Dear Editor and Reviewers:

Thank you for your letter and for the reviewers' comments concerning our manuscript Entitled 'Eight key lncRNAs predict hepatitis virus positive hepatocellular carcinoma as the prognostic targets' (Manuscript NO: 47823). Those comments are all valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our researches. We have studied comments carefully and have made correction which we hope meet with approval. The main corrections in the paper and the responds to the reviewer's comments are as flowing:

Responds to the reviewer's comments:

Reviewer 1

This manuscript of Huang ZL and coworkers aims at identifying a signature of long non-coding RNAs (LncRNAs) in hepatocellular carcinoma (HCC) associated with viral infections. Using the Cancer Genome Atlas (TCGA) datasets, they identified a signature of 8 LncRNAs as prognosis factors for HCC associated with hepatitis B infection in particular. Some complementary information is needed to clarify this study.

Major points:

1. Some figure legends are lacking (Figures 3-5-6), which rendered the paper difficult to read. In particular for figure 3, a detailed legend for the LASSO analysis is required: what is the optimal lambda value selected? In figure 3A, which curves correspond to the eight LncRNAs?

Response: We have re-written this part according to the Reviewer's suggestion. We have added key lines in Figure 3. It is really true as Reviewer suggested that optimal lambda value is key point in LASSO analysis. Figure 3A showed variables become less when lambda value enlarge. Figure 3B was the cross-validation for tuning parameter selection in the LASSO model. In figure 3A, the solid vertical line indicated the value chosen by cross-validation. In figure 3B, the vertical lines were drawn at the optimal values by minimum criteria and 1-s.e. criteria. Both the vertical lines could be selected in LASSO model, and the solid vertical line representing 8 lncRNAs was suggested by

staff of Epidemiology department of our cancer center (because total number of variables usually ranges from 3 to 10 after LASSO).

We have revised figures' legend as following,

Figure 3 The plots of regression coefficient based on Least absolute shrinkage and selection operator (LASSO) regression. ***(A)*** *LASSO coefficient profiles of the fractions of the 406 significant lncRNAs in univariate Cox regression analysis. The solid vertical line implied the optimal lambda value determined by cross-validation.* ***(B)*** *Cross-validation for tuning parameter selection in the LASSO model. The vertical lines were drawn at the optimal values by minimum criteria and 1-standard error criteria. And the solid vertical line representing 8 lncRNAs was finally determined.*

Figure 5 Conditional interference tree for prediction. *The nodes of the conditional inference tree were calculated automatically based on the variables' influence power. The first node of the tree was risk_score, while the second was stage. Node 3, 4, or 5 was the corresponding group a certain patient would be assigned to. For example, a stage I patient with risk_score <= 0.855 would be divided into group node 3, which was related to the best overall survival.*

Figure 6 Overall survival nomogram predicting 1-, 3- and 5-year survival for hepatocellular carcinoma (A), and calibrated curves to predict the 1- (B), 3- (C), and 5-year (D) patient survival. *In calibrated curves, the X-axis represented the survival prediction by nomogram, and the Y-axis represented the actual observation which showed a good agreement with prediction.*

2. Nothing is said about the 149 selected samples (number of accession…). A table will be informative. Did the authors verify that all the samples were comparable if they were extracted from different datasets?

Response: Each patient in TCGA has an ID number. The virus condition of each patient was recorded as well. And the virus-related patients was summarized in the table below,

	HBV+HCV	HCV alone	HBV alone
Total nubmer	84	9	60

In detail, HBV+HCV were,

TCGA-UB-A7MA, TCGA-ED-A7XO, TCGA-CC-A1HT, TCGA-2Y-A9GZ, TCGA-5R-AA1D, TCGA-5R-AAAM, TCGA-2Y-A9H3, TCGA-CC-A8HT, TCGA-CC-5258, TCGA-CC-A7IF, TCGA-G3-AAUZ, TCGA-CC-A5UE, TCGA-CC-5259, TCGA-ED-A627, TCGA-UB-AA0U, TCGA-ED-A82E, TCGA-WJ-A86L, TCGA-DD-A114, TCGA-G3-AAV0, TCGA-2Y-A9H2, TCGA-DD-A4NQ, TCGA-DD-A4NO, TCGA-2Y-A9H4, TCGA-UB-A7MB, TCGA-DD-A4NE, TCGA-5R-AA1C, TCGA-CC-A9FV, TCGA-G3-A7M5, TCGA-FV-A3R2, TCGA-ZS-A9CF, TCGA-CC-A9FW, TCGA-DD-A4NF, TCGA-CC-A7IL, TCGA-ZS-A9CD, TCGA-CC-A3MA, TCGA-CC-A7II, TCGA-CC-A3MC, TCGA-CC-A3MB, TCGA-CC-5264, TCGA-CC-A7IJ, TCGA-CC-A9FS, TCGA-G3-A6UC, TCGA-G3-AAV2, TCGA-G3-AAV5, TCGA-HP-A5MZ, TCGA-ZS-A9CG, TCGA-CC-A9FU, TCGA-4R-AA8I, TCGA-UB-AA0V, TCGA-CC-5260, TCGA-CC-A7IG, TCGA-MR-A520, TCGA-2Y-A9GU, TCGA-2Y-A9GV, TCGA-CC-5263, TCGA-CC-A7IK, TCGA-CC-A5UC, TCGA-2Y-A9H7, TCGA-2Y-A9GT, TCGA-CC-A123, TCGA-ZS-A9CE, TCGA-ED-A8O5, TCGA-CC-5262, TCGA-CC-A7IH, TCGA-G3-A7M9, TCGA-UB-A7MF, TCGA-UB-A7MC, TCGA-MR-A8JO, TCGA-CC-A3M9, TCGA-G3-A7M8, TCGA-CC-A7IE, TCGA-CC-A8HS, TCGA-ED-A8O6, TCGA-DD-A1EE, TCGA-G3-A7M7, TCGA-G3-AAV1, TCGA-FV-A3I1, TCGA-CC-A5UD, TCGA-CC-A8HU, TCGA-CC-5261, TCGA-2Y-A9GS, TCGA-DD-A73D, TCGA-UB-A7MD, and TCGA-CC-A8HV

HCV alone were,

TCGA-WX-AA44, TCGA-2Y-A9HA, TCGA-GJ-A3OU, TCGA-WX-AA47, TCGA-RG-A7D4, TCGA-KR-A7K0, TCGA-2Y-A9H1, TCGA-2Y-A9H5, and TCGA-WX-AA46

HBV alone were,

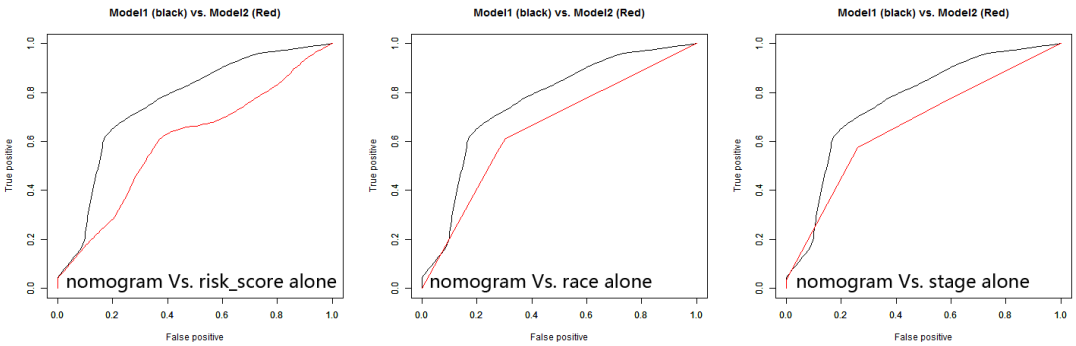
TCGA-DD-A4NR, TCGA-DD-A118, TCGA-DD-A4NH, TCGA-DD-A1EI, TCGA-DD-A1EK, TCGA-DD-A116, TCGA-DD-A4NP, TCGA-ED-A66X, TCGA-DD-A3A0, TCGA-DD-A3A7, TCGA-DD-A3A6, TCGA-ED-A7PX, TCGA-DD-A73B, TCGA-DD-A4NV, TCGA-ED-A7PZ, TCGA-ED-A5KG, TCGA-ED-A459, TCGA-DD-A4NJ, TCGA-DD-A11A, TCGA-DD-A4NK, TCGA-DD-A1EA, TCGA-DD-A119, TCGA-DD-A73E, TCGA-BC-A10W, TCGA-ED-A97K, TCGA-G3-AAV7, TCGA-DD-A113, TCGA-G3-AAV6, TCGA-DD-A4NG, TCGA-DD-A1E9, TCGA-2Y-A9GX, TCGA-ZP-A9CZ, TCGA-DD-A1EG, TCGA-UB-A7ME, TCGA-DD-A4NS, TCGA-DD-A11C, TCGA-DD-A4NA, TCGA-BD-A2L6, TCGA-DD-A39Y, TCGA-DD-A1EL, TCGA-DD-A1EJ, TCGA-DD-A73A, TCGA-DD-A1EH, TCGA-DD-AA3A, TCGA-DD-A1EF, TCGA-ED-A7PY, TCGA-DD-A115, TCGA-ED-A7XP, TCGA-ED-A4XI, TCGA-DD-A4NL, TCGA-DD-A3A8, TCGA-DD-A11B, TCGA-DD-A1EB, TCGA-DD-A39X, TCGA-BD-A3ER, TCGA-DD-A4NI, TCGA-ED-A66Y, TCGA-DD-A73G, TCGA-DD-A1ED, and TCGA-DD-A73C.

All the samples were extracted from one dataset (the TCGA project). TCGA is a landmark cancer genomics program, characterizing 33 cancer types including HCC. Here, we focused on virus-related HCC specially.

3. In figure 6, the authors compiled the three risk factors (TNM stage, RACE, LncRNA risk score) for the risk of recurrence. Have they obtained better results for the three than the risk of recurrence obtained for one risk factor independently analyzed?

Response: Thank you for your positive comments. We conducted univariate and multivariate Cox analysis (table below), and found three factors (TNM stage, RACE, LncRNA risk score) were associated with survival. Then a nomogram containing these three factors was constructed. Furthermore, considering your positive suggestion, we compared the nomogram model with models constructed with one risk factor independently. We found that nomogram obtained best result (figure below, black line indicated nomogram).

Variable	**Univariate analysis**			**Multivariate analysis**		
	HR	**95% CI**	***P*-value**	**HR**	**95% CI**	***P*-value**
Risk_score	1.14	1.096-1.185	4.46 e-11	1.944	1.613-2.343	2.94 e-12
Age	1.002	0.986-1.019	0.807	1.004	0.986-1.023	0.662
Gender	1.288	0.775-2.143	0.329	0.832	0.480-1.442	0.512
Race	1.686	1.155-2.463	0.007	1.531	1.005-2.330	0.047
Grade	1.287	0.912-1.818	0.151	1.018	0.702-1.477	0.925
Stage	1.728	1.289-2.318	2.59 e-4	1.584	1.167-2.151	0.003



4. It will be interesting to precise which LncRNAs among the 8 are involved in the Top 10 deregulated pathways in figure 8.

Response: We have revised this figure according to the Reviewer's suggestion. Corresponding lncRNAs were also presented in the new figure.

Figure 8 Functional annotations for the eight lncRNAs. *The 10 most remarkable pathways enriched based on the co-expressed mRNAs of eight lncRNAs. Corresponding*

lncRNAs involved in the deregulated pathways were also presented. A, AC005722.2; B, AC107959.3; C, AL353803.1; D, AL589182.1; E, AP000844.2; F, AP002478.1; G, FLJ36000; H, NPSR1-AS1.

Minor points:

1. The cut-off value used to discriminate high risk and low risk will be informative in figure 4 A and B.

Response: We have added the cut-off value to Figure 4 legend.

Figure 4 Establishment and verification of the lncRNA signature for predicting the prognosis for VHCC using the cut-off value 0.798. (A) *Distribution of the lncRNA risk scores.* ***(B)*** *Vital status as well as OS.* ***(C)*** *Kaplan-Meier curve regarding the OS of high-risk as well as low-risk group classified based on median risk score.* ***(D)*** *ROC curves regarding the 1-, 3-, and 5-year survival discriminated by the lncRNA signature.*

2. The figure 2A is unreadable.

Response: We are very sorry, and a clear figure has been uploaded.

Reviewer 2

Dear sir, thank you to select me to review manuscript Zi-Lin Huang et al. Eight key lncRNAs predict hepatitis virus positive hepatocellular carcinoma as the prognostic targets. The aim of this intersesting study was to construct long non-coding RNAs (lncRNAs) signature to predict overall survival (OS) in viral hepatitis hepatocellular cancer (VHCC). Authors analyzed data of 149 VHCC patients from The Cancer Genome Atlas (TCGA) database. A total of 1420 differentially expressed lncRNA (DElncRNAs) were identified, among which, 406 were significant in univariate Cox regression analysis. LASSO regression confirmed had 8 out of the 406 lncRNAs.the area under curve (AUC) of the prognostic model constructed based on the eight-lncRNA biomarkers were 0.73, 0.758, and 0.788 at 1-, 3-, and 5-year OS, respectively.

Authors also constructed nomogram at the base the eight-lncRNA risk_score, TNM stage and race, nomogram predicted OS in VHCC patients. Paper is well made, but some changes are needed.

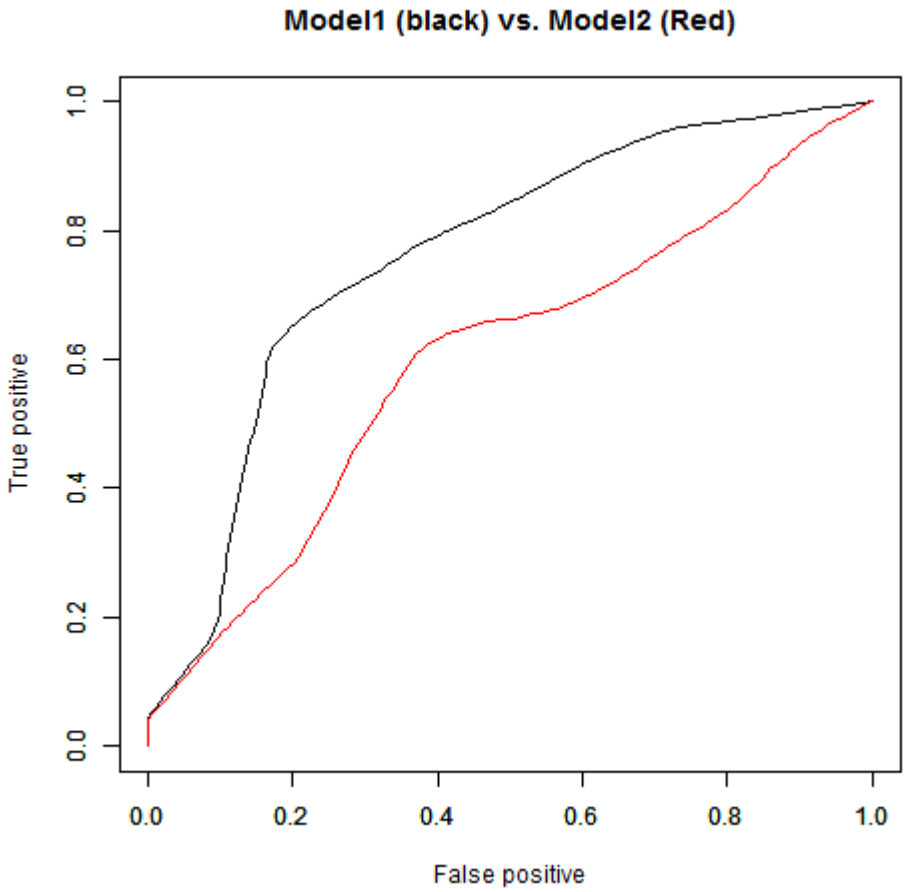
1) *Better clinical characteristics of patients are needed - duration of viral hepatitis, HBeAg/anti HBe status, HBV DNA at the time of HCC diagnosis, Milano or metro ticket travel criteria for HCC, treatment of HCC, antiviral treatment for HBV or HCV, etc. If data are unavialable, please add this information to the limitation of the study*.

Response: Thank you for your positive comments. We have added these limitation in our manuscript as following,

Secondly, data regarding more valuable clinical features are not available, including viral hepatitis duration, state of virus, Milan standards, HCC treatment, and anti-virus treatment.

2) *Is the nomogram better for prediction of OS than prognostic model constructed from 8 lncRNAs?*

Response: Considering your positive suggestion, we compared the nomogram model with 8 lncRNAs' model. The black line indicated nomogram, while red line indicated 8 lncRNAs' model. We found that nomogram had better result (figure below).



3) *Figures 2A and 6 are unreadable.*

Response: We are very sorry. We have re-uploaded this part according to the Reviewer's suggestion.

4) *Figure 8 - please explain the role of 8 lncRNAs in HCC pathogenesis, add the information in the figure. My decision is minor revision.*

Response: Corresponding lncRNAs information have been added in the new figure 8. The hypothesis of 8 lncRNAs were presented in fourth paragraph of discussion part.

We tried our best to improve the manuscript and made some changes in the manuscript. These changes will not influence the content and framework of the paper. We appreciate

for Editors/Reviewers' warm work earnestly, and hope that the correction will meet with approval.

Once again, thank you very much for your comments and suggestions.

I look forward to hearing from you soon.

With best wishes,

Yours sincerely,