

15-lncRNAs-based classifier-clinicopathologic nomogram improves the prediction of recurrence in patients with hepatocellular carcinoma

Qiong Zhang

Guangzhou Medical University Affiliated Cancer Hospital

Gang Ning

Third Affiliated Hospital of Sun Yat-Sen University

Hongye Jiang

Shunde Hospital of Southern Medical University

Yanlin Huang

Third Affiliated Hospital of Sun Yat-Sen University

Jinsong Piao

Guangzhou Medical University Affiliated Cancer Hospital

Chen Zhen

Shunde Hospital of Southern Medical University

Xiaojun Tan

Guangzhou Medical University Affiliated Cancer Hospital

Genglong Liu (✉ lglong3@mail2.sysu.edu.cn)

Affiliated Cancer Hospital of Guangzhou Medical <https://orcid.org/0000-0002-9163-4396>

Research article

Keywords: hepatocellular carcinoma; lncRNAs classifier; nomogram; prediction; recurrence-free survival;

Posted Date: October 5th, 2019

DOI: <https://doi.org/10.21203/rs.2.15723/v1>

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

[Read Full License](#)

Abstract

Background: Our study aims to develop an lncRNAs-based classifier and a nomogram incorporating the genomic signature and clinicopathologic factors to help to improve the accuracy of recurrence prediction for hepatocellular carcinoma (HCC) patients. **Methods:** The lncRNAs profiling data of 374 HCC patients and 50 normal healthy controls were downloaded from the Cancer Genome Atlas (TCGA). Using univariable Cox regression and Least absolute shrinkage and selection operator (LASSO) analysis, we developed 15-lncRNAs-based classifier and compared our classifier to existing six-lncRNAs signature. Besides, a nomogram incorporating the genomic classifier and clinicopathologic factors was also developed. The predictive accuracy and discriminative ability of the genomic-clinicopathologic nomogram were determined by a concordance index (C-index) and calibration curve and were compared with TNM staging system by C-index, receiver operating characteristic (ROC) analysis. Decision curve analysis (DCA) was performed to estimate clinical value of our nomogram. **Results:** Fifteen relapse-free survival (RFS) -related lncRNAs were identified and the classifier, consisting of the identified 15 lncRNAs, could effectively classify patients into high-risk and low-risk subgroup. The prediction accuracy of the 15-lncRNAs-based classifier for predicting 2-year and 5-year RFS were 0.791 and 0.834 in the training set and 0.684 and 0.747 in the validation set, which was better than the existing six-lncRNAs signature. Moreover, the AUC of genomic-clinicopathologic nomogram in predicting RFS were 0.837 in the training set and 0.753 in the validation set, and the C-index of the genomic-clinicopathologic nomogram was 0.78 (0.72-0.83) in the training set and 0.71 (0.65-0.76) in the validation set, which was better than traditional TNM stage and 15-lncRNAs-based classifier. Decision curve analysis further demonstrated that our nomogram had larger net benefit than TNM stage and 15-lncRNAs-based classifier. **Conclusion:** Compared to TNM stage, the 15-lncRNAs-based classifier-clinicopathologic nomogram is a more effective and valuable tool to identify HCC recurrence and may aid in clinical decision-making.

Background

Hepatocellular carcinoma (HCC) is one of most malignant cancers which represents the fourth leading cause of cancer-related death and the most common cause of mortality of cirrhotic patients. Every year, there are 841,000 patients developed HCC and 782,000 HCC patients died [1]. Till now, only two therapeutic treatments, including surgical resection and liver transplantation, is recommended as the first-line therapy to potentially cure HCC [2]. However, surgical resection is hampered as more than 70% of HCC patients experience disease recurrence approximately at 5 years after resection [3], and donor organ shortages render the large-scale application of liver transplantation. Therefore, identifying reliable and accurate predictive markers/models to screen out which subset of patients with HCC is vulnerable to develop recurrence is urgently needed.

Long non-coding RNAs (lncRNAs) are newly-discovered RNA transcripts which were found to play an important role in cell differentiation and development by transcriptionally or post-transcriptionally regulating biological processes [4; 5; 6]. Besides, increasing number of studies have reported the

Loading [MathJax]/jax/output/CommonHTML/jax.js nt and progression of cancers, including HCC [7; 8]. These

studies suggested that lncRNAs may be developed as potential useful biomarkers in diagnosis and prognosis of HCC. For example, overexpression of lncRNA LOC90784, RUSC1-AS-N, RNA AWPPH have been associated with poor clinical features and poorer overall survival in HCC patients [9; 10; 11]. Moreover, prognostic signature based on lncRNAs has also been found to improve the prognosis prediction of HCC [12; 13], but the predictive value of lncRNAs-based signature in recurrence of HCC remains poorly evaluated.

In the present study, we aimed to develop a lncRNAs-based classifier and a nomogram incorporating the genomic signature and clinicopathologic factors to help to improve the accuracy of recurrence for HCC patients after surgery. We identified lncRNAs that were significantly associated with relapse free survival (RFS) of HCC patients from the Cancer Genome Atlas (TCGA) and then used them to constructed lncRNAs-based classifier in the training set. A nomogram incorporating the lncRNAs-based classifier and clinicopathologic factors was also developed for predicting RFS. Finally, the predictive ability of the nomogram was evaluated and validated in an internal validation set.

Methods

Ethics Statement

All the data was obtained from TCGA, and the informed consent had been attained from the patients before our study.

Collection of lncRNAs data and clinical characteristics of HCC patients from TCGA

The lncRNAs profiling data of 374 HCC patients and 50 normal healthy controls were downloaded from TCGA. Then clinical parameters, such as age, gender, family history, alcohol consumption, mutation count, fraction genome altered, BMI, APF, platelet, albumin, creatinine, cirrhosis, histologic grade, T stage, TNM stage, Eastern Cooperative Oncology Group (ECOG), and RFS time were also downloaded from TCGA. Eighty-one HCC patients were excluded due to RFS time < 1 month or the unavailability of lncRNAs data. So, 293 HCC patients with available lncRNAs data and clinical characteristics were finally included in our study. Subsequently, 293 HCC patients were randomly assigned to a training set (N=147) and a validation set (N=146) by R software.

Construction and validation of lncRNAs-based classifier for RFS

First, moderated t-statistics method and Benjamini–Hochberg procedure were used to identify distinct differential lncRNAs between normal tissues and HCC tissues. The cut-off criteria of distinct differential lncRNAs was $P < 0.05$ and the false discovery rate (FDR) < 0.05 . Next, univariable Cox regression analysis

Loading [MathJax]/jax/output/CommonHTML/jax.js the training set. Least absolute shrinkage and selection

operator (LASSO) analysis was used to further narrow down the RFS-related lncRNAs and a signature consisting these well-selected lncRNAs was developed [14]. Lasso analysis is a popular estimation procedure in multiple linear regression when underlying design has a sparse structure, as it could set some regression coefficients exactly equal to 0. In our study, we developed a perturbation bootstrap method and established its validity in approximating the distribution of the Lasso in heteroscedastic linear regression. The underlying covariates were allowed to be either random or non-random and the proposed bootstrap method was proved to work irrespective of the nature of the covariates. Simulation study also justified our method in finite samples. In order to obtain accurate estimate stability of our model, cross validation was performed, as it could provide unbiased estimation [15]. In the present study, a 5-fold cross validation was carried out, the 293 HCC patients were divided into 5 subsets of equal size and trained for 5 times, each time leaving out a subsets as validation data. The accuracy were 90.3%, 91.1%, 90.8%, 90.4% and 91% , respectively, indicating the good stability of our 15-lncRNAs-based classifier. By this classifier, we calculated the risk scores of HCC patients were and then divided patients into high-risk patients and low-risk subgroup based on the best cut-of value, which was a point when the Youden index (sensitivity +specificity-1) reached the maximum value in the training set. The RSF difference between high-risk patients and low-risk patients were further compared by the Kaplan-Meier analysis. Log-rank test was used to compare subgroups. The flowchart of the present study was shown in **Figure 1**.

Receiver operating characteristic (ROC)

To further evaluate the predictive accuracy of lncRNAs-based classifier, ROC analysis was performed in the training set and validation set. We calculated area under ROC curve (AUC) of lncRNAs-based classifier for predicting 2-year and 5-year RFS and compared the predictive ability of our lncRNAs-based signature with other published lncRNAs signature [16] for RFS which also developed from HCC patients of TCGA.

Genomic-clinicopathologic nomogram

In order to make lncRNAs-based classifier to be more applicable for clinicians, a lncRNAs-based classifier related nomogram was constructed. First, univariate and multivariate Cox regression analysis were used to identify clinical risk parameters associated with RFS in the training set. Next, the lncRNAs-based classifier, together with the risk parameters, were used to develop a genomic-clinicopathologic nomogram in the training set.

Model performance was evaluated by determining the calibration and discrimination. Discrimination is the models ability to differentiate between patients who recur from HCC and patients who will not. Discrimination was calculated through the concordance index (C-index). We also illustrated

discrimination by dividing the dataset into three groups based on the score generated by the nomogram. We plotted a Kaplan–Meier curve for all three groups.

Calibration of the nomogram was assessed by plotting the observed RFS rate (the mean Kaplan-Meier estimate for patients in each octile) against the nomogram 2- and 5-year predicted RFS probability (ie, the mean nomogram predicted probability for patients in each octile). A perfectly accurate nomogram prediction model would result in a plot in which the observed and predicted probabilities for given groups would fall along the 45-degree line. The distance between the pairs and the 45-degree line was a measure of the absolute error of the nomogram's prediction.

ROC analysis was used to evaluate and compare the discrimination ability of the nomogram with lncRNAs-based classifier and TNM stage. Then, decision curve analysis (DCA) was performed to evaluate the clinical usefulness of the genomic-clinicopathologic nomogram [17; 18]. DCA was performed by calculating the net benefit for a range of threshold probabilities, which place benefits and harms on the same scale. This analysis determined whether clinical decision-making based on a model would do more good than harm. DCA provided straightforward information about the clinical value of a model, in contrast to traditional measures such as sensitivity or specificity, which were abstract statistical concepts.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of the 15-lncRNAs-based classifier

To explore the biological function and pathways of the 15-lncRNAs-based classifier, GO and KEGG analysis were conducted. First, Pearson correlation algorithm was performed between these 15 lncRNAs and the protein-coding genes (mRNAs) and correlation coefficient >0.4 , $p < 0.001$ were considered as significant correlation. Then, potential biological processes of these lncRNAs target genes were further investigated by Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis in DAVID, a common bioinformatics tool (<http://david.abcc.ncifcrf.gov/>, version 6.8) [19].

Statistical analysis

SPSS statistics 22.0 and R software (R version 3.5.2) were used to conduct the statistical analysis. Univariate and multivariate Cox regression analysis was performed to identify potential predictors associated with RFS. If there were missed data in some of the potential predictors, these missing data would be imputed, as full case analysis would improve the statistical power and reduce potentially biased results. Multiple imputation was used to impute the missing data as the missing data were considered missing at random after analyzing patterns of them. Multiple imputation was performed MI with the Markov Chain Monte Carlo function and 5 iterations were used to account for possible simulation errors.

LASSO analysis was performed with “glmnet” packages, cross validation was conducted with “caret” packages, and ROC analysis was done with “survivalROC” packages. The nomogram and calibration plots were generated with “rms” packages, and DCA was performed with the “stdca.R”. A two sided $P < 0.05$ would be recognized as statistically significant.

Results

Demographic parameters and RFS outcome of HCC patients

In the present study, 293 HCC patients with available lncRNAs data and clinical characteristics were included. The basic clinical characteristics of these HCC patients were summarized in **Table 1**. The median RFS was 20.99 months (range: 1.22- 120.73 months). Of all the 293 HCC patients, 170 (57.9%) patients developed recurrence during follow-up and the 2-year and 5-year RFS rates were 46.4% and 29.1%, respectively. It was also of note that among the recurrent patients, 143 patients (84.1%) experienced recurrence during the first two years after resection.

Development and validation of lncRNAs-based classifier

First, 1292 distinct differential lncRNAs between normal tissues and HCC tissues were got basing on the filter criteria described on the section of Methods (**supplementary material 1**). Then, the top 30 RFS-related lncRNAs were identified by univariable Cox regression analysis in the training set (**supplementary material 2**). Next, LASSO analysis were used to further narrow down the RFS-related lncRNAs (**Figure 2**) and we finally selected 15 RFS-related lncRNAs, which were AC012625.1, AC068481.1, AC092675.1, AC093772.1, AC109779.1, AC118653.1, AC246785.3, AL121985.1, AL121985.1, AP002478.1, ARHGEF7.AS1, GACAT3, LINC00462, LINC01700 and LINC02429 (**Table 2**). On the basis of the coefficients weighted by LASSO analysis, a classifier was developed, and the risk score was as follows: risk score=(0.05174*AC012625.1)+(0.10870*AC068481.1)+(0.05747*AC092675.1)+(0.05770*AC093772.1)+ (-0.32522*AC109779.1)+(0.06050*AC118653.1)+ (0.05021*AC246785.3)+(0.03112*AL121985.1)+(0.12532*AL512604.2)+(0.06333 *AP002478.1)+(0.10057*ARHGEF7.AS1)+(0.26230*GACAT3)+(0.25478* LINC00462)+(0.21032*LINC01700) + (0.12948*LINC02429). With this classifier, risk score for every HCC patient would be calculated and then they were classified into high-risk patients and low-risk patients according to the best cut-off (described in Method). As was shown at **Figure 3**, patients with high risk score were more likely to develop recurrence and had shorter RFS than those with low risk score in the training set (9.89 vs 67.58 months, HR=3.96, 95%CI: 2.5-6.3, $P < 0.0001$). Verification analysis were further performed in the validation set and in the total cohort. Similarly, the 15-lncRNAs-based classifier could also classify patients into the high-risk and the low-risk subgroup by the same cut-off value. The median RFS time of high-risk patients was shorter than low-risk patients in the validation set (14.22 vs 27.2 months, HR=1.941, 95%CI: 1.28-2.94, $P < 0.0001$) **A**). Besides, the 15-lncRNAs-based classifier could also

classify patients into the high-risk with shorter overall survival (OS) time and the low-risk subgroup with longer OS time no matter in the training set, the validation set or the total cohort (all $P < 0.0001$, Figure S2A, Figure S2C, Figure S1C). Taken together, these results suggested that the 15-lncRNAs-based classifier could effectively classify HCC patients into two distinct subgroups with high risk or low risk of recurrence or OS.

Predictive value of 15-lncRNAs-based classifier and comparison with other lncRNAs-based classifier from TCGA

Recently, a six-lncRNAs-based signature was developed and validated by Gu et al to predict RFS with the data of HCC patients from TCGA [16]. To compare the predictive value of our 15-lncRNAs-based classifier with the six-lncRNAs-based signature, ROC curve analysis was performed. As was shown at **Figure 4**, the AUC of 15-lncRNAs-based classifier for predicting 2-year and 5-year RFS in the training set were 0.791 and 0.834, respectively, while that of six-lncRNAs-based signature were 0.545 and 0.612, respectively (**Figure 4A, 4C**). Similar results were also found in the validation set. The AUC of 15-lncRNAs-based classifier for predicting 2-year and 5-year RFS were 0.684 and 0.747, respectively, and the AUC of the six-lncRNAs-based signature were 0.65 and 0.624, respectively (**Figure 4B, 4D**). Besides, the 15-lncRNAs-based classifier also had good prognostic value for RFS of the 293 HCC patients (total cohort) as the AUC of 15-lncRNAs-based classifier for predicting 2-year and 5-year DFS were 0.753 and 0.814 respectively (Figure S1B). What's more, we also examined the prognostic value of the 15-lncRNAs-based classifier for OS. The AUC of 15-lncRNAs-based classifier for predicting 2-year and 5-year OS were 0.846 and 0.816 in the training set (Figure S2B), 0.791 and 0.804 in the validation set (Figure S2D), and 0.828 and 0.805 in the total cohort (Figure S1D). These results suggested that 15-lncRNAs-based classifier demonstrated better performance in predicting RFS than the six-lncRNAs-based signature.

Development of genomic-clinicopathologic nomogram

To make lncRNAs-based classifier to be more applicable for clinicians, a 15-lncRNAs-based classifier-clinicopathologic nomogram was developed to predict the 2-year and 5-year RFS in HCC patients. Potential predictors associated with RFS were identified by univariate and multivariate Cox regression analysis in the training set. Univariable Cox regression analysis showed that mutation count, BMI, AFP, liver cirrhosis, tumor stage, TNM stage, ECOG and the 15-lncRNAs-based classifier was related with RFS, and multivariate Cox regression analysis further showed that mutation count, AFP, T stage, ECOG and the 15-lncRNAs-based classifier were independent predictors of RFS (**Table 3**). So, these five predictors were used to develop the genomic-clinicopathologic nomogram (**Figure 5**), which would help clinicians to preoperatively predict the recurrence risk in HCC patients. The C-index of the genomic-clinicopathologic nomogram was 0.78 (0.72-0.83) (**Table 4**) and the calibration plots exhibited good consistency between **Figure 6A, 6C**). Likewise, consistent results were also found in the

validation set. The C-index of the genomic-clinicopathologic nomogram in the validation set was 0.71 (0.65-0.76) (**Table 4**) and also showed good consistency between the predicted RFS and the actual RFS (**Figure 6B, 6D**). Additionally, the tertiles of all the

total points were used to divide the patients into high-, intermediate- and low-risk groups with distinct RFS time or OS time. The Kaplan-Meier analysis (Log-rank $P < 0.0001$) of the three risk subgroups indicated the great utility of the genomic-clinicopathologic nomogram for in training set (Figure S3A, Figure S4A), in validation set (Figure S3B, Figure S4B), and in the total cohort (Figure S5A, S5B). All these results indicated the perfect performance of our genomic-clinicopathologic nomogram.

To further evaluate the predictive ability of the genomic-clinicopathologic nomogram, we compared the C-index and ROC analysis results of genomic-clinicopathologic nomogram with AJCC TNM stage and the 15-lncRNAs-based classifier in the training set and validation set. As was shown at **Table 4**, the C-index of genomic-clinicopathologic nomogram was higher than that of TNM stage (0.64 (0.58-0.69) in the training set, and 0.56 (0.51-0.60) in the validation set) and the 15-lncRNAs-based classifier (0.74 (0.68-0.80) in the training set, and 0.61 (0.55-0.68) in the validation set). Likelihood ratio test, linear trend χ^2 test and akaike information criterion all demonstrated that the genomic-clinicopathologic nomogram had higher prediction efficiency than the TNM stage or 15-lncRNAs-based classifier alone. Similar to C-index, ROC analysis also indicated that the genomic-clinicopathologic nomogram (AUC 0.837 for the training set, and 0.753 for the validation set) was better than TNM stage (AUC 0.661 for the training set, and 0.585 for the validation set) or 15-lncRNAs-based classifier (AUC 0.791 for the training set, and 0.684 for the validation set) alone in predicting RFS (**Figure 7A, Figure 7B**).

Finally, the clinical usefulness of the genomic-clinicopathologic nomogram was assessed by the DCA, an abstract statistical concept, which gave visualized information on the clinical value of a model. As were presented in **Figure 8A and Figure 8B**, DCA results showed that HCC recurrence-associated treatment decision based on the genomic-clinicopathologic nomogram resulted more net benefit than treatment decision based on TNM stage or 15-lncRNAs-based classifier, or treating either all patients or none in the training set and the validation set.

Biological function and pathways of 15-lncRNAs-based classifier

To explore the biological function and pathways of 15-lncRNAs-based classifier, GO and KEGG analysis were performed. As were shown at Figure S6, biological processes such as GO:0048037 (cofactor binding), GO:0050662 (coenzyme binding), GO:0016614 (oxidoreductase activity, acting on CH-OH group of donors), GO:0009055 (electron transfer activity) and GO:0016616 (oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor) were mainly regulated by 15-lncRNAs-based classifier (Figure S6A). KEGG analysis also showed that carbon metabolism, peroxisome, fatty acid degradation, valine, leucine and PPAR signaling pathway were mainly affected by 15-lncRNAs-based classifier. These results indicated an important role played by abnormal metabolism in

hepatocellular carcinogenesis. Our findings were in accordance with the well-known viewpoint that alterations in cellular metabolism were hallmarks of cancer and various works suggested that lipid and glucose metabolism took an active role in hepatocellular carcinogenesis [20], which provided evidence for the rationality and molecular basis of 15-lncRNAs-based classifier.

Discussion

Based on lncRNAs profiling data and clinical parameters of 293 HCC patients from TCGA, we identified 15 lncRNAs related to RFS. On the basis of these lncRNAs, we developed and validated a classifier, which could effectively classify patients into high-risk patients with shorter RFS and low-risk patients with longer RFS. The 15-lncRNAs-based classifier is significantly associated with tumor recurrence and provides accurate prediction in predicting 2-year and 5-year RFS of HCC patients. More importantly, we also constructed and validated an lncRNAs-clinicopathologic nomogram incorporating clinicopathologic parameters of mutation count, AFP, T stage and ECOG, which might contribute to individual evaluation of RFS in HCC patients after surgical resection.

Increasing numbers of studies have found that lncRNAs may be exploited as potential effective biomarkers in diagnosis and prognosis of HCC [9; 11]. Yan et al reported that seven lncRNAs, including AC009005.2, RP11-363N22.3, RP11-93209.10, RP11-57206.1, RP11-190C22.8, RP11-388C12.8, and ZFPM2-AS1 were associated with OS of HCC patients and the seven-lncRNA signature could also divide patients into high-risk and low-risk groups with significantly different OS [13]. Wu et al found that MIR22HG, CTC-297N7.9, CTD-2139B15.2, RP11-589N15.2, RP11-343N15.5, and RP11-479G22.8 were independent predictors of HCC patients' OS, and a signature consisting of these six lncRNAs can effectively classify patients into high-risk patients with shorter survival time and low-risk patients with longer survival, and the six-lncRNA signature exhibited superior predictive capacity than TNM stage [12]. These studies suggested the potential clinical implications of lncRNA-based signature in improving the prognosis prediction of HCC. However, it should be noted that OS is more likely to be influenced by post-recurrence treatment and liver function, RFS could more accurately reflect the biologic behavior for HCC, thus in the present study, we tried to identify RFS-related lncRNAs, developed and validated a classifier, which may be more valuable for HCC patients management.

Recently, Gu et al developed and validated a six-lncRNAs-based signature to predict RFS also with the data of HCC patients from TCGA [16]. They demonstrated that MSC-AS1, POLR2J4, EIF3J-AS1, SERHL, RMST, and PVT1 were significantly up-regulated in tumor samples compared to non-tumor samples and were significantly associated with RFS. Different from the abovementioned six lncRNAs, we identified another 15 RFS-related lncRNAs, including AC012625.1, AC068481.1, AC092675.1, AC093772.1, AC109779.1, AC118653.1, AC246785.3, AL121985.1, AL121985.1, AP002478.1, ARHGEF7.AS1, GACAT3, LINC00462, LINC01700 and LINC02429 in the present study. The reason for the differences between our findings and other studies may be attributable to the difference in training datasets. TCGA dataset was used as the training dataset in present study, while GSE76427 was used in the study by Gu et al. The training dataset

Loading [MathJax]/jax/output/CommonHTML/jax.js

for further investigation. However, it should be noted that the

performance of a classifier lay on its prediction sensitivity and specificity. The prediction accuracy of the 15-lncRNAs-based classifier for predicting 2- year and 5-year RFS were 0.791 and 0.834 in the training set and 0.684 and 0.747 in the validation set, while the AUC of 6-lncRNAs-based signature were 0.545 and 0.612 in the training set and 0.65 and 0.624 in the validation set, suggesting that our 15-lncRNAs-based classifier exhibited better efficiency in predicting RFS.

Except for the six-lncRNAs-based signature, Gu et al also developed another lncRNAs-based signature for HCC recurrence in patients with small HCC (maximum tumor diameter ≤ 5 cm) [21]. In his study, a 3-lncRNAs-based signature, consist of LOC101927051, LINC00667 and NSUN5P2, was developed and validated for predicting OS and RFS in patients with small HCC. This 3-lncRNAs-based signature was more suitable for patients with higher serum levels of AFP (>20 ng/ml) and relatively lower levels of albumin (<4.0 g/dl) or Asian patients with no family history of HCC or history of alcohol consumption. Similar to our study, Gu et al also constructed a lncRNAs–clinicopathologic nomogram incorporating clinicopathologic parameter of liver cirrhosis and ECOG for predicting the 1-year and 3-year RFS with a C-index of 0.633. Compared to the 3-lncRNAs-based clinicopathologic nomogram, our 15-lncRNAs–clinicopathologic nomogram demonstrated better predictive ability (C-index of 0.78 in the training set and 0.71 in the validation set). Additionally, the nomogram by Gu et al was suitable for patients with small HCC, while our nomogram was developed for predicting recurrence of all HCC patients, potentially contributing to the prediction of recurrence in a broader patients with HCC. Secondly, the nomogram by Gu et al was used for predicting the 1-year and 3-year RFS, while our nomogram was developed for predicting 2- year and 5-year RFS, which was more fit with early recurrence (recurrence within 2 years) and long-term recurrence (recurrence within 5 years). Finally, we also validated our nomogram in an internal validation set to validate the stability of our nomogram, which was not performed in the study by Gu et al. Moreover, our nomogram appear to be a favorable predictive model compared to TNM stage with a C-index of 0.64 in the training set and 0.56 in the validation set, indicating that our genomic-clinicopathologic nomogram may represent a reliable and accurate model for HCC recurrence prediction, which is helpful for individualized therapeutic treatment decision and post-treatment follow-up decision making.

Consistent with previous studies, mutation count, AFP, T stage, ECOG were found to be significantly associated with HCC recurrence in the present study [21; 22; 23; 24]. In addition to these clinical factors, as expected, the 15-lncRNAs-based signature, was also found to be significantly related with HCC recurrence in our study. Till now, only a few validated clinical nomograms for HCC recurrence have been reported [25; 26; 27]. For example, a nomogram consisting of 7 clinical factors, including age, AFP, PT, magnitude of hepatectomy, postoperative complication, number of tumor nodules and microvascular invasion was developed and validated using data of 617 HCC patients. However, the results limited its use for HCC patients beyond the Milan criteria [26]. Another nomogram incorporating sex, log of calculated tumor volume, ALB, platelet count, and microvascular invasion, was well-constructed and validated by Shim et al with data from 1085 HCC patients. However, this model could only be applied for early-stage HCC patients and the authors only evaluated the prediction accuracy for predicting 2- year

Loading [MathJax]/jax/output/CommonHTML/jax.js

cribed above, we developed and validated a nomogram

incorporating clinicopathologic parameter and genomic data, which may help to improve the stability and accuracy of prediction probability of the nomogram [28].

Among the 15 RFS-related lncRNAs, AP002478.1, GACAT3 and LINC00462 have been previously reported to be related with cancers. AP002478.1 has been reported to be potential prognostic biomarkers for HCC patients and gastric cancers patients [29; 30]. GACAT3 is the first to be found to significantly overexpress in gastric cancer tissues and gastric cancer cell line MGC-803. Higher expression was of GACAT3 significantly associated with tumor size, distant metastasis, TNM stages and shorter OS [31; 32]. One mechanism study showed that GACAT3 knockdown could significantly inhibit proliferation, colony formation, migration, and invasion of GC cells by regulating miR-497, while down-regulation of GACAT3 decreased its tumorigenesis [33]. Moreover, GACAT3 was also found to similar tumorigenesis in breast cancer, glioma and colorectal cancer [34; 35; 36]. Up-regulation of LINC00462 was found to be associated with larger tumor size, poorer tumor differentiation, TNM stage and metastasis of pancreatic cancer patients. Mechanism study indicated that LINC00462 could promoted proliferation, migration invasion of by regulating miR-665 [37]. Notably, LINC00462 was also significantly over-expressed in HCC tissues and knockdown of LINC00462 inhibited the aggressive oncogenic phenotype in HCC cells by regulating PI3K/AKT signaling pathway, suggesting LINC00462 may be a potential and promising therapeutic target for HCC [38]. Therefore, further research on the biological function of these identified lncRNAs may shed light on HCC recurrence.

Although our genomic-clinicopathologic nomogram demonstrated impressive performance in HCC recurrence prediction, limitation of this study should also be noted. First, our nomogram is limited by the retrospective collection of data and fails to include some already recognized RFS-related factors (eg, liver cirrhosis, vascular invasion) and some important molecular factors (eg, TP53 mutation). Further efforts to incorporate more geographic and molecular factors will potentially help to improve the performance of the present model. Second, there is no external or prospective validation for the genomic-clinicopathologic nomogram in the present study, so external and multicenter prospective cohorts with large sample sizes are still needed to validate the clinical application of our model. Final, we do not explore the underlying biological function and pathways of the genomic classifier, so further mechanism study are needed to uncover the related mechanisms.

In conclusion, we develop an lncRNAs-clinicopathologic nomogram and demonstrate that it appear to be a more effective tool for HCC recurrence prediction, compared to TNM stage and other lncRNA-based signature from TCGA. The lncRNAs-clinicopathologic nomogram may help clinicians to make more fitly individualized therapeutic strategies for HCC patients.

Abbreviations

HCC: hepatocellular carcinoma; TCGA: lncRNAs: long non-coding RNAs; OS: overall survival; RFS: relapse-free survival; TCGA: the Cancer Genome Atlas; ECGO: eastern cooperative oncology group; FDR:false

discovery rate; ROC: receiver operating characteristic; LASSO: least absolute shrinkage and selection operator; AUC: area under curve; DCA: decision curve analysis.

Declarations

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

GLL,GN, QZ, and XJT conceived and designed the study. QZ, GN, HYJ, and JSP drafted the manuscript. GLL, GN, and XJT analyzed and interpreted all the data. GLL, GN, and XJT prepared the figures and tables. ZC, HYJ, JSP, and YLH reviewed and revised the manuscript. All authors have read and approved the manuscript for publication.

Authors' information

1 Department of Pathology, Affiliated Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, 510095, Guangdong Province, PR China

2 Department of Infectious Diseases, the Third Affiliated Hospital of Sun Yat-Sen University, Guangzhou, 510630, Guangdong Province, PR China.

3 Department of Clinical Laboratory, Shunde Hospital, Southern Medical University (the First people's hospital of Shunde), Foshan, 528300, Guangdong Province, PR China.

Acknowledgements

Not applicable.

References

Loading [MathJax]/jax/output/CommonHTML/jax.js

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, and A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 68 (2018) 394-424.
- [2] J.M. Llovet, A.M. Di Bisceglie, J. Bruix, B.S. Kramer, R. Lencioni, A.X. Zhu, M. Sherman, M. Schwartz, M. Lotze, J. Talwalkar, and G.J. Gores, Design and endpoints of clinical trials in hepatocellular carcinoma. *Journal of the National Cancer Institute* 100 (2008) 698-711.
- [3] P. Tabrizian, G. Jibara, B. Shrager, M. Schwartz, and S. Roayaie, Recurrence of hepatocellular cancer after resection: patterns, treatments, and prognosis. *Annals of surgery* 261 (2015) 947-55.
- [4] B.R. Nelson, C.A. Makarewich, D.M. Anderson, B.R. Winders, C.D. Troupes, F. Wu, A.L. Reese, J.R. McAnally, X. Chen, E.T. Kavalali, S.C. Cannon, S.R. Houser, R. Bassel-Duby, and E.N. Olson, A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* 351 (2016) 271-5.
- [5] D.M. Anderson, K.M. Anderson, C.L. Chang, C.A. Makarewich, B.R. Nelson, J.R. McAnally, P. Kasaragod, J.M. Shelton, J. Liou, R. Bassel-Duby, and E.N. Olson, A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* 160 (2015) 595-606.
- [6] A. Fatica, and I. Bozzoni, Long non-coding RNAs: new players in cell differentiation and development. *Nature reviews. Genetics* 15 (2014) 7-21.
- [7] M.G. Fernandez-Barrena, M.J. Perugorria, and J.M. Banales, Novel lncRNA T-UCR as a potential downstream driver of the Wnt/beta-catenin pathway in hepatobiliary carcinogenesis. *Gut* 66 (2017) 1177-1178.
- [8] W. Li, Z. Zhang, X. Liu, X. Cheng, Y. Zhang, X. Han, S. Liu, J. Yang, B. Xu, L. He, L. Sun, J. Liang, and Y. Shang, The FOXN3-NEAT1-SIN3A repressor complex promotes progression of hormonally responsive breast cancer. *The Journal of clinical investigation* 127 (2017) 3421-3440.
- [9] J.H. Xu, W.H. Chang, H.W. Fu, W.Q. Shu, T. Yuan, and P. Chen, Upregulated long non-coding RNA LOC90784 promotes cell proliferation and invasion and is associated with poor clinical features in HCC. *Biochemical and biophysical research communications* 490 (2017) 920-926.
- [10] X. Zhao, Y. Liu, and S. Yu, Long noncoding RNA AWPPH promotes hepatocellular carcinoma progression through YBX1 and serves as a prognostic biomarker. *Biochimica et biophysica acta. Molecular basis of disease* 1863 (2017) 1805-1816.
- [11] R. Tang, J.C. Wu, L.M. Zheng, Z.R. Li, K.L. Zhou, Z.S. Zhang, D.F. Xu, and C. Chen, Long noncoding RNA RUSC1-AS-N indicates poor prognosis and increases cell viability in hepatocellular carcinoma. *European review for medical and pharmacological sciences* 22 (2018) 388-396.

- [12] Y. Wu, P.S. Wang, B.G. Wang, L. Xu, W.X. Fang, X.F. Che, X.J. Qu, Y.P. Liu, and Z. Li, Genomewide identification of a novel six-LncRNA signature to improve prognosis prediction in resectable hepatocellular carcinoma. *Cancer medicine* 7 (2018) 6219-6233.
- [13] J. Yan, C. Zhou, K. Guo, Q. Li, and Z. Wang, A novel seven-lncRNA signature for prognosis prediction in hepatocellular carcinoma. *Journal of cellular biochemistry* 120 (2019) 213-223.
- [14] J. Gao, P.W. Kwan, and D. Shi, Sparse kernel learning with LASSO and Bayesian inference algorithm. *Neural networks : the official journal of the International Neural Network Society* 23 (2010) 257-64.
- [15] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, and D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal* 13 (2015) 8-17.
- [16] J.X. Gu, X. Zhang, R.C. Miao, X.H. Xiang, Y.N. Fu, J.Y. Zhang, C. Liu, and K. Qu, Six-long non-coding RNA signature predicts recurrence-free survival in hepatocellular carcinoma. *World journal of gastroenterology* 25 (2019) 220-232.
- [17] A.J. Vickers, A.M. Cronin, E.B. Elkin, and M. Gonen, Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC medical informatics and decision making* 8 (2008) 53.
- [18] A.J. Vickers, B. Van Calster, and E.W. Steyerberg, Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 352 (2016) i6.
- [19] W. Huang da, B.T. Sherman, and R.A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* 37 (2009) 1-13.
- [20] N. Berndt, J. Eckstein, N. Heucke, R. Gajowski, M. Stockmann, D. Meierhofer, and H.G. Holzhutter, Characterization of Lipid and Lipid Droplet Metabolism in Human HCC. *Cells* 8 (2019).
- [21] J. Gu, X. Zhang, R. Miao, X. Ma, X. Xiang, Y. Fu, C. Liu, W. Niu, and K. Qu, 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 (2018) 1627-1639.
- [22] J.H. Choi, M.J. Kim, Y.K. Park, J.Y. Im, S.M. Kwon, H.C. Kim, H.G. Woo, and H.J. Wang, Mutations acquired by hepatocellular carcinoma recurrence give rise to an aggressive phenotype. *Oncotarget* 8 (2017) 22903-22916.
- [23] A. Notarpaolo, R. Layese, P. Magistri, M. Gambato, M. Colledan, G. Magini, L. Miglioresi, A. Vitale, G. Vennarecci, C.D. Ambrosio, P. Burra, F. Di Benedetto, S. Faggioli, M. Colasanti, G. Maria Ettore, A. Andreoli, U. Cillo, A. Laurent, S. Katsahian, E. Audureau, F. Roudot-Thoraval, and C. Duvoux, Validation of the AFP model as a predictor of HCC recurrence in patients with viral hepatitis-related cirrhosis who had received a Loading [MathJax]/jax/output/CommonHTML/jax.js *Journal of Hepatology* 66 (2017) 552-559.

- [24] R.H. Yuan, Y.M. Jeng, R.H. Hu, P.L. Lai, P.H. Lee, C.C. Cheng, and H.C. Hsu, Role of p53 and beta-catenin mutations in conjunction with CK19 expression on early tumor recurrence and prognosis of hepatocellular carcinoma. *Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract* 15 (2011) 321-9.
- [25] S.Y. Lee, I.T. Konstantinidis, A.A. Eaton, M. Gonen, T.P. Kingham, M.I. D'Angelica, P.J. Allen, Y. Fong, R.P. DeMatteo, and W.R. Jarnagin, Predicting recurrence patterns after resection of hepatocellular cancer. *HPB : the official journal of the International Hepato Pancreato Biliary Association* 16 (2014) 943-53.
- [26] K.W. Ma, W.H. She, T.T. Cheung, A.C.Y. Chan, W.C. Dai, J.Y.Y. Fung, C.M. Lo, and K.S.H. Chok, Validated nomogram for the prediction of disease-free survival after hepatectomy for hepatocellular carcinoma within the Milan criteria: individualizing a surveillance strategy. *Surgery today* (2019).
- [27] J.H. Shim, M.J. Jun, S. Han, Y.J. Lee, S.G. Lee, K.M. Kim, Y.S. Lim, and H.C. Lee, Prognostic nomograms for prediction of recurrence and survival after curative liver resection for hepatocellular carcinoma. *Annals of surgery* 261 (2015) 939-46.
- [28] S.X. Wu, J. Huang, Z.W. Liu, H.G. Chen, P. Guo, Q.Q. Cai, J.J. Zheng, H.D. Qin, Z.S. Zheng, X. Chen, R.Y. Zhang, S.L. Chen, and T.X. Lin, A Genomic-clinicopathologic Nomogram for the Preoperative Prediction of Lymph Node Metastasis in Bladder Cancer. *EBioMedicine* 31 (2018) 54-65.
- [29] Y. Liu, J. Zhu, X. Ma, S. Han, D. Xiao, Y. Jia, and Y. Wang, ceRNA network construction and comparison of gastric cancer with or without *Helicobacter pylori* infection. *Journal of cellular physiology* 234 (2019) 7128-7140.
- [30] C. Yue, Y. Ren, H. Ge, C. Liang, Y. Xu, G. Li, and J. Wu, Comprehensive analysis of potential prognostic genes for the construction of a competing endogenous RNA regulatory network in hepatocellular carcinoma. *OncoTargets and therapy* 12 (2019) 561-576.
- [31] H.P. Xian, Z.L. Zhuo, Y.J. Sun, B. Liang, and X.T. Zhao, Circulating long non-coding RNAs HULC and ZNF1-AS1 are potential biomarkers in patients with gastric cancer. *Oncology letters* 16 (2018) 4689-4698.
- [32] C. Xu, Y. Shao, T. Xia, Y. Yang, J. Dai, L. Luo, X. Zhang, W. Sun, H. Song, B. Xiao, and J. Guo, lncRNA-AC130710 targeting by miR-129-5p is upregulated in gastric cancer and associates with poor prognosis. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* 35 (2014) 9701-6.
- [33] L. Feng, Y. Zhu, Y. Zhang, and M. Rao, lncRNA GACAT3 promotes gastric cancer progression by negatively regulating miR-497 expression. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* 97 (2018) 136-142.

- [34] B. Pan, M. Zhao, and L. Xu, Long noncoding RNA gastric cancer-associated transcript 3 plays oncogenic roles in glioma through sponging miR-3127-5p. *Journal of cellular physiology* 234 (2019) 8825-8833.
- [35] H. Zhong, J. Yang, B. Zhang, X. Wang, L. Pei, L. Zhang, Z. Lin, Y. Wang, and C. Wang, LncRNA GACAT3 predicts poor prognosis and promotes cell proliferation in breast cancer through regulation of miR-497/CCND2. *Cancer biomarkers : section A of Disease markers* 22 (2018) 787-797.
- [36] W. Zhou, L. Wang, Y. Miao, and R. Xing, Novel long noncoding RNA GACAT3 promotes colorectal cancer cell proliferation, invasion, and migration through miR-149. *OncoTargets and therapy* 11 (2018) 1543-1552.
- [37] B. Zhou, W. Guo, C. Sun, B. Zhang, and F. Zheng, Linc00462 promotes pancreatic cancer invasiveness through the miR-665/TGFBR1-TGFBR2/SMAD2/3 pathway. *Cell death & disease* 9 (2018) 706.
- [38] J. Gong, X. Qi, Y. Zhang, Y. Yu, X. Lin, H. Li, and Y. Hu, Long noncoding RNA linc00462 promotes hepatocellular carcinoma progression. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* 93 (2017) 40-47.

Tables

Table 1. Characteristics of patient in the training set and validation set

Variable	Category	Training set		Validation set	
		(n=147)	%	(n=146)	%
Age(years)	Median	60		62	
	Range(Years)	18-82		25-90	
Sex	NA	0	0	0	0
	Male	101	68.7	103	70.5
Race	NA	0	0	0	0
	Asian	69	46.9	63	43.2
	White	72	49.0	70	47.9
	Black or African American	4	2.7	8	5.5
Family history	NA	2	1.4	5	3.4
	Yes	44	29.9	45	30.8
HBV	NA	22	15.0	18	12.3
	Yes	49	33.3	44	30.1
HCV	NA	7	4.8	7	4.8
	Yes	19	12.9	23	15.8
Alcohol consumption	NA	7	4.8	7	4.8
	Yes	46	31.3	55	37.7
Mutation count	NA	7	4.8	7	4.8
	Median	79.5		83	
Mutation genome altered	Range	12-1323		1-685	
	NA	7	4.8	2	1.4
	Median	0.25		0.24	
BMI	Range	0-0.85	0	0-0.93	
	NA	3	2.0	2	1.4
	≤24	60	40.8		
AFP(ng/mL)	≤24	74	50.3		
	NA	13	8.8		
	≤20	59	40.1	71	48.6
Platelet(× 10 ⁹ /L)	≤20	59	40.1	42	28.8
	NA	29	19.7	33	22.6
	≤200	57	38.8	56	38.4
Albumin (g/dL)	≤200	74	50.3	65	44.5
	NA	16	10.9	25	17.1
	≤4.0	62	42.2	60	41.1
Creatinine (mg/dL)	≤4.0	63	42.9	61	41.8
	NA	22	15.0	25	17.1
	≤1.1	100	68	91	62.3
Liver cirrhosis	≥1.1	26	17.7	27	18.5
	NA	21	14.3	28	19.2
	Yes	51	34.7	67	45.9
Histological grade	NA	64	43.5	53	36.3
	G1-G2	83	56.5	97	66.4
	G3-G4	62	42.2	47	32.2
T stage	NA	2	1.4	2	1.4
	T1-T2	105	71.4	115	78.8
	T3-T4	41	27.9	30	20.5
TNM stage	NA	1	0.7	1	0.7
	StageI-II	101	68.7	105	71.9
	StageIII-IV	39	26.5	30	20.5
ECOG	NA	7	4.8	11	7.5
	0	79	53.7	68	46.6
	≥0	44	29.9	54	37.0
	NA	24	16.3	24	16.4

Table2.Multivariable Cox regression analysis of LncRNA with recurrence-free survival in the Training set.

LncRNA	Coef	exp(coef)	Lower.95	Upper.95	P value
AC012625.1	0.05174	1.0531	0.8952	1.2389	0.5325
AC068481.1	0.10870	1.1148	0.8413	1.4773	0.4492
AC092675.1	0.05747	1.0592	0.8497	1.3202	0.6092
AC093772.1	0.05770	1.0594	0.8869	1.2654	0.5245
AC109779.1	-0.32522	0.7224	0.5408	0.9649	0.0277*
AC118653.1	0.06050	1.0624	0.7956	1.4187	0.6818
AC246785.3	0.05021	1.0515	0.8937	1.2371	0.5449
AL121985.1	0.03112	1.0316	0.9067	1.1737	0.6364
AL512604.2	0.12532	1.1335	0.9463	1.3577	0.1736
AP002478.1	0.06333	1.0654	0.9115	1.2452	0.4262
ARHGEF7.AS1	0.10057	1.1058	0.9219	1.3264	0.2786
GACAT3	0.26230	1.2999	1.0439	1.6187	0.0191*
LINC00462	0.25478	1.2902	1.0717	1.5531	0.0071**
LINC01700	0.21032	1.2341	0.9076	1.6779	0.1797
LINC02429	0.12948	1.1382	0.8675	1.4935	0.3503

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table3. Univariable and multivariable Cox regression analysis for prediction of RFS.

Factors	Subgroup	Univariable analysis		Multivariable analysis	
		HR(95%CI)	P	HR(95%CI)	P
Age		1.00 (0.99-1.02)	0.673	NA	NA
Sex	Female	1			
	Male	0.89(0.57-1.39)	0.621	NA	NA
Race	Asian	1			
	White	1.15(0.75-1.78)	0.522	NA	NA
	Black or African American	0.96(0.34-2.70)	0.932	NA	NA
Family history	No	1			
	Yes	0.68(0.42-1.09)	0.109	NA	NA
HBV	No	1			
	Yes	0.80 (0.51-1.28)	0.359	NA	NA
HCV	No	1			
	Yes	1.64 (0.98-2.64)	0.058	NA	NA
Alcohol consumption	No	1			
	Yes	1.04(0.67-1.62)	0.863	NA	NA
Mutation count		1.02(1.01-1.03)	0.007*	1.02(1.00-1.03)	0.032*
Genome altered		2.86(0.97-8.42)	0.057	NA	NA
BMI	≤24	1			
	>24	0.57(0.37-0.89)	0.014*	0.76(0.47-1.04)	0.084
AFP	≤20	1		1	
	>20	1.79(1.14-2.82)	0.012*	1.82(1.14-2.90)	0.012*
Platelet	≤200	1			
	>200	1.26(0.79-2.00)	0.328	NA	NA
Albumin	≤4.0	1			
	>4.0	0.96(0.60-1.52)	0.855	NA	NA
Creatinine	≤1.1	1			
	≥1.1	0.81(0.47-1.40)	0.445	NA	NA
Liver cirrhosis	No	1			
	Yes	2.17(1.14-4.11)	0.018*	1.56(0.87-2.12)	0.122

Histological grade	G1-G2	1			
	G3-G4	1.07(0.68-1.67)	0.772	NA	NA
T stage	T1-T2	1			
	T3-T4	3.12 (1.98-4.92)	0.000*	2.50 (1.54-4.05)	0.000*
TNM stage	StageI-II	1			
	StageIII-IV	2.71(1.71-4.30)	0.000*	1.12 (0.47-2.69)	0.801
ECOG	0	1			
	≥0	2.41(1.54-3.77)	0.000*	2.00(1.23-3.23)	0.005*
cRNA classifier		1.13(1.09-1.18)	0.000*	1.13(1.09-1.18)	0.000*

Abbreviations: HR=hazard ratio, CI =confidence intervals,

NOTE: NA, not available. These variables were eliminated in the multivariate logistic regression model, so the HR and *P* values were not available. **P* < 0.05

Table 4. Assessing the prognostic performance of the AJCC stage and nomogram in training set and validation set

Port	Model	Homogeneity monotonicity and discriminatory ability			
		Likelihood ratio(LR) test*	Linear trend χ^2 test**	C-index (95% CI) ***	Akaike information criterion(AIC)****
Training set	TNM stage	16.4	19.6	0.64 (0.58-0.69)	663
	Nine lncRNA classifier	25.6	82.7	0.74 (0.68-0.80)	654
	Nomogram	62.7	112.1	0.78 (0.72-0.83)	625
Validation set	TNM stage	5.8	6.6	0.56 (0.51-0.60)	765
	Nine lncRNA classifier	12.2	21.4	0.61 (0.55-0.68)	758
	Nomogram	41.6	59.2	0.71(0.65-0.76)	737

*Higher homogeneity likelihood ratio indicates a smaller difference within the staging system, it means better homogeneity

**Higher discriminatory ability linear trend indicates a higher linear trend between staging system, it means better discriminatory ability and gradient monotonicity

***A higher c-index means better discriminatory ability.

****Smaller AIC values indicate better optimistic prognostic stratification

Figures

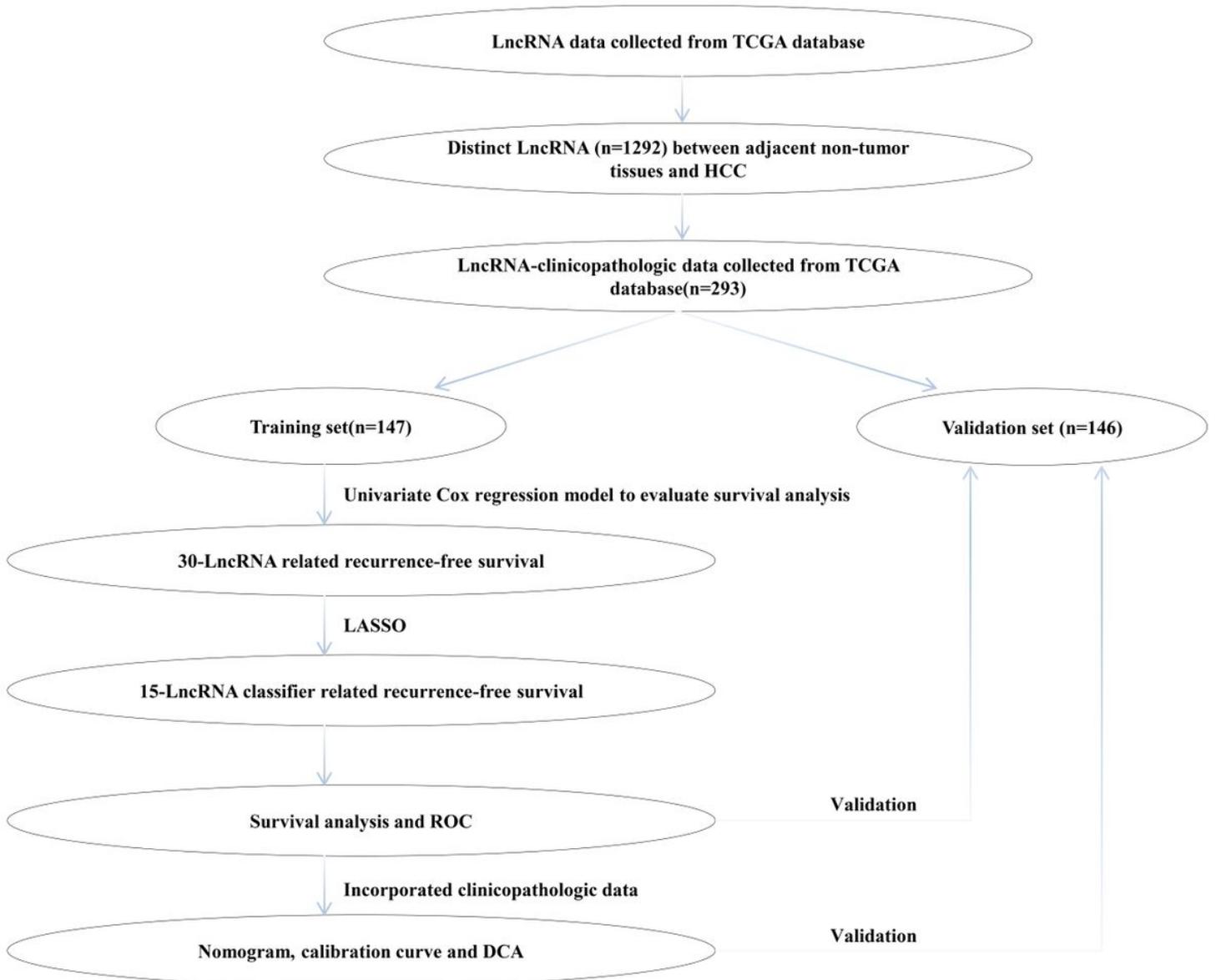


Figure 1

The flowchart of study design. LASSO: least absolute shrinkage and selection operator.

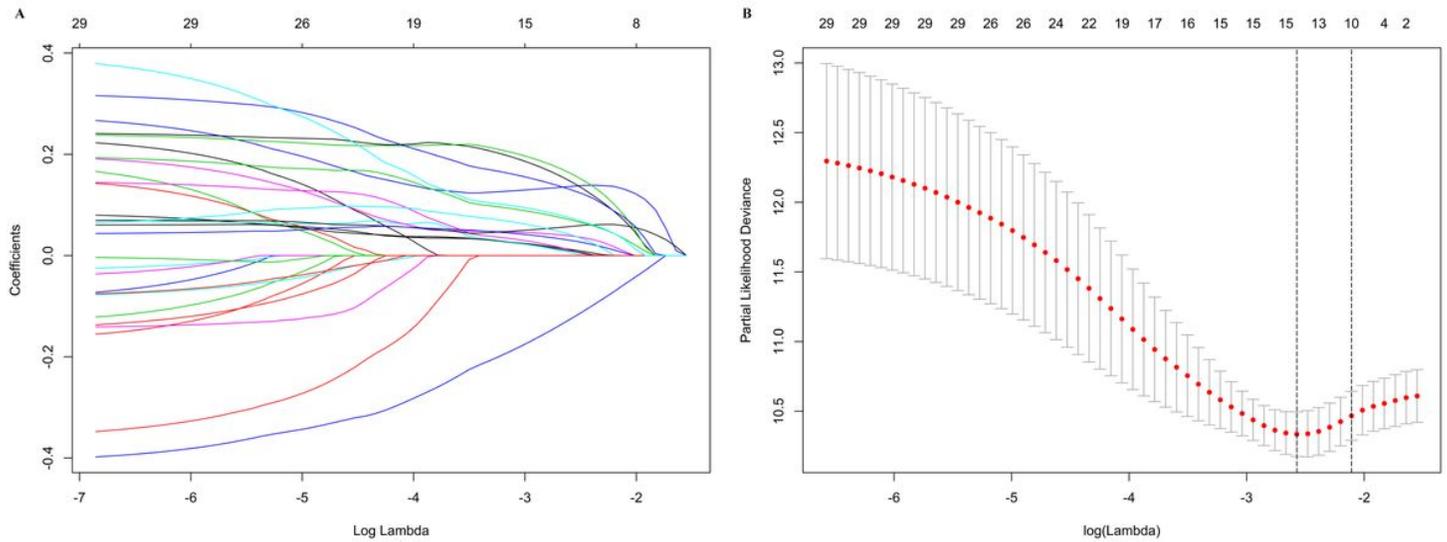


Figure 2

Selection of RFS-associated lncRNAs by LASSO analysis in the training set. A: LASSO coefficient profiles of the 30 lncRNAs. A vertical line is drawn at the optimal value by minimum criteria and results in fifteen non-zero coefficients. Fifteen lncRNAs—AC012625.1, AC068481.1, AC092675.1, AC093772.1, AC109779.1, AC118653.1, AC246785.3, AL121985.1, AL121985.1, AP002478.1, ARHGEF7.AS1, GACAT3, LINC00462, LINC01700 and LINC02429, respectively, were selected in the LASSO Cox regression model. B: fifteen lncRNAs selected by LASSO Cox regression analysis. The two dotted vertical lines are drawn at the optimal values by minimum criteria (left) and 1 - s.e. criteria (right).

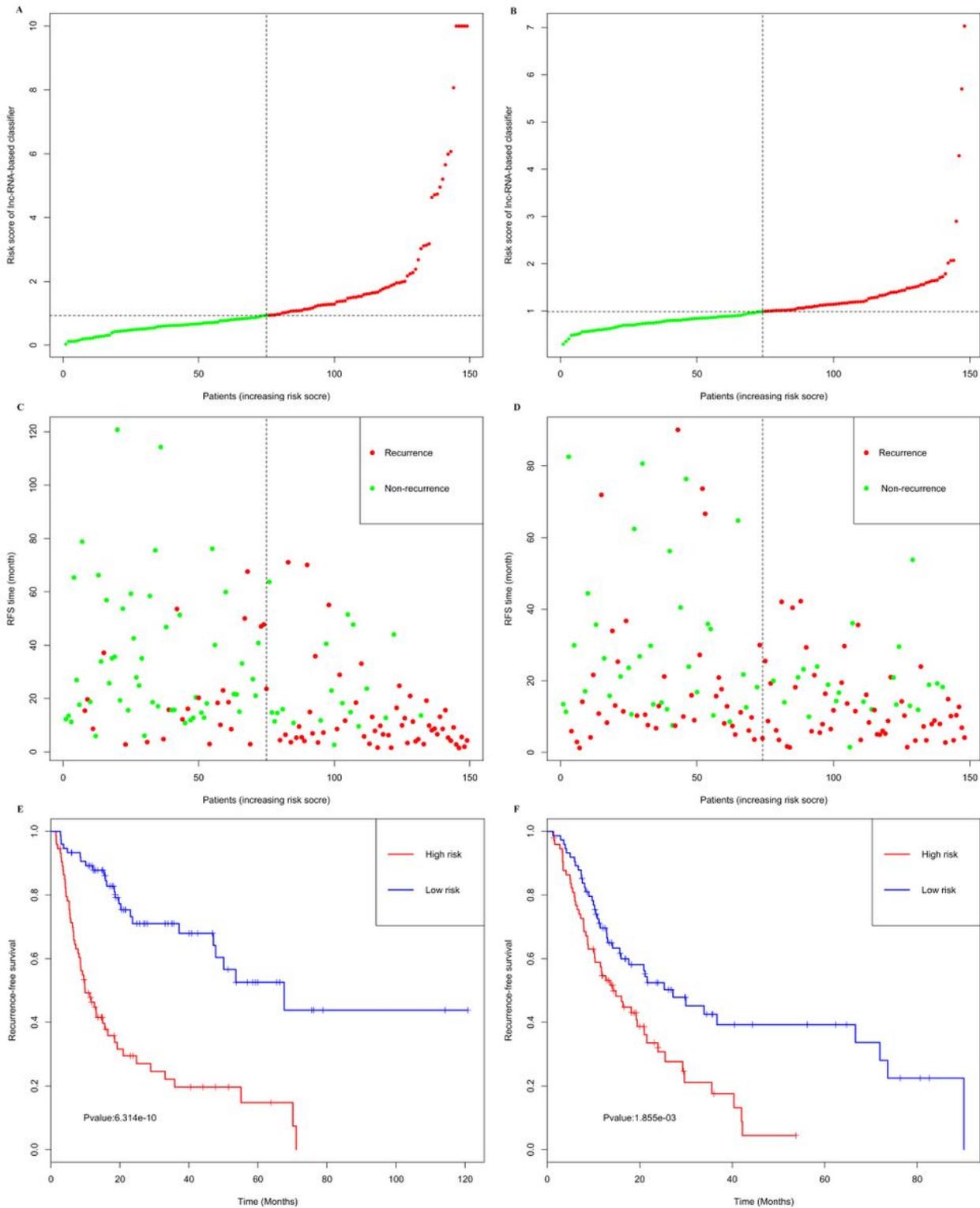


Figure 3

Development and validation of 15-lncRNAs-based classifier for HCC recurrence. A and C: Distribution of lncRNAs-based classifier risk score

\in *thetra* \in \in *gset*. B and D: Distribution of lncRNAs-based classifier risk score in the validation set.

E: Kaplan-Meier analysis of RFS time for the high-risk and low-risk patients of the training set. F: Kaplan-

and low-risk patients of the validation set.

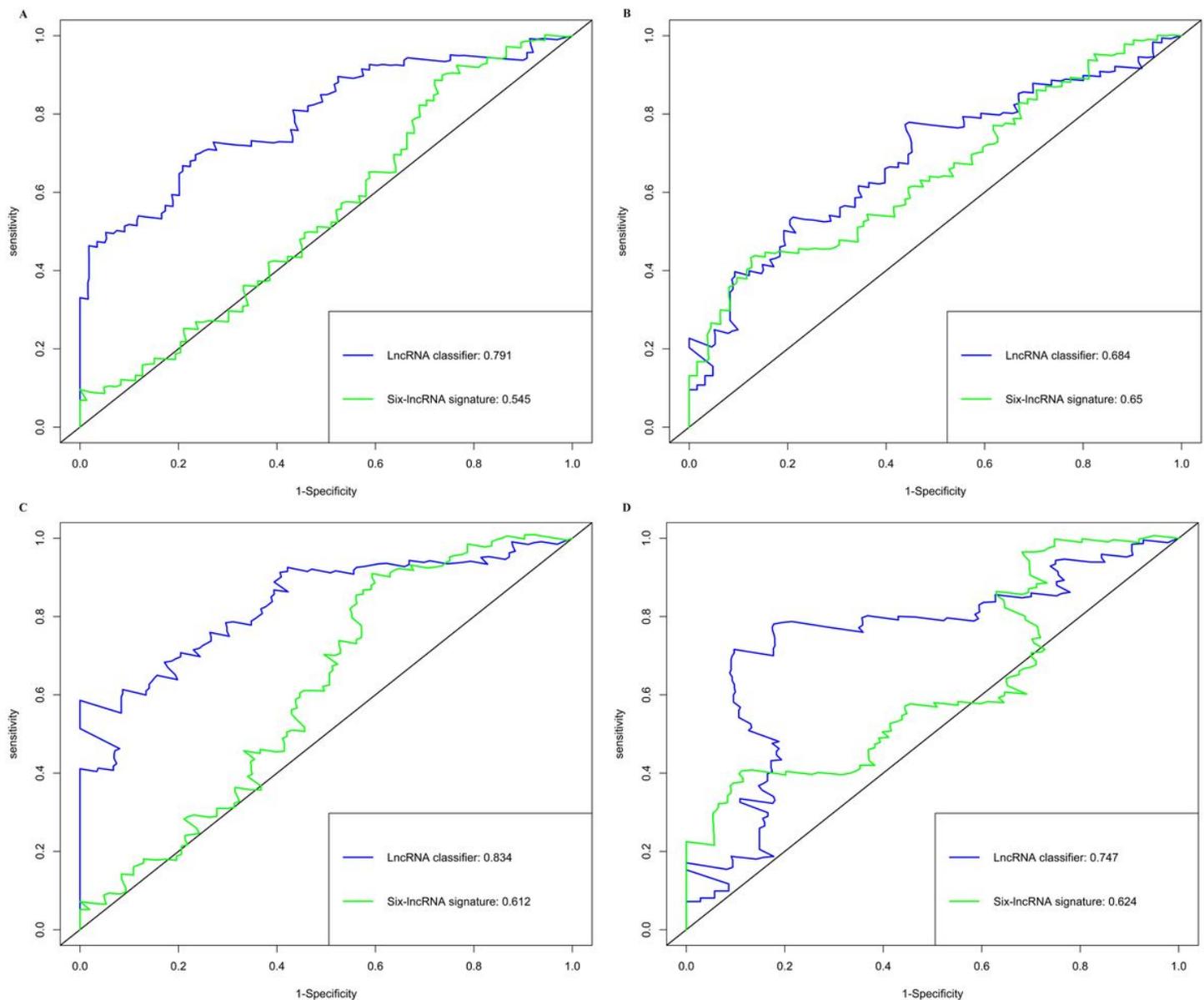


Figure 4

Comparison the predictive value of 15-lncRNAs-based classifier with other lncRNAs -based signature from TCGA. ROC analysis was performed to assess the predictive value of 15-lncRNAs-based classifier and the 6-lncRNAs-based signature. A: The AUC of 15-lncRNAs-based classifier in predicting 2-year RFS time; B: The AUC of 6-lncRNAs-based signature in predicting 2-year RFS time; C: The AUC of 15-lncRNAs-based classifier in predicting 5-year RFS time; D: The AUC of 6-lncRNAs-based signature in predicting 2-year RFS time.

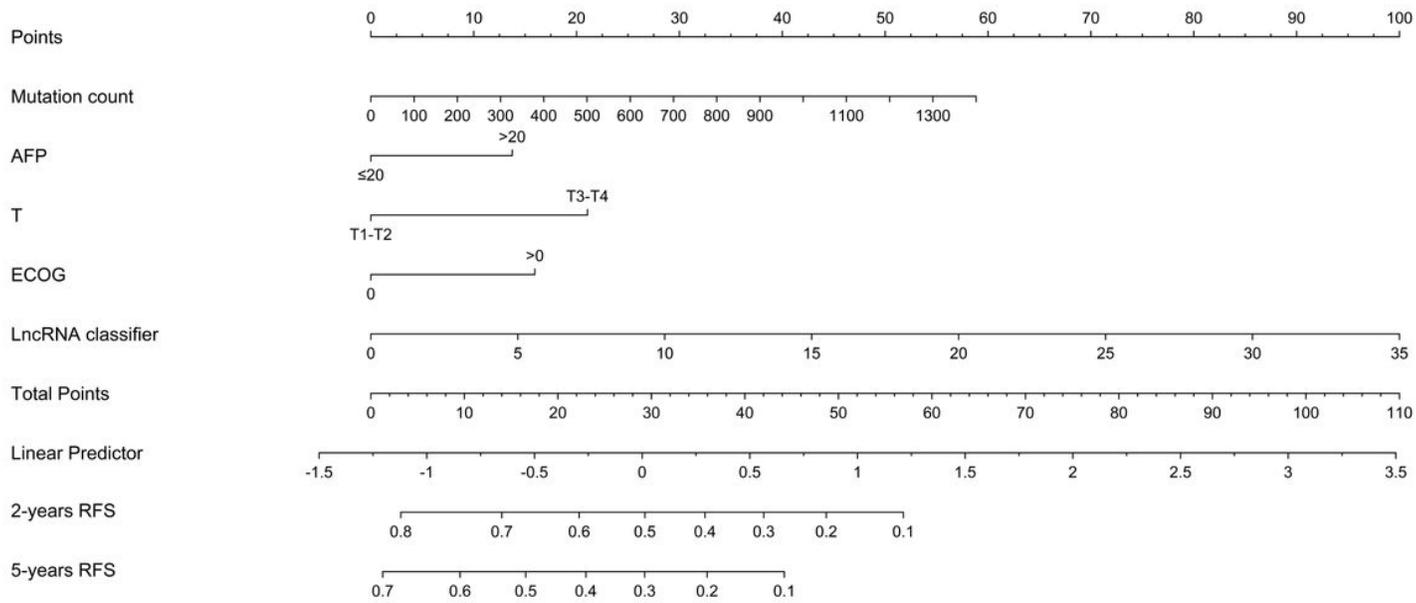
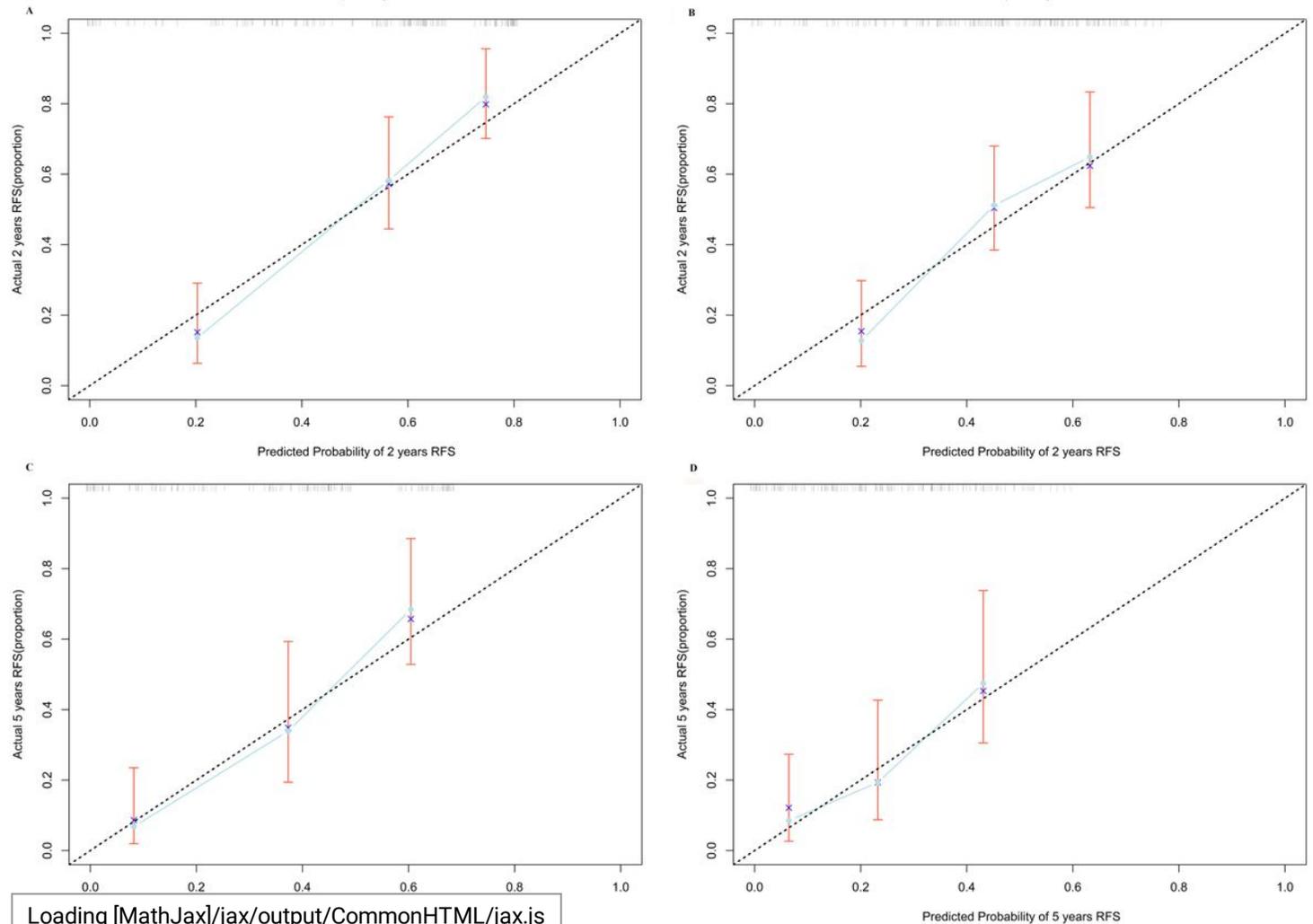


Figure 5

Development of 15-lncRNAs-based classifier-clinicopathologic nomogram for predicting 2-year and 5-year RFS.



Loading [MathJax]/jax/output/CommonHTML/jax.js

Figure 6

Calibration curve for 15-lncRNAs-based classifier-clinicopathologic nomogram in predicting 2-year (A) and 5-year (C) RFS time in the training set and the 2-year (B) and 5-year (D) RFS time in the validation set.

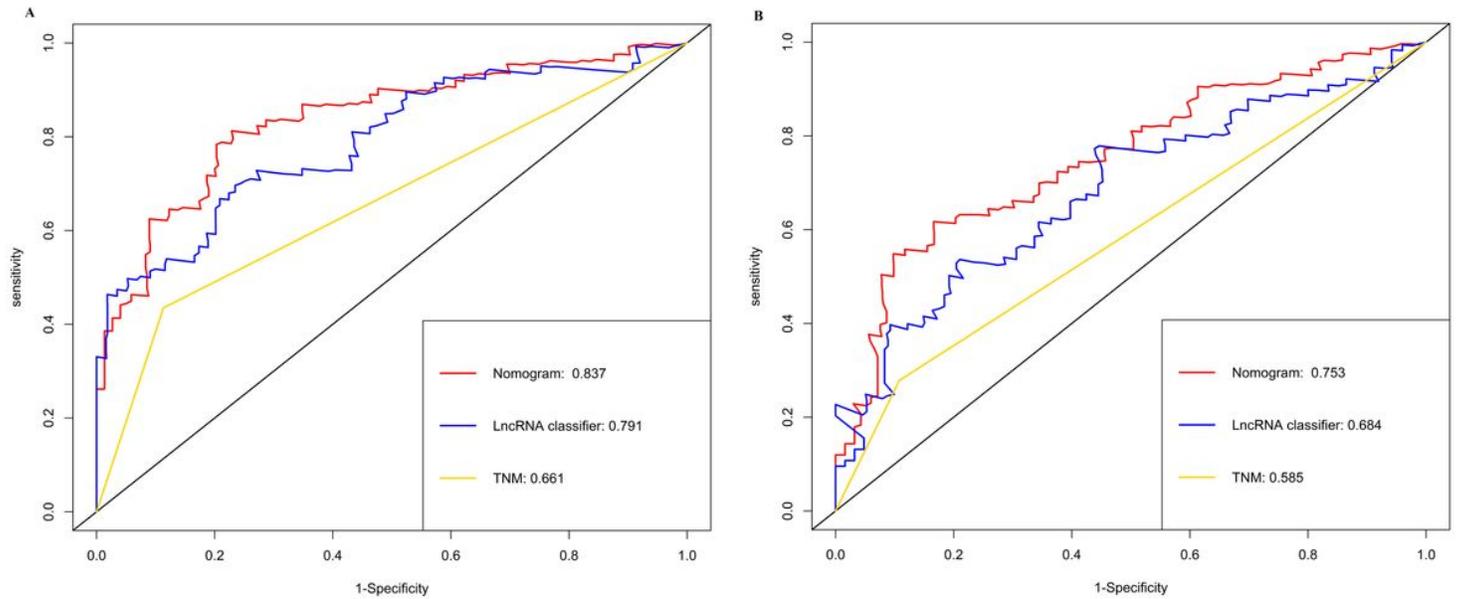


Figure 7

Comparison the predictive value of 15-lncRNAs-based classifier-clinicopathologic nomogram with TNM stage and 15-lncRNAs-based classifier. ROC analysis was used to evaluate the predictive accuracy of the three models in predicting RFS time of the training set (A) and the validation set (B).

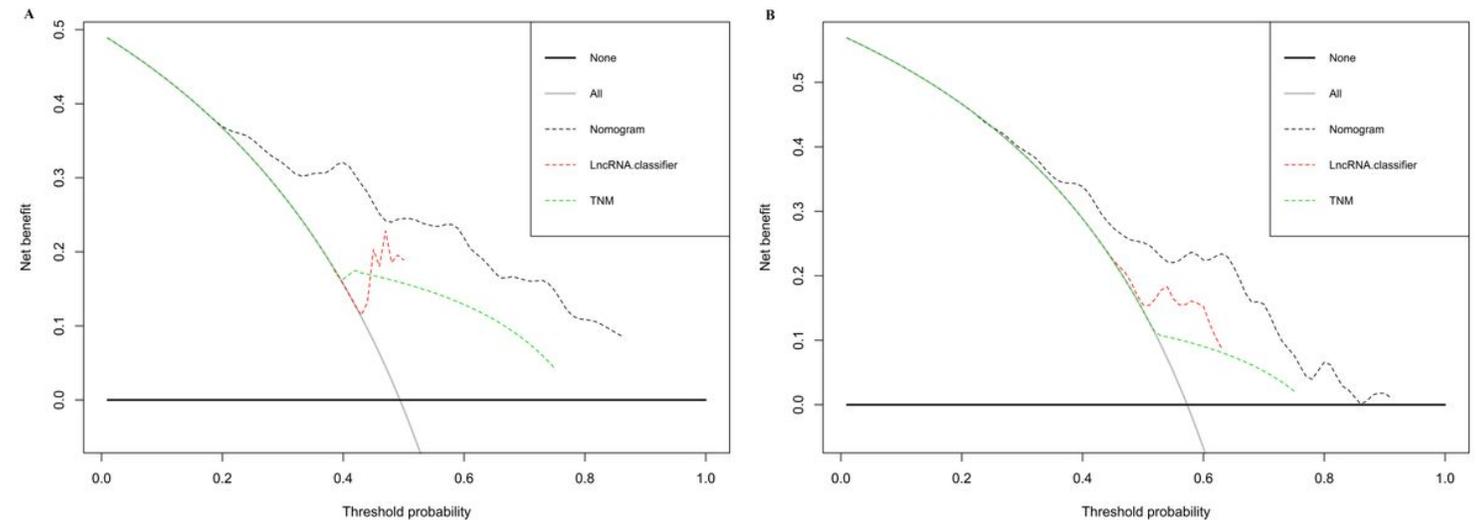


Figure 8

Decision curve analysis (DCA) analysis for 15-lncRNAs-based classifier-clinicopathologic nomogram with TNM stage and 15-lncRNAs-based classifier in the training set (A) and the validation set (B).

Supplementary Files

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

- [FigureS1.tif](#)
- [FigureS3.tif](#)
- [FigureS6.tif](#)
- [FigureS4.tif](#)
- [FigureS5.tif](#)
- [FigureS2.tif](#)