2991-2997.
 43. Dookeran KA, Zhang W, Stayner L, Argos M. Associations of two-pore domain potassium channels and triple negative breast cancer subtype in The Cancer Genome Atlas: systematic evaluation of gene expression and methylation. *BMC Res Notes* 2017, 10(1):475.
 44. Mohamoud HS, Ahmed S, Jelani M, Alrayes N, Childs K, Vadgama N, Almramhi MM, Al-Aama JY, Goodbourn S, Nasir J. A missense mutation in TRAPPC6A leads to build-up of the protein, in patients with a neurodevelopmental syndrome and dysmorphic features. *Sci Rep* 2018, 8(1):2053.
 45. Skaug B, Chen J, Du F, He J, Ma A, Chen ZJ. Direct, noncatalytic mechanism of IKK inhibition by A20. *Mol Cell* 2011, 44(4):559-571.
 46. Martens A, Priem D, Hoste E, Vettters J, Rennen S, Catrysse L, Voet S, Deelen L, Sze M, Vikkula H, Slowicka K, Hochepped T, Iliaki K, Wullaert A, Janssens S, Lamkanfi M, Beyaert R, Armaka M, Bertrand MJM, van Loo G. Two distinct ubiquitin-binding motifs in A20 mediate its anti-inflammatory and cell-protective activities. *Nat Immunol* 2020, 21(4):381-387.
 47. Wertz IE, Newton K, Seshasayee D, Kusam S, Lam C, Zhang J, Popovych N, Helgason E, Schoeffler A, Jeet S, Ramamoorthi N, Kategaya L, Newman RJ, Horikawa K, Dugger D, Sandoval W, Mukund S, Zindal A, Martin F, Quan C, Tom J, Fairbrother WJ, Townsend M, Warming S, DeVoss J, Liu J, Dueber E, Caplazi P, Lee WP, Goodnow CC, Balazs M, Yu K, Kolumam G, Dixit VM. Phosphorylation and linear ubiquitin direct A20 inhibition of inflammation. *Nature* 2015, 528(7582):370-375.
 48. Draber P, Kupka S, Reichert M, Draberova H, Lafont E, de Miguel D, Spilgies L, Surinova S, Taraborrelli L, Hartwig T, Rieser E, Martino L, Rittinger K, Walczak H. LUBAC-Recruited CYLD and A20 Regulate Gene Activation and Cell Death by Exerting Opposing Effects on Linear Ubiquitin in Signaling Complexes. *Cell Rep* 2015, 13(10):2258-2272.
 49. Yamaguchi N, Yamaguchi N. The seventh zinc finger motif of A20 is required for the suppression of TNF-alpha-induced apoptosis. *FEBS Lett* 2015, 589(12):1369-1375.
 50. Karin M. Nuclear factor-kappaB in cancer development and progression. *Nature* 2006, 441(7092):431-436.
 51. Nishimura N, Gotoh T, Oike Y, Yano M. TMEM65 is a mitochondrial inner-membrane protein. *PeerJ* 2014, 2:e349.
 52. Nazli A, Safdar A, Saleem A, Akhtar M, Brady LI, Schwartzentruber J, Tarnopolsky MA. A mutation in the TMEM65 gene results in mitochondrial myopathy with severe neurological manifestations. *Eur J Hum Genet* 2017, 25(6):744-751.

Tables

Table 1 The relationship between clinical characteristic and overall survival of TCGA cohort.

	Testing set	Training set	p value
N	73	111	
Age (mean (SD))	58.55 (13.53)	59.79 (10.45)	0.483
Gender = male (%)	56 (76.7)	82 (73.9)	0.794
Race (%)			0.434
Asian	41 (56.2)	61 (55.0)	
Black or African American	4 (5.5)	4 (3.6)	
Not reported	4 (5.5)	2 (1.8)	
White	24 (32.9)	44 (39.6)	
AFP (mean (SD))	6296.75 (41597.65)	32503.79 (220471.63)	0.385
BMI (mean (SD))	25.46 (6.76)	25.36 (5.92)	0.924
Inflammation (%)			0.098
Mild	23 (31.5)	47 (42.3)	
None	28 (38.4)	25 (22.5)	
Severe	19 (26.0)	30 (27.0)	
Tumor Grade (%)	3 (4.1)	9 (8.1)	0.19
Grade 1	1 (1.4)	0 (0.0)	
Grade 2	10 (13.7)	10 (9.0)	
Grade 3	29 (39.7)	56 (50.5)	
Grade 4	31 (42.5)	37 (33.3)	
Grade 4	2 (2.7)	8 (7.2)	
Tumor stage (%)			0.302
not reported	4 (5.5)	4 (3.6)	
Stage I	30 (41.1)	61 (55.0)	
Stage II	23 (31.5)	22 (19.8)	
Stage IIIA	15 (20.5)	20 (18.0)	
Stage IIIB	0 (0.0)	2 (1.8)	
Stage IIIC	1 (1.4)	2 (1.8)	
Residual tumor (%)			0.928
R0	1 (1.4)	2 (1.8)	
R1	67 (91.8)	99 (89.2)	
RX	3 (4.1)	5 (4.5)	
RX	2 (2.7)	5 (4.5)	
Vascular tumor invasion (%)			0.684
Macro	12 (16.4)	19 (17.1)	
Micro	3 (4.1)	3 (2.7)	
None	22 (30.1)	26 (23.4)	
None	36 (49.3)	63 (56.8)	
TMB (mean (SD))	6.46 (4.41)	6.51 (6.31)	0.953

Table 2 Univariate and multivariate Cox regression analyses of gene signature and overall survival of HCC in TCGA cohort.

Characteristics	Hazard. Ratio	CI 95	P. value	Hazard. Ratio	CI 95	P. value
AFP(>=400 vs <400)	1.1062	0.67-1.84	0.697			
Age(>=60 vs <60)	1.1606	0.81-1.67	0.419			
BMI(>=25 vs <25)	0.8099	0.55-1.19	0.285			
Gender(Male vs Female)	0.7974	0.55-1.15	0.226			
Residual_tumor(R1/2 vs R0)	1.8877	0.95-3.74	0.068			
Risk(Low vs High)	0.2415	0.16-0.37	1.82E-10	0.3225	0.20-0.52	3.78E-06
TMB(>= 10 vs <10)	1.5874	0.97-2.59	0.0643			
Tumor grade(G3/4 vs G1/2)	1.0238	0.70-1.49	0.903			
Tumor stage(III/IV vs I/II)	2.6608	1.81-3.91	5.71E-07	2.1295	1.42-3.19	0.000245
Vascular tumor invasion(Macro/Micro vs None)	1.5169	0.99-2.33	0.057			
Inflammation(Mild/Severe vs None)	1.2809	0.76-2.12	0.334			

Table 3 Significantly enriched hallmarks in the TCGA cohort by GSEA.

NAME	SIZE	NES	NOM p-val	FDR q-val	Signals Score
HALLMARK_BILE_ACID_METABOLISM	112	2.35	0.00E+00	7.00E-04	Low
HALLMARK_XENOBIOTIC_METABOLISM	200	2.24	0.00E+00	7.00E-04	Low
HALLMARK_COAGULATION	138	2.05	0.00E+00	7.19E-03	Low
HALLMARK_FATTY_ACID_METABOLISM	158	2.03	1.84E-03	6.52E-03	Low
HALLMARK_PEROXISOME	103	1.94	1.89E-03	1.45E-02	Low
HALLMARK_ADIPOGENESIS	200	1.93	3.74E-03	1.24E-02	Low
HALLMARK_HEME_METABOLISM	197	1.62	8.13E-03	1.06E-01	Low
HALLMARK_G2M_CHECKPOINT	197	-2.27	0.00E+00	4.52E-03	High
HALLMARK_E2F_TARGETS	200	-2.14	0.00E+00	5.00E-03	High
HALLMARK_MYC_TARGETS_V1	199	-2.09	2.04E-03	6.13E-03	High
HALLMARK_MTORC1_SIGNALING	200	-1.94	6.21E-03	2.81E-02	High
HALLMARK_MYC_TARGETS_V2	58	-1.87	2.08E-02	4.07E-02	High
HALLMARK_DNA_REPAIR	149	-1.83	9.94E-03	4.81E-02	High
HALLMARK_MITOTIC_SPINDLE	198	-1.69	3.41E-02	1.09E-01	High
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	112	-1.68	2.29E-02	1.05E-01	High
HALLMARK_UV_RESPONSE_UP	158	-1.62	2.52E-03	1.32E-01	High
HALLMARK_SPERMATOGENESIS	133	-1.54	2.91E-02	1.79E-01	High

Figure legends

Fig. 1 Flow chart of multi-gene signature identification and validation.

Fig. 2 Time dependent ROC analysis of the 7-gene signature for prediction overall survival at 1-, 3-, and 5- years in the training set **(a)**, the testing set **(b)**, the TCGA discovery cohort **(c)** and the ICGC validation cohort **(d)**.

Fig. 3 Kaplan-Meier analysis to compare patients between high-risk and low-risk groups based on 7-gene signature in the training set **(a)**, the testing set **(b)**, the TCGA discovery cohort **(c)** and **(D)** the ICGC validation cohort.

Fig. 4 The mRNA expression level of seven genes between high-risk and low-risk groups of HCC in TCGA cohort and ICGC cohort.

Fig. 5 The nomogram to predict probability of overall survival (OS) for HCC in TCGA cohort. The nomogram for predicting the proportion of patients with OS **(a)**. The calibration plot of the nomogram for the probability of OS at 1-, 3- and 5- years **(b)**. The DCA curve of the nomograms compared for 1-, 3-, and 5- year OS, respectively **(c)**.

Fig. 6 The significantly enriched KEGG pathways in TCGA by GSEA. Ten representative KEGG pathways in high-risk patients **(a)**. Seven representative KEGG pathways in low-risk patients **(b)**.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. The relationship between clinical characteristic and overall survival of TCGA cohort.

Figures

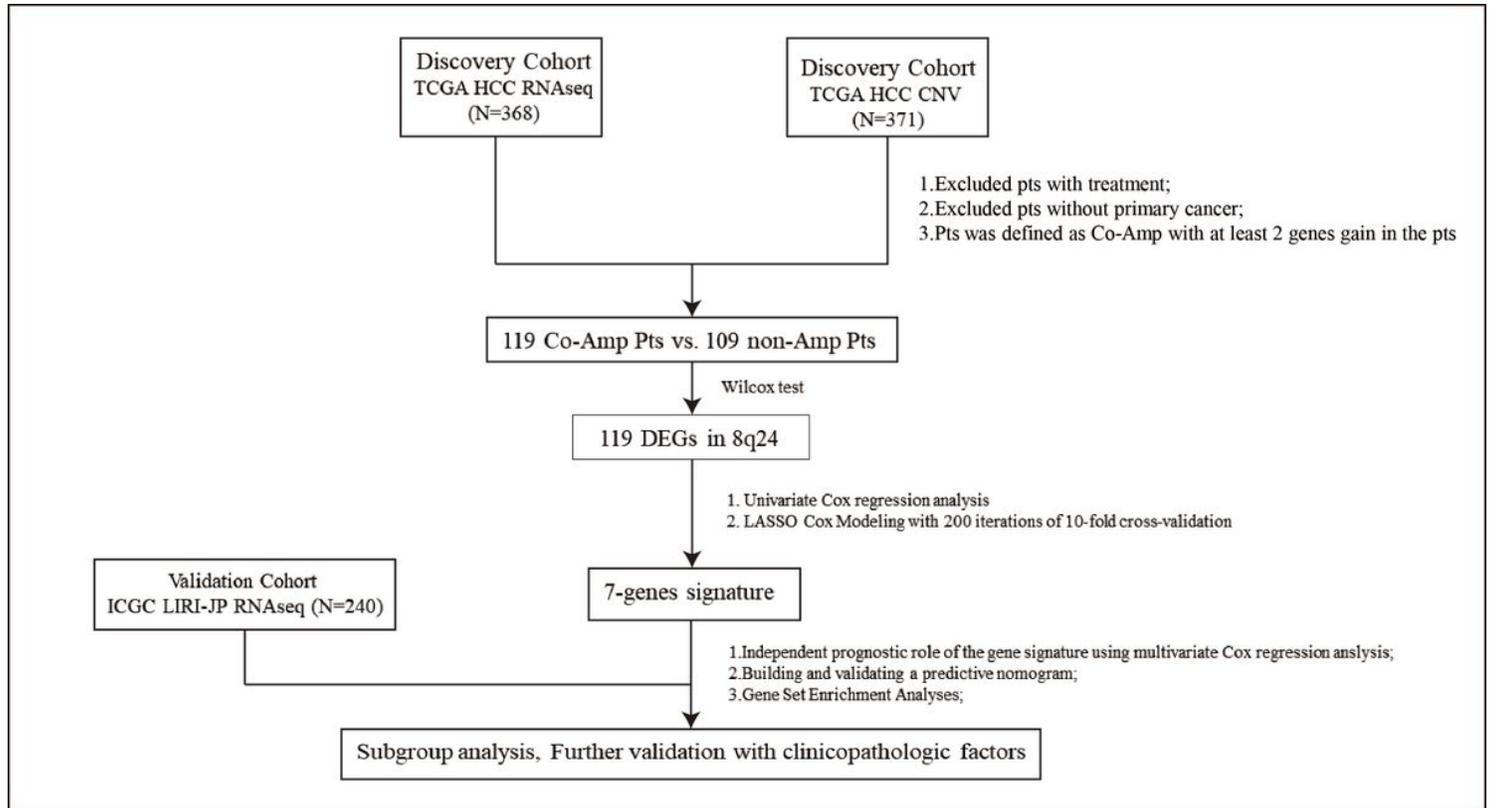


Figure 1

Flow chart of multi gene signature identification and validation

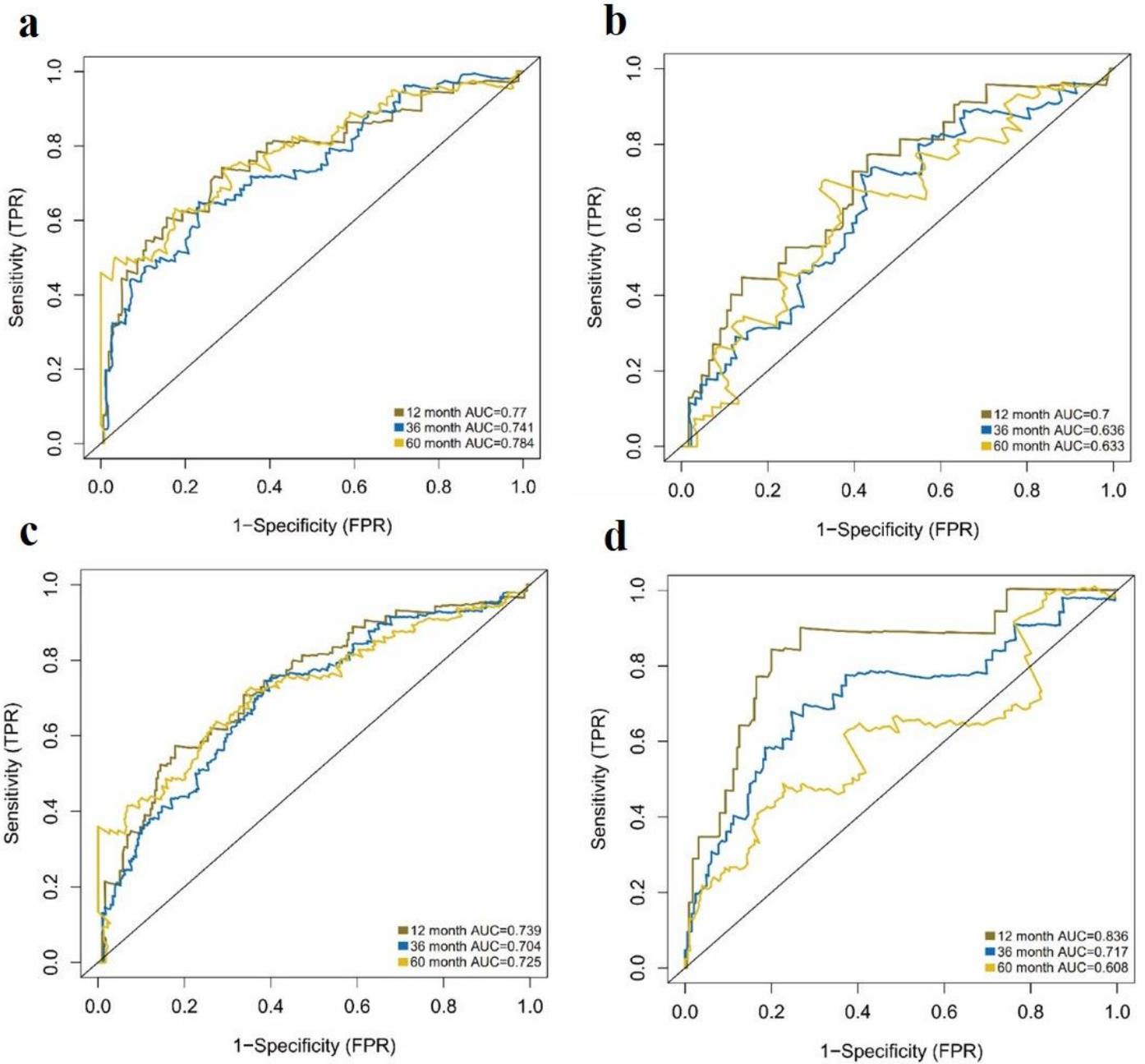


Figure 2

Time dependent ROC analysis of the 7 gene signature for prediction overall survival at 1, 3-, and 5- years in the training set (a), the testing set (b), the TCGA discovery cohort (c) and the ICGC validation cohort (d).

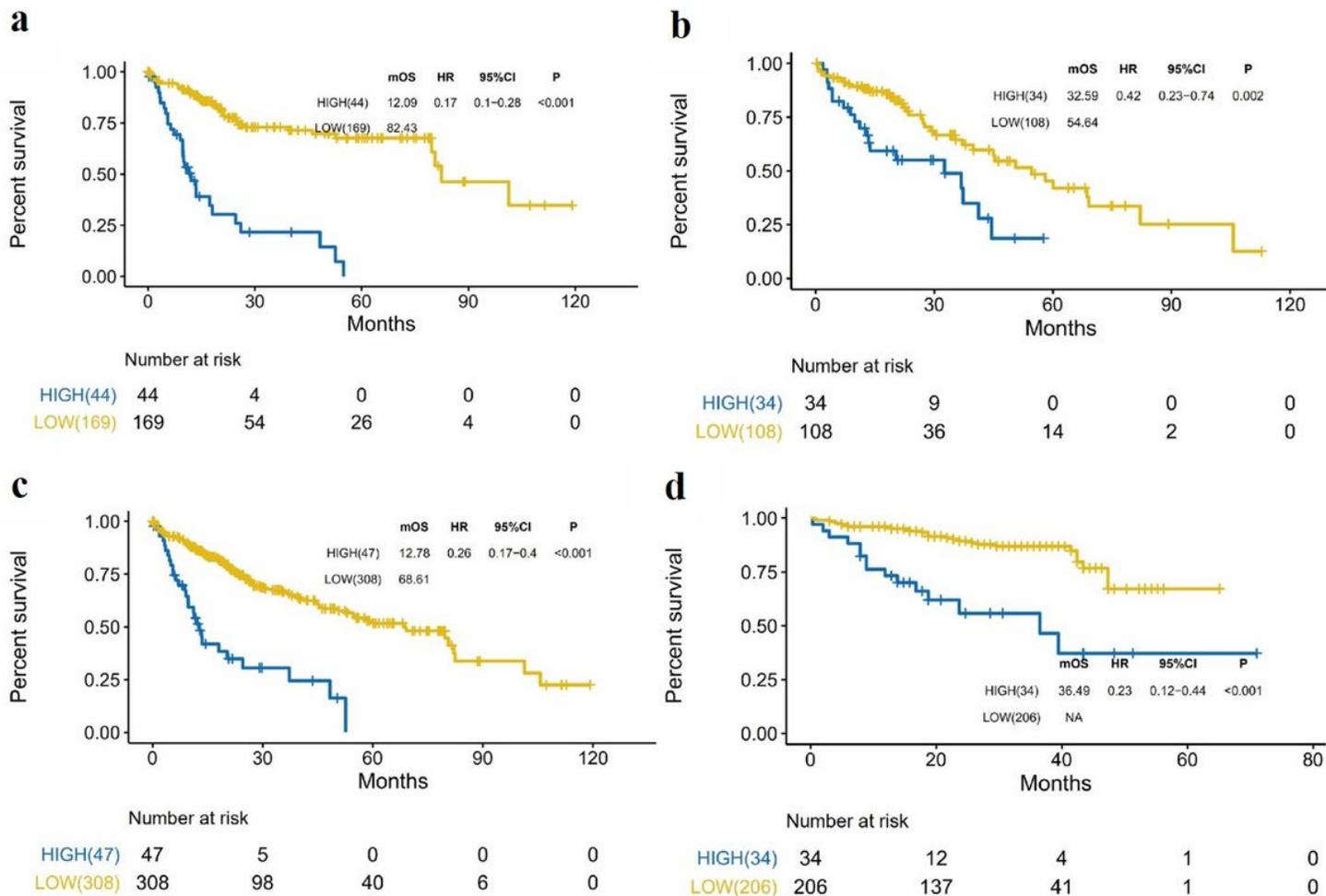


Figure 3

Kaplan Meier analysis to compare patients between high risk and low risk groups based on 7 gene signature in the training set (a), the testing set (b), the TCGA discovery cohort (c) and (D) the ICGC validation cohort.

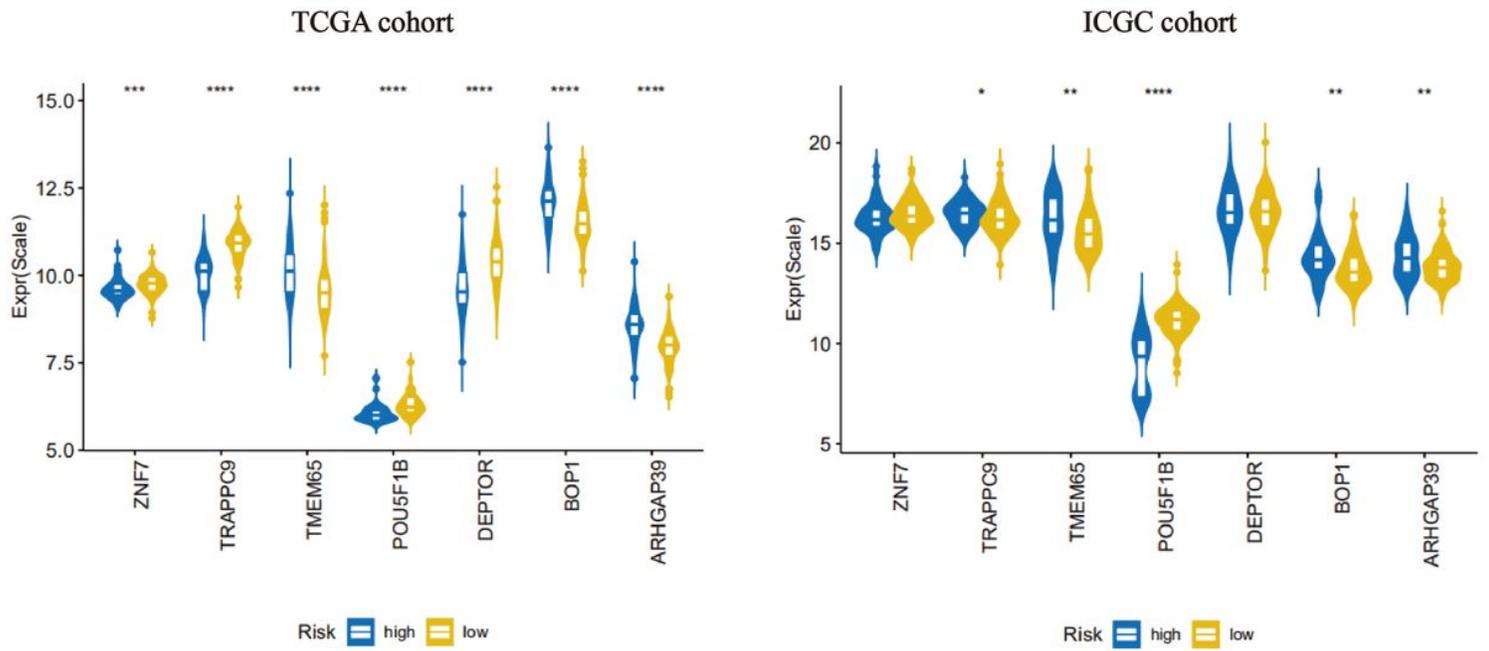


Figure 4

The mRNA expression level of seven genes between high risk and low risk groups of HCC in TCGA cohort and ICGC cohort.

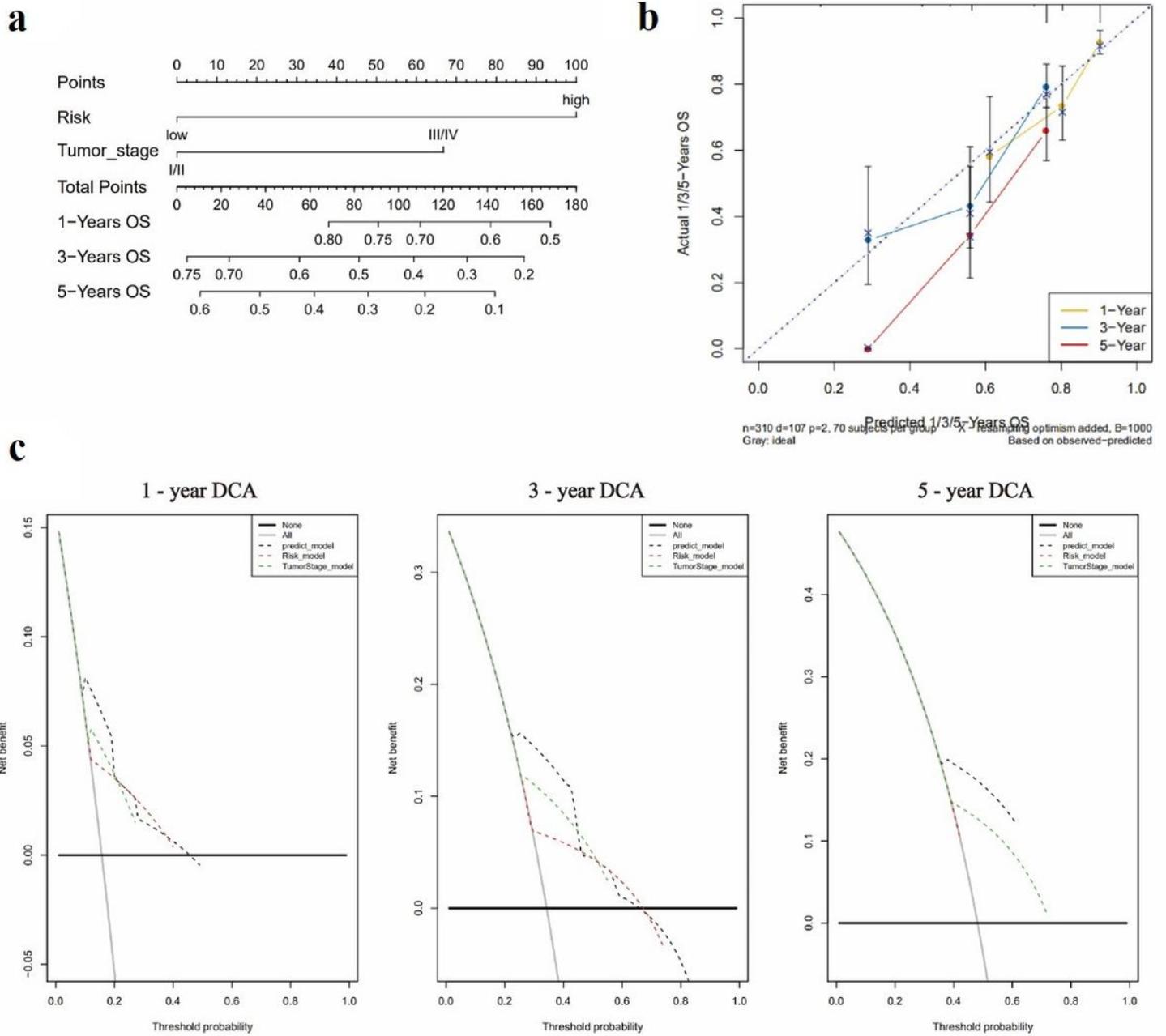


Figure 5

The nomogram to predict probability of overall survival (OS) for HCC in TCGA cohort. The nomogram for predicting the proportion of patients with OS (a). The calibration plot of the nomogram for the probability of OS at 1-, 3- and 5- years (b). The DCA curve of the nomograms compared for 1-, 3-, and 5- year OS, respectively (c).

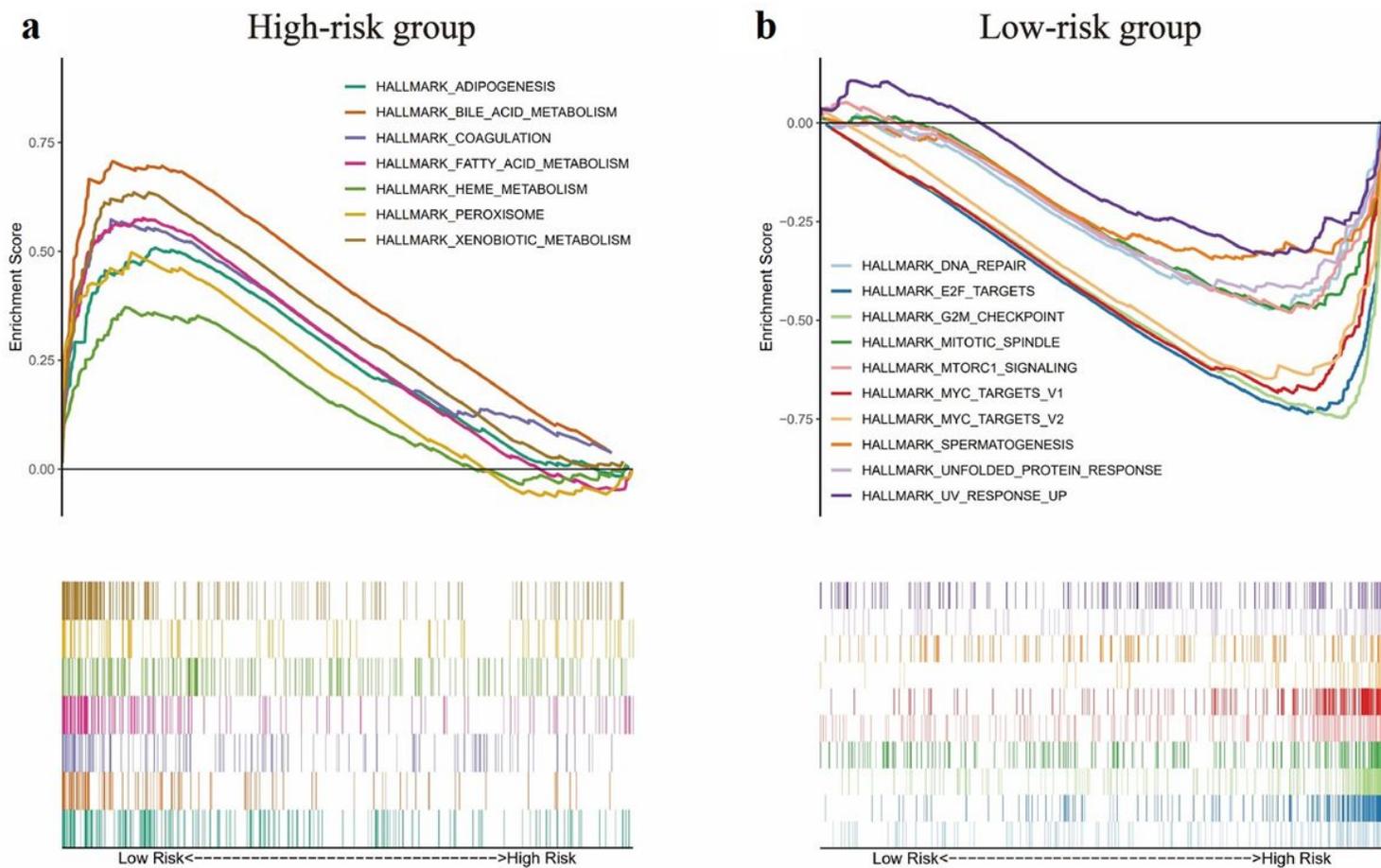


Figure 6

The significantly enriched KEGG pathways in TCGA by GSEA. Ten representative KEGG pathways in high risk patients (a). Seven representative KEGG pathways in low risk patients (b).

Supplementary Files

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

- [SupportingInformation.docx](#)