

# Identification of Multi-omics Biomarkers and Construction of the Novel Prognostic Model for Hepatocellular Carcinoma

Xiaokai Yan (✉ [YXK11011@163.com](mailto:YXK11011@163.com))

Zunyi Medical University <https://orcid.org/0000-0002-8507-3473>

Chiyang Xiao

Zunyi Medical University

Kunyan Yue

Zunyi Medical University

Min Chen

Zunyi Medical University

Hang Zhou

Zunyi Medical University

Xiao Liu

Zunyi Medical University

---

## Research

**Keywords:** hepatocellular carcinoma, prognostic model, multi-omics, TCGA, ICGC

**Posted Date:** January 6th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-452644/v3>

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

[Read Full License](#)

---

# Abstract

Genome changes play a crucial role in carcinogenesis, and many biomarkers can be used as effective prognostic indicators in various tumours. Although previous studies have constructed many predictive models for hepatocellular carcinoma (HCC) based on molecular signatures, the performance is unsatisfactory. To fill this shortcoming, we hope to build a more accurate predictive model to guide prognostic assessments of HCC. We use the TCGA to identify crucial biomarkers and construct single-omic prognostic models through difference analysis, univariate Cox, and LASSO/stepwise Cox analysis. The performances of single-omic models were evaluated and validated through survival analysis, Harrell's concordance index (C-index), and receiver operating characteristic (ROC) curve. A multi-omics model was built and evaluated by decision curve analysis (DCA), C-index, and ROC analysis. Multiple mRNAs, lncRNAs, miRNAs, CNV genes, and SNPs were significantly associated with the prognosis of HCC. Five single-omic models were constructed, and the mRNA and lncRNA models showed good performance with c-indexes over 0.70. The multi-omics model presented a quite predictive solid ability with a c-index over 0.80. In this study, we identified many biomarkers that may help study underlying carcinogenesis mechanisms in HCC. In addition, we constructed multiple single-omic models and an integrated multi-omics model that may provide practical and reliable guides for prognosis assessment and treatment decision-making.

## Background

Liver cancer is one of the most prevalent human malignancies globally, seriously threatening people's lives and health<sup>1</sup>. Hepatocellular carcinoma (HCC) is the predominant liver cancer, accounts for 70–85% of cases<sup>2</sup>. In different populations, the 5-year survival rate varies greatly, with an average rate of about 35%<sup>3-6</sup>. HCC is a highly heterogeneous tumor, and its pathogenesis is quite complicated. Besides, the patients' outcome is influenced by many factors, such as heredity, environment, and infection. These make the prognosis prediction very challenging<sup>7,8</sup>. Therefore, it is necessary and urgent to develop a robust and practical prognostic evaluation model for HCC.

Previous research has shown that genome changes play an essential role in various tumour-related biological processes such as cellular proliferation and differentiation, angiogenesis, stemness, cancer metabolism, immune response, migration, invasion and metastasis<sup>8,9</sup>. Besides, many biomarkers exhibited good prognostic predictive value<sup>10,11</sup>. For example, LMO1 was a critical oncogene that promotes neuroblastoma initiation, progression, and widespread metastatic dissemination<sup>12</sup>. lncRNA SNHG10 was associated with poor overall survival of HCC, while influenced the cell proliferation, invasion, migration, cell cycle and epithelial-mesenchymal transition<sup>13</sup>. miR-487a could enhance the proliferation and metastasis of HCC cells by directly binding to sprouty-related EVH1 domain containing 2 (SPRED2) or phosphoinositide-3-Kinase regulatory subunit 1 (PIK3R1) and can be used as a potential prognostic marker<sup>14</sup>. [Vladimir Bezrookove](#) et al. have proved the vital role of PHIP copy-number elevation as a prognostic and progression marker for cutaneous melanoma<sup>15</sup>. SNP in 3'UTR of RAS-related

proteins (RAP1A) was significantly associated with esophageal squamous cell carcinoma risk and metastasis<sup>16</sup>. With the continuous discovery of vital biomarkers in various cancers and to make up for the inadequacy of traditional predictive models based on clinicopathological characteristics, an increasing number of studies are devoted to building predictive models based on genomics.

At present, many HCC prediction models based on biomarkers have been reported<sup>17-23</sup>. However, most of them are single-omic models built with RNA-sequence or DNA methylation, which may decrease prediction performance due to the lack of other omic information. For example, Long J et al.' study<sup>21</sup> reported a four-gene-based prognostic model for HCC with a C-index of 0.65. Even adding the age and pathologic stage information, the C-index is less than 0.70 (0.69). Such a prediction ability is not excellent. To our knowledge, only Chaudhary K et al.<sup>24</sup> built a multi-omics predictive model with mRNA, miRNA and DNA methylation, which showed a better power than Long J et al.' research. Therefore, the predictive model build by multi-omics data should be more robust and efficient. However, Chaudhary K et al.' model is still not very prominent, with a C-index of only 0.70.

The underlying mechanisms of cancer are often complex, and the multi-omics approach is valuable for revealing the pathogenic mechanism and evaluating the prognosis<sup>25</sup>. Our study tried to construct novel and accurate models through omics features analysis based on mRNA, lncRNA, miRNA, SNP and CNV. This study might promote the prognostic prediction and treatment decision-making of HCCs.

## Results

### 2.1 Construction and validation of mRNA model

320 HCC samples with complete mRNA expression profiling and survival information were kept as a training set. 267 DE-mRNAs (including 184 up-regulated and 83 down-regulated mRNAs in HCC) (Figure S1A, 1B, and Table S1) were selected for univariate Cox regression analysis. Among these DE-mRNAs, 124 mRNAs were significantly associated with OS (Table S2). Then 91 mRNAs with HR > 1 and up-regulated in HCC, and 31 mRNAs with HR < 1 and down-regulated in HCC were analysed with LASSO Cox (Figure 2A). Parameter  $\log(\lambda) = -3.154$  ( $\lambda = 0.04269$ ) chosen by the 10-fold cross-validation method with minimum criteria was regarded as the best value (Figure S1C). Sixteen key mRNAs with nonzero coefficients (Figure 2B) were selected to build the mRNA model (Figure S1F). All were associated with OS (Figure S1D) and significantly changed in HCC samples (Figure S1E1, S1E2). The mRNA risk score for each patient was computed: mRNA risk score =  $\sum \beta_i \times \text{exp-mRNA}$ , where exp-mRNA is the expression level of key mRNA and  $\beta$  is the regression coefficient derived from the LASSO COX analysis (Table S9). The mRNA model was evaluated with C-index, ROC curve, and survival analysis (Figure 2C, 2D), which showed a relatively good predictive ability (C-index = 0.752). 160 HCC samples were randomly selected as a test set to validate the mRNA model, and good performance was observed (C-index = 0.744) (Figure 2E, 2F). The mRNA model was externally verified in the LIRI-JP with survival analysis, which also showed decent performance (Figure 2G).

## 2.2 Construction and validation of lncRNA model

320 HCC samples with complete lncRNA expression profiling and survival information were retained as a training set. 540 up-regulated and 27 down-regulated lncRNAs in HCC (Figure S2A, 2B, and Table S3) were used for Cox regression analysis (Table S4). 146 key lncRNAs were selected for LASSO COX analysis (Figure 3A). Parameter  $\log(\lambda) = -2.99$  ( $\lambda = 0.05027$ ) chosen by the 10-fold cross-validation method with minimum criteria was regarded as the best value (Figure S2C). Twenty key lncRNAs with nonzero coefficients (Figure 3B) were associated with OS (Figure S2D) and significantly changed in HCC samples (Figure S2E1, S2E2) and were used to build the lncRNA model (Figure S2F). The lncRNA risk score for each patient was computed:  $\text{lncRNA risk score} = \sum \beta_i \times \text{exp-lncRNA}$ , where  $\text{exp-lncRNA}$  is the expression level of key lncRNA and  $\beta$  is the regression coefficient derived from the LASSO Cox analysis (Table S9). In the training set, the AUC of the lncRNA model at 1, 3, and 5 years OS was 0.854, 0.796, and 0.806, respectively, while the C-index was 0.753 (Figure 3C). In the test set, the AUC at 1, 3, and 5 years OS was 0.807, 0.785, and 0.767, respectively, while the C-index was 0.727 (Figure 3E). In addition, the log-rank analysis revealed that scoring using the lncRNA risk score could discriminate the risk groups in the training set and test set ( $p$ -value  $< 0.0001$ ) (Figure 3D, 3F).

## 2.3 Construction and validation of miRNA model

321 HCC samples with complete miRNA and survival information were retained as a training set. Sixteen up-regulated, and seventy down-regulated miRNAs in HCC were identified (Figure S3A, 3B, and Table S5). Ten miRNAs in HCC were used (Figure 4A, Table S6) for the stepwise Cox analysis, and five key miRNAs were selected to build the miRNA model (Figure 4B, S3C, and S3D). The miRNA risk score for each patient was computed:  $\text{miRNA risk score} = \sum \beta_i \times \text{exp-miRNA}$ , where  $\text{exp-miRNA}$  is the expression level of key miRNA and  $\beta$  is the regression coefficient derived from the stepwise Cox analysis (Table S9). Survival analysis showed that the high risk group has a poor outcome in the training set ( $p$ -value = 0.00037) and test set ( $p$ -value = 0.027) (Figure 4D, 4F). Besides, in the training and test set, the AUC values of the miRNA model at 1, 3, and 5-year points were all more than 0.68, and the C-index values were over 0.65 (Figure 4C, 4E).

## 2.4 Construction and validation of CNV model

324 HCC samples with complete CNV and survival information were retained as a training set. 5006 genes with different copy number alteration (Figure 5A, Table S7) were selected to perform univariate Cox regression analysis. 357 CNV genes significantly associated with OS were identified (Table S8). Then we performed LASSO Cox analysis for key CNV genes selection. Parameter  $\log(\lambda) = -2.634269$  ( $\lambda = 0.07177142$ ) chosen by the 10-fold cross-validation method with minimum criteria was regarded as the best value (Figure S4A). Five key CNV genes with nonzero coefficients (Figure 5B) were significantly different in HCC samples (Figure S4B) and associated with OS (Figure S4C), which were used to build the CNV model (Figure S4D). The CNV risk score for each patient was computed:  $\text{CNV risk score} = \sum \beta_i \times \text{CNV}$

gene status, where  $\beta$  is the regression coefficient derived from the LASSO Cox analysis (Table S9). The CNV model was evaluated with survival analysis in the training set (Figure 5D) and test set (Figure 5F), which showed a worse prognosis in the high risk group (at least one key CNV gene with copy number alteration). Moreover, the AUC values of the CNV model at 1,3 and 5 years OS were all over 0.65, and the C-index values were more than 0.63(Figure 5C, 5E).

## 2.5 Construction and validation of SNP model

313 HCC samples with complete SNP and survival information were retained as a training set. Eighty-five high-frequency SNPs (Figure 6A, Figure S5A) in HCC were selected to perform univariate Cox analysis, and ten high-frequency SNPs significantly associated with OS were identified (Figure S5B). Seven key SNPs were selected to build the SNP model through stepwise Cox analysis (Figure S5C). The SNP risk score for each patient was computed:  $\text{SNP risk score} = \sum \beta_i \times \text{SNP status}$ , where  $\beta$  is the regression coefficient derived from the stepwise Cox analysis (Table S9). In the training set, the AUC of the SNP model at 1, 3, and 5 years OS was 0.799, 0.703, and 0.745, respectively, while the C-index was 0.709(Figure 6B). In the test set, the AUC at 1, 3, and 5 years OS was 0.745, 0.660, and 0.737, respectively, while the C-index was 0.683(Figure 6D). In addition, survival analysis showed that the high risk group (at least one key SNP with non-synonymous mutation) has a poor prognosis in the training set (p-value < 0.0001), test set (p-value < 0.0001), and external validation set (p-value = 0.029) (Figure 6C, 6E, and 6F).

## 2.6 Construction and validation of multi-omics model

302 HCC samples with complete mRNA, lncRNA, miRNA, CNV, SNP, and survival information were retained as a training set. The five single-omic models were integrated through multiple Cox regression analysis to construct a multi-omics model and visualized as a nomogram (Figure 7A). A fairly good agreement was observed between the expected and observed outcomes for 1, 3, and 5 years OS in the calibration curves (Figure 7B). Whether in the training set or the test set, the AUC values of the multi-omics model at 1,3, and 5-year points were all over 0.840, while the C-index values were more than 0.800 (Figure 7C, 7H), which were significantly greater than those of the five single-omic models (all p values are less than 0.05) (Figure 7G). DCA analysis showed that the multi-omics model has a better performance in predicting prognosis than the five single-omic models (Figure 7E, 7F). In addition, we stratified patients into low, medium and high risk groups based on the total points of the nomogram (cut-off points were selected at each tertile point). We found that scoring using the nomogram effectively discriminated the risk groups in the training set and test set (p-value < 0.0001) (Figure 7D, 7I).

## Discussion

With the development of molecular biology techniques, the therapeutic, diagnostic, and predictive value of molecular targets in cancer is gradually becoming evident<sup>26</sup>. Traditional predictive models, such as TNM system<sup>27</sup>, BCLC<sup>28</sup> and CLIP stage<sup>29</sup>, mainly reflect the clinicopathological characteristics but not

paying attention to the genome changes, which are gradually unable to meet the clinical needs in prognosis evaluation. At present, many HCC prediction models based on biomarkers have been reported<sup>17-23</sup>. However, most of them are single-omic models, with C-indexes range from 0.65 to 0.72. Such predictive ability is not satisfactory. Therefore, a more accurate predictive model is needed.

This research identified sixteen key mRNAs, twenty key lncRNAs, five key miRNAs, five key CNV genes, and seven key SNPs significantly associated with the HCC prognosis. Previous researches have demonstrated that some of these critical molecules play essential roles in the occurrence, development, metastasis, and prognosis of HCC. Chen H et al.<sup>30</sup> have found that APLN could activate phosphatidylinositol 3-kinase/protein kinase B (PI3K/Akt) pathway via APLN receptor to promote HCC growth. The APLN overexpression is significantly associated with poorer survival of patients, which is in line with our analysis (Figure S1D). MYCN overexpress in clinical HCC samples, and the knockdown of which can inhibit HCC cell growth and invasion<sup>31</sup>. Also, our research further proves that overexpression of MYCN predicts a poor outcome (Figure S1D). lncRNA LINC01554 is a novel tumour suppressor that could suppress tumorigenicity in HCC via Akt/mTOR signalling pathway<sup>32</sup>. The down-regulation of LINC01554 significantly predicts worse survival<sup>33</sup> (Figure S2D). Chen S et al. have demonstrated that knockdown of PCAT6 could inhibit cell proliferation and migration in HCC<sup>34</sup>, and our analysis further found that PCAT6 was a significant prognostic risk factor (Figure S2D). miR-452-5p could be sponged by circular RNA-ABCB10, elevating expressions of neuropilin-1 (NRP1) and ABL related gene (ABL2) and leading inhibition of the HCC progression<sup>35</sup>. Therefore, miR-452-5p may act as an oncogene and was associated with poor survival (Figure 4B). PCDH9 could inhibit HCC cell proliferation by inducing cell cycle arrest at the G0/G1 phase, and the frequent deletion was observed in Lv J et al.'s<sup>36</sup> and our study (Figure 5A). Meanwhile, survival analysis in the current study further proves the tumour suppressor function of PCDH9 in HCC (Figure S4C). TP53 mutation is one of the most common mutations in human tumors<sup>37</sup>, and our research also clarified this point in HCC (Figure 6A). Moreover, the prognostic impact of TP53 mutation in HCC has been proved by the previous and current researches<sup>38</sup> (Figure S5B). All these findings above greatly enhanced the reliability of our analysis results. However, the roles of many vital molecules (e.g., GNAZ, PZP, ZNNT1, LINC00862, CNV of CCNA1, ARID1A mutation, etc.) in HCC are still unclear, and further cell and animal experiments to reveal their underlying mechanism is warranted.

Next, we constructed five single-omic predictive models, including mRNA, lncRNA, miRNA, CNV, and SNP. The performance of each single-omic model in prognostic prediction was not bad, with c-index values ranged from 0.63 to 0.75 in the training and test set. Meanwhile, we demonstrated in the separate external validation set that the mRNA and SNP risk scores are significantly prognostic factors (Figure 2G, 6F), which significantly increased the credibility and universality of our analysis results. Of course, compared with other models reported previously<sup>17-23</sup>, our single-omic models have no advantages in prognosis evaluation. Besides, we could not perform the external validation for the lncRNA, miRNA, and CNV models due to the lack of independent external public datasets, and this is a shortcoming of our study.

Given that the predictive ability of our single-omic models is not satisfactory, we constructed an integrated multi-omics model based on mRNA, lncRNA, miRNA, CNV, and SNP. The results showed that our multi-omics model has more accurate predictive power than the single-omic models. To the best of our knowledge, our multi-omics model has the most potent predictive ability compared with the previous models based on molecular markers, with a c-index over 0.80 and all AUC values at 1, 3, 5-years more than 0.84 (Figure 7C, 7H). Of course, the lack of external verification is the weakness of this model. To increase the reliability of our research findings, the collection of clinical HCC samples for verification will be the focus of our future work. However, just like the discussion above, previous research has demonstrated that many biomarkers we identified play essential roles in HCC. The mRNA and SNP models were also effective in external datasets. Therefore, our multi-omics model may have potential application value in guiding prognostic assessments and treatment decision-making.

Besides, our multi-omics model contains more than fifty biomarkers and seems difficult to apply in the clinic. This creates a problem: too many markers will lead to application difficulties, while few will reduce the accuracy. Now more and more patients are willing to use sequencing technology to understand their disease status. Therefore, we believe that this model can also be used clinically in the future.

By way of conclusion, the current study identified sixteen key mRNAs, twenty key lncRNAs, five key miRNAs, five key CNV genes, and seven key SNPs that are significantly associated with HCC prognosis, which may be helpful in studying underlying [carcinogenesis mechanisms](#) in HCC. The predictive models we constructed showed potential prognostic values, which may better guide clinicians to make prognosis assessment and treatment decision-making for HCC patients.

## Materials And Methods

### 4.1 Data acquisition

#### 4.1.1 The genome Cancer atlas (TCGA)

The [transcriptome profiling](#) (including RNA-Seq and miRNA-Seq), the masked copy number segment profiles (CNV, [copy number variation](#)), and the [single nucleotide polymorphism \(SNP\)](#) data detected with VarScan2 were obtained from TCGA (<https://cancergenome.nih.gov/>). The corresponding HCC prognostic information was obtained from cBioPortal for cancer genomics (<https://www.cbioportal.org>). For RNA-Seq data, the raw count data were normalized with the TPM (Transcripts per million) method.

#### 4.1.2 LIRI-JP dataset (Liver Cancer-RIKEN, JP project)

The mRNA expression data and corresponding prognostic information of the LIRI-JP dataset were from the ICGC data portal (The International Cancer Genome Consortium, <https://icgc.org/>). The mRNA expression data were normalized with the TPM method.

#### 4.1.3 LICA-FR dataset (Liver Cancer-FR project)

The SNP data and corresponding prognostic information of the LICA-FR dataset were from the ICGC data portal.

## 4.2 Construction and validation of prognostic models

### 4.2.1 Model based on mRNA expression

All HCC samples of the TCGA dataset were used as a training set. mRNAs expressed in over 95% of samples were retained, and the zero values in the expression matrix were replaced with the minimum non-zero value of the corresponding gene. Then the expression data were log<sub>2</sub> transformed. Differentially expressed mRNAs (DE-mRNAs) between HCC and non-tumor samples were identified via 'limma' package<sup>39</sup>, and p-value < 0.0001 and |logFC (log fold change)| > 3 were set as the cut-off criterions. Univariate Cox regression analysis was performed to identify mRNAs significantly associated with OS (Overall survival) among DE-mRNAs, and p-value < 0.05 was considered statistically significant. mRNAs with HR (hazard rate ratio) > 1 and up-regulated in HCC, as well as mRNAs with HR < 1 and down-regulated in HCC, were used for key mRNAs selection with LASSO (least absolute shrinkage and selection operator) COX analysis. 10-fold cross-validation with minimum criteria was applied to identify the optimal parameter  $\lambda$  in LASSO Cox analysis<sup>40,41</sup>, depending on which the regression coefficients of most features were shrunk towards zero, and the key mRNAs can be identified. The key mRNAs were used to build a predictive model for HCC. The mRNA risk score for each patient was computed according to the summation of mRNA expression value multiplied by the corresponding coefficient from the LASSO Cox analysis. The performance of the mRNA model in predicting OS was evaluated through survival analysis, Harrell's concordance index (C-index)<sup>42</sup>, and receiver operating characteristic (ROC) curve.

50% HCC samples in TCGA were randomly selected as a test set. Survival analysis, C-index, and ROC analysis were performed to validate the predictive ability of the mRNA model.

The LIRI-JP was used as independent, external cohorts to assess the generalizability and accuracy of the mRNA model.

### 4.2.2 Model based on lncRNA expression

The methods to construct, evaluate and validate lncRNA mode are similar to those in the mRNA model above. However, it is needed to pay attention that to get enough differently expressed lncRNAs (DE-lncRNAs) to establish a stable model, p-value < 0.0001 and |logFC| > 1.5 were set as the cut-off criterions. Meanwhile, due to the lack of an external dataset with complete lncRNA expression and corresponding prognostic information, the external verification of the lncRNA mode cannot be approached.

### 4.2.3 Model based on miRNA expression

In the TCGA training set, miRNAs expressed in over 80% of samples were retained, the zero values were processed in the same way mentioned above. 'limma' package was performed to identify differentially

expressed miRNAs (DE-miRNAs), with p-value < 0.01 and  $|\log\text{FC}| > 1.5$ . Univariate Cox regression analysis was used to identify miRNAs significantly associated with OS among DE-miRNAs, with a p-value < 0.05. Due to few miRNAs significantly associated with OS were obtained, and LASSO Cox is suitable for analysing high-dimensional data<sup>43</sup>, we used the backward stepwise Cox proportional hazard analysis<sup>44</sup> to screen critical miRNAs. Then the same methods used in the lncRNA model were performed to construct, evaluate, and validate the miRNA model. For the same reason, we cannot complete the external verification of miRNA model.

#### 4.2.4 Model based on CNV

In the TCGA training set, the segment mean value is used to reflect the CNV of DNA fragments. A segment is called a gain or loss if the segment mean value is more or less than zero. According to the GENCODE v34 annotation file (downloaded from <https://www.gencodegenes.org>) and segment mean value of DNA fragments, we identified genes with copy number variation (CNV genes) in each sample. Chi-square analysis was used to compare the statistical difference of CNV genes between HCC and non-tumour samples. Then we used the univariate Cox regression analysis to identify CNV genes significantly associated with OS. The LASSO Cox analysis was used to screen essential CNV genes and construct the CNV model. The CNV risk score for each patient was computed according to the summation of CNV gene status (non-CNV = 0; CNV = 1) multiplied by the corresponding coefficient from the LASSO Cox analysis. The evaluation and validation methods of the CNV model are the same as those in the lncRNA model. Similarly, for the same reason, we could not perform external validation for the CNV model.

#### 4.2.5 Model based on SNP

In the TCGA training set, the high-frequency SNPs (not including synonymous mutation) in HCC samples were selected to perform univariate Cox analysis. Due to few SNPs significantly associated with OS were obtained, we performed the backward stepwise Cox proportional hazard analysis to identify critical SNPs and build the SNP model. The SNP risk score for each patient was computed according to the summation of SNP status (wild = 0; mutation = 1) multiplied by the corresponding coefficient from stepwise Cox analysis. Then the same methods we used in the mRNA model were performed to evaluate and validate the SNP model. The LICA-FR dataset was used as an independent, external cohort to assess the generalizability and accuracy of the SNP model through survival analysis.

#### 4.2.6 Model based on multi-omics

We built a multi-omics model based on the mRNA, lncRNA, miRNA, SNV, and SNP risk score through multiple Cox regression analyses. Nomogram was used for the visualisation of the prediction model. We performed the survival analysis, calibration plot, C-index, ROC, and decision curve analysis (DCA) to evaluate and compare the predictive ability of the multi-omics model with the five single-omic models. We performed the same methods in the mRNA model to validate the multi-omics model in the test set. The entire workflow is shown in Figure 1.

## 4.3 Statistical analysis

We performed data processing and statistical analysis with R (<https://www.r-project.org/>, v 3.6.0). Chi-square or Fisher's exact test was used to assess differences in categorical variables. Student t-test or non-parametric Mann-Whitney test was used to detect differences in continuous variables. Volcano, box and histogram plots were performed with the R package "ggplot2". Heatmap was plotted with the R package "gplots". The survival analysis and Cox proportional hazard regression analysis were carried out on the R package "survival". The C-index, stepwise Cox analysis, and nomogram were performed with the R package "rms". LASSO Cox analysis was performed using the R package "glmnet". The ROC curve was plotted using the R package "qROC". The DAC analysis was performed using the R package "stdca.R". The summarized mutation plots were constructed using the R package "GenVisR". The circus graph was drawn using the "RCircos" package. In the TCGA, LIRI-JP, LICA-FR and GSE1898 datasets, patients died within three months, and non-HCC patients were removed.

## Declarations

## Acknowledgments

We thank the TCGA and ICGC working groups for generating public data.

## Author contributions

Conception and design of the study (XY, HZ, and XL), collection of the data used in this study (CX), data analysis and interpretation (CX, KY), writing of the initial paper (XY, MC). All authors read and approved the final manuscript.

## Funding

This work was supported by the Natural Science Foundation of Guizhou Province of China (Grant No. QKHCG [2019] 4433, Grant No. QKHJC-ZK [2021] YB468 and Grant No. QKPTRC [2019]-036).

## Disclosure statement

The authors report no conflict of interest.

# Data availability statement

The data that support the findings of this study are openly available in the TCGA (<https://cancergenome.nih.gov/>) and ICGC data portal (<https://icgc.org/>).

## References

1. Siegel, R.L., Miller, K.D., and Jemal, A., Cancer statistics, 2020. *Ca-a Cancer Journal for Clinicians*.**70**,7-30(2020).
2. Bakiri, L., et al., Liver carcinogenesis by FOS-dependent inflammation and cholesterol dysregulation. *Journal of Experimental Medicine*.**214**,1387-1409(2017).
3. Nguyen, V.T., Law, M.G., and Dore, G.J., Hepatitis B-related hepatocellular carcinoma: epidemiological characteristics and disease burden. *J Viral Hepat*.**16**,453-463(2009).
4. Chen, C.H., et al., Long-term trends and geographic variations in the survival of patients with hepatocellular carcinoma: analysis of 11,312 patients in Taiwan. *Journal of gastroenterology and hepatology*.**21**,1561-1566(2006).
5. Kulik, L. and El-Serag, H.B., Epidemiology and Management of Hepatocellular Carcinoma. *Gastroenterology*.**156**,477-491(2019).
6. McGlynn, K.A., Petrick, J.L., and El-Serag, H.B., Epidemiology of Hepatocellular Carcinoma. *Hepatology (Baltimore, Md.)*.**73 Suppl 1**,4-13(2021).
7. Marrero, J.A., Kudo, M., and Bronowicki, J.P., The challenge of prognosis and staging for hepatocellular carcinoma. *The oncologist*.**15 Suppl 4**,23-33(2010).
8. Colagrande, S., et al., Challenges of advanced hepatocellular carcinoma. *World journal of gastroenterology*.**22**,7645-7659(2016).
9. Zhao, L., Zhao, Y., He, Y., Li, Q., and Mao, Y., The functional pathway analysis and clinical significance of miR-20a and its related lncRNAs in breast cancer. *Cellular signalling*.**51**,152-165(2018).
10. Yan, X., et al., Importance of gene expression signatures in pancreatic cancer prognosis and the establishment of a prediction model. *Cancer management and research*.**11**,273-283(2019).
11. Cao, W. and Lee, H., Multi-faceted epigenetic dysregulation of gene expression promotes esophageal squamous cell carcinoma. **11**,3675(2020).
12. Zhu, S., et al., LMO1 Synergizes with MYCN to Promote Neuroblastoma Initiation and Metastasis. *Cancer cell*.**32**,310-323.e5(2017).

13. Lan, T., et al., KIAA1429 contributes to liver cancer progression through N6-methyladenosine-dependent post-transcriptional modification of GATA3. *Molecular cancer*.**18**,186(2019).
14. Chang, R.M., et al., miRNA-487a Promotes Proliferation and Metastasis in Hepatocellular Carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*.**23**,2593-2604(2017).
15. Bezrookove, V., et al., Role of Elevated PHIP Copy Number as a Prognostic and Progression Marker for Cutaneous Melanoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*.**24**,4119-4125(2018).
16. Wang, K., et al., MiR-196a binding-site SNP regulates RAP1A expression contributing to esophageal squamous cell carcinoma risk and metastasis. *Carcinogenesis*.**33**,2147-2154(2012).
17. Liu, G.M., Xie, W.X., Zhang, C.Y., and Xu, J.W., Identification of a four-gene metabolic signature predicting overall survival for hepatocellular carcinoma. *Journal of cellular physiology*.**235**,1624-1636(2020).
18. Liu, G.M., Zeng, H.D., Zhang, C.Y., and Xu, J.W., Identification of a six-gene signature predicting overall survival for hepatocellular carcinoma. *Cancer cell international*.**19**,138(2019).
19. Long, J., et al., DNA methylation-driven genes for constructing diagnostic, prognostic, and recurrence models for hepatocellular carcinoma. *Theranostics*.**9**,7251-7267(2019).
20. Long, J., et al., Development and validation of a TP53-associated immune prognostic model for hepatocellular carcinoma. *EBioMedicine*.**42**,363-374(2019).
21. Long, J., et al., A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma. *Journal of cellular and molecular medicine*.**22**,5928-5938(2018).
22. Wang, X., et al., Identification of prognostic biomarkers for patients with hepatocellular carcinoma after hepatectomy. *Oncology reports*.**41**,1586-1602(2019).
23. Wang, Z., et al., Development and validation of a novel immune-related prognostic model in hepatocellular carcinoma. *Journal of translational medicine*.**18**,67(2020).
24. Chaudhary, K., Poirion, O.B., Lu, L., and Garmire, L.X., Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*.**24**,1248-1259(2018).
25. Hu, W., et al., Multi-omics Approach Reveals Distinct Differences in Left- and Right-Sided Colon Cancer. *Molecular cancer research : MCR*.**16**,476-485(2018).

26. Connor, A.A., et al., Integration of Genomic and Transcriptional Features in Pancreatic Cancer Reveals Increased Cell Cycle Progression in Metastases. *Cancer cell*.**35**,267-282.e7(2019).
27. Lei, H.J., et al., Prognostic value and clinical relevance of the 6th Edition 2002 American Joint Committee on Cancer staging system in patients with resectable hepatocellular carcinoma. *Journal of the American College of Surgeons*.**203**,426-435(2006).
28. Llovet, J.M., Brú, C., and Bruix, J., Prognosis of hepatocellular carcinoma: the BCLC staging classification. *Seminars in liver disease*.**19**,329-338(1999).
29. A new prognostic system for hepatocellular carcinoma: a retrospective study of 435 patients: the Cancer of the Liver Italian Program (CLIP) investigators. *Hepatology (Baltimore, Md.)*.**28**,751-755(1998).
30. Chen, H., et al., APLN promotes hepatocellular carcinoma through activating PI3K/Akt pathway and is a druggable target. *Theranostics*.**9**,5246-5260(2019).
31. Yasukawa, K., et al., MicroRNA-493-5p-mediated repression of the MYCN oncogene inhibits hepatic cancer cell growth and invasion. **111**,869-880(2020).
32. Zheng, Y.L., et al., LINC01554-Mediated Glucose Metabolism Reprogramming Suppresses Tumorigenicity in Hepatocellular Carcinoma via Downregulating PKM2 Expression and Inhibiting Akt/mTOR Signaling Pathway. *Theranostics*.**9**,796-810(2019).
33. Ding, Y., et al., Down-regulation of Long Non-coding RNA LINC01554 in Hepatocellular Cancer and its Clinical Significance. *Journal of Cancer*.**11**,3369-3374(2020).
34. Chen, S., et al., Gene amplification derived a cancer-testis long noncoding RNA PCAT6 regulates cell proliferation and migration in hepatocellular carcinoma. **8**,3017-3025(2019).
35. Yang, W., Ju, H.Y., and Tian, X.F., Circular RNA-ABCB10 suppresses hepatocellular carcinoma progression through upregulating NRP1/ABL2 via sponging miR-340-5p/miR-452-5p. *European review for medical and pharmacological sciences*.**24**,2347-2357(2020).
36. Lv, J., et al., PCDH9 acts as a tumor suppressor inducing tumor cell arrest at G0/G1 phase and is frequently methylated in hepatocellular carcinoma. *Molecular medicine reports*.**16**,4475-4482(2017).
37. Toufektchan, E., et al., Germline mutation of MDM4, a major p53 regulator, in a familial syndrome of defective telomere maintenance. *Science advances*.**6**,eaay3511(2020).
38. Shen, T., et al., TP53 R249S mutation detected in circulating tumour DNA is associated with Prognosis of hepatocellular carcinoma patients with or without hepatectomy. *Liver international : official journal of the International Association for the Study of the Liver*.**40**,2834-2847(2020).

39. Ritchie, M.E., et al., limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*.**43**,e47(2015)
40. Lao, J., Chen, Y., and Li, Z.C., A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme. **7**,10353(2017).
41. Lohavanichbutr, P., et al., A 13-gene signature prognostic of HPV-negative OSCC: discovery and external validation. *Clinical cancer research : an official journal of the American Association for Cancer Research*.**19**,1197-1203(2013).
42. Huitzil-Melendez, F., et al., Advanced hepatocellular carcinoma: which staging systems best predict prognosis? *J. Clin. Oncol.***28**,2889-2895(2010).
43. Wei, J.H., et al., A CpG-methylation-based assay to predict survival in clear cell renal cell carcinoma. *Nature communications*.**6**,8699(2015).
44. Wu, J., et al., Nomogram integrating gene expression signatures with clinicopathological features to predict survival in operable NSCLC: a pooled analysis of 2164 patients. *J. Exp. Clin. Cancer Res.***36**,4(2017).

## Figures

### Figure 1

**Overall workflow.** We used all HCCs in TCGA as a training set and 50% HCCs as a test set. In the training set, we performed the limma analysis to identify DE-mRNAs, DE-lncRNAs, and DE-miRNAs. Chi-square analysis was used to screen abnormal CNV genes. The high-frequency SNPs (Top SNPs) in HCC were selected for further research. The univariate Cox regression analysis, LASSO Cox analysis, and backward stepwise Cox proportional hazard analysis were used to identify critical markers. We constructed five single-omic models (mRNA, lncRNA, miRNA, CNV, and SNP model) through LASSO Cox analysis or stepwise Cox. The multi-omics model was constructed based on the five single-omic models through multiple Cox regression analysis. These models were evaluated and verified in the training set and test set, respectively. Moreover, we externally validated the mRNA and SNP models in the LIRI-JP, and LICA-FR, respectively. HCC, hepatocellular carcinoma; TCGA, The Genome Cancer Atlas; LASSO, Least absolute shrinkage and selection operator; OS, Overall survival; DE-mRNAs, Differentially expressed mRNAs; DE-lncRNAs, Differently expressed lncRNAs; DE-miRNAs, Differentially expressed miRNA; CNV, [Copy number variation](#); SNP, [Single nucleotide polymorphism](#).

### Figure 2

**Construction and validation of the mRNA model.** (A) Selection of mRNAs with HR > 1 and up-regulation, and mRNAs with HR < 1 and down-regulation in HCC. (B) LASSO coefficients of the sixteen key mRNAs. The dotted vertical line is drawn at the  $\lambda$  value chosen by the minimum criteria. L1 Norm represents the summation of absolute nonzero coefficients at each  $\lambda$ . Y-axis represents the values of nonzero coefficients at each  $\lambda$ . (C) The evaluation of the mRNA model via the ROC curve and C-index in the TCGA training set. (D) Kaplan-Meier survival analysis of the different risk groups stratified with the trisection of the mRNA risk score in the TCGA training set. (E) The verification of the mRNA model via the ROC curve and C-index in the TCGA test set. (F) The verification of the mRNA model with Kaplan-Meier survival analysis in the TCGA test set. (G) The external validation of the mRNA model with Kaplan-Meier survival analysis in the LIRI-JP dataset. HCC, hepatocellular carcinoma; TCGA, The Genome Cancer Atlas; C-index, Harrell's concordance index; ROC, Receiver operating characteristic; AUC, Area under the curve; LASSO, Least absolute shrinkage and selection operator; HR, Hazard rate ratio.

### Figure 3

**Construction and validation of the lncRNA model.** (A) Selection of lncRNAs with HR > 1 and up-regulation, and lncRNAs with HR < 1 and down-regulation in HCC. (B) LASSO coefficients of the twenty key lncRNAs. (C) The evaluation of the lncRNA model via the ROC curve and C-index in the TCGA training set. (D) Kaplan-Meier survival analysis of the different risk groups stratified with the trisection of the lncRNA risk score in the TCGA training set. (E) The verification of the lncRNA model via the ROC curve and C-index in the TCGA test set. (F) The validation of the lncRNA model with Kaplan-Meier survival analysis in the TCGA test set. HCC, hepatocellular carcinoma; TCGA, The Genome Cancer Atlas; C-index, Harrell's concordance index; ROC, Receiver operating characteristic; AUC, Area under the curve; LASSO, Least absolute shrinkage and selection operator; HR, Hazard rate ratio.

### Figure 4

**Construction and validation of the miRNA model.** (A) Selection of miRNAs with HR > 1 and up-regulation, and miRNAs with HR < 1 and down-regulation in HCC. (B) Univariate Cox regression analysis of the five key miRNAs. (C) The evaluation of the miRNA model via the ROC curve and C-index in the TCGA training set. (D) Kaplan-Meier survival analysis of the different risk groups stratified with the trisection of the miRNA risk score in the TCGA training set. (E) The verification of the miRNA model via the ROC curve and C-index in the TCGA test set. (F) The validation of the miRNA model with Kaplan-Meier survival analysis in the TCGA test set. HCC, hepatocellular carcinoma; TCGA, The Genome Cancer Atlas; C-index, Harrell's concordance index; ROC, Receiver operating characteristic; AUC, Area under the curve; HR, Hazard rate ratio.

## Figure 5

**Construction and validation of the CNV model.** (A) Circos plot shows genes with different copy number alterations between HCC and non-tumour samples. The blue dots represent genes with copy number loss, and the black dots represent genes with copy number gain. (B) LASSO coefficients of the five key CNV genes. (C) The evaluation of the CNV model via ROC curve and C-index in the TCGA training set. (D) Kaplan-Meier survival analysis of the different risk groups stratified with the CNV risk score in the TCGA training set. Patients with no copy number alteration of the five key CNV genes were attributed to the low risk group, and the others to the high risk group. (E) The verification of the CNV model via the ROC curve and C-index in the TCGA test set. (F) The validation of the CNV model with Kaplan-Meier survival analysis in the TCGA test set. HCC, hepatocellular carcinoma; TCGA, The Genome Cancer Atlas; C-index, Harrell's concordance index; ROC, Receiver operating characteristic; AUC, Area under the curve; LASSO, Least absolute shrinkage and selection operator; CNV, Copy number variation.

## Figure 6

**Construction and validation of the SNP model.** (A) Distributions of various mutation types of the sixteen high-frequency SNPs. The histogram at the top indicates the sum of non-synonymous and synonymous mutations in every case. The histogram on the right stands for the sample number suffering from a gene mutation. In the heat map, the various colours stand for various mutation types, whereas the white represents no mutation. (B) The evaluation of the SNP model via the ROC curve and C-index in the TCGA training set. (C) Kaplan-Meier survival analysis of the different risk groups stratified with the SNP risk score in the TCGA training set. Patients with no mutation of the seven key SNPs were attributed to the low risk group, and the others were attributed to the high risk group. (D) The verification of the SNP model via the ROC curve and C-index in the TCGA test set. (E) The validation of the SNP model with Kaplan-Meier survival analysis in the TCGA test set. (F) The external validation of the SNP model with Kaplan-Meier survival analysis in the LICA-FR dataset. HCC, hepatocellular carcinoma; TCGA, The Genome Cancer Atlas; C-index, Harrell's concordance index; ROC, Receiver operating characteristic; AUC, Area under the curve; SNP, Single nucleotide polymorphism.

## Figure 7

**Construction and validation of the multi-omics model.** (A) Nomogram of the multi-omics model for predicting 1-, 3-, and 5-year OS in the TCGA training set. (B) Calibration plot for 1-, 3-, and 5-year OS of the multi-omics model in the TCGA training set. (C) The evaluation of the multi-omics model via the ROC curve and C-index in the TCGA training set. (D) Kaplan-Meier survival analysis of the different risk groups stratified with the trisection of the total point of the proposed nomogram in the TCGA training set. (E, F) Decision curve analysis for the multi-omics model and the 5 single-omic models at 1- and 3-year points in

the TCGA training set. **(G)** Comparison of the predictive power of different models with C-index and ROC analysis in the TCGA training set. **(H)** The verification of the multi-omics model via the ROC curve and C-index in the TCGA test set. **(I)** The validation of the multi-omics model with Kaplan-Meier survival analysis in the TCGA test set. TCGA, The Genome Cancer Atlas; C-index, Harrell's concordance index; ROC, Receiver operating characteristic; AUC, Area under the curve; DCA, Decision curve analysis; OS, Overall survival; CNV, Copy number variation; SNP, Single nucleotide polymorphism.

## Supplementary Files

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

- [SupplementaryTables.xlsx](#)
- [Supplementaryfigures.pdf](#)