Identification of a six-gene signature predicting overall survival for hepatocellular carcinoma

Gao-Min Liu*[iD], Hua-Dong Zeng, Cai-Yun Zhang and Ji-Wei Xu*

Abstract

Background: Hepatocellular carcinoma (HCC) remains a major challenge for public health worldwide. Considering the great heterogeneity of HCC, more accurate prognostic models are urgently needed. To identify a robust prognostic gene signature, we conduct this study.

Materials and methods: Level 3 mRNA expression profiles and clinicopathological data were obtained in The Cancer Genome Atlas Liver Hepatocellular Carcinoma (TCGA-LIHC). GSE14520 dataset from the gene expression omnibus (GEO) database was downloaded to further validate the results in TCGA. Differentially expressed mRNAs between HCC and normal tissue were investigated. Univariate Cox regression analysis and lasso Cox regression model were performed to identify and construct the prognostic gene signature. Time-dependent receiver operating characteristic (ROC), Kaplan–Meier curve, multivariate Cox regression analysis, nomogram, and decision curve analysis (DCA) were used to assess the prognostic capacity of the six-gene signature. The prognostic value of the gene signature was further validated in independent GSE14520 cohort. Gene Set Enrichment Analyses (GSEA) was performed to further understand the underlying molecular mechanisms. The performance of the prognostic signature in differentiating between normal liver tissues and HCC were also investigated.

Results: A novel six-gene signature (including CSE1L, CSTB, MTHFR, DAGLA, MMP10, and GYS2) was established for HCC prognosis prediction. The ROC curve showed good performance in survival prediction in both the TCGA HCC cohort and the GSE14520 validation cohort. The six-gene signature could stratify patients into a high- and low-risk group which had significantly different survival. Cox regression analysis showed that the six-gene signature could independently predict OS. Nomogram including the six-gene signature was established and shown some clinical net benefit. Furthermore, GSEA revealed several significantly enriched oncological signatures and various metabolic process, which might help explain the underlying molecular mechanisms. Besides, the prognostic signature showed a strong ability for differentiating HCC from normal tissues.

Conclusions: Our study established a novel six-gene signature and nomogram to predict overall survival of HCC, which may help in clinical decision making for individual treatment.

Keywords: Hepatocellular carcinoma, TCGA, GEO, Prognosis, Gene signature

Background

Hepatocellular carcinoma (HCC) is the fifth leading cause of malignant cancer and the third most common cause of cancer-related death worldwide [1]. Despite the great improvement in earlier diagnosis and multidisciplinary cancer management, the long-term prognosis remains poor. Thus, an effective prognostic model that identify patients with a high risk of recurrence and metastasis could guide clinical management. Conventional models utilizing clinical tumor-node-metastasis (TNM) staging, vascular invasion, and other parameters help predict HCC prognosis [2]. However, considering

*Correspondence: liugaom88@foxmail.com; javee@126.com
Department of Hepatobiliary Surgery, Meizhou People's Hospital, No. 38
Huangtang Road, Meizhou 514000, China

Liu *et al. Cancer Cell Int*   (2019) 19:138

Page 2 of 13

the great heterogeneity of HCC, the predictive efficacy of conventional models is still far from satisfying. It's important to take molecular markers into key account when establishing novel predictive tools.

With the advance of genome-sequencing technologies, accumulating evidence shown that gene signatures at mRNA level had great potential in predicting HCC prognosis. For example, Long et al. established a four-gene-based prognostic model (including gene CENPA, SPP1, MAGEB6, and HOXD9) that accurately predicted overall survival OS using data from The Cancer Genome Atlas-Liver Hepatocellular Carcinoma Dataset (TCGA-LIHC) [3]. Similarly, Zheng et al. identified another four-gene-based signature (including gene SPINK1, TXNRD1, LCAT, and PZP) for predicting the prognosis of HCC using data from the TCGA-LIHC and gene expression omnibus (GEO) database [4]. Deep mining of publicly available genomic data tends to be an efficient method to identify novel robust gene prognostic signatures to guide patients' prognostic stratification and personalized therapy.

In this study, we conduct univariate and lasso Cox regression analysis to identify novel prognostic biomarkers and established a prognostic six-gene signature using data from TCGA. Multivariate Cox regression analysis confirmed the independent prognostic role of our six-gene signature. Nomogram was established to predict HCC prognosis. Gene set enrichment analysis was performed to help explain the intrinsic mechanisms. In addition, the prognostic value of our six-gene signature was further validated in GSE14520 dataset from GEO database. Besides, the prognostic signature showed a strong ability for differentiating HCC from normal tissues. Collectively, our results suggest the six-gene signature and nomogram might help effectively predict overall survival of HCC patients.

Materials and methods
Data collection
Level 3 mRNA expression and clinical data from 374 LIHC and 50 normal control samples were obtained from TCGA-LIHC and cBioportal for Cancer Genomics [5, 6]. Data were downloaded from the publicly available database hence it was not applicable for additional ethical approval.

Identification of differentially expressed mRNA in HCC
The raw count data were firstly normalized with transcripts per million (TPM) method and underwent a log2 transformation. Then 19654 protein-coding genes were annotated. The differentially expressed mRNA (DEMs) were calculated using the Limma version 3.36.2 R package [7]. DEMs with an absolute log2 fold change (FC) > 1 and an adjusted P value of < 0.05 were considered for subsequent analysis.

Establishment of the prognostic gene signature
Only patients with a follow-up period longer than 1 month were included for survival analysis. Univariate Cox regression analysis was performed to identify prognostic genes, and genes were considered significant with a cut-off of P < 0.001. Then patients were randomly separated into a training set and testing set. Lasso-penalized Cox regression analysis was conducted to further select prognostic genes for OS in patients with HCC [8]. Then a prognostic gene signature was constructed based on a linear combination of the regression coefficient derived from the lasso Cox regression model coefficients (β) multiplied with its mRNA expression level. The risk score $= (\beta_{mRNA1} *$ expression level of mRNA1) $+ (\beta_{mRNA2} *$ expression level of mRNA2) $+ (\beta_{mRNA3} *$ expression level of mRNA3) $+ \cdots + (\beta_{mRNAn} *$ expression level of mRNAn). The optimal cut-off value was investigated by The R package "survival" [7] and "survminer" and two-sided log-rank test. Patients were classified into a high-risk and low-risk cohort according to the threshold. The time-dependent receiver operating characteristic (ROC) curve was drawn to evaluate the predictive value of the prognostic gene signature for overall survival using the R package "survivalROC" [9]. The Kaplan–Meier survival curve combined with a log-rank test was used to compare the survival difference in the high- and low-risk group using the R package "survival". Then the predictive value of the prognostic gene signature was further investigated in the testing cohort and the whole cohort.

External validation of the prognostic gene signature and gene expression pattern
GSE14520 dataset from the GEO database was downloaded [10]. The risk score for each included patient was calculated with the same prognostic gene-signature based model. Next, the ROC curve and the Kaplan–Meier curve were used to test the predictive value of the prognostic gene signature. The mRNA expression of the genes in the prognostic gene signature was analyzed in HCC and non-tumor tissue using the Wilcoxon signed-rank test. The two-sided P < 0.05 was considered statistically significant. The protein expression of the genes in the prognostic gene signature was explored in the Human Protein Atlas (http://www.proteinatlas.org) online database.

Independent prognostic role of the gene signature
To investigate whether the prognostic gene signature could be independent of other clinical parameters [including gender, age, body mass index (BMI), alpha-fetoprotein (AFP), tumor grade, inflammation, vascular

Liu *et al. Cancer Cell Int*   (2019) 19:138

Page 3 of 13

tumor invasion, and TNM stage], univariate and multivariate analyses were performed using the Cox regression model method with forwarding stepwise procedure. P < 0.05 were considered as statistically significant.

Building and validating a predictive nomogram

Nomogram is widely used to predict cancer prognosis [11]. All independent prognostic factors identified by multivariate Cox regression analysis were included to build a nomogram to investigate the probability of 1-, 3-, and 5-OS of HCC. Validation of the nomogram was assessed by discrimination and calibration. The concordance index (C-index) was calculated to assess the discrimination of the nomogram by a bootstrap method with 1000 resamples. The calibration curve of the nomogram was plotted to observe the nomogram prediction probabilities against the observed rates. Subsequently, we compared the nomogram including all with those including only one independent prognostic factor using the time-ROC curve, C-index, and the decision curve analysis (DCA) [12]. DCA was used to calculate the clinical net benefit of each model compared to all or none strategies. The best model is the one with the highest net benefit as calculated.

Gene Set Enrichment Analyses

To explore the potential molecular mechanisms underlying our constructed prognostic gene signature, GSEA (Gene Set Enrichment Analyses) [13, 14] was performed to find enriched terms predicted to have a correlation with the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway in C2; in C5, a gene set that contain genes annotated by the same gene ontology (GO) term; and in C6, oncogenic signatures of gene sets often dysregulated in cancer. P < 0.01 and FDR (false discovery rate) q < 0.05 were considered statistically significant.

Differentiating performance of the prognostic signature

Boxplot and ROC curve was used to explore the difference of the risk score and the differentially diagnostic capability of the risk score between normal liver and HCC, respectively. P < 0.05 was considered statistically significant.

Statistical analysis

Statistical analyses were performed using R software v3.5.0 (R Foundation for Statistical Computing, Vienna, Austria) and GraphPad Prism v7.00 (GraphPad Software Inc., USA). Qualitative variables were analyzed using the Pearson $\chi 2$ test or Fisher's exact test; quantitative variables were analyzed using a t-test for paired samples or a non-parametric Wilcoxon rank-sum test for unpaired samples as appropriate. Multiple groups of normalized data were analyzed using one-way ANOVA. If not specified above, P < 0.05 was considered statistically significant.
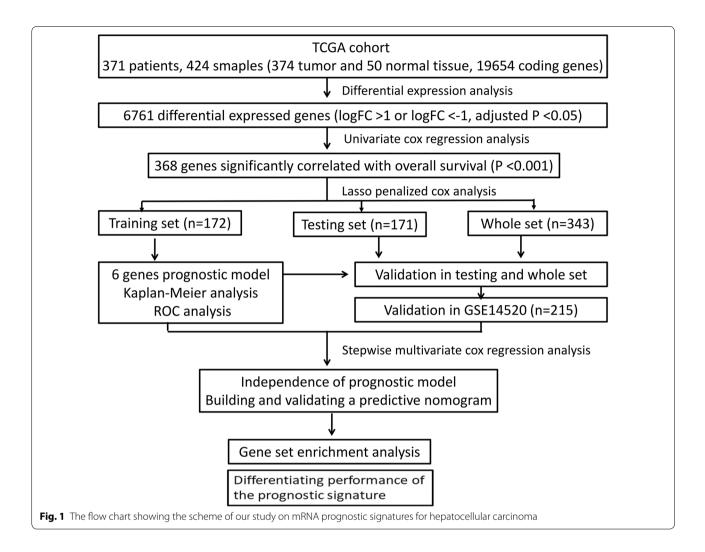
Results

DEMs identification

We conducted our study as described in the flow chart (Fig. 1). A total of 6761 genes were identified as differentially expressed at mRNA level in tumor tissues (n = 374) when compared with that of normal tissues (n = 50). The heatmap of the DEMs was shown in Additional file 1: Figure S1. mRNA of 5822 genes were found to be significantly up-regulated, while that of 939 genes were found to significantly down-regulated (Additional file 2: Figure S2).

Establishment of the six-gene-based prognostic gene signature

343 patients with a follow-up period longer than 1 month were included for subsequent survival analysis. Then patients were randomly separated into a training set (n = 172) and testing set (n = 171). The baseline characteristics were summarized in Additional file 3: Table S1. No clinical parameters except adjacent hepatic tissue inflammation type were significantly different in the training set and testing set. Univariate Cox regression model identified 368 genes that significantly associated with OS. Then, lasso-penalized Cox analysis was performed in the training set (n = 171) to further narrow the mRNAs (Additional file 4: Figure S3). Six genes were identified and subsequently used to construct a prognostic gene-signature. The six genes identified were chromosome segregation 1-like (CSE1L), cystatin B (CSTB), methylenetetrahydrofolate reductase (MTHFR), diacylglycerol lipase alpha (DAGLA), matrix metalloproteinase 10 (MMP10), and glycogen synthase 2 (GYS2). The risk score $= 0.0606 * \text{Expression}_{CSE1L} + 0.0257 * \text{Expression}_{CSTB} + 0.1177 * \text{Expression}_{MTHFR} + 0.1912 * \text{Expression}_{DAGLA} + 0.4324 * \text{Expression}_{MMP10} + (- 0.1003) * \text{Expression}_{GYS2}$. We then calculated the six-gene based risk score for each patient and used the Survminer R package to find the optimal cut-off for the risk score. Time-dependent ROC and Kaplan–Meier curve were used to assess the prognostic capacity of the six-gene signature. Similar procedures were performed in the testing set and the whole set. The AUCs (Area under the ROC curve) for 1-year, 3-year, and 5-year OS were 0.832, 0.850, 0.768, and 0.712, 0.591, 0.602, and 0.773, 0.702, 0.673 for training

Liu *et al. Cancer Cell Int*   (2019) 19:138

Page 4 of 13

TCGA cohort
371 patients, 424 smaples (374 tumor and 50 normal tissue, 19654 coding genes)

↓ Differential expression analysis

6761 differential expressed genes (logFC >1 or logFC <-1, adjusted P <0.05)

↓ Univariate cox regression analysis

368 genes significantly correlated with overall survival (P <0.001)

↓ Lasso penalized cox analysis

Training set (n=172)   Testing set (n=171)   Whole set (n=343)

6 genes prognostic model
Kaplan-Meier analysis
ROC analysis   →   Validation in testing and whole set

Validation in GSE14520 (n=215)

↓ Stepwise multivariate cox regression analysis

Independence of prognostic model
Building and validating a predictive nomogram

↓

Gene set enrichment analysis

Differentiating performance of
the prognostic signature

Fig. 1 The flow chart showing the scheme of our study on mRNA prognostic signatures for hepatocellular carcinoma

set, testing set, and whole set, respectively. Patients in the high-risk group shown significantly poorer OS than patients in the low-risk group (all P < 0.001) (Fig. 2a–c). Compared with other six signatures [3, 4, 15–18], our signature showed a middle C-index and comparable AUCs for 1-, 3-, 5-year OS prediction (Additional file 5: Figure S4; Additional file 6: Table S2). Collectively, our results indicated a good performance of the six-gene signature for survival prediction.

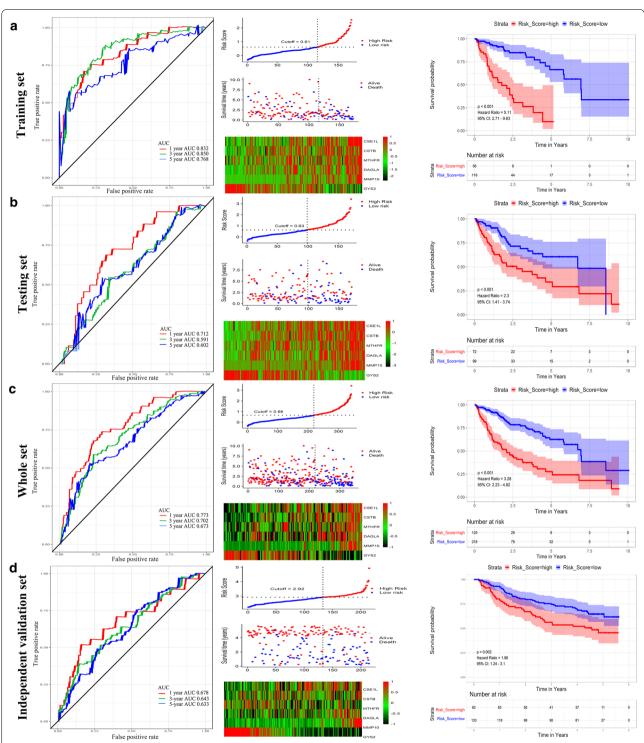#### External validation of the prognostic gene signature

To validate the predictive value of the six-gene signature, we calculated risk score with the same formula for patients in GSE14520. Consistent with the results in the TCGA cohort, patients in the high-risk group shown significantly poorer OS than patients in the low-risk group (P = 0.002). The AUCs for 1-year, 3-year, and 5-year OS were 0.678, 0.643, and 0.633, respectively (Fig. 2d). Taking together, the six-gene signature was capable of predicting OS in HCC.
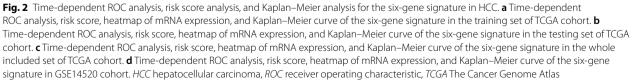
External validation of the genetic alteration and expression of the six gene

Among the 366 patients included in cBioportal for Cancer Genomics database, 36 (10%) shown genetic alterations in the six genes. Missense mutation was the most common genetic alteration (Fig. 3a). Consistent with the results in the TCGA cohort, the mRNA expression of CSE1L, CSTB, MMP10 was significantly up-regulated in HCC while GYS2 was significantly down-regulated when compared with non-tumor tissues. Yet the upregulation of MHTFR and DAGLA was not found in GSE14520 cohort (Fig. 3b). We further explored the protein expression of the six genes in the Human Protein Profiles and shown the characteristic pictures of them in Fig. 3c. However, we did not find MTHFR protein expression in the database.
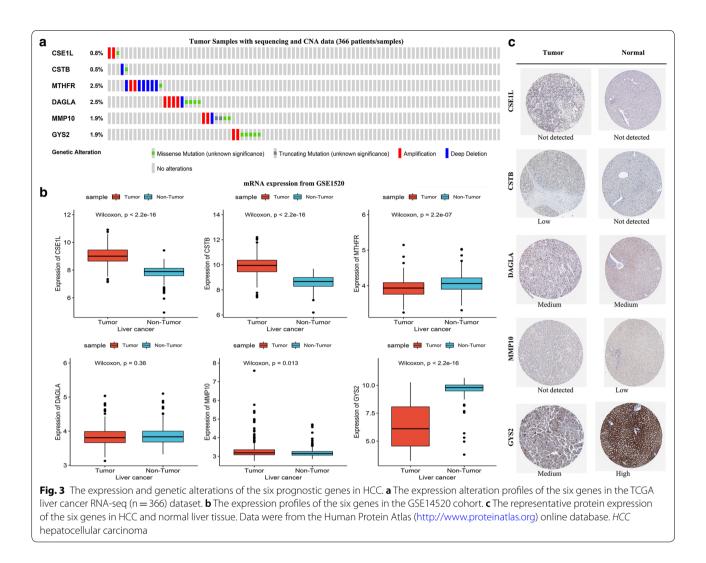
Independent prognostic role of the gene signature

165 patients with complete information including gender, age, BMI, AFP, tumor grade, inflammation, vascular

Liu *et al. Cancer Cell Int*   (2019) 19:138

Page 5 of 13

Fig. 2 Time-dependent ROC analysis, risk score analysis, and Kaplan–Meier analysis for the six-gene signature in HCC. **a** Time-dependent ROC analysis, risk score, heatmap of mRNA expression, and Kaplan–Meier curve of the six-gene signature in the training set of TCGA cohort. **b** Time-dependent ROC analysis, risk score, heatmap of mRNA expression, and Kaplan–Meier curve of the six-gene signature in the testing set of TCGA cohort. **c** Time-dependent ROC analysis, risk score, heatmap of mRNA expression, and Kaplan–Meier curve of the six-gene signature in the whole included set of TCGA cohort. **d** Time-dependent ROC analysis, risk score, heatmap of mRNA expression, and Kaplan–Meier curve of the six-gene signature in GSE14520 cohort. *HCC* hepatocellular carcinoma, *ROC* receiver operating characteristic, *TCGA* The Cancer Genome Atlas

Liu *et al. Cancer Cell Int* (2019) 19:138

Page 6 of 13

Fig. 3 The expression and genetic alterations of the six prognostic genes in HCC. **a** The expression alteration profiles of the six genes in the TCGA liver cancer RNA-seq (n = 366) dataset. **b** The expression profiles of the six genes in the GSE14520 cohort. **c** The representative protein expression of the six genes in HCC and normal liver tissue. Data were from the Human Protein Atlas (http://www.proteinatlas.org) online database. *HCC* hepatocellular carcinoma

tumor invasion, and TNM stage were included for further analysis. Univariate and multivariate Cox regression analysis indicated that vascular tumor invasion, TNM stage, and risk score calculated from the six-gene signature were independent prognostic factors for OS (Fig. 4).
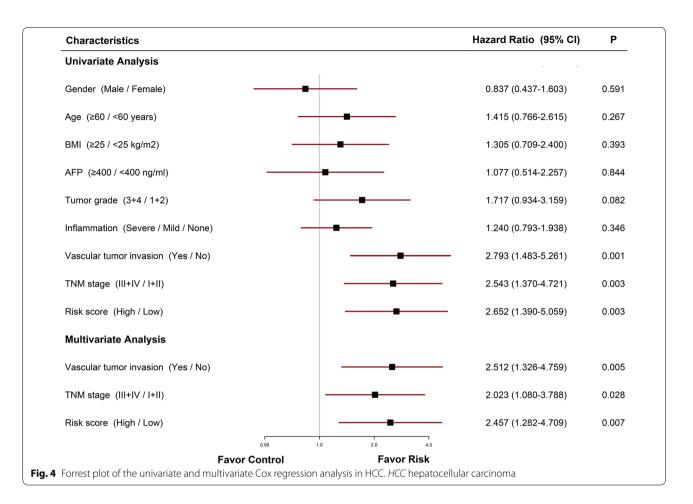
Building and validating a predictive nomogram

We then built a nomogram to predict 1-year, 3-year, and 5-year OS in the 165 HCC patients using three independent prognostic factors including vascular tumor invasion, TNM stage, and risk score. Calibration plots showed that the nomogram (combined model) might under-estimate or over-estimate the mortality (Fig. 5). The C-index for vascular tumor invasion, TNM stage, risk score, and the combined model was 0.66 (95% confidence interval [CI] 0.58–0.74), 0.61(95% CI 0.54–0.61), 0.72 (95% CI 0.62–0.82), and 0.77 (95% CI 0.67–0.86), respectively. The AUCs of the nomogram were 0.87 (95% confidence interval [CI] 0.80–0.95), 0.78 (95% CI 0.66–0.88), and

0.71 (95% CI 0.58–0.85) for 1-year, 3-year, and 5-year OS, respectively (Table 1). Compared with nomogram including only the vascular, TNM, or prognostic gene signature, the combined model shown the largest AUC for 1-year and 3-year OS but not for 5-year OS (Table 1, Fig. 6a–c). DCA demonstrated that the combined model showed the best net benefit for 1-year and 3-year OS but not for 5-year OS as well (Fig. 6d–f). Taking together, these results indicated that compared with nomograms built with a single prognostic factor, the nomogram built with the combined model might be the best nomogram for predicting short-term survival (1-year and 3-year) but not for long-term survival (such as 5-year) for patients with HCC, which might help clinical management.

Gene Set Enrichment Analyses

To explore the underlying molecular mechanisms of the signature, we conduct GSEA comparing the high-risk

Liu *et al. Cancer Cell Int*   (2019) 19:138

Page 7 of 13

Characteristics	Hazard Ratio (95% CI)	P
Univariate Analysis		
Gender (Male / Female)	0.837 (0.437-1.603)	0.591
Age (≥60 / <60 years)	1.415 (0.766-2.615)	0.267
BMI (≥25 / <25 kg/m2)	1.305 (0.709-2.400)	0.393
AFP (≥400 / <400 ng/ml)	1.077 (0.514-2.257)	0.844
Tumor grade (3+4 / 1+2)	1.717 (0.934-3.159)	0.082
Inflammation (Severe / Mild / None)	1.240 (0.793-1.938)	0.346
Vascular tumor invasion (Yes / No)	2.793 (1.483-5.261)	0.001
TNM stage (III+IV / I+II)	2.543 (1.370-4.721)	0.003
Risk score (High / Low)	2.652 (1.390-5.059)	0.003
Multivariate Analysis		
Vascular tumor invasion (Yes / No)	2.512 (1.326-4.759)	0.005
TNM stage (III+IV / I+II)	2.023 (1.080-3.788)	0.028
Risk score (High / Low)	2.457 (1.282-4.709)	0.007

0.50   1.0   2.0   4.0

Favor Control   Favor Risk

Fig. 4 Forrest plot of the univariate and multivariate Cox regression analysis in HCC. *HCC* hepatocellular carcinoma

group with the low-risk group in 343 TCGA patients of the whole set. In the high-risk group, 4 oncological signatures including Early serum response (CSR), E2F Transcription Factor 1 (E2F1), Rb-P107, and granule cell neuron precursors (GCNP) were enriched; however, no KEGG or GO terms were significantly enriched. In the low-risk group, the enriched KEGG pathways and GO terms were mainly focused on various metabolism process (including fatty acid, retinol, tyrosine, butanoate and so on). However, no oncological signatures were significantly enriched (Additional file 7: Table S3).
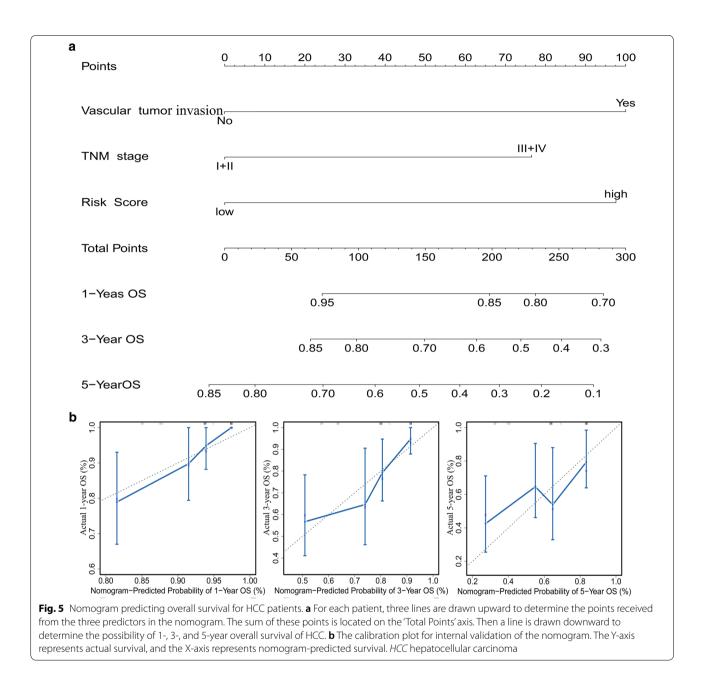
Differentiating performance of the prognostic signature

The risk score was then compared between normal liver and HCC to explore the differentially diagnostic capability of the prognostic signature. The risk score was found to be significantly higher in HCC when compared with normal control. The risk score was also found to be significantly higher in patients with advanced TNM stage and tumor grade (Fig. 7). The AUC of the risk score was 0.93 in both cohorts, indicating a strong diagnostic ability for HCC. Furthermore, the subgroup analysis of different

stages and grades also showed modest diagnostic capability (Fig. 8). Taking together, these results also suggest great potential of the signature in the differential diagnosis of HCC.
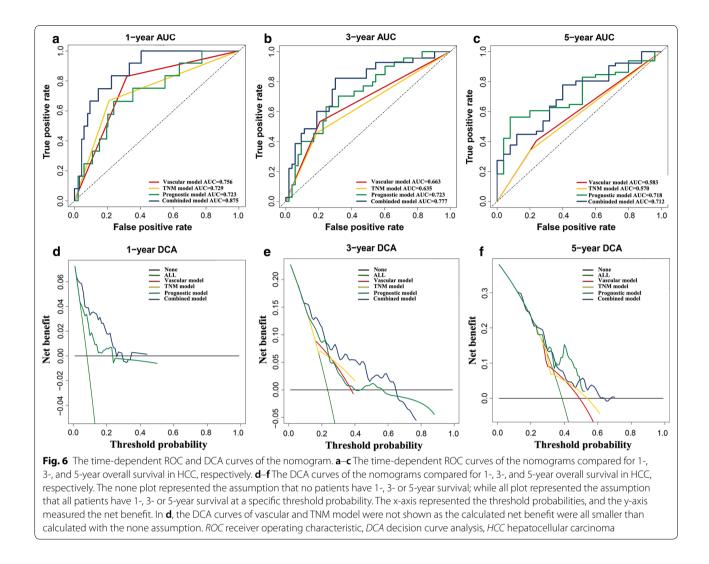
Discussion

HCC remains a major challenge for public health worldwide. Conventional parameters such as TNM staging, vascular invasion, and AFP help predict HCC prognosis in some degree. However, considering the great heterogeneity of HCC, identification of novel prognostic biomarkers and establishment of more accurate prognostic models are urgently needed. And the combination of the prognostic gene signature with conventional clinical parameters may have better predictive efficacy than a single biomarker. Recently, gene-signatures based on aberrant mRNA have gained much attention and shown great potential in prognosis prediction of cancer [3, 16, 17, 19].

In this study, we established a novel six-gene signature (including CSE1L, CSTB, MTHFR, DAGLA, MMP10, and GYS2) for HCC prognosis prediction. While CSE1L, CSTB, MTHFR, DAGLA, and MMP10 were found to
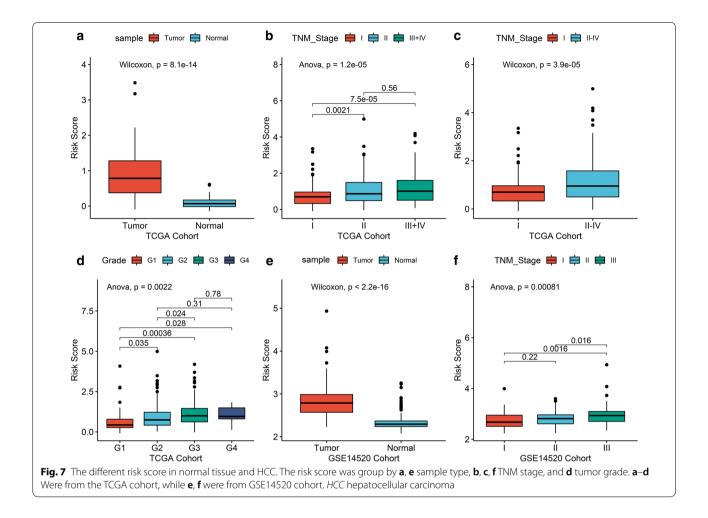
Liu *et al. Cancer Cell Int*   (2019) 19:138

Page 8 of 13

Fig. 5 Nomogram predicting overall survival for HCC patients. **a** For each patient, three lines are drawn upward to determine the points received from the three predictors in the nomogram. The sum of these points is located on the 'Total Points' axis. Then a line is drawn downward to determine the possibility of 1-, 3-, and 5-year overall survival of HCC. **b** The calibration plot for internal validation of the nomogram. The Y-axis represents actual survival, and the X-axis represents nomogram-predicted survival. *HCC* hepatocellular carcinoma

Table 1 Comparison of the nomogram with vascular, TNM stage, prognostic model, and the combined model

Models	1-year AUC (95% CI)	P-value	3-year AUC (95% CI)	P-value	5-year ACU (95% CI)	P-value
Vascular model	0.76 (0.64–0.87)	–	0.66 (0.56–0.77)	–	0.58 (0.47–0.70)	–
TNM model	0.73 (0.59–0.87)	–	0.63 (0.53–0.74)	–	0.57 (0.46–0.68)	–
Prognostic model	0.73 (0.58–0.87)	–	0.72 (0.61–0.84)	–	0.72 (0.57–0.86)	–
Nomogram (combined) model	0.87 (0.80–0.95)	–	0.78 (0.66–0.88)	–	0.71 (0.58–0.85)	–
Nomogram vs. vascular model		0.05		0.027		0.01
Nomogram vs. TNM model		0.01		0.01		0.03
Nomogram vs. prognostic model		0.01		0.29		0.93

AUC area under curve, *CI* confidence interval, *TNM* tumor-node-metastasis

Liu *et al. Cancer Cell Int*   (2019) 19:138

Page 9 of 13

Fig. 6 The time-dependent ROC and DCA curves of the nomogram. **a–c** The time-dependent ROC curves of the nomograms compared for 1-, 3-, and 5-year overall survival in HCC, respectively. **d–f** The DCA curves of the nomograms compared for 1-, 3-, and 5-year overall survival in HCC, respectively. The none plot represented the assumption that no patients have 1-, 3- or 5-year survival; while all plot represented the assumption that all patients have 1-, 3- or 5-year survival at a specific threshold probability. The x-axis represented the threshold probabilities, and the y-axis measured the net benefit. In **d**, the DCA curves of vascular and TNM model were not shown as the calculated net benefit were all smaller than calculated with the none assumption. *ROC* receiver operating characteristic, *DCA* decision curve analysis, *HCC* hepatocellular carcinoma

be negative prognostic genes, GYS2 was found to do the opposite. The prognosis predictive performance of the signature was good not only in the TCGA HCC cohort but also in the GSE14520 cohort, and comparable with six previously reported models. The six-gene risk was an independent prognostic factor of HCC and patients in the high-risk group shown significantly poorer survival than patients in the low-risk group. ROC and DCA demonstrated that the nomogram combining the six-gene signature and conventional clinical prognostic factors performed the best in predicting short-term survival (1-year and 3-year) but not in long-term survival (such as 5-year) for patients with HCC. All these results indicated that the risk model developed from the six genes could be a useful indicator for HCC survival. Furthermore, GSEA revealed several significantly enriched oncological signatures and various metabolic process, which might help explain the underlying molecular mechanisms of the signature. And we found the risk score shown a strong

ability in differentiating HCC from normal tissues, suggesting a great potential of utilizing the signature in HCC differential diagnosis.

CSE1L, also named as CAS (cellular apoptosis susceptibility protein), has been reported as an oncogene in several cancers [20–22]. CSE1L is a multifunctional gene that participates in apoptosis, chromosome assembly, nucleocytoplasmic transport, microvesicle formation, chemo-resistance, and cancer progression [20, 23, 24]. However, the role and mechanism of aberrant CSE1L in HCC remains poorly defined. CSTB is a reversible endogenous inhibitor of lysosomal cysteine proteinases [25]. Mutations of CSTB leads to progressive myoclonus epilepsy (EPM1), which is an inherited and lethal autosomal disease [26]. Dysregulated expression of CSTB has been implicated to be a useful biomarker in various cancers such as ovarian cancer [27], esophageal cancer [28] and breast cancer [29]. Especially, CSTB was found to be overexpressed in most HCCs and was elevated in

Fig. 7 The different risk score in normal tissue and HCC. The risk score was group by **a**, **e** sample type, **b**, **c**, **f** TNM stage, and **d** tumor grade. **a–d** Were from the TCGA cohort, while **e**, **f** were from GSE14520 cohort. *HCC* hepatocellular carcinoma

the serum of most HCC patients [30]. DAGL (Diacylglycerol lipase) hydrolyzes diacylglycerol to 2-arachidonoylglycerol (2-AG) and free fatty acid (FFA) [31]. Disruption of DAGL activity influenced the development of the central nervous system [32]. Recently, Okubo et al. reported that DAGLA promoted tumorigenesis in oral squamous cell carcinomas by regulating cell-cycle [33]. Roy et al. indicated that DAGLA participated in ovarian progression caused by loss of the endosulfatase HSulf-1 [34]. Nevertheless, the role of DAGLA in HCC remains unclear. MTHFR catalyzes the 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate, a co-substrate for homocysteine re-methylation to methionine. Methionine is the forebody of *S*-adenosylmethionine (SAM), and SAM is the direct methyl donor for the DNA methylation [35]. Abnormal MTHFR activity leads to abnormal gene methylation, gene instability and finally cancer [36]. Accumulating studies demonstrate that MTHFR polymorphism affects the susceptibility of various cancer, especially HCC [37–41]. Matrix metalloproteinases (MMPs) are widely accepted as critical modulators for

tumor microenvironment [42]. MMP10 promoted HCC by involving in tumor angiogenesis, growth, and dissemination [43]. Decreased glycogen concentration negatively correlated with tumor growth [44]. GYS, the rate-limiting enzyme of glycogen synthesis, consists of two isoforms including GYS1 and GYS2. Loss of GYS2 caused glycogen storage disease type 0 [45]. A very recent study revealed that decreased expression of GYS2 reduced glycogen and indicated unfavorable clinical outcomes of HCC. Mechanically, GYS2 suppressed tumor growth in HBV-related HCC via a negative feedback loop with p53 [46].

To our knowledge, the six-gene signature related prognostic model and nomogram have not been reported previously and could be a useful prognostic and diagnostic classification tool of HCC. The risk score was based on mRNA expression but not somatic mutations or methylation status of only six prognostic genes. It could be more routine and cost-effective in practice as it decreased the necessity of whole-genome sequencing for all patients. Nomogram combining our signature

Liu *et al. Cancer Cell Int*   (2019) 19:138

Page 11 of 13

Fig. 8 The ROC curve of the risk score in normal tissue and HCC. The ROC curve of the risk score showing its capacity in differentiating between normal and HCC (**a**, **e**), between HCC of different TNM stage (**b**, **c**, **f**), and between HCC of different tumor grade (**d**). *ROC* receiver operating characteristic, *HCC* hepatocellular carcinoma, *TNM* tumor-node-metastasis

with conventional clinical parameters like TNM stage shown significantly improved performance, especially in predicting short-term survival (1-year or 3-year), indicating a more accurate reflection of the great heterogeneity of HCC. However, several limitations of our study should be taken into consideration. Firstly, our study was mainly based on data from TCGA in which most patients were White or Asian. Extending our findings to other ethnic patients should be with great caution. Secondly, external validation of the six-gene signature and prognostic nomogram in more independent cohorts is necessary. Thirdly, the expression and the prognostic role of the six genes at protein level warrant further investigation. Forth, calibration plots showed that the nomogram (combined model) might under-estimate or over-estimate the mortality, efforts should be made to further improve the prediction performance. Fifth, all mechanical analysis in our study was descriptive, further functional experiments are needed to clarify the underlying mechanism of the six genes. Sixth, except its excellent performance in differentiating HCC from normal liver, the performance of our signature in differentiating between the normal liver, liver adenomas, focal nodular hyperplasia, and hepatocellular carcinomas remains to be further elucidated.

Conclusions

Our study established a novel six-gene signature and nomogram to predict overall survival of HCC, which may help in clinical decision making for individual treatment.

Additional files

Additional file 1: Figure S1. The heatmap of the differentially expressed mRNA in HCC when compared with normal tissue.

Additional file 2: Figure S2. Volcano plot shown the expression change in HCC when compared with normal tissue. An absolute log2 fold change (FC) > 1 and an adjusted P value of < 0.05 cutoff was used to defined differentially expressed mRNAs. The red represented significantly up-regulated mRNAs. The blue represented significantly down-regulated mRNAs. The black represented not differentially expressed mRNAs.

Additional file 3: Table S1. Clinical features of HCC patients in training set, testing set, and overall set.

Additional file 4: Figure S3. LASSO profiles of the 368 prognostic genes in HCC. (A) LASSO coefficient profiles of the 368 prognostic genes in HCC. (B) Lasso deviance profiles of the 368 prognostic genes in HCC.

Additional file 5: Figure S4. Comparison of our signature with six previous models using time-dependent ROC analyses.

Additional file 6: Table S2. Comparison of our signature with six other signatures reported previously.

Additional file 7: Table S3. Gene set enrichment analyses between the high- and low-risk group in 343 TCGA HCC.

Liu *et al. Cancer Cell Int*   (2019) 19:138

Page 12 of 13

Abbreviations

Acknowledgements

Authors' contributions

Funding

Availability of data and materials

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61(2):69–90.
2. Bruix J, Reig M, Sherman M. Evidence-based diagnosis, staging, and treatment of patients with hepatocellular carcinoma. Gastroenterology. 2016;150(4):835–53.
3. 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):5928–38.
4. Zheng Y, Liu Y, Zhao S, Zheng Z, Shen C, An L, Yuan Y. Large-scale analysis reveals a novel risk score to predict overall survival in hepatocellular carcinoma. Cancer Manag Res. 2018;10:6079–96.
5. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401–4.
6. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269):pl1.
7. Diboun I, Wernisch L, Orengo CA, Koltzenburg M. Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. BMC Genomics. 2006;7:252.
8. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med. 1997;16(4):385–95.
9. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics. 2000;56(2):337–44.
10. Stephanie R, Hu-Liang J, Anuradha B, Marshonna F, Qing-Hai Y, Ju-Seog L, Thorgeirsson SS, Zhongtang S, Zhao-You T, Lun-Xiu Q. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Can Res. 2010;70(24):10202–12.
11. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol. 2008;26(8):1364–70.
12. Eb E. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565–74.
13. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003;34:267.
14. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102(43):15545–50.
15. 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.
16. 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.
17. 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;10:5291–302.
18. 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.
19. Liu S, Miao C, Liu J, Wang CC, Lu XJ. Four differentially methylated gene pairs to predict the prognosis for early stage hepatocellular carcinoma patients. J Cell Physiol. 2018;233(9):6583–90.
20. Ma S, Yang D, Liu Y, Wang Y, Lin T, Li Y, Yang S, Zhang W, Zhang R. LncRNA BANCR promotes tumorigenesis and enhances adriamycin resistance in colorectal cancer. Aging. 2018;10(8):2062–78.
21. Cheng DD, Lin HC, Li SJ, Yao M, Yang QC, Fan CY. CSE1L interaction with MSH6 promotes osteosarcoma progression and predicts poor patient survival. Sci Rep. 2017;7:46238.
22. Lee WR, Shen SC, Wu PR, Chou CL, Shih YH, Yeh CM, Yeh KT, Jiang MC. CSE1L Links cAMP/PKA and Ras/ERK pathways and regulates the expressions and phosphorylations of ERK1/2, CREB, and MITF in melanoma cells. Mol Carcinog. 2016;55(11):1542–52.
23. Jiang MC. CAS (CSE1L) signaling pathway in tumor progression and its potential as a biomarker and target for targeted therapy. Tumour Biol. 2016;37(10):13077–90.
24. Tai CJ, Hsu CH, Shen SC, Lee WR, Jiang MC. Cellular apoptosis susceptibility (CSE1L/CAS) protein in cancer metastasis and chemotherapeutic drug-induced apoptosis. J Exp Clin Cancer Res. 2010;29:110.
25. Keppler D. Towards novel anti-cancer strategies based on cystatin function. Cancer Lett. 2006;235(2):159–76.
26. Pennacchio LA, Lehesjoki AE, Stone NE, Willour VL, Virtaneva K, Miao J, D'Amato E, Ramirez L, Faham M, Koskiniemi M. Mutations in the gene encoding cystatin B in progressive myoclonus epilepsy (EPM1). Science. 1996;271(5256):1731–4.
27. Takaya A, Peng WX, Ishino K, Kudo M, Yamamoto T, Wada R, Takeshita T, Naito Z. Cystatin B as a potential diagnostic biomarker in ovarian clear cell carcinoma. Int J Oncol. 2015;46(4):1573–81.
28. Yan Y, Zhou K, Wang L, Wang F, Chen X, Fan Q. Clinical significance of serum cathepsin B and cystatin C levels and their ratio in the prognosis of patients with esophageal cancer. Onco Targets Ther. 2017;10:1947–54.
29. Butinar M, Prebanda MT, Rajkovic J, Jeric B, Stoka V, Peters C, Reinheckel T, Kruger A, Turk V, Turk B, et al. Stefin B deficiency reduces tumor growth via sensitization of tumor cells to oxidative stress in a breast cancer model. Oncogene. 2014;33(26):3392–400.
30. Lee MJ, Yu GR, Park SH, Cho BH, Ahn JS, Park HJ, Song EY, Kim DG. Identification of cystatin B as a potential serum marker in hepatocellular carcinoma. Clin Cancer Res. 2008;14(4):1080–9.

Liu *et al. Cancer Cell Int*   (2019) 19:138

Page 13 of 13

31. Bisogno T, Howell F, Williams G, Minassi A, Cascio MG, Ligresti A, Matias I, Schiano-Moriello A, Paul P, Williams EJ, Gangadharan U, et al. Cloning of the first sn1-DAG lipases points to the spatial and temporal regulation of endocannabinoid signaling in the brain. J Cell Biol. 2003;163(3):463–8.

32. Smith DR, Stanley CM, Foss T, Boles RG, McKernan K. Rare genetic variants in the endocannabinoid system genes CNR1 and DAGLA are associated with neurological phenotypes in humans. PLoS ONE. 2017;12(11):e0187926.

33. Okubo Y, Kasamatsu A, Yamatoji M, Fushimi K, Ishigami T, Shimizu T, Kasama H, Shiiba M, Tanzawa H, Uzawa K. Diacylglycerol lipase alpha promotes tumorigenesis in oral cancer by cell-cycle progression. Exp Cell Res. 2018;367(1):112–8.

34. Roy D, Mondal S, Wang C, He X, Khurana A, Giri S, Hoffmann R, Jung DB, Kim SH, Chini EN, et al. Loss of HSulf-1 promotes altered lipid metabolism in ovarian cancer. Cancer Metab. 2014;2:13.

35. Kraus D, Yang Q, Kong D, Banks AS, Zhang L, Rodgers JT, Pirinen E, Pulinilkunnil TC, Gong F, Wang YC, Cen Y. Nicotinamide *N*-methyl-transferase knockdown protects against diet-induced obesity. Nature. 2014;508(7495):258–62.

36. Larsson SC, Giovannucci E, Wolk A. Folate intake, MTHFR polymorphisms, and risk of esophageal, gastric, and pancreatic cancer: a meta-analysis. Gastroenterology. 2006;131(4):1271–83.

37. Wang C, Xie H, Lu D, Ling Q, Jin P, Li H, Zhuang R, Xu X, Zheng S. The MTHFR polymorphism affect the susceptibility of HCC and the prognosis of HCC liver transplantation. Clin Transl Oncol. 2018;20(4):448–56.

38. Qiao K, Zhang S, Trieu C, Dai Q, Huo Z, Du Y, Lu W, Hou W. Genetic polymorphism of MTHFR C677T influences susceptibility to HBV-related hepatocellular carcinoma in a Chinese population: a case-control study. Clin Lab. 2017;63(4):787–95.

39. Ventura P, Venturelli G, Marcacci M, Fiorini M, Marchini S, Cuoghi C, Pietrangelo A. Hyperhomocysteinemia and MTHFR C677T polymorphism in patients with portal vein thrombosis complicating liver cirrhosis. Thromb Res. 2016;141:189–95.

40. Kwak SY, Kim UK, Cho HJ, Lee HK, Kim HJ, Kim NK, Hwang SG. Methylene-tetrahydrofolate reductase (MTHFR) and methionine synthase reductase (MTRR) gene polymorphisms as risk factors for hepatocellular carcinoma in a Korean population. Anticancer Res. 2008;28(5a):2807–11.

41. Yuan JM, Lu SC, Van Den Berg D, Govindarajan S, Zhang ZQ, Mato JM, Mimi CY. Genetic polymorphisms in the methylenetetrahydrofolate reductase and thymidylate synthase genes and risk of hepatocellular carcinoma. Hepatology. 2007;46(3):749–58.

42. Zhang JJ, Zhu Y, Xie KL, Peng YP, Tao JQ, Tang J, Li Z, Xu ZK, Dai CC, Qian ZY, et al. Yin Yang-1 suppresses invasion and metastasis of pancreatic ductal adenocarcinoma by downregulating MMP10 in a MUC4/ErbB2/p38/MEF2C-dependent mechanism. Mol Cancer. 2014;13:130.

43. Garcia-Irigoyen O, Latasa MU, Carotti S, Uriarte I, Elizalde M, Urtasun R, Vespasiani-Gentilucci U, Morini S, Benito P, Ladero JM, et al. Matrix metal-loproteinase 10 contributes to hepatocarcinogenesis in a novel crosstalk with the stromal derived factor 1/C-X-C chemokine receptor 4 axis. Hepatology. 2015;62(1):166–78.

44. Rousset M, Zweibaum A, Fogh J. Presence of glycogen and growth-related variations in 58 cultured human tumor cell lines of various tissue origins. Can Res. 1981;41(3):1165.

45. Szymańska E, Rokicki D, Wątrobinska U, Ciara E, Halat P, Płoski R, Tylki-Szymańka A. Pediatric patient with hyperketotic hypoglycemia diagnosed with glycogen synthase deficiency due to the novel homozygous mutation in GYS2. Mol Genet Metab Rep. 2015;4(C):83–6.

46. Chen SL, Zhang CZ, Liu LL, Lu SX, Pan YH, Wang CH, He YF, Lin CS, Yang X, Xie D, et al. A GYS2/p53 negative feedback loop restricts tumor growth in HBV-related hepatocellular carcinoma. Cancer Res. 2019;79(3):534–45.

Publisher's Note