

PRIMARY RESEARCH

Open Access



Identification of DNA repair-related genes predicting pathogenesis and prognosis for liver cancer

Wenjing Zhu¹ , Qiliang Zhang³, Min Liu¹, Meixing Yan⁴, Xiao Chu^{1*} and Yongchun Li^{2*}

Abstract

Background: Liver cancer (LC) is one of the most fatal cancers throughout the world. More efficient and sensitive gene signatures that could accurately predict survival in LC patients are vitally needed to promote a better individualized and effective treatment.

Material/methods: 422 LC and adjacent normal tissues with both RNA-Seq and clinical data in TCGA were embedded in our study. Gene set enrichment analysis (GSEA) was applied to identify genes and hallmark gene sets that are more valuable for liver cancer therapy. Cox regression analysis was used to identify genes related to overall survival (OS) and build the prediction model. cBioPortal database was used to examine the alterations of the panel mRNA signature. ROC curves and Kaplan–Meier curves were used to validate the prediction model. Besides, the expression of the genes in the model were validated using quantitative real-time PCR in clinical tissue specimens.

Results: The panel of DNA repair-related mRNA signature consisted of seven mRNAs: RFC4 (replication factor C subunit 4), ZWINT (ZW10 interacting kinetochore protein), UPF3B (UPF3B regulator of nonsense mediated mRNA decay), NCBP2 (nuclear cap binding protein subunit 2), ADA (adenosine deaminase), SF3A3 (splicing factor 3a subunit 3) and GTF2H1 (general transcription factor IIH subunit 1). On-line analysis of cBioPortal database found that the expression of the panel mRNA has a wide variation ranging from 7 to 10%. All the mRNAs were significantly upregulated in LC tissues compared to normal tissues ($P < 0.05$). The risk model is closely related to the OS of LC patients. The hazard ratio (HR) is 2.184 [95% CI (confidence interval) 1.523–3.132] and log-rank P -value < 0.0001 . For clinical specimen validation, we found that all of the genes in the model upregulated in liver cancer tissues versus normal liver tissues, which was consistent with the results predicted.

Conclusions: Our study demonstrated a mRNA signature including seven mRNA for prognosis prediction of LC. This panel gene signature provides a new criterion for accurate diagnosis and therapeutic target of LC.

Keywords: Liver cancer, mRNA, Biomarker, Prognosis, TCGA

Background

Liver cancer (LC) is one of the most fatal cancers throughout the world. It was estimated about 841,080 new cases and 781,631 deaths of liver cancer worldwide in 2020 according to the American Cancer Society [1]. Moreover, rates of both incidence and mortality are 2 to 3 times higher among male than female in most world regions according to the global cancer statistics 2018 in 185 countries [1]. Globally, incidence and mortality of

*Correspondence: chux1970@163.com; liych797@163.com

¹ Department of Pharmacy, School of Medicine, Qingdao Municipal Hospital, Qingdao University, Qingdao 266011, Shandong, China

² Department of Pulmonary Medicine, School of Medicine, Qingdao Municipal Hospital, Qingdao University, Qingdao 266011, Shandong, China

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



© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

liver cancer in developing countries ranks higher than those in developed countries [2]. The therapeutic effect of liver cancer depends largely on the time interval from diagnosis to treatment, especially for early-stage patients with liver cancer [3]. Some improved treatment methods, such as liver transplantation, hepatectomy and early radiofrequency therapy, have therapeutic value for patients with liver cancer in early stage [4]. However, about more than 70% of patients with liver cancer are diagnosed at advanced stage, which limited the application of conventional therapies [5].

Genomic instability is an important hallmark of cancer, and facilitates the transformation of cancer [6]. Some circulating markers that can predict the response to therapy and the survival of cancer patients have been demonstrated. As reported, miR423-5p can be used as a useful tool to predict response to sorafenib in HCC patients [7]. Moreover, M Caraglia et al. demonstrated that the oxidative stress status and pERK activity in peripheral blood mononuclear cells had high value in the prediction of the response to sorafenib + octreotide therapy in HCC patients [8]. Therefore, understanding the genetic and epigenetic alterations, which are important to hepatocarcinogenesis, is an urgent problem to be solved for providing novel therapeutic targets for HCC [9]. Recent epidemiological studies have reported that two-thirds of cancers are caused by mistakes in DNA replication [10]. Recently, a multi-cohort retrospective analysis revealed that 138 DNA repair genes had prognostic significance in 16 cancer types [11]. DNA repair capacity has a significant correlation with the lymphatic invasion in colorectal cancer patients [12], and DNA Repair promotes drug resistance in ovarian cancer by different hallmark gene sets [13]. Therefore, DNA repair acts as an essential role in maintaining genome stability and cancer development. The deregulation of DNA repair-associated molecules could enhance the resistance of cancer cells to chemotherapy. Genes and proteins related to DNA repair have become therapeutic targets in prostate cancer and ovarian cancer [14, 15]. For the recent study, plenty of gene sets have been reported as a guidance in optimizing the treatment and play an important role as a biomarker in various cancers [16–20]. Multigene prognostic signatures from tumor tissue of patients can predict the prognosis of cancer patients more accurately than a single gene. Especially, multigene prognostic signatures from messenger RNA (mRNA) could provide a better accuracy in cancer prognosis than the non-coding prognostic genes, which enable a better individualized treatment and more effective treatment [21, 22]. And, identifying the mRNAs and hallmark gene sets becomes a prerequisite for clinical application and treatment progress in cancer. However, there is a lack of research on the prognosis of liver

cancer with combined markers of mRNAs. Therefore, it is still an urgent problem to find more efficient and sensitive gene sets of mRNAs biomarker for liver cancer.

In this study, Gene set enrichment analysis (GSEA) was used to identify genes that are more valuable for liver cancer therapy and proceeded with further analysis. Finally, 141 DNA repair-related mRNAs were screened out and a seven-gene set that can accurately predict the prognosis of patients with liver cancer was established.

Methods

Data sets

The data of patients with LC in TCGA, including mRNA-Seq of transcriptome profiling data and clinical data of the LC patients, was downloaded by the GDC data portal: <https://portal.gdc.cancer.gov/>. All cases of liver cancer in this study were exclusively hepatocellular carcinoma. The detailed clinicopathological parameters of patients with liver cancer, including age, gender, TNM stage, stage, grade, cancer status and family history of liver cancer, were listed in Table 1.

Data sets of GSE101685 and GSE101728 was downloaded from GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). The website of KM plots online analysis (http://kmplot.com/analysis/index.php?p=service&cancer=liver_rnaseq).

Patients and clinical specimens

We recruited 5 pairs of matched liver cancer tissues and adjacent normal tissues from Chinese Institution. Among of them, one patient had 2 HCC nodules, and the other patients had 0, 0, 1 and 1 HCC nodule respectively. The tissue samples and corresponding clinical pathology data were from Qingdao Municipal Hospital. This study was approved by Institutional Review Board of Qingdao Municipal Hospital. The number of the approval of this study by the ethical committee is No.109. And the approval document was approved on December 7th, 2019.

RNA isolation and quantitative real-time PCR (qRT-PCR)

For tissue RNA isolation, 1 mL TRIzol (Invitrogen) was added to 50 mg of tissue and total RNA samples were extracted according to the manufacturer's instructions. Purified RNA was quantified using NanoDrop 2000 (Thermo Scientific).

cDNAs were synthesized from total RNAs by using ReverTra Ace qPCR RT Kit (Toyobo Co., LTD, Japan). qRT-PCR of GAPDH (glyceraldehyde-3-phosphate dehydrogenase), RFC4, ZWINT, UPF3B, NCBP2, ADA, SF3A3, and GTF2H1 was performed with the SYBR qPCR Mix (Toyobo Co., LTD, Japan). 10 μ L reaction system was set up according to the manufacturer's

Table 1 Clinical pathological parameters of liver cancer patients in TCGA database

Clinical pathological parameters	N	%
Age		
≤ 61	195	52
> 61	181	48
Tissue		
Adjacent noncancerous tissue	50	12
Liver cancer	377	88
Gender		
Male	255	68
Female	122	32
TNM stage		
T1 + T2	280	75
T3 + T4	94	25
Stage		
Stage1 + stage2	262	74
Stage3 + stage4	91	26
Grade		
Grade1 + grade2	235	63
Grade3 + grade4	137	37
Cancer status		
Tumor free	236	68
With tumor	113	32
Family history of liver cancer		
No	212	65
Yes	114	35
Paired samples from China		
0 HCC nodule	2	40
1 HCC nodule	2	40
2 HCC nodules	1	20

instructions and amplified for 40 cycles. The expression levels were normalized by GAPDH. Relative expression was calculated using the method of $2^{-\Delta\Delta C_t}$. Primer names and primer sequences are listed in table.

Primer name	Primer sequence
GAPDH forward	CAGGAGGCATTGCTGATGAT
GAPDH reverse	GAAGGCTGGGGCTCATT
RFC4 forward	TAAGTCTCCTGGGCCGTTA
RFC4 reverse	TGCATGGTACTTCACCCAGT
ZWINT forward	GTGGCAGCTACAACAGGAGA
ZWINT reverse	CAGCTTACCCTCTGCAGCTT
UPF3B forward	ATCGAATAAGAAACAAGGATCGTC
UPF3B reverse	AGGCCATCTGGACTTATCACT
NCBP2 forward	CTCTGCACTATGTCGGGTGG
NCBP2 reverse	TGGCGTTTTCCGCATAGCTT
ADA forward	CCAAAGAGGGCGTGGTGTAT
ADA reverse	GTGAGGTCCCTTCAGCCT

Primer name	Primer sequence
SF3A3 forward	CCATGCAAGATATCTGTGTGCC
SF3A3 reverse	TCTGTGAACTCCACCAAGTTTTG
GTF2H1 forward	GGAACCTCGACCGGATCCAAC
GTF2H1 reverse	ATGGTGGCTAGAAGGTGCAAT

Statistical analysis

GSEA was used to screen out hallmark gene sets with $|NES| > 1$, NOM P -val < 0.05 and FDR q -val < 0.25 . Univariate Cox regression analysis was used to screen out prognostic genes with P values of < 0.001 . Next, a prognostic risk score formula was established according to the results of multivariate Cox regression model. Risk score of each LC patients was calculated by the risk score formula = β_1 *expression of gene 1 + β_2 *expression of gene 2 + β_3 *expression of gene 3 + ... + β_n *expression of gene n. Paired and unpaired t test were used to compare gene expression in LC cancer tissues and normal tissues. LC patients were divided into high-risk group and low-risk group using the median risk score as the cutoff. Receiver operating characteristic (ROC) curves and KM plot curves were used to validate the prediction model.

The Log-rank (Mantel–Cox) test was used for survival analysis by GraphPad Prism 7.0. SPSS Statistics 24.0 software calculated HRs and 95% confidence intervals (CI). Paired and unpaired t test were used to assess the difference of genes' expression by GraphPad Prism 7.0. Differences were considered statistically significant when the P -value was < 0.05 .

Results

Primary screening of hallmark gene sets by GSEA

In TCGA, 422 LC and adjacent normal tissues performed RNA-Seq, and in a total of 29,226 mRNA have expression data. GSEA was used to screen out valuable hallmark gene sets for LC patients. Finally, fourteen abnormal hallmark gene sets in tumor tissue versus adjacent normal tissue were screened out, and three hallmark gene sets, including E2F targets, G2M checkpoint and DNA repair hallmark gene sets, with $|NES| > 1$, NOM p -val < 0.05 and FDR q -val < 0.25 (Fig. 1a–c and Table 2). Genomic instability is an important hallmark of cancer, and facilitates the transformation of cancer [6]. Recent epidemiological studies have reported that two-thirds of cancers are caused by mistakes in DNA replication [23, 24]. Thus, we select the DNA repair hallmark gene set, which include 141 mRNAs, for further research.

Identification of prognostic genes related to DNA repair

The results of GSEA showed that a total of 141 genes were obtained in the DNA REPAIR gene sets. We used

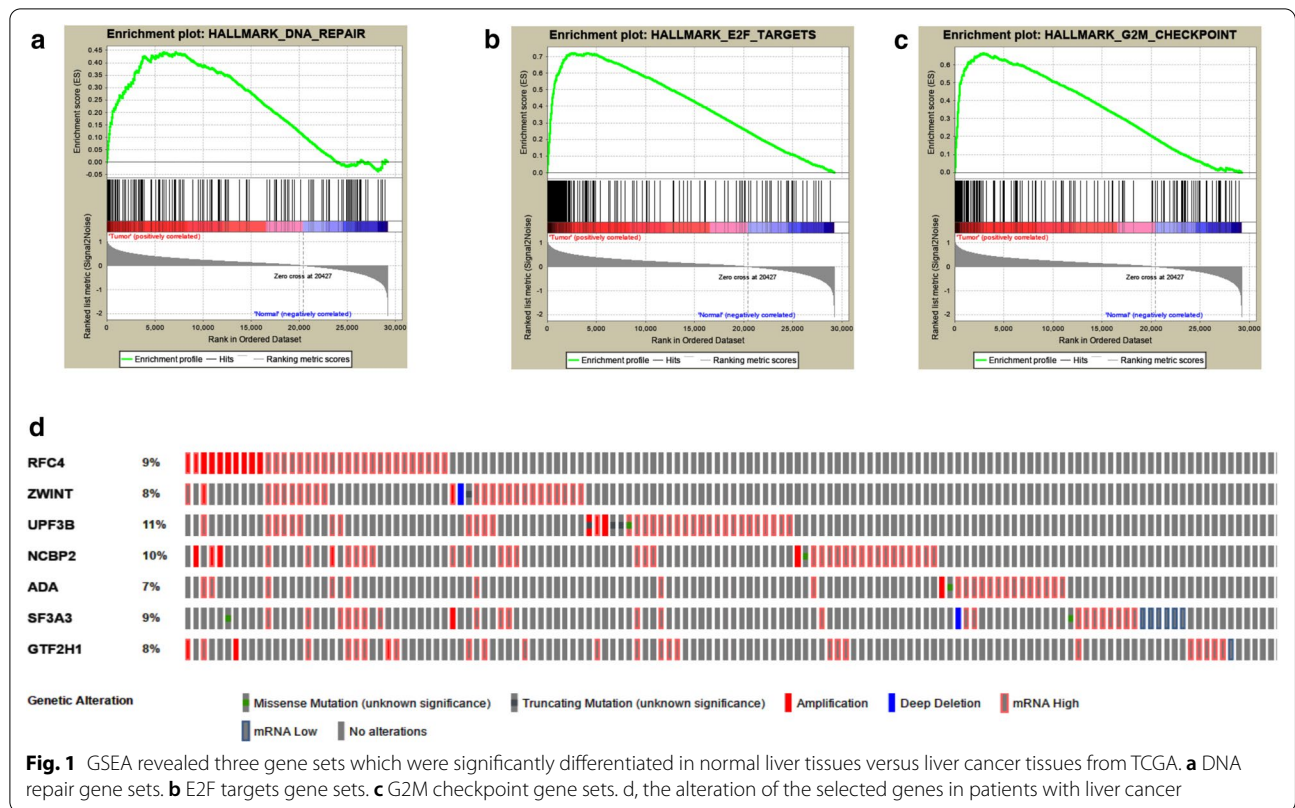


Table 2 Hallmark gene sets were enriched in normal liver tissues versus liver cancer tissues from TCGA

GS follow link to MSigDB	SIZE	NES	NOM p-val	FDR q-val	RANK AT MAX
E2F TARGETS	197	2.071552	0.001961	0.007023	4039
G2M CHECKPOINT	195	2.028344	0.005693	0.005554	2901
DNA REPAIR	141	1.649191	0.033333	0.128221	7117

univariate Cox regression analysis to select prognostic genes with P value of <0.001 from DNA REPAIR hallmark gene sets, and 12 mRNA were screened out. Seven independent LC prognostic indicators, including RFC4, ZWINT, UPF3B, NCBP2, ADA, SF3A3 and GTF2H1, were confirmed by multivariate Cox regression analysis (Table 3). And, some of them have been reported to affect the development of various tumors. ADA encodes an enzyme that catalyzes the hydrolysis of adenosine to inosine in the purine catabolic pathway and elevated in various diseases [25]. RFC4 is involved in DNA replication as a clamp loader, and its dysregulation related to the prognosis of patients with liver cancer [26, 27]. Overexpression of ZWINT predicts poor prognosis and promotes the proliferation of hepatocellular carcinoma by regulating cell-cycle-related proteins [28]. SF3A3 encodes subunit 3 of the splicing factor 3a protein complex and act as a novel DNA repair-related prognostic signature

Table 3 The detailed information of multivariate COX survival analysis of the selected prognostic mRNA

gene	Ensemble ID	HR	B(COX)	P value
RFC4	ENSG00000163918	1.406482	-0.3093	0.000546
ZWINT	ENSG00000122952	1.359246	0.2752	5.79E-05
UPF3B	ENSG00000125351	1.655998	0.3289	0.000148
NCBP2	ENSG00000114503	2.396939	0.6654	2.02E-05
ADA	ENSG00000196839	1.41584	0.2153	0.000757
SF3A3	ENSG00000183431	2.234263	0.5587	6.27E-05
GTF2H1	ENSG00000110768	2.573995	0.3868	5.18E-05

in patients with hepatocellular carcinoma [29]. However, in our study, the results of multivariate Cox regression analysis showed that GTF2H1 has the greatest impact on OS of patients with liver cancer. Moreover, it has been

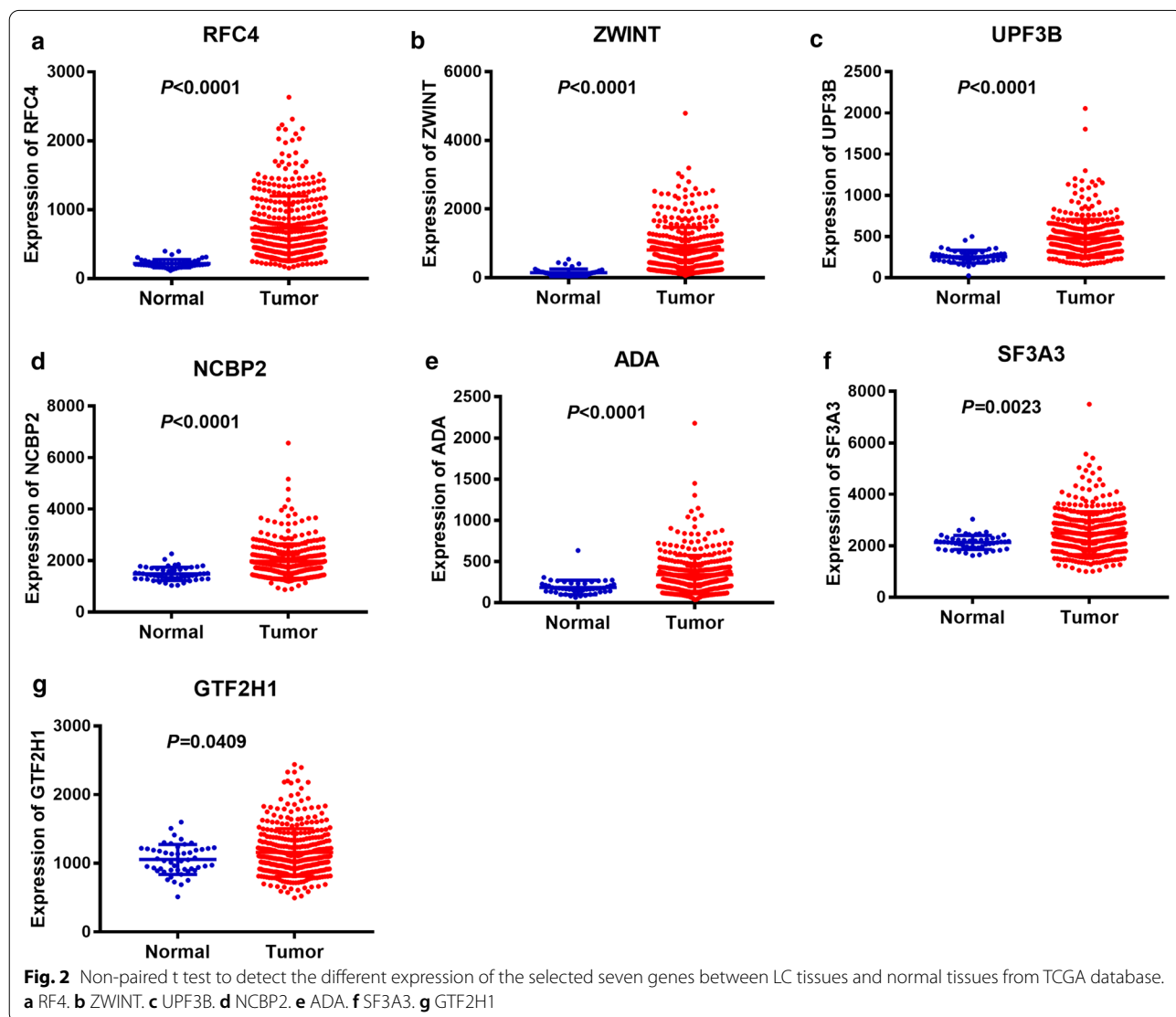
reported to strengthen the genetic susceptibility of lung cancer [30].

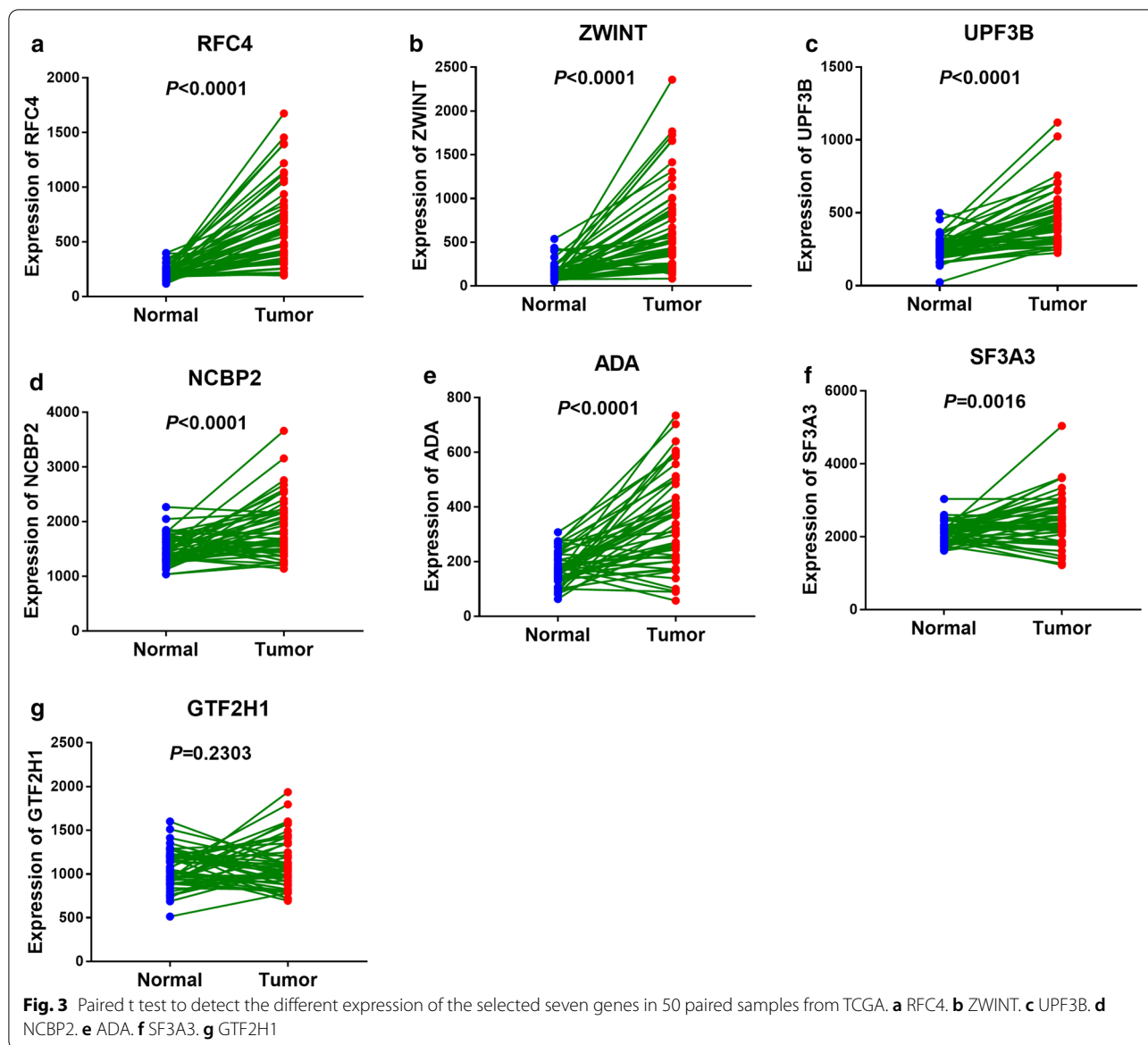
Further, KM plot online analysis was adopted to analyze the effect of the seven mRNA on the prognosis of patients with liver cancer. The results of KM plot online analysis in accord with Cox regression analysis, LC patients with high expression of these seven genes have more poorly prognosis (P value < 0.05, Additional file 1: Figure S1).

The alternation of the selected panel mRNA signature was detected by cBioPortal database. Results showed that the 7 genes screened by us dysregulated in 36% (131 out of 363) of LC patients. UPF3B has the highest rate among all the alternations (Fig. 1d). UPF3B encodes a protein that is part of a post-splicing multi-protein complex involved in both mRNA nuclear export and mRNA surveillance, and it has been reported to

have an effect on the prognosis of patients with liver cancer [29].

Furthermore, we analyzed the differential expression of the seven genes between LC tissues and normal tissues resort to TCGA and GEO database. Unpaired t test showed that all the seven genes we screened have a higher expression in LC tissues versus adjacent normal tissues with a significant differentiation (P < 0.05, Fig. 2 and Additional file 2: Figure S2). In addition, we detected the expression of the seven genes in 50 paired samples from TCGA and 7 paired samples from GEO by paired t test. The results of paired t test showed that all the genes' expression between LC tissues and adjacent normal tissues had statistical differences except GTF2H1 (P < 0.05, Fig. 3 and Additional file 3: Figure S3).





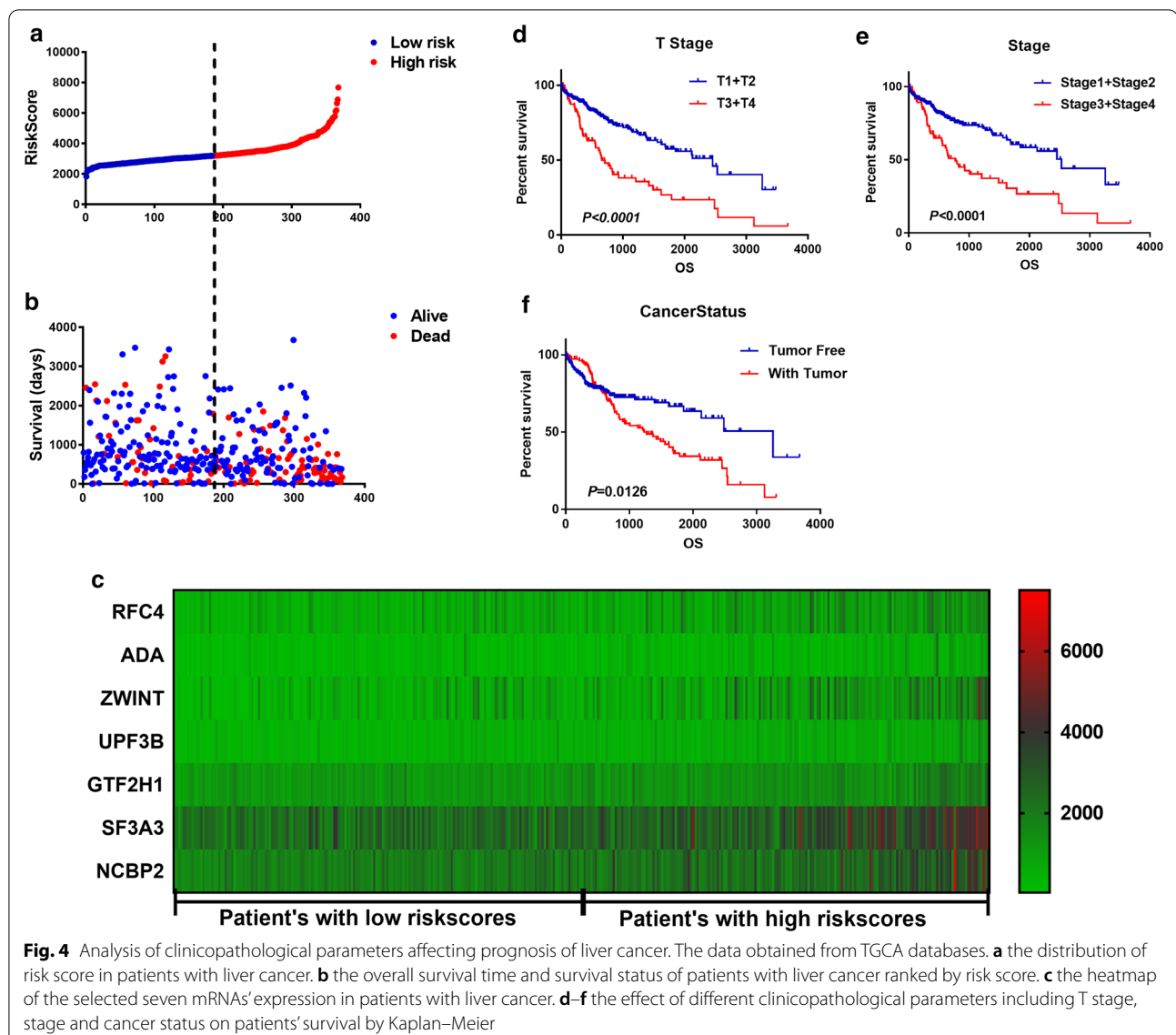
Construction of seven mRNA signature panel for predicting the prognosis of LC patients

The expression level of the selected genes is weighted and then linearly integrated with the regression coefficients obtained by multivariate Cox regression analysis. Risk score = $-0.3093 \times \text{expression of RFC4} + 0.2153 \times \text{expression of ADA} + 0.3289 \times \text{expression of UPF3B} + 0.2752 \times \text{expression of ZWINT} + 0.3868 \times \text{expression of GTF2H1} + 0.5587 \times \text{expression of SF3A3} + 0.6654 \times \text{expression of NCBP2}$. The unique risk value for each LC patient was calculated based on the risk score formula. Among of them, NCBP2 has the maximum weighting coefficient and has been reported

to interfere with the drug sensitivity of platinum in non-small cell lung cancer [31]. Then, LC patients were divided into low-risk group and high-risk group according to the median value of the risk score (Fig. 4a), and we found that the low-risk group had longer overall survival time and fewer deaths than the high-risk group (Fig. 4b). Moreover, the expression of the selected seven genes in LC patients ascended along with the increase of risk score (Fig. 4c).

Risk score act as an independent prognostic factor for LC

Univariate and multivariate COX regression analysis were performed to confirm the factors affected the prognosis of LC patients. Univariate analysis showed



that risk score, new event time, T stage, stage and cancer status were significantly associated with the overall survival time of LC patients. In addition, multivariate COX regression analysis showed that risk score and new event time were independent predictors for the overall survival time of LC patients. Risk score is an important clinicopathological parameter affecting the prognosis of LC patients. It is not only an independent prognostic factor for LC patients, moreover, the mortality rate of patients with high-risk score is 2.184 times as high as that of patients with low-risk score (Table 4). Besides, Kaplan–Meier curves and the log-rank analysis were then performed to verify the results of COX regression analysis. According to the curves, patients who has a high-risk score, who had later T stage and clinical stage and who

are with tumor had poorer prognosis (Figs. 4d–f, 5b). Consequently, the results were self-consistent, and this observation further proved the accuracy of our analysis.

Validation of the risk score derived from mRNAs signature

In order to verify the validity of the risk score derived from the seven selected mRNAs, ROC curve and K-M plot curves were adopted. LC patients were divided into two groups according to the median value, then perform a statistical analysis and constructed the ROC curve. The area under the ROC curve is 0.788, and this result indicates that the risk score has high sensitivity and specificity in predicting the prognosis of LC patients (Fig. 5a). Further, we analyzed the survival of the two groups by K-M plot curve, and the prognosis of patients with a

Table 4 Univariate and multivariate Cox analysis to confirm the factors affected the prognosis of patients with liver cancer in TCGA

Factor	Univariate analysis		P value	Multivariate analysis		P value
	HR	95% CI of HR		HR	95% CI of HR	
Risk score (Low/High)	2.184	1.523–3.132	<0.0001	2.161	1.302–3.586	0.003
Age	1.012	0.999–1.026	0.076			
Gender (male/female)	0.82	0.575–1.168	0.271			
New event time	0.998	0.997–0.998	<0.0001	0.998	0.997–0.998	<0.0001
Tumor topography (T1 + T2/T3 + T4)	2.524	1.774–3.591	<0.0001	1.251	0.162–9.635	0.83
Stage (stage1 + 2/stage3 + 4)	2.432	1.678–3.525	<0.0001	0.806	0.105–6.168	0.835
Cancer status (tumor free/with tumor)	1.577	1.099–2.262	0.013	0.611	0.355–1.051	0.075

low-risk score is much better than that of high-risk score ($P < 0.0001$, Fig. 5b).

Further, a stratified analysis of LC patients' prognosis was performed to confirm the validity of the risk score. As shown in the K–M plot curves, including clinicopathological parameters of gender (male or female), age (≤ 61 or > 67), person neoplasm cancer status (tumor free or with tumor), new tumor event (no or yes), relative family cancer history (no or yes) and race (Asian or White), patients in high-risk group had poorer prognosis than that in low-risk group ($P < 0.05$, Fig. 5c–h). Besides, the expression of the key genes in the risk model was validated by quantitative real-time PCR in clinical tissue specimens, and we found that all of the genes in the model upregulated in liver cancer tissues versus normal liver tissues, which was consistent with the results predicted (Fig. 5i). These results strongly demonstrate that the risk score derived from the seven mRNAs was a stable and accurate prognostic marker for LC.

Discussion

Liver cancer (LC) is one of the most common cause of cancer-related death throughout the world [1]. Early diagnosis and novel systemic therapies, including drugs, gene and immune therapies are key factors affecting the prognosis of patients with liver cancer. Previous studies reported that the combination between sorafenib and long acting octreotide is active and well tolerated in patients with advanced hepatocellular carcinoma [32]. Recently, many prognostic non-coding gene expression signatures have been applied to diagnose and predict prognosis of liver cancer [33]. Establishing a specific model by combining multiple genes can identify genes with highly stable diagnostic and prognostic characteristics, and it is a valuable therapeutic target [34]. Moreover, multiple mRNA prognostic sets showed to have a better accuracy in cancer prognosis than non-coding prognostic genes [21, 22]. Consequently, the establishment of

prognostic characteristics panel by combining multiple genes is an important step in achieving precise targeted therapy for liver cancer. However, there are few reports on the analysis of prognostic-related mRNA signatures in liver cancer. Hence, we analyzed the expression of mRNA in liver cancer patients and identified valuable hallmark gene sets, including E2F targets, G2M checkpoint and DNA repair hallmark gene sets, by GSEA.

Recent epidemiological studies have shown that two-thirds of cancers are caused by DNA replication errors [10, 23, 24]. Especially, errors in the replication of mRNA, such as mutations in the suppressor gene P53, are particularly important for the development of cancer [35, 36]. Consequently, mutations of DNA repair genes are particularly important in the development of cancer. For example, recent study showed that DNA Repair inhibitor DT01 acts as a novel therapeutic target in colorectal liver metastasis [37]. In our study, we found that DNA repair gene set plays an important role in liver cancer and further analyzed this gene set. In DNA repair gene set, 12 mRNAs are confirmed with $P < 0.001$ by univariate Cox regression analysis. Moreover, among of them, RFC4, ZWINT, UPF3B, NCBP2, ADA, SF3A3 and GTF2H1 are independent prognostic indicators of liver cancer. Unpaired and paired t test showed that all the seven screened genes have a higher expression in LC tissues versus normal tissues with a significant differentiation except GTF2H1 ($P < 0.05$, Figs. 2 and 3).

A marker of multiple genes signature has a more powerful and precise prognostic ability to predict LC patients' prognosis and maybe an effective classification tool for patients with liver cancer [38, 39]. Therefore, we established a prognostic model for liver cancer using the seven selected genes and divided the patients into two groups by the risk score derived from the prognostic model. We found that the expression of the selected seven genes ascends along with the increase of the risk score and patients in low-risk score group had longer

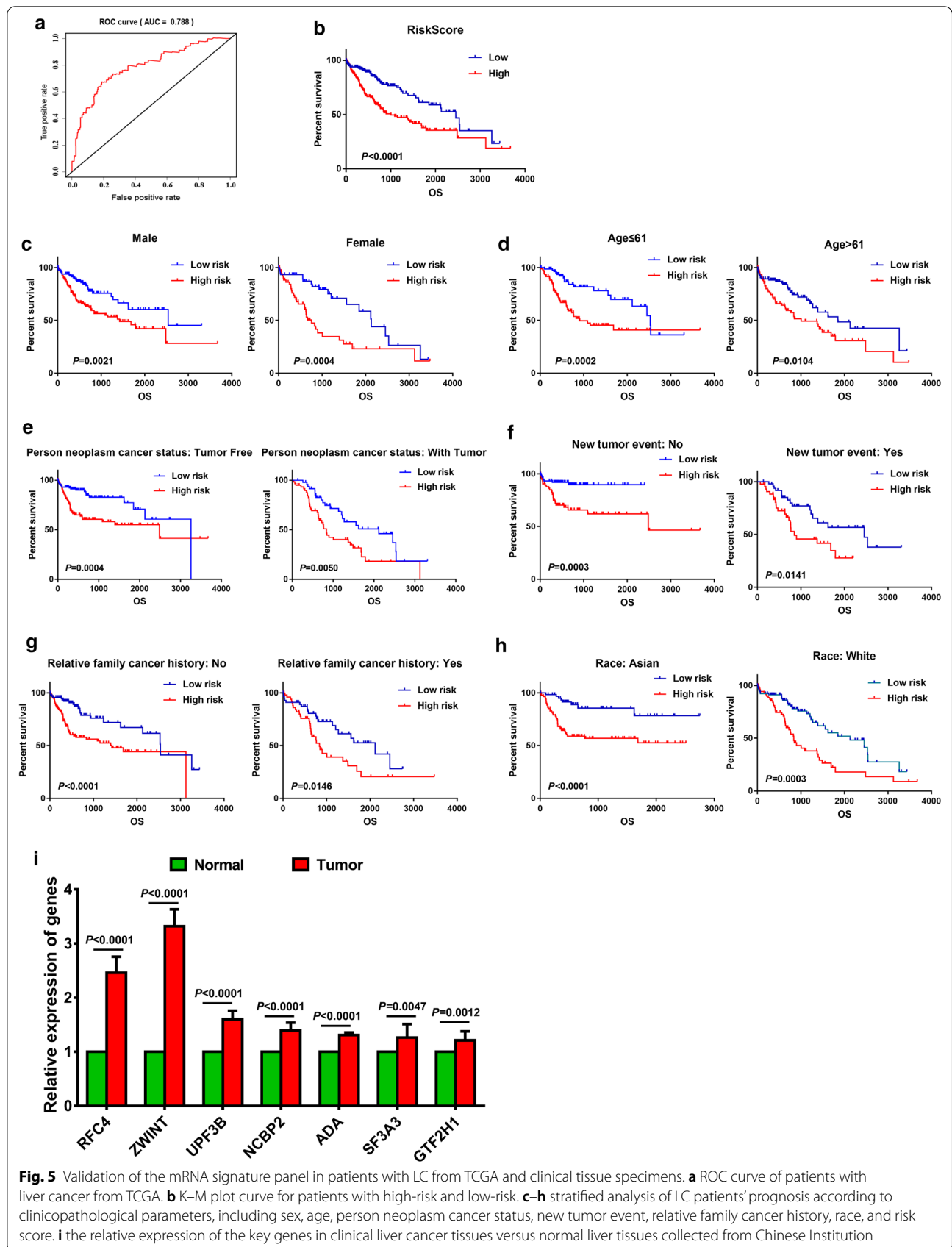


Fig. 5 Validation of the mRNA signature panel in patients with LC from TCGA and clinical tissue specimens. **a** ROC curve of patients with liver cancer from TCGA. **b** K-M plot curve for patients with high-risk and low-risk. **c-h** stratified analysis of LC patients' prognosis according to clinicopathological parameters, including sex, age, person neoplasm cancer status, new tumor event, relative family cancer history, race, and risk score. **i** the relative expression of the key genes in clinical liver cancer tissues versus normal liver tissues collected from Chinese Institution

overall survival time and fewer deaths than in high-risk group (Fig. 4b, c).

In addition, the prognostic model of the seven mRNAs signatures was verified by ROC curve and stratified survival analysis. The stratified analysis of different pathological parameters, including sex, age, person neoplasm cancer status, new tumor event, relative family cancer history and race, showed that the risk score derived from the seven mRNA signatures could make a distinction between LC patients with different pathological parameters. Moreover, LC patients in low-risk score group have a longer overall survival time. The results in the stratified analysis are all in accord with integral analysis, which prove the reliability of the risk score derived from the seven mRNA signatures for predicting liver cancer prognosis.

Conclusions

In summary, the signature of the seven DNA repair-related genes is closely related to the prognosis of patients with liver cancer. And, the model based on the seven genes can be act as an effective tool to predict the prognosis of the LC patients and provide a reference for clinical risk level of the LC patients (Additional files 4, 5, 6).

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12935-021-01779-1>.

Additional file 1: Figure S1. KM plot curves of the selected seven gene by KM plot online analysis. a, RF4. b, ADA. c, ZWINT. d, UPF3B. e, GTF2H1. f, SF3A3. g, NCBP2.

Additional file 2: Figure S2. Non-paired t test to detect the different expression of the selected seven genes between LC tissues and normal tissues from GEO database. a, RF4. b, ZWINT. c, UPF3B. d, NCBP2. e, ADA. f, SF3A3. g, GTF2H1.

Additional file 3: Figure S3. Paired t test to detect the different expression of the selected seven genes in 7 paired samples from GEO. a, RF4. b, ZWINT. c, UPF3B. d, NCBP2. e, ADA. f, SF3A3. g, GTF2H1.

Additional file 4. The raw data of GSE101685 from GEO database.

Additional file 5. The raw data of GSE101728 from GEO database.

Additional file 6. The raw data from TCGA database and the raw data of qRT-PCR.

Abbreviations

LC: Liver cancer; TCGA: The Cancer Genome Atlas; GSEA: Gene set enrichment analysis; OS: Overall survival; ROC: Receiver operating characteristic; K-M: Kaplan–Meier; HR: Hazard ratio; mRNA: Messenger RNA; RFC4: Replication factor C subunit 4; ZWINT: ZW10 interacting kinetochore protein; UPF3B: UPF3B regulator of nonsense mediated mRNA decay; NCBP2: Nuclear cap binding protein subunit 2; ADA: Adenosine deaminase; SF3A3: Splicing factor 3a subunit 3; GTF2H1: General transcription factor IIH subunit 1; GAPDH: Glyceraldehyde-3-phosphate dehydrogenase.

Acknowledgements

We would like to acknowledge TCGA for free use.

Authors' contributions

ML, QLZ and XC downloaded the LC data from TCGA; WJZ, YCL and MX Y analyzed the data and drew the plot; WJZ and YCL wrote the manuscript. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The datasets analyzed during the current study are available in the TCGA repository, <https://cancergenome.nih.gov/>.

Ethics approval and consent to participate

All patients consented to the institutional review board which allows comprehensive analysis of tumor specimens (Ethics committee of the Qingdao Municipal Hospital, Ref.: 2019 LSZ No.109).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Department of Pharmacy, School of Medicine, Qingdao Municipal Hospital, Qingdao University, Qingdao 266011, Shandong, China. ² Department of Pulmonary Medicine, School of Medicine, Qingdao Municipal Hospital, Qingdao University, Qingdao 266011, Shandong, China. ³ Department of Orthopedics and Sports Medicine and Joint Surgery, Qingdao Municipal Hospital, Qingdao, Shandong, China. ⁴ Department of Pharmacy, Women and Children's Hospital, Qingdao, Shandong, China.

Received: 21 July 2020 Accepted: 20 January 2021

Published online: 30 January 2021

References

- Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424.
- Fitzmaurice C, Dicker D, Pain A, et al. The Global Burden of Cancer 2013. *JAMA Oncol*. 2015;1(4):505–27.
- Bai L, Liu Z, Fang Q, et al. The trends and projections in the incidence and mortality of liver cancer in urban Shanghai: a population-based study from 1973 to 2020. *Clin Epidemiol*. 2018;10:277–88.
- Njei B, Rotman Y, Ditch I, et al. Emerging trends in hepatocellular carcinoma incidence and mortality. *Hepatology*. 2015;61(1):191–9.
- Chen X, Liu HP, Li M, et al. Advances in non-surgical management of primary liver cancer. *World J Gastroenterol*. 2014;20(44):16630–8.
- Kitao H, Iimori M, Kataoka Y, et al. DNA replication stress and cancer chemotherapy. *Cancer Sci*. 2018;109(2):264–71.
- Stiuso P, Potenza N, Lombardi A, et al. MicroRNA-423-5p promotes autophagy in cancer cells and is increased in serum from hepatocarcinoma patients treated with sorafenib. *Mol Ther Nucleic Acids*. 2015;4:e233.
- Caraglia M, Giuberti G, Marra M, et al. Oxidative stress and ERK1/2 phosphorylation as predictors of outcome in hepatocellular carcinoma patients treated with sorafenib plus octreotide LAR. *Cell Death Dis*. 2011;2(4):e150.
- Lombardi A, Grimaldi A, Zappavigna S, et al. Hepatocarcinoma: genetic and epigenetic features. *Minerva Gastroenterol Dietol*. 2018;64(1):14–27.
- Tomasetti C, Li L. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science*. 2017;355(6331):1330–4.
- Chang WH, Lai AG. Transcriptional landscape of DNA repair genes underpins a pan-cancer prognostic signature associated with cell cycle dysregulation and tumor hypoxia. *DNA Repair (Amst)*. 2019;78:142–53.

12. Tolan HK, Tozan-Beceran A, Sardas S, et al. DNA repair capacity of the colorectal cancer patients and the correlation between the pathological parameters. *Niger J Clin Pract.* 2019;22(2):194–200.
13. Damia G, Broggini M. Platinum resistance in ovarian cancer: role of DNA repair. *Cancers.* 2019;11(1):119.
14. Chen X, Chen F, Ren Y, et al. IL-6 signaling contributes to radioresistance of prostate cancer through key DNA repair-associated molecules ATM, ATR, and BRCA 1/2. *J Cancer Res Clin Oncol.* 2019;145:1471–84.
15. Lopez-Camarillo C, Rincon DG, Ruiz-Garcia E, et al. DNA repair proteins as therapeutic targets in ovarian cancer. *Curr Protein Pept Sci.* 2019;20(4):316–23.
16. Wang R, Ye XH, Zhao XL, et al. Development of a five-gene signature as a novel prognostic marker in ovarian cancer. *Neoplasma.* 2018;66:343–9.
17. Zheng T, Pang Z, Zhao Z. A gene signature predicts response to neoadjuvant chemotherapy in triple-negative breast cancer patients. *Biosci Rep.* 2019. <https://doi.org/10.1042/BSR20190414>.
18. Wei H, Li J, Xie M, et al. Comprehensive analysis of metastasis-related genes reveals a gene signature predicting the survival of colon cancer patients. *PeerJ.* 2018;6:e5433.
19. Kim K, Jeon S, Kim TM, et al. Immune gene signature delineates a subclass of papillary thyroid cancer with unfavorable clinical outcomes. *Cancers.* 2018;10(12):494.
20. Cai W, Li Y, Huang B, et al. Esophageal cancer lymph node metastasis-associated gene signature optimizes overall survival prediction of esophageal cancer. *J Cell Biochem.* 2019;120(1):592–600.
21. Zuo S, Dai G, Ren X. Identification of a 6-gene signature predicting prognosis for colorectal cancer. *Cancer Cell Int.* 2019;19(6):1–15.
22. Sun LL, Wu JY, Wu ZY, et al. A three-gene signature and clinical outcome in esophageal squamous cell carcinoma. *Int J Cancer.* 2015;136(6):E569–77.
23. Gold B. Somatic mutations in cancer: Stochastic versus predictable. *Mutat Res.* 2017;814:37–46.
24. Perduca V, Alexandrov LB, Kelly-Irving M, et al. Stem cell replication, somatic mutations and role of randomness in the development of cancer. *Eur J Epidemiol.* 2019;34(5):439–45.
25. Nakajima Y, Kuribayashi K. Adenosine deaminase in pleural effusion and its relationship with clinical parameters in patients with malignant pleural mesothelioma. *Cancer Investig.* 2020;38(6):356–64.
26. Kong J, Wang T, Zhang Z, et al. Five core genes related to the progression and prognosis of hepatocellular carcinoma identified by analysis of a coexpression network. *DNA Cell Biol.* 2019;38(12):1564–76.
27. Xiang J, Fang L, Luo Y, et al. Levels of human replication factor C4, a clamp loader, correlate with tumor progression and predict the prognosis for colorectal cancer. *J Transl Med.* 2014;12:320.
28. Ying H, Xu Z, Chen M, et al. Overexpression of Zwint predicts poor prognosis and promotes the proliferation of hepatocellular carcinoma by regulating cell-cycle-related proteins. *Onco Targets Ther.* 2018;11:689–702.
29. Li N, Zhao L, Guo C, et al. Identification of a novel DNA repair-related prognostic signature predicting survival of patients with hepatocellular carcinoma. *Cancer Manag Res.* 2019;11:7473–84.
30. Wu W, Liu H, Lei R, et al. Genetic variants in GTF2H1 and risk of lung cancer: a case-control analysis in a Chinese population. *Lung Cancer.* 2009;63(2):180–6.
31. Nastase A, Lupo A, Laszlo V, et al. Platinum drug sensitivity polymorphisms in stage III non-small cell lung cancer with invasion of mediastinal lymph nodes. *Cancer Genomics Proteomics.* 2020;17(5):587–95.
32. Prete SD, Montella L, Caraglia M, et al. Sorafenib plus octreotide is an effective and safe treatment in advanced hepatocellular carcinoma: multicenter phase II SoLAR study. *Cancer Chemother Pharmacol.* 2010;66(5):837–44.
33. Itzel T, Spang R, Maass T, et al. Random gene sets in predicting survival of patients with hepatocellular carcinoma. *J Mol Med (Berl).* 2019;97(6):879–88.
34. Lee JS, Chu IS, Heo J, et al. Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. *Hepatology.* 2004;40(3):667–76.
35. Poulos RC, Wong YT, Ryan R. Analysis of 7815 cancer exomes reveals associations between mutational processes and somatic driver mutations. *PLoS Genet.* 2018;14(11):e1007779.
36. Levine AJ, Puzio-Kuter AM, Chan CS, et al. The role of the p53 protein in stem-cell biology and epigenetic regulation. *Cold Spring Harb Perspect Med.* 2016;6(9):a026153.
37. Herath NI, Devun F, Lienafa MC, et al. The DNA repair inhibitor DT01 as a novel therapeutic strategy for chemosensitization of colorectal liver metastasis. *Mol Cancer Ther.* 2016;15(1):15–22.
38. Yan J, Zhou C, Guo K, et al. A novel seven-lncRNA signature for prognosis prediction in hepatocellular carcinoma. *J Cell Biochem.* 2019;120(1):213–23.
39. Kancherla V, Abdullazade S, Matter MS, et al. Genomic analysis revealed new oncogenic signatures in TP53-mutant hepatocellular carcinoma. *Front Genet.* 2018;9:2.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

