A genomic-clinical nomogram predicting recurrence-free survival for patients diagnosed with hepatocellular carcinoma

Junjie Kong 1, Tao Wang 1, Shu Shen 1, Zifei Zhang 1, Xianwei Yang 1, Wentao Wang Corresp. 1

1 Department of Liver Surgery & Liver Transplantation Center, West China Hospital of Sichuan University, Chengdu, Sichuan Province, China

Corresponding Author: Wentao Wang
Email address: wwt0510@163.com

Liver resection surgery is the most commonly used treatment strategy for patients diagnosed with hepatocellular carcinoma (HCC). However, there is still a chance for recurrence in these patients despite the survival benefits of this procedure. This study aimed to explore recurrence-related genes (RRGs) and establish a genomic-clinical nomogram for predicting postoperative recurrence in HCC patients. 123 differently expressed genes (DEGs) and 3 RRGs (PZP, SPP2, and PRC1) were identified from online databases via Cox regression and LASSO logistic regression analyses and a gene-based risk model containing RRGs was then established. The Harrell’s concordance index (C-index), receiver operating characteristic (ROC) curves and calibration curves showed that the model performed well. Finally, a genomic-clinical nomogram incorporating the gene-based risk model, AJCC staging system, and Eastern Cooperative Oncology Group performance status (ECOG-PS) was constructed to predict the 1-, 2-, and 3-year recurrence-free survival rates (RFS) for HCC patients. The C-index, ROC analysis, and decision curve analysis (DCA) were good indicators of the nomogram’s performance. In conclusion, we identified 3 reliable RRGs associated with the recurrence of cancer and constructed a nomogram that performed well in predicting RFS for HCC patients. These findings could enrich our understanding of the mechanisms for HCC recurrence, help surgeons predict patients’ prognosis, and promote HCC treatment.
A genomic-clinical nomogram predicting recurrence-free survival for patients diagnosed with hepatocellular carcinoma

Junjie Kong¹, Tao Wang¹, Shu Shen¹, Zifei Zhang¹, Xianwei Yang¹, Wentao Wang¹

¹Department of Liver Surgery & Liver Transplantation Center, West China Hospital of Sichuan University, Chengdu, P. R. China

Corresponding Author:
Wentao Wang
No.37 Guoxue Xiang, Wuhou District, Chengdu, 610041, Sichuan Province, China.
E-mail: wwt0510@163.com

Abstract
Liver resection surgery is the most commonly used treatment strategy for patients diagnosed with hepatocellular carcinoma (HCC). However, there is still a chance for recurrence in these patients despite the survival benefits of this procedure. This study aimed to explore recurrence-related genes (RRGs) and establish a genomic-clinical nomogram for predicting postoperative recurrence in HCC patients. 123 differently expressed genes (DEGs) and 3 RRGs (PZP, SPP2,
and PRC1) were identified from online databases via Cox regression and LASSO logistic regression analyses and a gene-based risk model containing RRGs was then established. The Harrell’s concordance index (C-index), receiver operating characteristic (ROC) curves and calibration curves showed that the model performed well. Finally, a genomic-clinical nomogram incorporating the gene-based risk model, AJCC staging system, and Eastern Cooperative Oncology Group performance status (ECOG-PS) was constructed to predict the 1-, 2-, and 3-year recurrence-free survival rates (RFS) for HCC patients. The C-index, ROC analysis, and decision curve analysis (DCA) were good indicators of the nomogram’s performance. In conclusion, we identified 3 reliable RRGs associated with the recurrence of cancer and constructed a nomogram that performed well in predicting RFS for HCC patients. These findings could enrich our understanding of the mechanisms for HCC recurrence, help surgeons predict patients’ prognosis, and promote HCC treatment.

Introduction

Hepatocellular carcinoma (HCC) is the most common primary liver malignancy and a contributor to the third population of cancer-related deaths, ranking sixth among the most frequent malignancies worldwide (Forner et al., 2018). Liver resection is the most commonly used therapeutic strategy for HCC, and accounts for an overall 5-year survival rate of approximately 70% (Orcutt & Anaya, 2018; Akoad & Pomfret, 2015). However, for these patients, survival is impacted by postoperative tumor recurrence (Tabrizian et al., 2015). Previous studies have suggested that more than half of HCC patients would suffer from a
disease relapse after a hepatectomy (Akoad & Pomfret, 2015). Disease recurrence was divided into two types depending on their different etiologies: those who relapsed within 2 years after surgery (early recurrence) and those whose relapse occurred more than 2 years after the operation (late recurrence) (Sherman, 2008). Many clinical characteristics, such as the AFP level, tumor size, vascular invasion (including microscopic and macroscopic), and HBsAg level proved to be risk factors for recurrence and several predictive models were established to predict the postoperative recurrence of HCC (Lee et al., 2014a; Qiu et al., 2017; He et al., 2018; Chan et al., 2018a). However, these studies and models only focused on the relationships between clinical traits and HCC recurrence with little focus on the crucial role of molecular data.

Advances in biomedical research and high-throughput technologies have greatly contributed to the identification of molecular biomarkers related to cancer development, recurrence, and prognosis in the past few decades (Teufel, 2015). A large number of mRNAs, microRNAs and other kinds of biomarkers have been identified and regarded as potential targets for cancer diagnosis and treatment (Lee et al., 2014b). Biomarkers could reflect the molecular mechanisms of tumor recurrence and, as a result, the identification of reliable biomarkers could improve the accuracy in predicting cancer recurrence, thereby contributing to cancer treatment (Long et al., 2018; Gu et al., 2018). Consequently, gene signatures for recurrence and prognosis prediction are receiving more attention.

In this study we explored differently expressed genes (DEGs) between the HCC and non-tumor samples using data obtained from the Gene Expression Omnibus (GEO) and the Cancer Genome Atlas (TCGA) databases. 3 recurrence-related genes (RRGs) were then identified and a
gene-based risk model was established from the data of the TCGA dataset. Furthermore, a genomic-clinical nomogram containing the gene-based risk model and clinical characteristics was constructed to predict the 1-, 2-, and 3-year recurrence-free survival (RFS) for HCC patients. We also assessed the performance of the nomogram using the Harrell’s concordance index (C-index), calibration curves, receiver operating characteristic (ROC) curves, and decision curves analysis (DCA).

**Materials & Methods**

**Dataset collection**

Using “hepatocellular carcinoma” as our keyword, we searched gene expression profiles of HCC from the GEO database (https://www.ncbi.nlm.nih.gov/gds/). “Homo sapiens” was the term chosen for the organism parameter. The following criteria were used to screen datasets: 1) datasets were mRNA expression profiled by array; 2) datasets with more than 100 samples; 3) datasets with comparison between tumor and non-tumor samples; and 4) profiles with available expression information. Finally, to obtain more accurate results and to avoid individual heterogeneity, the top 4 GEO datasets with the largest sample numbers, GSE25097 (Sung et al., 2012), GSE76427 (Grinchuk et al., 2018), GSE36376 (Lim et al., 2013) and GSE14520 (Roessler et al., 2010), were selected. The details of the datasets were shown in Table S1. Meanwhile, we downloaded the RNA-seq of HCC form the TCGA database (https://cancergenome.nih.gov/), which contained 374 HCC samples and 50 non-tumor samples. The related clinical files of HCC were also obtained for further analysis. Log2 transformation...
was used for all of the expression data.

**Exploration of differently expressed genes (DEGs) and bioinformatic analysis**

R software (Version 3.5.3) and the related packages “limma” and “edgR” were used to screen DEGs. The cut-off criteria with a significance of \( p < 0.05 \) and \(|\log_{2}FC| > 1\) were used in the identification of DEGs. Overlapping analysis was performed to find DEGs among all of the 4 GEO datasets and the TCGA dataset and the Venn plot was completed using the online tool entitled “Calculate and draw custom Venn diagrams” ([http://bioinformatics.psb.ugent.be/webtools/Venn/](http://bioinformatics.psb.ugent.be/webtools/Venn/)). In addition, the online biological tools Database for Annotation, Visualization and Integrated Discovery (DAVID, [https://david.ncifcrf.gov/](https://david.ncifcrf.gov/)) and KOBAS 3.0 ([http://kobas.cbi.pku.edu.cn/anno_iden.php](http://kobas.cbi.pku.edu.cn/anno_iden.php)) were used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enrichment analysis, respectively. Furthermore, the search tool for the Retrieval of Interacting Genes (STRING, [https://string-db.org/cgi/input.pl](https://string-db.org/cgi/input.pl)) and Cytoscape software (Version 3.6.1) were used to establish a protein-protein interaction (PPI) network complex to visualize the correlations among DEGs.

**Identification of recurrence-related genes (RRGs)**

In this progress, 3 datasets (TCGA, GSE14520 and GSE76427 datasets) with detailed information about tumor recurrence were divided into 2 types of sets: the discovery set (TCGA dataset) and the validation set (GSE14520 and GSE76427 datasets). In the discovery set, univariate, LASSO and multivariate Cox regression analysis were employed to screen RRGs from DEGs. Using the “survival” R package, univariate analysis was used to make a primary
selection of DEGs and those with \( p<0.05 \) were regarded as potential candidates associated with the recurrence of HCC. Afterwards, we employed the LASSO logistic regression model using the “glmnet” R package to further select genes from the potential candidates. In the LASSO-penalized regression, the optimal lambda was determined using 10-fold cross-validation and the L1 penalty was used to shrink regression coefficients towards zero. Many variables were excluded based on the principle that a larger penalty led to fewer predictive factors (Goeman, 2010). Consequently, candidate RRGs were considered as DEGs with nonzero coefficients. After subsampling the dataset with 1000 iterations, the seed DEGs were shrunk into various sets and those containing DEGs with nonzero coefficients were defined as models with potential prognostic ability. DEGs with at least 900 occurrence frequencies were considered as candidate RRGs. Finally, stepwise multivariate regression analysis was conducted to identify RRGs and to construct a gene-based risk model which could predict HCC recurrence after liver resection. Using X-tile software (Camp et al., 2004), the optimal cut-off points of the risk scores and the expression levels of RRGs were found. The Kaplan-Meier (KM) analysis was used to evaluate the performance of RRGs and the risk model. Furthermore, ROC analysis and calibration curves were used in validating the performance of the risk model and the “pROC” and “rms” R packages were used, respectively. In addition, the same methods were used to assess the performance of the RRGs and the risk model in the 2 validation datasets. In these statistic processes, a two-sided \( P<0.05 \) was considered statistically significant. Finally, using “pheatmap” and “forestplot” R packages, the results were visualized by heatmap plots and forest plots, respectively.
Development and validation of nomogram

Univariate and multivariate regression analyses using the RNA-seq and clinical traits of the TCGA dataset combined with the gene-based risk model were used to find clinical and molecular factors independently associated with HCC recurrence. Based on the results of multivariate analysis, a genomic-clinical nomogram was established using the “rms” R package to predict 1-, 2-, and 3-year RFS for HCC patients (Wang et al., 2013). The performance of the nomogram was evaluated using calibration curves. The calibration was conducted with 1000 bootstrap samples to reduce the bias. The discrimination of the nomogram was assessed using C-index and area under the curve (AUC), which measured the predictive ability of the nomogram for patients with different outcomes. The larger the C-index and a bigger AUC indicated a higher accuracy of prognostic prediction (Huitzil-Melendez et al., 2010). Finally, using an “rmda” R package, DCA was conducted to assess the clinical usefulness of the nomogram, which could estimate the net benefits for different prediction models across all possible risk thresholds (Vickers et al., 2008).

Results

Exploration of DEGs

Figure 1 reveals the workflow of this study. First, 4 GEO datasets, GSE25097, GSE36376, GSE14520, and GSE 76427 containing 1812 samples (1037 HCC samples and 775 non-tumor samples) were obtained to screen DEGs and a total of 1454, 413, 1088 and 398 genes were found dysregulated in the tumor tissues, respectively (Fig. S1A-E). Meanwhile, 9314 dysregulated
genes were identified in the TCGA dataset (Fig. S1F). After an overlapping analysis, 123 genes containing 12 up-regulated genes and 111 down-regulated genes were regarded as DEGs aberrantly expressed in HCC samples compared to non-tumor samples and were selected for further analysis (Fig. 2A, B).

**Functional enrichment analysis and construction of PPI network complex**

We performed GO and KEGG analysis to elucidate the functional characteristics of the DEGs. The GO analysis indicated that the DEGs were significantly enriched in the extracellular exosome, extracellular region, extracellular space, oxidation-reduction process, cell adhesion, and serine-type endopeptidase activity. In the KEGG pathways analysis, we could find that the DEGs were mainly enriched in metabolism, metabolic pathways, metabolism of lipids and lipoproteins, biological oxidations, metabolism of amino acids and derivatives, and the immune system (Fig. 2C, D). Finally, a PPI network complex, containing 121 nodes and 773 edges, was constructed to elucidate the relationships among the DEGs (Fig. S2).

**Identification of recurrence-related genes (RRGs)**

After a rigorous screening, 252 cases from the TCGA dataset were selected for the identification of RRGs from the discovery set. Using univariate analysis, we found that 79 DEGs were significantly related to HCC recurrence with a p-value <0.05. LASSO logistic regression analysis was used to find candidates associated with recurrence and the optimal tuning parameters related to the minimum generalization error were determined from 10-fold cross-validation (Fig. 3). Consequently, 8 DEGs (PZP, C1RL, PRC1, PTTG1, UBE2C, AFM, SPP2, HGFAC) were screened. Finally, we performed stepwise multivariate regression analysis and
found that 3 genes, PZP, SPP2, and PRC1, were biomarkers independently associated with HCC recurrence and were included in the construction of a gene-based risk model. Using X-tile software, the optimal cut-off values of the 3 RRGs’ expression levels were found in the discovery and validation sets, and the HCC patients were divided into 2 groups (low-risk group and high-risk group). The KM analysis showed that the 3 RRGs were significantly related to HCC recurrence among all of the 3 datasets (Fig. 4).

According to the relative coefficient in the regression model, a risk score could be calculated to evaluate the risk of recurrence for each patient based on the expression levels of RRGs: risk score = (-0.0793 * PZP expression level) + (0.2295 * PRC1 expression level) + (-0.0662 * SPP2 expression level). We then used X-tile software to select the optimal cut-off value of the risk score for recurrence, and the patients were divided into a low-risk group or a high-risk group in the discovery and validation sets, respectively (Fig. S3). The RFS in the low-risk group was significantly longer than that of the high-risk group among all of the 3 datasets (Fig. 5A-C). The C-index of the risk model in the TCGA, GSE14520, and GSE76427 datasets was 0.663 (95% CI, 0.615-0.711), 0.571 (95% CI, 0.519-0.623) and 0.599 (95% CI, 0.492-0.706), respectively. The ROC curves also suggested that the risk model had a good predictive efficiency among all of the 3 datasets (Fig. 5D-F). Finally, the calibration curves suggested that the risk model had a good predictive efficiency compared to the observed outcomes for 1-, 2- and 3-year postoperative recurrence (Fig. 5G-I).

**Identification of genomic and clinical factors associated with recurrence**

The RNA-seq and clinical characteristics of the TCGA dataset were used for exploring
genomic and clinical factors related to recurrence. After removing patients who were diagnosed
with fibrolamellar carcinoma and combined hepatocellular-cholangiocarcinoma, and who did not
have intact clinical information (including age, height, weight, race, hepatitis B virus status,
alcohol consumption, tumor grade, AJCC stage, APF level, platelet level, albumin level, Eastern
Cooperative Oncology Group performance status (ECOG-PS), and sex), a total of 162 HCC
patients with both intact genomic and clinical information were included for further analysis.
Univariate and multivariate cox regression analyses were used to find factors associated with
HCC recurrence. As shown in Figure 6, gene-based risk model, race, AJCC staging system and
ECOG-PS were found to be potential factors related to HCC recurrence in univariate analysis.
Furthermore, the multivariate analysis indicated that the gene-based risk model, AJCC staging
system and ECOG-PS were factors associated with HCC recurrence (Fig. 6).

Construction and validation of genomic-clinical nomogram

Based on the results of the multivariate analysis, a genomic-clinical nomogram including
the gene-based risk model, AJCC staging system, and ECOG-PS was established to predict the
1-, 2-, and 3-year RFS for HCC patients (Fig. 7A). The calibration curves of the nomogram for
predicting the 1-, 2-, and 3-year probability of recurrence indicated that it performed well (Fig.
7B). The C-index of the nomogram was 0.678 (95% CI, 0.618-0.738).

The DCA results showed that the nomogram was more clinically useful than using the gene-
based risk model, AJCC staging system, or ECOG-PS alone (Fig. 7C-E). Afterwards, we
performed ROC analysis to assess the discrimination ability of the genomic-clinical nomogram
and the results suggested that the AUC of the nomogram in predicting the 1-, 2-, and 3-year
probability of recurrence was greater than that of using the gene-based risk model or AJCC staging system alone (Fig. 7F-H). Finally, according to the total points of each patient in the nomogram, the patients were divided into 2 groups (low-risk group and high-risk group) using X-tile software, and KM analysis demonstrated that the discrimination ability of the nomogram was satisfactory (p<0.0001) (Fig. 8).

**Discussion**

As the sixth most frequently occurring malignancy and the third leading cause of cancer-related deaths, HCC remains a significant medical problem in the world (Forner et al., 2018). Although advances in therapeutic strategies such as liver resection, transplantation, radiofrequency ablation, and transcatheater arterial chemoembolization have improved the chances of survival for certain patients, the prognosis of these patients remains unsatisfactory (Zamora-Valdes et al., 2017; Bailey & Sydnor, 2019; Tanwar et al., 2009; Sun et al., 2019). Recurrence is a significant challenge for HCC patients after liver resection and usually leading to a poor prognosis (Chan et al., 2018b). Therefore, it is vitally important to find reliable biomarkers in the recurrent progression of the disease and to explore the mechanisms and risk factors closely associated with HCC recurrence.

In this study we identified 123 DEGs in HCC samples compared to non-tumor samples using data obtained from GEO and TCGA databases. Univariate, LASSO, and multivariate regression analysis were used to find DEGs related to recurrence. Consequently, 3 RRGs were identified. A gene-based risk model was then constructed using the relative coefficients in the
multivariate model and the expression levels of the 3 RRGs. The risk model had a good predictive efficiency for HCC recurrence. Furthermore, based on the clinical information obtained from the TCGA database, a genomic-clinical nomogram, including the gene-based risk model, AJCC staging system, and ECOG-PS was established, the C-index of the nomogram was 0.678, the calibration curves fitted well and the ROC analysis and DCA also showed the nomogram performed well in discerning those patients with a high probability of recurrence.

Three RRGs, PZP, SPP2, and PRC1, were identified as factors associated with HCC recurrence. PRC1, also known as protein regulator of cytokinesis, is a microtubule binding protein that plays crucial roles in mitosis (Subramanian et al., 2010). Due to its essential function in the cell cycle, dysregulation of PRC1 could be regarded as an important factor for carcinogenesis. The development of various cancers, such as breast cancer (Shimo et al., 2007), lung adenocarcinoma (Zhan et al., 2017) and pancreatic cancer (Nakamura et al., 2004), was associated with the overexpression of PRC1. Meanwhile, the dysregulation of PRC1 was demonstrated to be related to the early recurrence and chemoresistance of HCC (Chen et al., 2016; Wang et al., 2017). PRC1 could contribute to the early recurrence of HCC via the regulation of the Wnt/β-catenin signaling pathway, which has been proven to be a crucial pathway in hepatocarcinogenesis and a potential therapeutic target (Pez et al., 2013).

Furthermore, Liu et al. suggested that knockdown PRC1 using siRNA could block cytokines and limit the proliferation of HCC, which could bring new insights into HCC treatment (Liu et al., 2018). Although previous studies have demonstrated that the other two RRGs, PZP and SPP2, were related to the prognosis of HCC, there was no study discussing their potential functions in
HCC recurrence (Yang et al., 2018; Zheng et al., 2018). Robust experimental results also suggested that SPP2 and PZP were associated with the stability of spliceosome and chromatin (Klein et al., 2016; Warkocki et al., 2015), and their dysregulation might affect the progression of cell proliferation and apoptosis, which could lead to carcinogenesis and disease progression.

In addition, in validation of RRGs’ performance, the overexpression of PZP was found to be negatively associated with recurrence in the TCGA dataset, while positively related to recurrence in the GSE14520 and GSE76427 datasets. Further exploration with larger HCC and non-tumor samples should be performed to validate this issue. In summary, the mechanisms of the 3 RRGs in HCC development and recurrence have not been clear until now and they should be more closely followed as these could be prospective therapeutic targets for identifying the potential for HCC recurrence.

The genomic-clinical nomogram contained 2 clinical traits, the AJCC staging system and the ECOG-PS, aside from the gene-based risk model. The AJCC staging system is a conventional staging system used for evaluating the characteristics of cancers, selecting optimal therapeutic strategies, and predicting prognosis (Chun et al., 2018). In this study, patients with AJCC stage II and III had a higher risk for HCC recurrence. According to the 8th AJCC staging system the tumors classified into stage II and III were those with vascular invasion, bigger tumor size, multiple tumor numbers, or the invasion into adjacent organs compared with the stage-I tumors,. Tumors with these characteristics were more likely to have a greater invasive ability and a later tumor stage and, as a result, these patients were easier to determine as having HCC recurrence and were given higher points in the nomogram. ECOG-PS was used to assess the
patients’ functional capability of self-care (Sorensen et al., 1993). Previous studies demonstrated that ECOG-PS was closely related to the survival of cancer patients. In the BCLC staging system, ECOG-PS was listed as an independent factor for selecting therapeutic strategies and predicting prognosis (Forner et al., 2018). Recently, a Chinese study discovered that patients with higher ECOG-PS were more likely to have poor molecular phenotypes in colorectal cancer, which could lead to a higher probability of cancer recurrence (Chen et al., 2016). It is clear that more focus should be placed on the functions of ECOG-PS in the recurrence of HCC, which could provide new insights into cancer treatment.

Functional enrichment analysis showed that the DEGs were mainly enriched in extracellular parts, such as the extracellular exosome, extracellular region, and extracellular space. Furthermore, metabolism-related pathways and immune-related pathways were significantly enriched by DEGs. It is well known that the liver is associated with various metabolic processes, which could help to maintain the microenvironmental stability of various tissues (Rui, 2014). As a result, the aberrant metabolism could provide for the unique needs of the tumor cells, contain macromolecular biosynthesis, increase energy production, and maintain the redox balance, which could provide a selective advantage for the proliferation, growth and survival of tumor cells (De Matteis et al., 2018). Recently, the advances of metabolomics have brought new insights into the mechanisms of HCC development, recurrence and prognosis, and the related findings could be therapeutic targets for HCC treatment (Shang et al., 2016). One of the most widely-known special mechanisms of carcinogenesis is immune invasion (Aerts et al., 2016). However, until the discovery of immune checkpoints, including programmed cell death protein 1 (PD-1) and
cytotoxic T-lymphocyte protein 4 (CTLA-4), there had not been any breakthroughs in immunotherapy of HCC. The immune checkpoint inhibitors have been proven to possess significant clinical functions and are promising strategies for the systemic treatment of HCC (Inarrairaegui et al., 2018). The metabolism-related pathways and immune-related pathways played crucial roles in the recurrence of HCC and could certainly be potential therapeutic targets in the future.

Several other studies also reported the establishment of models to predict RFS for HCC after liver resection (Cui et al., 2017; Gu et al., 2019). For instance, using the gene expression file GSE76427, Gu et al. (Gu et al., 2019) discovered that six long non-coding RNA (MSC-AS1, POLR2J4, EIF3J-AS1, SERHL, RMST, and PVT1) were associated with a poor prognosis of HCC and constructed a nomogram including the six-lncRNA signature, TNM stage and ECOG to predict RFS for patients with small HCC. The C-index of their nomogram was 0.684 (95%CI: 0.635-0.733), which was comparable with our nomogram (C-index 0.678; 95% CI, 0.618-0.738). Furthermore, besides the gene-based signature, TNM stage and ECOG were also integrated into their nomogram, which was robust evidence for our nomogram (including the AJCC staging system and ECOG). However, Gu’s study only analyzed one gene expression file with a relatively small number of samples, which could make the results inaccurate to some extent. Our study analyzed a total of 1645 samples from four GEO datasets, containing 870 HCC samples and 775 normal samples. Consequently, the RRGs found in our study were more reliable. Combined with rigorous validations, it was clear that our nomogram was could performed well in the prediction of RFS for HCC patients after liver resection.
However, limitations existed in the current study. First of all, the expression of RRGs were only explored and validated among online databases; no experiments were conducted to confirm them. Consequently, experiments about the dysregulation of the RRGs and their mechanisms in HCC recurrence should be explored in the future. Secondly, we failed to make a comparison of the genomic-clinical nomogram to other widely accepted staging systems, such as the BCLC staging system. Lastly, we did not perform an external validation for the genomic-clinical nomogram, and further validations with additional data should be performed to validate the performance of our nomogram.

Conclusions

Three new biomarkers were found that were related to postoperative recurrence for HCC patients after hepatectomy and a reliable genomic-clinical nomogram was established to predict the 1-, 2-, and 3-year RFS. The C-index, ROC analysis, and DCA showed that the nomogram performed well. These findings could bring new insights into molecular mechanisms for HCC recurrence, contribute to HCC treatment, and help surgeons to predict the likelihood of recurrence and the prognosis for HCC patients postoperatively.

Acknowledgments

None.

References:
Aerts, M., Benteyn, D., Van Vlierberghe, H., Thielemans, K., and Reynaert, H. 2016. Current status and perspectives of immune-based therapies for hepatocellular carcinoma. World J Gastroenterol 22:253-261. 10.3748/wjg.v22.i1.253

Akoad, M.E., and Pomfret, E.A. 2015. Surgical resection and liver transplantation for hepatocellular carcinoma. Clinics in Liver Disease 19:381-399. 10.1016/j.cld.2015.01.007

Bailey, C.W., and Sydnor, M.J. 2019. Current State of Tumor Ablation Therapies. Dig Dis Sci. 10.1007/s10620-019-05514-9

Camp, R.L., Dolled-Filhart, M., and Rimm, D.L. 2004. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clinical Cancer Research 10:7252-7259. 10.1158/1078-0432.CCR-04-0713

Chan, A., Zhong, J., Berhane, S., Toyoda, H., Cucchetti, A., Shi, K., Tada, T., Chong, C., Xiang, B.D., Li, L.Q., Lai, P., Mazzaferro, V., Garcia-Finana, M., Kudo, M., Kumada, T., Roayaie, S., and Johnson, P.J. 2018a. Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection. Journal of Hepatology 69:1284-1293. 10.1016/j.jhep.2018.08.027

Chen, D., Wen, X., Song, Y.S., Rhee, Y.Y., Lee, T.H., Cho, N.Y., Han, S.W., Kim, T.Y., and
Kang, G.H. 2016. Associations and prognostic implications of Eastern Cooperative Oncology Group performance status and tumoral LINE-1 methylation status in stage III colon cancer patients. *Clinical Epigenetics* 8:36. 10.1186/s13148-016-0203-8

Chen, J., Rajasekaran, M., Xia, H., Zhang, X., Kong, S.N., Sekar, K., Seshachalam, V.P., Deivasigamani, A., Goh, B.K.P., Ooi, L.L., Hong, W., and Hui, K.M. 2016. The microtubule-associated protein PRC1 promotes early recurrence of hepatocellular carcinoma in association with the Wnt/β-catenin signalling pathway. *Gut* 65:1522-1534. 10.1136/gutjnl-2015-310625

Chun, Y.S., Pawlik, T.M., and Vauthey, J.N. 2018. 8th Edition of the AJCC Cancer Staging Manual: Pancreas and Hepatobiliary Cancers. *Annals of Surgical Oncology* 25:845-847. 10.1245/s10434-017-6025-x

Cui, H., Zhang, Y., Zhang, Q., Chen, W., Zhao, H., and Liang, J. 2017. A comprehensive genome-wide analysis of long noncoding RNA expression profile in hepatocellular carcinoma. *Cancer Med* 6:2932-2941. 10.1002/cam4.1180

De Matteis, S., Ragusa, A., Marisi, G., De Domenico, S., Casadei Gardini, A., Bonafè, M., and Giudetti, A.M. 2018. Aberrant Metabolism in Hepatocellular Carcinoma Provides Diagnostic and Therapeutic Opportunities. *Oxidative Medicine and Cellular Longevity* 2018:1-13. 10.1155/2018/7512159

Forner, A., Reig, M., and Bruix, J. 2018. Hepatocellular carcinoma. *Lancet* 391:1301-1314. 10.1016/S0140-6736(18)30010-2

Goeman, J.J. 2010. L1 penalized estimation in the Cox proportional hazards model. *Biom J* 52:70-84. 10.1002/bimj.200900028
Grinchuk, O.V., Yenamandra, S.P., Iyer, R., Singh, M., Lee, H.K., Lim, K.H., Chow, P.K., and Kuznetsov, V.A. 2018. Tumor-adjacent tissue co-expression profile analysis reveals pro-oncogenic ribosomal gene signature for prognosis of resectable hepatocellular carcinoma. *Molecular Oncology* 12:89-113. 10.1002/1878-0261.12153

Gu, J., Zhang, X., Miao, R., Ma, X., Xiang, X., Fu, Y., Liu, C., Niu, W., and Qu, K. 2018. A three-long non-coding RNA-expression-based risk score system can better predict both overall and recurrence-free survival in patients with small hepatocellular carcinoma. *Aging* 10:1627-1639. 10.18632/aging.101497

Gu, J.X., Zhang, X., Miao, R.C., Xiang, X.H., Fu, Y.N., Zhang, J.Y., Liu, C., and Qu, K. 2019. Six-long non-coding RNA signature predicts recurrence-free survival in hepatocellular carcinoma. *World J Gastroenterol* 25:220-232. 10.3748/wjg.v25.i2.220

He, W., Peng, B., Tang, Y., Yang, J., Zheng, Y., Qiu, J., Zou, R., Shen, J., Li, B., and Yuan, Y. 2018. Nomogram to Predict Survival of Patients With Recurrence of Hepatocellular Carcinoma After Surgery. *Clin Gastroenterol Hepatol* 16:756-764. 10.1016/j.cgh.2017.12.002

Huitzil-Melendez, F.D., Capanu, M., O'Reilly, E.M., Duffy, A., Gansukh, B., Saltz, L.L., and Abou-Alfa, G.K. 2010. Advanced hepatocellular carcinoma: which staging systems best predict prognosis? *Journal of Clinical Oncology* 28:2889-2895. 10.1200/JCO.2009.25.9895

Inarrairaegui, M., Melero, I., and Sangro, B. 2018. Immunotherapy of Hepatocellular Carcinoma: Facts and Hopes. *Clinical Cancer Research* 24:1518-1524. 10.1158/1078-0432.CCR-17-0289

Klein, B.J., Muthurajan, U.M., Lalonde, M.E., Gibson, M.D., Andrews, F.H., Hepler, M., Machida, S., Yan, K., Kurumizaka, H., Poirier, M.G., Cote, J., Luger, K., and Kutateladze, T.G.
2016. Bivalent interaction of the PZP domain of BRPF1 with the nucleosome impacts chromatin dynamics and acetylation. *Nucleic Acids Research* 44:472-484. 10.1093/nar/gkv1321

Lee, S.C., Tan, H.T., and Chung, M.C. 2014a. Prognostic biomarkers for prediction of recurrence of hepatocellular carcinoma: current status and future prospects. *World J Gastroenterol* 20:3112-3124. 10.3748/wjg.v20.i12.3112

Lee, S.C., Tan, H.T., and Chung, M.C. 2014b. Prognostic biomarkers for prediction of recurrence of hepatocellular carcinoma: current status and future prospects. *World J Gastroenterol* 20:3112-3124. 10.3748/wjg.v20.i12.3112

Lim, H.Y., Sohn, I., Deng, S., Lee, J., Jung, S.H., Mao, M., Xu, J., Wang, K., Shi, S., Joh, J.W., Choi, Y.L., and Park, C.K. 2013. Prediction of disease-free survival in hepatocellular carcinoma by gene expression profiling. *Annals of Surgical Oncology* 20:3747-3753. 10.1245/s10434-013-3070-y

Liu, X., Li, Y., Meng, L., Liu, X.Y., Peng, A., Chen, Y., Liu, C., Chen, H., Sun, S., Miao, X., Zhang, Y., Zheng, L., and Huang, K. 2018. Reducing protein regulator of cytokinesis 1 as a prospective therapy for hepatocellular carcinoma. *Cell Death & Disease* 9:534. 10.1038/s41419-018-0555-4

Long, J., Zhang, L., Wan, X., Lin, J., Bai, Y., Xu, W., Xiong, J., and Zhao, H. 2018. A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma. *Journal of Cellular and Molecular Medicine* 22:5928-5938. 10.1111/jcmm.13863

Nakamura, T., Furukawa, Y., Nakagawa, H., Tsunoda, T., Ohigashi, H., Murata, K., Ishikawa, O., Ohgaki, K., Kashimura, N., Miyamoto, M., Hirano, S., Kondo, S., Katoh, H., Nakamura, Y., and
Katagiri, T. 2004. Genome-wide cDNA microarray analysis of gene expression profiles in pancreatic cancers using populations of tumor cells and normal ductal epithelial cells selected for purity by laser microdissection. *Oncogene* 23:2385-2400. 10.1038/sj.onc.1207392

Orcutt, S.T., and Anaya, D.A. 2018. Liver Resection and Surgical Strategies for Management of Primary Liver Cancer. *Cancer Control* 25:1145179885. 10.1177/1073274817744621

Pez, F., Lopez, A., Kim, M., Wands, J.R., Caron, D.F.C., and Merle, P. 2013. Wnt signaling and hepatocarcinogenesis: molecular targets for the development of innovative anticancer drugs. *Journal of Hepatology* 59:1107-1117. 10.1016/j.jhep.2013.07.001

Qiu, J.F., Ye, J.Z., Feng, X.Z., Qi, Y.P., Ma, L., Yuan, W.P., Zhong, J.H., Zhang, Z.M., Xiang, B.D., and Li, L.Q. 2017. Pre- and post-operative HBsAg levels may predict recurrence and survival after curative resection in patients with HBV-associated hepatocellular carcinoma. *Journal of Surgical Oncology* 116:140-148. 10.1002/jso.24628

Roessler, S., Jia, H.L., Budhu, A., Forgues, M., Ye, Q.H., Lee, J.S., Thorgeirsson, S.S., Sun, Z., Tang, Z.Y., Qin, L.X., and Wang, X.W. 2010. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Research* 70:10202-10212. 10.1158/0008-5472.CAN-10-2607

Rui, L. 2014. Energy metabolism in the liver. *Comprehensive Physiology* 4:177-197. 10.1002/cphy.c130024

Shang, R.Z., Qu, S.B., and Wang, D.S. 2016. Reprogramming of glucose metabolism in hepatocellular carcinoma: Progress and prospects. *World J Gastroenterol* 22:9933-9943. 10.3748/wjg.v22.i45.9933
Sherman, M. 2008. Recurrence of hepatocellular carcinoma. *N Engl J Med* 359:2045-2047. 10.1056/NEJMe0807581

Shimo, A., Nishidate, T., Ohta, T., Fukuda, M., Nakamura, Y., and Katagiri, T. 2007. Elevated expression of protein regulator of cytokinesis 1, involved in the growth of breast cancer cells. *Cancer Science* 98:174-181. 10.1111/j.1349-7006.2006.00381.x

Sorensen, J.B., Klee, M., Palshof, T., and Hansen, H.H. 1993. Performance status assessment in cancer patients. An inter-observer variability study. *Br J Cancer* 67:773-775.

Subramanian, R., Wilson-Kubalek, E.M., Arthur, C.P., Bick, M.J., Campbell, E.A., Darst, S.A., Milligan, R.A., and Kapoor, T.M. 2010. Insights into antiparallel microtubule crosslinking by PRC1, a conserved nonmotor microtubule binding protein. *Cell* 142:433-443. 10.1016/j.cell.2010.07.012

Sun, J.Y., Yin, T., Zhang, X.Y., and Lu, X.J. 2019. Therapeutic advances for patients with intermediate hepatocellular carcinoma. *Journal of Cellular Physiology*. 10.1002/jcp.28019

Sung, W.K., Zheng, H., Li, S., Chen, R., Liu, X., Li, Y., Lee, N.P., Lee, W.H., Ariyaratne, P.N., Tennakoon, C., Mulawadi, F.H., Wong, K.F., Liu, A.M., Poon, R.T., Fan, S.T., Chan, K.L., Gong, Z., Hu, Y., Lin, Z., Wang, G., Zhang, Q., Barber, T.D., Chou, W.C., Aggarwal, A., Hao, K., Zhou, W., Zhang, C., Hardwick, J., Buser, C., Xu, J., Kan, Z., Dai, H., Mao, M., Reinhard, C., Wang, J., and Luk, J.M. 2012. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nature Genetics* 44:765-769. 10.1038/ng.2295

Tabrizian, P., Jibara, G., Shrager, B., Schwartz, M., and Roayaie, S. 2015. Recurrence of Hepatocellular Cancer After Resection. *Annals of Surgery* 261:947-955.
Tanwar, S., Khan, S.A., Grover, V.P., Gwilt, C., Smith, B., and Brown, A. 2009. Liver transplantation for hepatocellular carcinoma. *World J Gastroenterol* 15:5511-5516.

Teufel, A. 2015. Bioinformatics and database resources in hepatology. *Journal of Hepatology* 62:712-719. 10.1016/j.jhep.2014.10.036

Vickers, A.J., Cronin, A.M., Elkin, E.B., and Gonen, M. 2008. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak* 8:53. 10.1186/1472-6947-8-53

Wang, Y., Li, J., Xia, Y., Gong, R., Wang, K., Yan, Z., Wan, X., Liu, G., Wu, D., Shi, L., Lau, W., Wu, M., and Shen, F. 2013. Prognostic nomogram for intrahepatic cholangiocarcinoma after partial hepatectomy. *Journal of Clinical Oncology* 31:1188-1195. 10.1200/JCO.2012.41.5984

Wang, Y., Shi, F., Xing, G.H., Xie, P., Zhao, N., Yin, Y.F., Sun, S.Y., He, J., Wang, Y., and Xuan, S.Y. 2017. Protein Regulator of Cytokinesis PRC1 Confers Chemoresistance and Predicts an Unfavorable Postoperative Survival of Hepatocellular Carcinoma Patients. *Journal of Cancer* 8:801-808. 10.7150/jca.17640

Warkocki, Z., Schneider, C., Mozaffari-Jovin, S., Schmitzova, J., Hobartner, C., Fabrizio, P., and Luhrmann, R. 2015. The G-patch protein Spp2 couples the spliceosome-stimulated ATPase activity of the DEAH-box protein Prp2 to catalytic activation of the spliceosome. *Genes Dev* 29:94-107. 10.1101/gad.253070.114

Yang, Y., Lu, Q., Shao, X., Mo, B., Nie, X., Liu, W., Chen, X., Tang, Y., Deng, Y., and Yan, J. 2018. Development Of A Three-Gene Prognostic Signature For Hepatitis B Virus Associated
Hepatocellular Carcinoma Based On Integrated Transcriptomic Analysis. *Journal of Cancer* 9:1989-2002. 10.7150/jca.23762

Zamora-Valdes, D., Taner, T., and Nagorney, D.M. 2017. Surgical Treatment of Hepatocellular Carcinoma. *Cancer Control* 24:1145164522. 10.1177/1073274817729258

Zhan, P., Zhang, B., Xi, G., Wu, Y., Liu, H., Liu, Y., Xu, W., Zhu, Q., Cai, F., Zhou, Z., Miu, Y., Wang, X., Jin, J., Li, Q., Qian, L., Lv, T., and Song, Y. 2017. PRC1 contributes to tumorigenesis of lung adenocarcinoma in association with the Wnt/β-catenin signaling pathway. *Molecular Cancer* 16. 10.1186/s12943-017-0682-z

Zheng, Y., Liu, Y., Zhao, S., Zheng, Z., Shen, C., An, L., and Yuan, Y. 2018. Large-scale analysis reveals a novel risk score to predict overall survival in hepatocellular carcinoma. *Cancer Management and Research* 10:6079-6096. 10.2147/CMAR.S181396
Figure 1

Workflow of exploring RRGs and establishing genomic-clinical nomogram.

Recurrence-related genes, RRGs; The Gene Expression Omnibus, GEO; The Cancer Genome Atlas, TCGA; Gene Ontology, GO; Kyoto Encyclopedia of Genes and Genomes, KEGG; Eastern Cooperative Oncology Group performance status, ECOG-PS; receiver operating characteristic, ROC; the Harrell’s concordance index, C-index.
Figure 2

Overlapping analysis and functional enrichment analysis.

(A) Overlapping analysis for the four GEO datasets; (B) Overlapping analysis for GEO and TCGA databases; (C) GO analysis for DEGs; (D) KEGG pathways enrichment analysis for DEGs. Differently expressed genes, DEGs; the Gene Expression Omnibus, GEO; The Cancer Genome Atlas, TCGA; Gene Ontology, GO; Kyoto Encyclopedia of Genes and Genomes, KEGG.
Figure 3

LASSO logistic regression analysis for selection of RRGs from DEGs.

(A) Selection of tuning parameter (\(\lambda\)) in the LASSO model via 10-fold cross-validation; (B) LASSO coefficients produced by the regression analysis (in A). Recurrence-related genes, RRGs; differently expressed genes, DEGs.
Figure 4

KM analysis for RRGs in TCGA, GSE14520 and GSE76427 datasets.

KM analysis for PRC1(A, D, G), PZP (B, E, H) and SPP2 (C, F, I) in TCGA (A, B, C), GSE14520 (D, E, F) and GSE76427 (G, H, I) datasets, respectively. Kaplan-Meier analysis, KM analysis; The Cancer Genome Atlas, TCGA.
Figure 5

Performance of the gene-based risk model in TCGA, GSE14520 and GSE76427 datasets.

KM analysis, AUC, and calibration curves for the gene-based risk model in TCGA (A, D, G), GSE14520 (B, E, H) and GSE76427 (C, F, I) datasets, respectively. Kaplan-Meier analysis, KM analysis; area under the curve, AUC; The Cancer Genome Atlas, TCGA.
Figure 6

Univariate and multivariate analysis for gene-based risk model and clinical characteristics with postoperative recurrence.

Body Mass Index, BMI; hepatitis B virus, HBV; alpha fetoprotein, AFP; Eastern Cooperative Oncology Group, ECOG.

| Characteristics          | Hazards Ratio (95% CI)       | P  |
|--------------------------|------------------------------|----|
| **Univariate analysis**  |                              |    |
| risk(high/low)           | 2.62(1.69–4.09)              | 0  |
| sex(female/male)         | 1.03(0.64–1.65)              | 0.909 |
| BMI                      | 0.99(0.95–1.04)              | 0.672 |
| race(other/white)        | 0.61(0.40–0.95)              | 0.028 |
| HBV(with/without)        | 0.84(0.54–1.31)              | 0.441 |
| alcohol(yes/no)          | 0.75(0.45–1.26)              | 0.273 |
| grade(III–IV/I–II)       | 1.04(0.67–1.60)              | 0.873 |
| AJCC(II/I)               | 1.96(1.61–3.32)              | 0.012 |
| AJCC(III/I)              | 2.39(1.39–4.12)              | 0.002 |
| AFP(=20/<20)             | 0.80(0.45–1.42)              | 0.446 |
| platelet(=200/<200)      | 1.29(0.83–1.99)              | 0.26 |
| albumin(=4.0/<4.0)       | 0.80(0.51–1.22)              | 0.286 |
| ECOG(1/0)                | 0.98(0.58–1.63)              | 0.926 |
| ECOG(2/0)                | 3.66(1.66–8.10)              | 0.001 |
| age                      | 1.00(0.98–1.01)              | 0.731 |
| **Multivariate analysis**|                              |    |
| risk(high/low)           | 2.27(1.43–3.59)              | 0  |
| race(other/white)        | 0.80(0.49–1.30)              | 0.363 |
| AJCC(II/I)               | 1.88(1.08–3.29)              | 0.027 |
| AJCC(III/I)              | 1.83(1.01–3.31)              | 0.045 |
| ECOG(1/0)                | 0.75(0.43–1.30)              | 0.304 |
| ECOG(2/0)                | 2.98(1.27–6.97)              | 0.012 |
Figure 7

Performance of the genomic-clinical nomogram in predicting postoperative recurrence in TCGA dataset.

(A) Nomogram for predicting 1-, 2- and 3-year probability of recurrence for HCC patients after liver resection; (B) Calibration curves for 1-, 2- and 3-year recurrence of the nomogram; (C-E) DCA to compare the clinical usefulness of the genomic-clinical nomogram (purple line), gene-based risk model (pink line), AJCC staging system (green line) and ECOG (turquoise line) at 1, 2 and 3 years; (F-H) ROC curves analysis to compare the predictive power of the genomic-clinical nomogram (red line), gene-based risk model (blue line) and AJCC staging system (green line) at 1, 2 and 3 years. Hepatocellular carcinoma, HCC; decision curve analysis, DCA; Eastern Cooperative Oncology Group, ECOG; receiver operating characteristic, ROC.
Figure 8

Kaplan-Meier analysis of risk groups stratified using total points of the proposed nomogram.

Red lines represent patients in high risk group and blue lines represent patients in low risk group.