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.

1 **Title page**

3 **A genomic-clinical nomogram predicting recurrence-free survival for patients diagnosed**

4 **with hepatocellular carcinoma**

6 Junjie Kong[1], Tao Wang[1], Shu Shen[1], Zifei Zhang[1], Xianwei Yang[1], Wentao Wang[1]

8 [1]Department of Liver Surgery & Liver Transplantation Center, West China Hospital of Sichuan

9 University, Chengdu, P. R. China

11 **Corresponding Author:**

12 Wentao Wang

13 No.37 Guoxue Xiang, Wuhou District, Chengdu, 610041, Sichuan Province, China.

14 E-mail: wwt0510@163.com

16 **Abstract**

17 Liver resection surgery is the most commonly used treatment strategy for patients diagnosed

18 with hepatocellular carcinoma (HCC). However, there is still a chance for recurrence in these

19 patients despite the survival benefits of this procedure. This study aimed to explore recurrence-

20 related genes (RRGs) and establish a genomic-clinical nomogram for predicting postoperative

21 recurrence in HCC patients. 123 differently expressed genes (DEGs) and 3 RRGs (PZP, SPP2,

22   and PRC1) were identified from online databases via Cox regression and LASSO logistic

23   regression analyses and a gene-based risk model containing RRGs was then established. The

24   Harrell's concordance index (C-index), receiver operating characteristic (ROC) curves and

25   calibration curves showed that the model performed well. Finally, a genomic-clinical nomogram

26   incorporating the gene-based risk model, AJCC staging system, and Eastern Cooperative

27   Oncology Group performance status (ECOG-PS) was constructed to predict the 1-, 2-, and 3-

28   year recurrence-free survival rates (RFS) for HCC patients. The C-index, ROC analysis, and

29   decision curve analysis (DCA) were good indicators of the nomogram's performance. In

30   conclusion, we identified 3 reliable RRGs associated with the recurrence of cancer and

31   constructed a nomogram that performed well in predicting RFS for HCC patients. These findings

32   could enrich our understanding of the mechanisms for HCC recurrence, help surgeons predict

33   patients' prognosis, and promote HCC treatment.

35   **Introduction**

36   Hepatocellular carcinoma (HCC) is the most common primary liver malignancy and a

37   contributor to the third population of cancer-related deaths, ranking sixth among the most

38   frequent malignancies worldwide (Forner et al., 2018). Liver resection is the most commonly

39   used therapeutic strategy for HCC, and accounts for an overall 5-year survival rate of

40   approximately 70% (Orcutt & Anaya, 2018; Akoad & Pomfret, 2015). However, for these

41   patients, survival is impacted by postoperative tumor recurrence (Tabrizian et al., 2015).

42   Previous studies have suggested that more than half of HCC patients would suffer from a

43   disease relapse after a hepatectomy (Akoad & Pomfret, 2015). Disease recurrence was divided

44   into two types depending on their different etiologies: those who relapsed within 2 years after

45   surgery (early recurrence) and those whose relapse occurred more than 2 years after the

46   operation (late recurrence) (Sherman, 2008). Many clinical characteristics, such as the AFP level,

47   tumor size, vascular invasion (including microscopic and macroscopic), and HBsAg level proved

48   to be risk factors for recurrence and several predictive models were established to predict the

49   postoperative recurrence of HCC (Lee et al., 2014a; Qiu et al., 2017; He et al., 2018; Chan et al.,

50   2018a). However, these studies and models only focused on the relationships between clinical

51   traits and HCC recurrence with little focus on the crucial role of molecular data.

52   Advances in biomedical research and high-throughput technologies have greatly

53   contributed to the identification of molecular biomarkers related to cancer development,

54   recurrence, and prognosis in the past few decades (Teufel, 2015). A large number of mRNAs,

55   microRNAs and other kinds of biomarkers have been identified and regarded as potential targets

56   for cancer diagnosis and treatment (Lee et al., 2014b). Biomarkers could reflect the molecular

57   mechanisms of tumor recurrence and, as a result, the identification of reliable biomarkers could

58   improve the accuracy in predicting cancer recurrence, thereby contributing to cancer treatment

59   (Long et al., 2018; Gu et al., 2018). Consequently, gene signatures for recurrence and prognosis

60   prediction are receiving more attention.

61   In this study we explored differently expressed genes (DEGs) between the HCC and non-

62   tumor samples using data obtained from the Gene Expression Omnibus (GEO) and the Cancer

63   Genome Atlas (TCGA) databases. 3 recurrence-related genes (RRGs) were then identified and a

64 gene-based risk model was established from the data of the TCGA dataset. Furthermore, a

65 genomic-clinical nomogram containing the gene-based risk model and clinical characteristics

66 was constructed to predict the 1-, 2-, and 3-year recurrence-free survival (RFS) for HCC patients.

67 We also assessed the performance of the nomogram using the Harrell's concordance index (C-

68 index), calibration curves, receiver operating characteristic (ROC) curves, and decision curves

69 analysis (DCA).

70

71 **Materials & Methods**

72 **Dataset collection**

73   Using "hepatocellular carcinoma" as our keyword, we searched gene expression profiles of

74 HCC from the GEO database (https://www.ncbi.nlm.nih.gov/gds/). "Homo sapiens" was the

75 term chosen for the organism parameter. The following criteria were used to screen datasets: 1)

76 datasets were mRNA expression profiled by array; 2) datasets with more than 100 samples; 3)

77 datasets with comparison between tumor and non-tumor samples; and 4) profiles with available

78 expression information. Finally, to obtain more accurate results and to avoid individual

79 heterogeneity, the top 4 GEO datasets with the largest sample numbers, GSE25097 (Sung et al.,

80 2012), GSE76427 (Grinchuk et al., 2018), GSE36376 (Lim et al., 2013) and GSE14520

81 (Roessler et al., 2010), were selected. The details of the datasets were shown in Table S1.

82 Meanwhile, we downloaded the RNA-seq of HCC form the TCGA database

83 (https://cancergenome.nih.gov/), which contained 374 HCC samples and 50 non-tumor samples.

84 The related clinical files of HCC were also obtained for further analysis. Log2 transformation

85   was used for all of the expression data.

Exploration of differently expressed genes (DEGs) and bioinformatic analysis

87   R software (Version 3.5.3) and the related packages "limma" and "edgR" were used to

88   screen DEGs. The cut-off criteria with a significance of $p < 0.05$ and $|log2FC| > 1$ were used in the

89   identification of DEGs. Overlapping analysis was performed to find DEGs among all of the 4

90   GEO datasets and the TCGA dataset and the Venn plot was completed using the online tool

91   entitled "Calculate and draw custom Venn diagrams"

92   (http://bioinformatics.psb.ugent.be/webtools/Venn/). In addition, the online biological tools

93   Database for Annotation, Visualization and Integrated Discovery (DAVID,

94   https://david.ncifcrf.gov/) and KOBAS 3.0 (http://kobas.cbi.pku.edu.cn/anno_iden.php) were

95   used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG)

96   pathways enrichment analysis, respectively. Furthermore, the search tool for the Retrieval of

97   Interacting Genes (STRING, https://string-db.org/cgi/input.pl) and Cytoscape software (Version

98   3.6.1) were used to establish a protein-protein interaction (PPI) network complex to visualize the

99   correlations among DEGs.

Identification of recurrence-related genes (RRGs)

101   In this progress, 3 datasets (TCGA, GSE14520 and GSE76427 datasets) with detailed

102   information about tumor recurrence were divided into 2 types of sets: the discovery set (TCGA

103   dataset) and the validation set (GSE14520 and GSE76427 datasets). In the discovery set,

104   univariate, LASSO and multivariate Cox regression analysis were employed to screen RRGs

105   from DEGs. Using the "survival" R package, univariate analysis was used to make a primary

106 selection of DEGs and those with p<0.05 were regarded as potential candidates associated with

107 the recurrence of HCC. Afterwards, we employed the LASSO logistic regression model using

108 the "glmnet" R package to further select genes from the potential candidates. In the

109 LASSO-penalized regression, the optimal lambda was determined using 10-fold cross-validation

110 and the L1 penalty was used to shrink regression coefficients towards zero. Many variables were

111 excluded based on the principle that a larger penalty led to fewer predictive factors (Goeman,

112 2010). Consequently, candidate RRGs were considered as DEGs with nonzero coefficients. After

113 subsampling the dataset with 1000 iterations, the seed DEGs were shrunk into various sets and

114 those containing DEGs with nonzero coefficients were defined as models with potential

115 prognostic ability. DEGs with at least 900 occurrence frequencies were considered as candidate

116 RRGs. Finally, stepwise multivariate regression analysis was conducted to identify RRGs and to

117 construct a gene-based risk model which could predict HCC recurrence after liver resection.

118 Using X-tile software (Camp et al., 2004), the optimal cut-off points of the risk scores and the

119 expression levels of RRGs were found. The Kaplan-Meier (KM) analysis was used to evaluate

120 the performance of RRGs and the risk model. Furthermore, ROC analysis and calibration curves

121 were used in validating the performance of the risk model and the "pROC" and "rms" R

122 packages were used, respectively. In addition, the same methods were used to assess the

123 performance of the RRGs and the risk model in the 2 validation datasets. In these statistic

124 processes, a two-sided P<0.05 was considered statistically significant. Finally, using "pheatmap"

125 and "forestplot" R packages, the results were visualized by heatmap plots and forest plots,

126 respectively.

127 **Development and validation of nomogram**

128   Univariate and multivariate regression analyses using the RNA-seq and clinical traits of the

129 TCGA dataset combined with the gene-based risk model were used to find clinical and molecular

130 factors independently associated with HCC recurrence. Based on the results of multivariate

131 analysis, a genomic-clinical nomogram was established using the "rms" R package to predict 1-,

132 2-, and 3-year RFS for HCC patients (Wang et al., 2013). The performance of the nomogram was

133 evaluated using calibration curves. The calibration was conducted with 1000 bootstrap samples

134 to reduce the bias. The discrimination of the nomogram was assessed using C-index and area

135 under the curve (AUC), which measured the predictive ability of the nomogram for patients with

136 different outcomes. The larger the C-index and a bigger AUC indicated a higher accuracy of

137 prognostic predication (Huitzil-Melendez et al., 2010). Finally, using an "rmda" R package,

138 DCA was conducted to assess the clinical usefulness of the nomogram, which could estimate the

139 net benefits for different prediction models across all possible risk thresholds (Vickers et al.,

140 2008).

141

142 **Results**

143 **Exploration of DEGs**

144 Figure 1 reveals the workflow of this study. First, 4 GEO datasets, GSE25097, GSE36376,

145 GSE14520, and GSE 76427 containing 1812 samples (1037 HCC samples and 775 non-tumor

146 samples) were obtained to screen DEGs and a total of 1454, 413, 1088 and 398 genes were found

147 dysregulated in the tumor tissues, respectively (Fig. S1A-E). Meanwhile, 9314 dysregulated

148   genes were identified in the TCGA dataset (Fig. S1F). After an overlapping analysis, 123 genes

149   containing 12 up-regulated genes and 111 down-regulated genes were regarded as DEGs

150   aberrantly expressed in HCC samples compared to non-tumor samples and were selected for

151   further analysis (Fig. 2A, B).

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

153   We performed GO and KEGG analysis to elucidate the functional characteristics of the DEGs.

154   The GO analysis indicated that the DEGs were significantly enriched in the extracellular

155   exosome, extracellular region, extracellular space, oxidation-reduction process, cell adhesion,

156   and serine-type endopeptidase activity. In the KEGG pathways analysis, we could find that the

157   DEGs were mainly enriched in metabolism, metabolic pathways, metabolism of lipids and

158   lipoproteins, biological oxidations, metabolism of amino acids and derivatives, and the immune

159   system (Fig. 2C, D). Finally, a PPI network complex, containing 121 nodes and 773 edges, was

160   constructed to elucidate the relationships among the DEGs (Fig. S2).

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

162   After a rigorous screening, 252 cases from the TCGA dataset were selected for the

163   identification of RRGs from the discovery set. Using univariate analysis, we found that 79 DEGs

164   were significantly related to HCC recurrence with a p-value <0.05. LASSO logistic regression

165   analysis was used to find candidates associated with recurrence and the optimal tuning

166   parameters related to the minimum generalization error were determined from 10-fold cross-

167   validation (Fig. 3). Consequently, 8 DEGs (PZP, C1RL, PRC1, PTTG1, UBE2C, AFM, SPP2,

168   HGFAC) were screened. Finally, we performed stepwise multivariate regression analysis and

169 found that 3 genes, PZP, SPP2, and PRC1, were biomarkers independently associated with HCC

170 recurrence and were included in the construction of a gene-based risk model. Using X-tile

171 software, the optimal cut-off values of the 3 RRGs' expression levels were found in the

172 discovery and validation sets, and the HCC patients were divided into 2 groups (low-risk group

173 and high-risk group). The KM analysis showed that the 3 RRGs were significantly related to

174 HCC recurrence among all of the 3 datasets (Fig. 4).

175  According to the relative coefficient in the regression model, a risk score could be

176 calculated to evaluate the risk of recurrence for each patient based on the expression levels of

177 RRGs: risk score = (-0.0793 * PZP expression level) + (0.2295 * PRC1 expression level) + (-

178 0.0662 * SPP2 expression level). We then used X-tile software to select the optimal cut-off value

179 of the risk score for recurrence, and the patients were divided into a low-risk group or a high-risk

180 group in the discovery and validation sets, respectively (Fig. S3). The RFS in the low-risk group

181 was significantly longer than that of the high-risk group among all of the 3 datasets (Fig. 5A-C).

182 The C-index of the risk model in the TCGA, GSE14520, and GSE76427 datasets was 0.663

183 (95% CI, 0.615-0.711), 0.571 (95% CI, 0.519-0.623) and 0.599 (95% CI, 0.492-0.706),

184 respectively. The ROC curves also suggested that the risk model had a good predictive efficiency

185 among all of the 3 datasets (Fig. 5D-F). Finally, the calibration curves suggested that the risk

186 model had a good predictive efficiency compared to the observed outcomes for 1-, 2- and 3-year

187 postoperative recurrence (Fig. 5G-I).

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

189  The RNA-seq and clinical characteristics of the TCGA dataset were used for exploring

190   genomic and clinical factors related to recurrence. After removing patients who were diagnosed

191   with fibrolamellar carcinoma and combined hepatocellular-cholangiocarcinoma, and who did not

192   have intact clinical information (including age, height, weight, race, hepatitis B virus status,

193   alcohol consumption, tumor grade, AJCC stage, APF level, platelet level, albumin level, Eastern

194   Cooperative Oncology Group performance status (ECOG-PS), and sex), a total of 162 HCC

195   patients with both intact genomic and clinical information were included for further analysis.

196   Univariate and multivariate cox regression analyses were used to find factors associated with

197   HCC recurrence. As shown in Figure **6**, gene-based risk model, race, AJCC staging system and

198   ECOG-PS were found to be potential factors related to HCC recurrence in univariate analysis.

199   Furthermore, the multivariate analysis indicated that the gene-based risk model, AJCC staging

200   system and ECOG-PS were factors associated with HCC recurrence (Fig. **6**).

201   **Construction and validation of genomic-clinical nomogram**

202   Based on the results of the multivariate analysis, a genomic-clinical nomogram including

203   the gene-based risk model, AJCC staging system, and ECOG-PS was established to predict the

204   1-, 2-, and 3-year RFS for HCC patients (Fig. **7**A). The calibration curves of the nomogram for

205   predicting the 1-, 2-, and 3-year probability of recurrence indicated that it performed well (Fig.

206   **7**B). The C-index of the nomogram was 0.678 (95% CI, 0.618-0.738).

207   The DCA results showed that the nomogram was more clinically useful than using the gene-

208   based risk model, AJCC staging system, or ECOG-PS alone (Fig. **7**C-E). Afterwards, we

209   performed ROC analysis to assess the discrimination ability of the genomic-clinical nomogram

210   and the results suggested that the AUC of the nomogram in predicting the 1-, 2-, and 3-year

211   probability of recurrence was greater than that of using the gene-based risk model or AJCC

212   staging system alone (Fig. **7**F-H). Finally, according to the total points of each patient in the

213   nomogram, the patients were divided into 2 groups (low-risk group and high-risk group) using

214   X-tile software, and KM analysis demonstrated that the discrimination ability of the nomogram

215   was satisfactory (p<0.0001) (Fig. 8).

216

217   **Discussion**

218   As the sixth most frequently occurring malignancy and the third leading cause of cancer-

219   related deaths, HCC remains a significant medical problem in the world (Forner et al., 2018).

220   Although advances in therapeutic strategies such as liver resection, transplantation,

221   radiofrequency ablation, and transcatheter arterial chemoembolization have improved the

222   chances of survival for certain patients, the prognosis of these patients remains unsatisfactory

223   (Zamora-Valdes et al., 2017; Bailey & Sydnor, 2019; Tanwar et al., 2009; Sun et al., 2019).

224   Recurrence is a significant challenge for HCC patients after liver resection and usually leading to

225   a poor prognosis (Chan et al., 2018b). Therefore, it is vitally important to find reliable

226   biomarkers in the recurrent progression of the disease and to explore the mechanisms and risk

227   factors closely associated with HCC recurrence.

228   In this study we identified 123 DEGs in HCC samples compared to non-tumor samples

229   using data obtained from GEO and TCGA databases. Univariate, LASSO, and multivariate

230   regression analysis were used to find DEGs related to recurrence. Consequently, 3 RRGs were

231   identified. A gene-based risk model was then constructed using the relative coefficients in the

232   multivariate model and the expression levels of the 3 RRGs. The risk model had a good

233   predictive efficiency for HCC recurrence. Furthermore, based on the clinical information

234   obtained from the TCGA database, a genomic-clinical nomogram, including the gene-based risk

235   model, AJCC staging system, and ECOG-PS was established, the C-index of the nomogram was

236   0.678, the calibration curves fitted well and the ROC analysis and DCA also showed the

237   nomogram performed well in discerning those patients with a high probability of recurrence.

238   Three RRGs, PZP, SPP2, and PRC1, were identified as factors associated with HCC

239   recurrence. PRC1, also known as protein regulator of cytokinesis, is a microtubule binding

240   protein that plays crucial roles in mitosis (Subramanian et al., 2010). Due to its essential function

241   in the cell cycle, dysregulation of PRC1 could be regarded as an important factor for

242   carcinogenesis. The development of various cancers, such as breast cancer (Shimo et al., 2007),

243   lung adenocarcinoma (Zhan et al., 2017) and pancreatic cancer (Nakamura et al., 2004), was

244   associated with the overexpression of PRC1. Meanwhile, the dysregulation of PRC1 was

245   demonstrated to be related to the early recurrence and chemoresistance of HCC (Chen et al.,

246   2016; Wang et al., 2017). PRC1 could contribute to the early recurrence of HCC via the

247   regulation of the Wnt/β-catenin signaling pathway, which has been proven to be a crucial

248   pathway in hepatocarcinogenesis and a potential therapeutic target (Pez et al., 2013).

249   Furthermore, Liu et al. suggested that knockdown PRC1 using siRNA could block cytokines and

250   limit the proliferation of HCC, which could bring new insights into HCC treatment (Liu et al.,

251   2018). Although previous studies have demonstrated that the other two RRGs, PZP and SPP2,

252   were related to the prognosis of HCC, there was no study discussing their potential functions in

253   HCC recurrence (Yang et al., 2018; Zheng et al., 2018). Robust experimental results also

254   suggested that SPP2 and PZP were associated with the stability of spliceosome and chromatin

255   (Klein et al., 2016; Warkocki et al., 2015), and their dysregulation might affect the progression

256   of cell proliferation and apoptosis, which could lead to carcinogenesis and disease progression.

257   In addition, in validation of RRGs' performance, the overexpression of PZP was found to be

258   negatively associated with recurrence in the TCGA dataset, while positively related to recurrence

259   in the GSE14520 and GSE76427 datasets. Further exploration with larger HCC and non-tumor

260   samples should be performed to validate this issue. In summary, the mechanisms of the 3 RRGs

261   in HCC development and recurrence have not been clear until now and they should be more

262   closely followed as these could be prospective therapeutic targets for identifying the potential for

263   HCC recurrence.

264   The genomic-clinical nomogram contained 2 clinical traits, the AJCC staging system and

265   the ECOG-PS, aside from the gene-based risk model. The AJCC staging system is a

266   conventional staging system used for evaluating the characteristics of cancers, selecting optimal

267   therapeutic strategies, and predicting prognosis (Chun et al., 2018). In this study, patients with

268   AJCC stage II and III had a higher risk for HCC recurrence. According to the 8th AJCC staging

269   system the tumors classified into stage II and III were those with vascular invasion, bigger tumor

270   size, multiple tumor numbers, or the invasion into adjacent organs compared with the stage-I

271   tumors,. Tumors with these characteristics were more likely to have a greater invasive ability and

272   a later tumor stage and, as a result, these patients were easier to determine as having HCC

273   recurrence and were given higher points in the nomogram. ECOG-PS was used to assess the

274   patients' functional capability of self-care (Sorensen et al., 1993). Previous studies demonstrated

275   that ECOG-PS was closely related to the survival of cancer patients. In the BCLC staging

276   system, ECOG-PS was listed as an independent factor for selecting therapeutic strategies and

277   predicting prognosis (Forner et al., 2018). Recently, a Chinese study discovered that patients

278   with higher ECOG-PS were more likely to have poor molecular phenotypes in colorectal cancer,

279   which could lead to a higher probability of cancer recurrence (Chen et al., 2016). It is clear that

280   more focus should be placed on the functions of ECOG-PS in the recurrence of HCC, which

281   could provide new insights into cancer treatment.

282   Functional enrichment analysis showed that the DEGs were mainly enriched in extracellular

283   parts, such as the extracellular exosome, extracellular region, and extracellular space.

284   Furthermore, metabolism-related pathways and immune-related pathways were significantly

285   enriched by DEGs. It is well known that the liver is associated with various metabolic processes,

286   which could help to maintain the microenvironmental stability of various tissues (Rui, 2014). As

287   a result, the aberrant metabolism could provide for the unique needs of the tumor cells, contain

288   macromolecular biosynthesis, increase energy production, and maintain the redox balance, which

289   could provide a selective advantage for the proliferation, growth and survival of tumor cells (De

290   Matteis et al., 2018). Recently, the advances of metabolomics have brought new insights into the

291   mechanisms of HCC development, recurrence and prognosis, and the related findings could be

292   therapeutic targets for HCC treatment (Shang et al., 2016). One of the most widely-known

293   special mechanisms of carcinogenesis is immune invasion (Aerts et al., 2016). However, until

294   the discovery of immune checkpoints, including programmed cell death protein 1 (PD-1) and

295   cytotoxic T-lymphocyte protein 4 (CTLA-4), there had not been any breakthroughs in

296   immunotherapy of HCC. The immune checkpoint inhibitors have been proven to possess

297   significant clinical functions and are promising strategies for the systemic treatment of HCC

298   (Inarrairaegui et al., 2018). The metabolism-related pathways and immune-related pathways

299   played crucial roles in the recurrence of HCC and could certainly be potential therapeutic targets

300   in the future.

301   Several other studies also reported the establishment of models to predict RFS for HCC

302   after liver resection (Cui et al., 2017; Gu et al., 2019). For instance, using the gene expression

303   file GSE76427, Gu et al. (Gu et al., 2019) discovered that six long non-coding RNA (MSC-AS1,

304   POLR2J4, EIF3J-AS1, SERHL, RMST, and PVT1) were associated with a poor prognosis of

305   HCC and constructed a nomogram including the six-lncRNA signature, TNM stage and ECOG

306   to predict RFS for patients with small HCC. The C-index of their nomogram was 0.684 (95%CI:

307   0.635-0.733), which was comparable with our nomogram (C-index 0.678; 95% CI, 0.618-0.738).

308   Furthermore, besides the gene-based signature, TNM stage and ECOG were also integrated into

309   their nomogram, which was robust evidence for our nomogram (including the AJCC staging

310   system and ECOG). However, Gu's study only analyzed one gene expression file with a

311   relatively small number of samples, which could make the results inaccurate to some extent. Our

312   study analyzed a total of 1645 samples from four GEO datasets, containing 870 HCC samples

313   and 775 normal samples. Consequently, the RRGs found in our study were more reliable.

314   Combined with rigorous validations, it was clear that our nomogram was could performed well

315   in the prediction of RFS for HCC patients after liver resection.

316   However, limitations existed in the current study. First of all, the expression of RRGs were

317 only explored and validated among online databases; no experiments were conducted to confirm

318 them. Consequently, experiments about the dysregulation of the RRGs and their mechanisms in

319 HCC recurrence should be explored in the future. Secondly, we failed to make a comparison of

320 the genomic-clinical nomogram to other widely accepted staging systems, such as the BCLC

321 staging system. Lastly, we did not perform an external validation for the genomic-clinical

322 nomogram, and further validations with additional data should be performed to validate the

323 performance of our nomogram.

324

325 **Conclusions**

326   Three new biomarkers were found that were related to postoperative recurrence for HCC

327 patients after hepatectomy and a reliable genomic-clinical nomogram was established to predict

328 the 1-, 2-, and 3-year RFS. The C-index, ROC analysis, and DCA showed that the nomogram

329 performed well. These findings could bring new insights into molecular mechanisms for HCC

330 recurrence, contribute to HCC treatment, and help surgeons to predict the likelihood of

331 recurrence and the prognosis for HCC patients postoperatively.

332

333 **Acknowledgments**

334 None.

335

336 **References:**

337   Aerts, M., Benteyn, D., Van Vlierberghe, H., Thielemans, K., and Reynaert, H. 2016. Current

338   status and perspectives of immune-based therapies for hepatocellular carcinoma. *World J*

339   *Gastroenterol* 22:253-261. 10.3748/wjg.v22.i1.253

340   Akoad, M.E., and Pomfret, E.A. 2015. Surgical resection and liver transplantation for

341   hepatocellular carcinoma. *Clinics in Liver Disease* 19:381-399. 10.1016/j.cld.2015.01.007

342   Bailey, C.W., and Sydnor, M.J. 2019. Current State of Tumor Ablation Therapies. *Dig Dis Sci*.

343   10.1007/s10620-019-05514-9

344   Camp, R.L., Dolled-Filhart, M., and Rimm, D.L. 2004. X-tile: a new bio-informatics tool for

345   biomarker assessment and outcome-based cut-point optimization. *Clinical Cancer Research*

346   10:7252-7259. 10.1158/1078-0432.CCR-04-0713

347   Chan, A., Zhong, J., Berhane, S., Toyoda, H., Cucchetti, A., Shi, K., Tada, T., Chong, C., Xiang,

348   B.D., Li, L.Q., Lai, P., Mazzaferro, V., Garcia-Finana, M., Kudo, M., Kumada, T., Roayaie, S.,

349   and Johnson, P.J. 2018a. Development of pre and post-operative models to predict early recurrence

350   of hepatocellular carcinoma after surgical resection. *Journal of Hepatology* 69:1284-1293.

351   10.1016/j.jhep.2018.08.027

352   Chan, A., Zhong, J., Berhane, S., Toyoda, H., Cucchetti, A., Shi, K., Tada, T., Chong, C., Xiang,

353   B.D., Li, L.Q., Lai, P., Mazzaferro, V., Garcia-Finana, M., Kudo, M., Kumada, T., Roayaie, S.,

354   and Johnson, P.J. 2018b. Development of pre and post-operative models to predict early recurrence

355   of hepatocellular carcinoma after surgical resection. *Journal of Hepatology* 69:1284-1293.

356   10.1016/j.jhep.2018.08.027

357   Chen, D., Wen, X., Song, Y.S., Rhee, Y.Y., Lee, T.H., Cho, N.Y., Han, S.W., Kim, T.Y., and

358   Kang, G.H. 2016. Associations and prognostic implications of Eastern Cooperative Oncology

359   Group performance status and tumoral LINE-1 methylation status in stage III colon cancer

360   patients. *Clinical Epigenetics* 8:36. 10.1186/s13148-016-0203-8

361   Chen, J., Rajasekaran, M., Xia, H., Zhang, X., Kong, S.N., Sekar, K., Seshachalam, V.P.,

362   Deivasigamani, A., Goh, B.K.P., Ooi, L.L., Hong, W., and Hui, K.M. 2016. The microtubule-

363   associated protein PRC1 promotes early recurrence of hepatocellular carcinoma in association with

364   the Wnt/ β -catenin signalling pathway. *Gut* 65:1522-1534. 10.1136/gutjnl-2015-310625

365   Chun, Y.S., Pawlik, T.M., and Vauthey, J.N. 2018. 8th Edition of the AJCC Cancer Staging

366   Manual: Pancreas and Hepatobiliary Cancers. *Annals of Surgical Oncology* 25:845-847.

367   10.1245/s10434-017-6025-x

368   Cui, H., Zhang, Y., Zhang, Q., Chen, W., Zhao, H., and Liang, J. 2017. A comprehensive genome-

369   wide analysis of long noncoding RNA expression profile in   hepatocellular carcinoma. *Cancer*

370   *Med* 6:2932-2941. 10.1002/cam4.1180

371   De Matteis, S., Ragusa, A., Marisi, G., De Domenico, S., Casadei Gardini, A., Bonafè, M., and

372   Giudetti, A.M. 2018. Aberrant Metabolism in Hepatocellular Carcinoma Provides Diagnostic and

373   Therapeutic Opportunities. *Oxidative Medicine and Cellular Longevity* 2018:1-13.

374   10.1155/2018/7512159

375   Forner, A., Reig, M., and Bruix, J. 2018. Hepatocellular carcinoma. *Lancet* 391:1301-1314.

376   10.1016/S0140-6736(18)30010-2

377   Goeman, J.J. 2010. L1 penalized estimation in the Cox proportional hazards model. *Biom J* 52:70-

378   84. 10.1002/bimj.200900028

379   Grinchuk, O.V., Yenamandra, S.P., Iyer, R., Singh, M., Lee, H.K., Lim, K.H., Chow, P.K., and

380   Kuznetsov, V.A. 2018. Tumor-adjacent tissue co-expression profile analysis reveals pro-

381   oncogenic ribosomal gene signature for prognosis of resectable hepatocellular carcinoma.

382   *Molecular Oncology* 12:89-113. 10.1002/1878-0261.12153

383   Gu, J., Zhang, X., Miao, R., Ma, X., Xiang, X., Fu, Y., Liu, C., Niu, W., and Qu, K. 2018. A three-

384   long non-coding RNA-expression-based risk score system can better predict both overall and

385   recurrence-free survival in patients with small hepatocellular carcinoma. *Aging* 10:1627-1639.

386   10.18632/aging.101497

387   Gu, J.X., Zhang, X., Miao, R.C., Xiang, X.H., Fu, Y.N., Zhang, J.Y., Liu, C., and Qu, K. 2019.

388   Six-long non-coding RNA signature predicts recurrence-free survival in hepatocellular carcinoma.

389   *World J Gastroenterol* 25:220-232. 10.3748/wjg.v25.i2.220

390   He, W., Peng, B., Tang, Y., Yang, J., Zheng, Y., Qiu, J., Zou, R., Shen, J., Li, B., and Yuan, Y.

391   2018. Nomogram to Predict Survival of Patients With Recurrence of Hepatocellular Carcinoma

392   After Surgery. *Clin Gastroenterol Hepatol* 16:756-764. 10.1016/j.cgh.2017.12.002

393   Huitzil-Melendez, F.D., Capanu, M., O'Reilly, E.M., Duffy, A., Gansukh, B., Saltz, L.L., and

394   Abou-Alfa, G.K. 2010. Advanced hepatocellular carcinoma: which staging systems best predict

395   prognosis? *Journal of Clinical Oncology* 28:2889-2895. 10.1200/JCO.2009.25.9895

396   Inarrairaegui, M., Melero, I., and Sangro, B. 2018. Immunotherapy of Hepatocellular Carcinoma:

397   Facts and Hopes. *Clinical Cancer Research* 24:1518-1524. 10.1158/1078-0432.CCR-17-0289

398   Klein, B.J., Muthurajan, U.M., Lalonde, M.E., Gibson, M.D., Andrews, F.H., Hepler, M.,

399   Machida, S., Yan, K., Kurumizaka, H., Poirier, M.G., Cote, J., Luger, K., and Kutateladze, T.G.

400   2016. Bivalent interaction of the PZP domain of BRPF1 with the nucleosome impacts chromatin

401   dynamics and acetylation. *Nucleic Acids Research* 44:472-484. 10.1093/nar/gkv1321

402   Lee, S.C., Tan, H.T., and Chung, M.C. 2014a. Prognostic biomarkers for prediction of recurrence

403   of hepatocellular carcinoma: current status and future prospects. *World J Gastroenterol* 20:3112-

404   3124. 10.3748/wjg.v20.i12.3112

405   Lee, S.C., Tan, H.T., and Chung, M.C. 2014b. Prognostic biomarkers for prediction of recurrence

406   of hepatocellular carcinoma: current status and future prospects. *World J Gastroenterol* 20:3112-

407   3124. 10.3748/wjg.v20.i12.3112

408   Lim, H.Y., Sohn, I., Deng, S., Lee, J., Jung, S.H., Mao, M., Xu, J., Wang, K., Shi, S., Joh, J.W.,

409   Choi, Y.L., and Park, C.K. 2013. Prediction of disease-free survival in hepatocellular carcinoma

410   by gene expression profiling. *Annals of Surgical Oncology* 20:3747-3753. 10.1245/s10434-013-

411   3070-y

412   Liu, X., Li, Y., Meng, L., Liu, X.Y., Peng, A., Chen, Y., Liu, C., Chen, H., Sun, S., Miao, X.,

413   Zhang, Y., Zheng, L., and Huang, K. 2018. Reducing protein regulator of cytokinesis 1 as a

414   prospective therapy for hepatocellular carcinoma. *Cell Death & Disease* 9:534. 10.1038/s41419-

415   018-0555-4

416   Long, J., Zhang, L., Wan, X., Lin, J., Bai, Y., Xu, W., Xiong, J., and Zhao, H. 2018. A four-gene-

417   based prognostic model predicts overall survival in patients with hepatocellular carcinoma.

418   *Journal of Cellular and Molecular Medicine* 22:5928-5938. 10.1111/jcmm.13863

419   Nakamura, T., Furukawa, Y., Nakagawa, H., Tsunoda, T., Ohigashi, H., Murata, K., Ishikawa, O.,

420   Ohgaki, K., Kashimura, N., Miyamoto, M., Hirano, S., Kondo, S., Katoh, H., Nakamura, Y., and

421   Katagiri, T. 2004. Genome-wide cDNA microarray analysis of gene expression profiles in

422   pancreatic cancers using populations of tumor cells and normal ductal epithelial cells selected for

423   purity by laser microdissection. *Oncogene* 23:2385-2400. 10.1038/sj.onc.1207392

424   Orcutt, S.T., and Anaya, D.A. 2018. Liver Resection and Surgical Strategies for Management of

425   Primary Liver Cancer. *Cancer Control* 25:1145179885. 10.1177/1073274817744621

426   Pez, F., Lopez, A., Kim, M., Wands, J.R., Caron, D.F.C., and Merle, P. 2013. Wnt signaling and

427   hepatocarcinogenesis: molecular targets for the development of   innovative anticancer drugs.

428   *Journal of Hepatology* 59:1107-1117. 10.1016/j.jhep.2013.07.001

429   Qiu, J.F., Ye, J.Z., Feng, X.Z., Qi, Y.P., Ma, L., Yuan, W.P., Zhong, J.H., Zhang, Z.M., Xiang,

430   B.D., and Li, L.Q. 2017. Pre- and post-operative HBsAg levels may predict recurrence and survival

431   after curative resection in patients with HBV-associated hepatocellular carcinoma. *Journal of*

432   *Surgical Oncology* 116:140-148. 10.1002/jso.24628

433   Roessler, S., Jia, H.L., Budhu, A., Forgues, M., Ye, Q.H., Lee, J.S., Thorgeirsson, S.S., Sun, Z.,

434   Tang, Z.Y., Qin, L.X., and Wang, X.W. 2010. A unique metastasis gene signature enables

435   prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Research*

436   70:10202-10212. 10.1158/0008-5472.CAN-10-2607

437   Rui, L. 2014. Energy metabolism in the liver. *Comprehensive Physiology* 4:177-197.

438   10.1002/cphy.c130024

439   Shang, R.Z., Qu, S.B., and Wang, D.S. 2016. Reprogramming of glucose metabolism in

440   hepatocellular carcinoma: Progress and prospects. *World J Gastroenterol* 22:9933-9943.

441   10.3748/wjg.v22.i45.9933

442   Sherman, M. 2008. Recurrence of hepatocellular carcinoma. *N Engl J Med* 359:2045-2047.

443   10.1056/NEJMe0807581

444   Shimo, A., Nishidate, T., Ohta, T., Fukuda, M., Nakamura, Y., and Katagiri, T. 2007. Elevated

445   expression of protein regulator of cytokinesis 1, involved in the growth of breast cancer cells.

446   *Cancer Science* 98:174-181. 10.1111/j.1349-7006.2006.00381.x

447   Sorensen, J.B., Klee, M., Palshof, T., and Hansen, H.H. 1993. Performance status assessment in

448   cancer patients. An inter-observer variability study. *Br J Cancer* 67:773-775.

449   Subramanian, R., Wilson-Kubalek, E.M., Arthur, C.P., Bick, M.J., Campbell, E.A., Darst, S.A.,

450   Milligan, R.A., and Kapoor, T.M. 2010. Insights into antiparallel microtubule crosslinking by

451   PRC1, a conserved nonmotor microtubule binding protein. *Cell* 142:433-443.

452   10.1016/j.cell.2010.07.012

453   Sun, J.Y., Yin, T., Zhang, X.Y., and Lu, X.J. 2019. Therapeutic advances for patients with

454   intermediate hepatocellular carcinoma. *Journal of Cellular Physiology*. 10.1002/jcp.28019

455   Sung, W.K., Zheng, H., Li, S., Chen, R., Liu, X., Li, Y., Lee, N.P., Lee, W.H., Ariyaratne, P.N.,

456   Tennakoon, C., Mulawadi, F.H., Wong, K.F., Liu, A.M., Poon, R.T., Fan, S.T., Chan, K.L., Gong,

457   Z., Hu, Y., Lin, Z., Wang, G., Zhang, Q., Barber, T.D., Chou, W.C., Aggarwal, A., Hao, K., Zhou,

458   W., Zhang, C., Hardwick, J., Buser, C., Xu, J., Kan, Z., Dai, H., Mao, M., Reinhard, C., Wang, J.,

459   and Luk, J.M. 2012. Genome-wide survey of recurrent HBV integration in hepatocellular

460   carcinoma. *Nature Genetics* 44:765-769. 10.1038/ng.2295

461   Tabrizian, P., Jibara, G., Shrager, B., Schwartz, M., and Roayaie, S. 2015. Recurrence of

462   Hepatocellular   Cancer   After   Resection.   *Annals   of   Surgery*   261:947-955.

463   10.1097/SLA.0000000000000710

464   Tanwar, S., Khan, S.A., Grover, V.P., Gwilt, C., Smith, B., and Brown, A. 2009. Liver

465   transplantation for hepatocellular carcinoma. *World J Gastroenterol* 15:5511-5516.

466   Teufel, A. 2015. Bioinformatics and database resources in hepatology. *Journal of Hepatology*

467   62:712-719. 10.1016/j.jhep.2014.10.036

468   Vickers, A.J., Cronin, A.M., Elkin, E.B., and Gonen, M. 2008. Extensions to decision curve

469   analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers.

470   *BMC Med Inform Decis Mak* 8:53. 10.1186/1472-6947-8-53

471   Wang, Y., Li, J., Xia, Y., Gong, R., Wang, K., Yan, Z., Wan, X., Liu, G., Wu, D., Shi, L., Lau,

472   W., Wu, M., and Shen, F. 2013. Prognostic nomogram for intrahepatic cholangiocarcinoma after

473   partial hepatectomy. *Journal of Clinical Oncology* 31:1188-1195. 10.1200/JCO.2012.41.5984

474   Wang, Y., Shi, F., Xing, G.H., Xie, P., Zhao, N., Yin, Y.F., Sun, S.Y., He, J., Wang, Y., and Xuan,

475   S.Y. 2017. Protein Regulator of Cytokinesis PRC1 Confers Chemoresistance and Predicts an

476   Unfavorable Postoperative Survival of Hepatocellular Carcinoma Patients. *Journal of Cancer*

477   8:801-808. 10.7150/jca.17640

478   Warkocki, Z., Schneider, C., Mozaffari-Jovin, S., Schmitzova, J., Hobartner, C., Fabrizio, P., and

479   Luhrmann, R. 2015. The G-patch protein Spp2 couples the spliceosome-stimulated ATPase

480   activity of the DEAH-box protein Prp2 to catalytic activation of the spliceosome. *Genes Dev*

481   29:94-107. 10.1101/gad.253070.114

482   Yang, Y., Lu, Q., Shao, X., Mo, B., Nie, X., Liu, W., Chen, X., Tang, Y., Deng, Y., and Yan, J.

483   2018. Development Of A Three-Gene Prognostic Signature For Hepatitis B Virus Associated

484   Hepatocellular Carcinoma Based On Integrated Transcriptomic Analysis. *Journal of Cancer*

485   9:1989-2002. 10.7150/jca.23762

486   Zamora-Valdes, D., Taner, T., and Nagorney, D.M. 2017. Surgical Treatment of Hepatocellular

487   Carcinoma. *Cancer Control* 24:1145164522. 10.1177/1073274817729258

488   Zhan, P., Zhang, B., Xi, G., Wu, Y., Liu, H., Liu, Y., Xu, W., Zhu, Q., Cai, F., Zhou, Z., Miu, Y.,

489   Wang, X., Jin, J., Li, Q., Qian, L., Lv, T., and Song, Y. 2017. PRC1 contributes to tumorigenesis

490   of lung adenocarcinoma in association with the Wnt/ β -catenin signaling pathway. *Molecular*

491   *Cancer* 16. 10.1186/s12943-017-0682-z

492   Zheng, Y., Liu, Y., Zhao, S., Zheng, Z., Shen, C., An, L., and Yuan, Y. 2018. Large-scale analysis

493   reveals a novel risk score to predict overall survival in hepatocellular carcinoma. *Cancer*

494   *Management and Research* 10:6079-6096. 10.2147/CMAR.S181396

495

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.
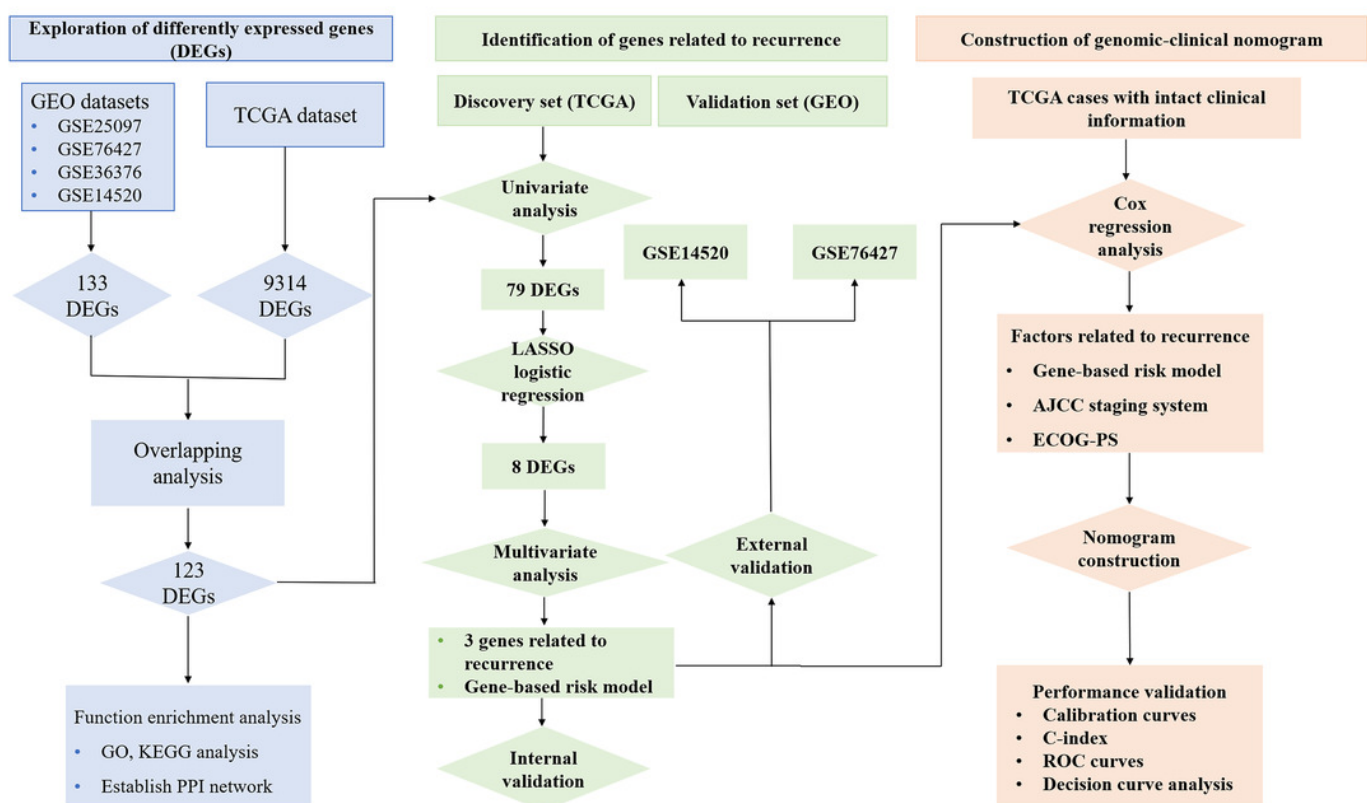
Figure 2

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.

A

B

C

D
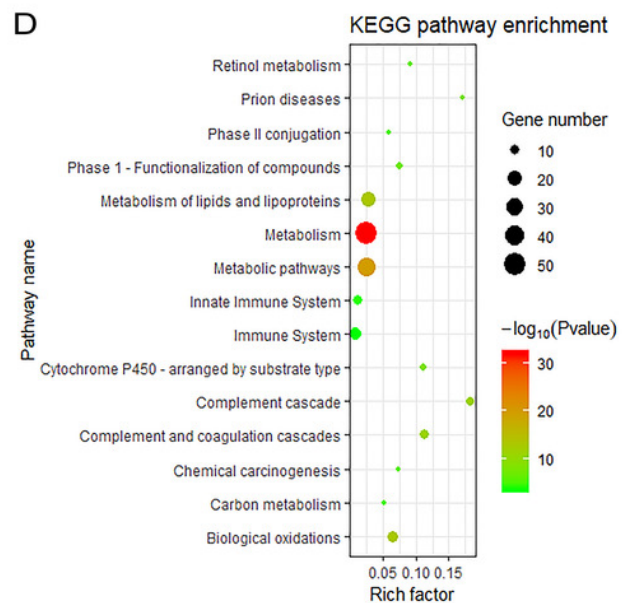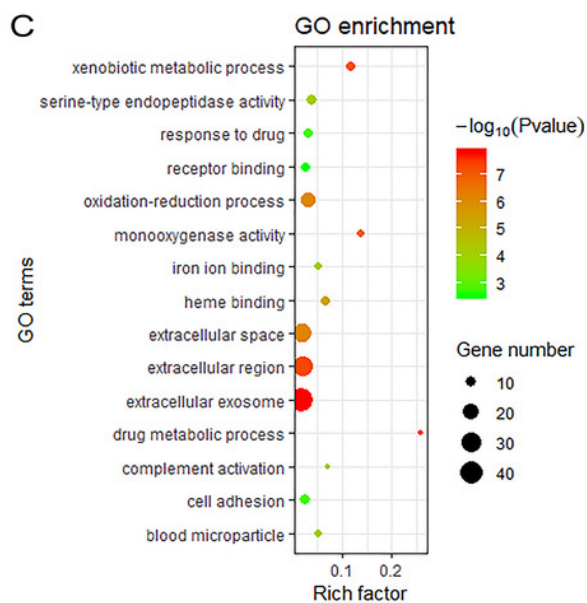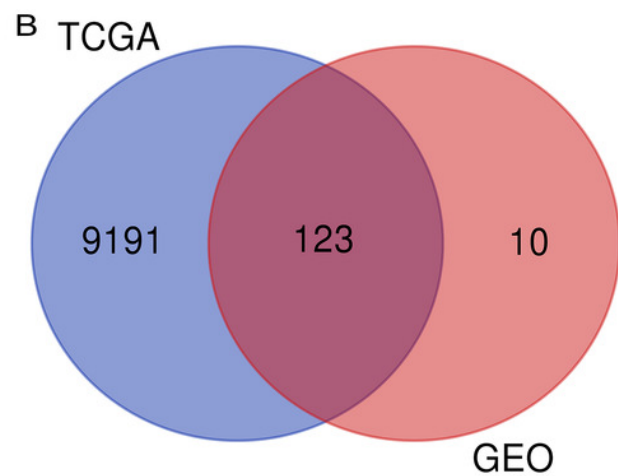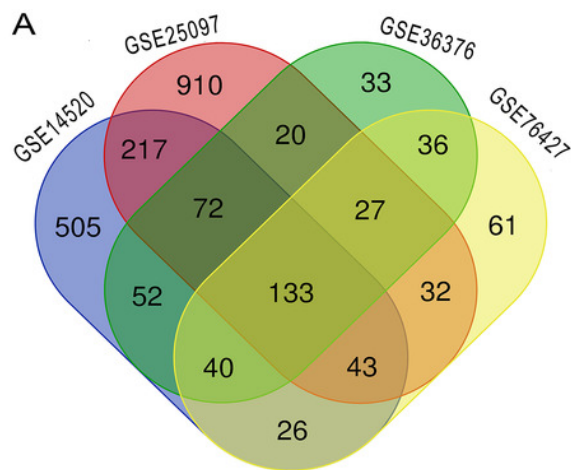
Figure 3

LASSO logistic regression analysis for selection of RRGs from DEGs.

(A) Selection of tuning parameter (λ) 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.
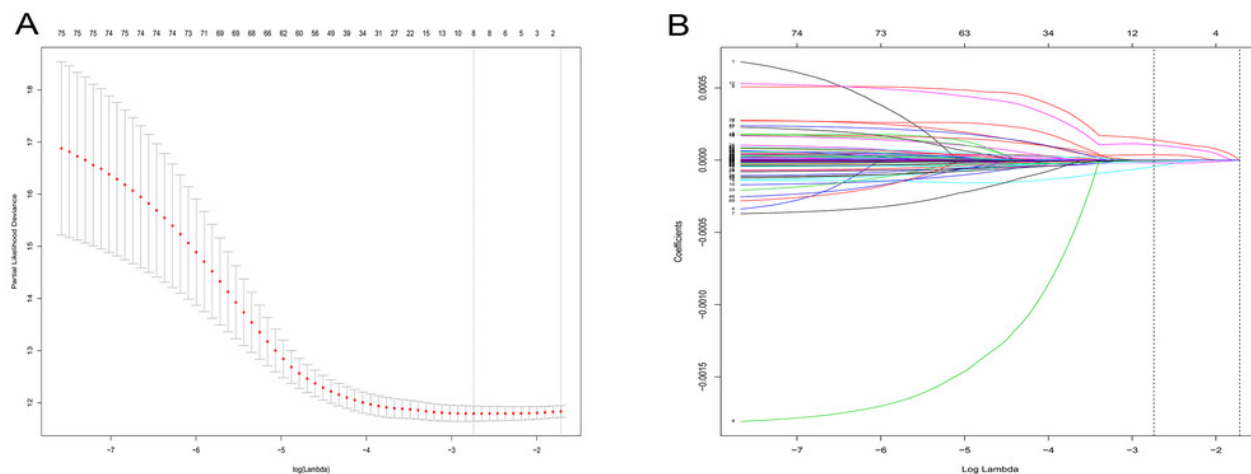
Figure 4

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.
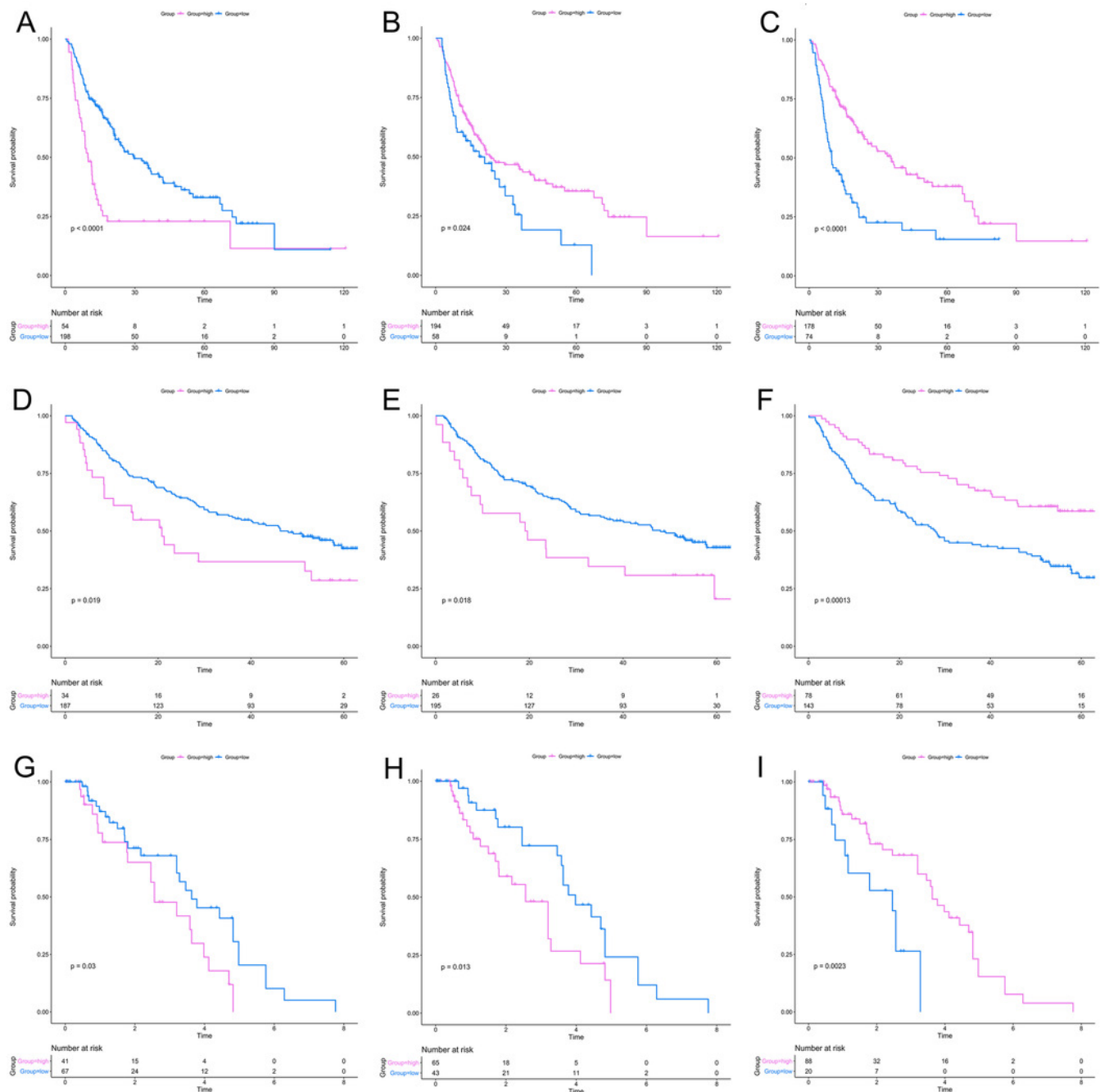
Figure 5

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.
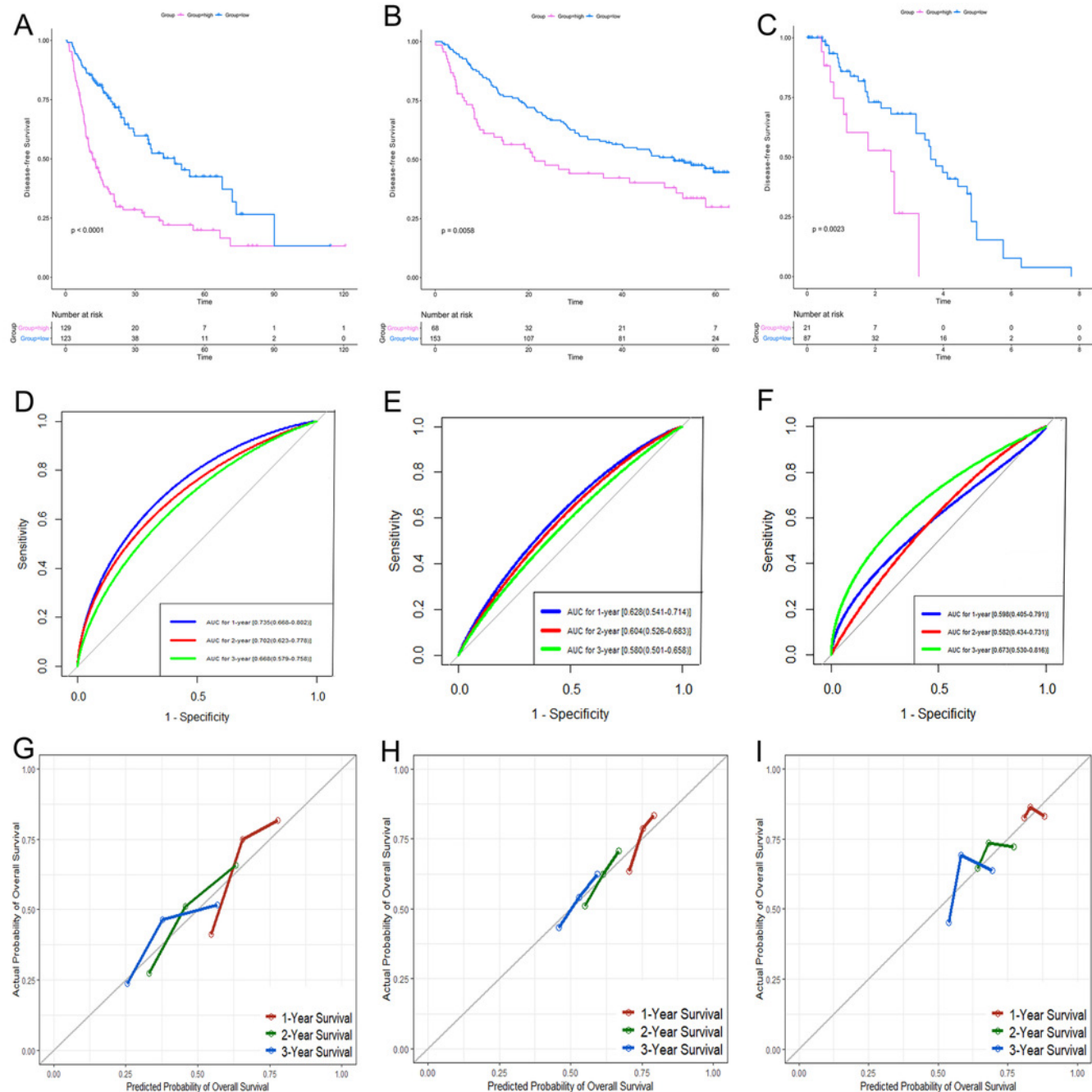
Figure 6

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.
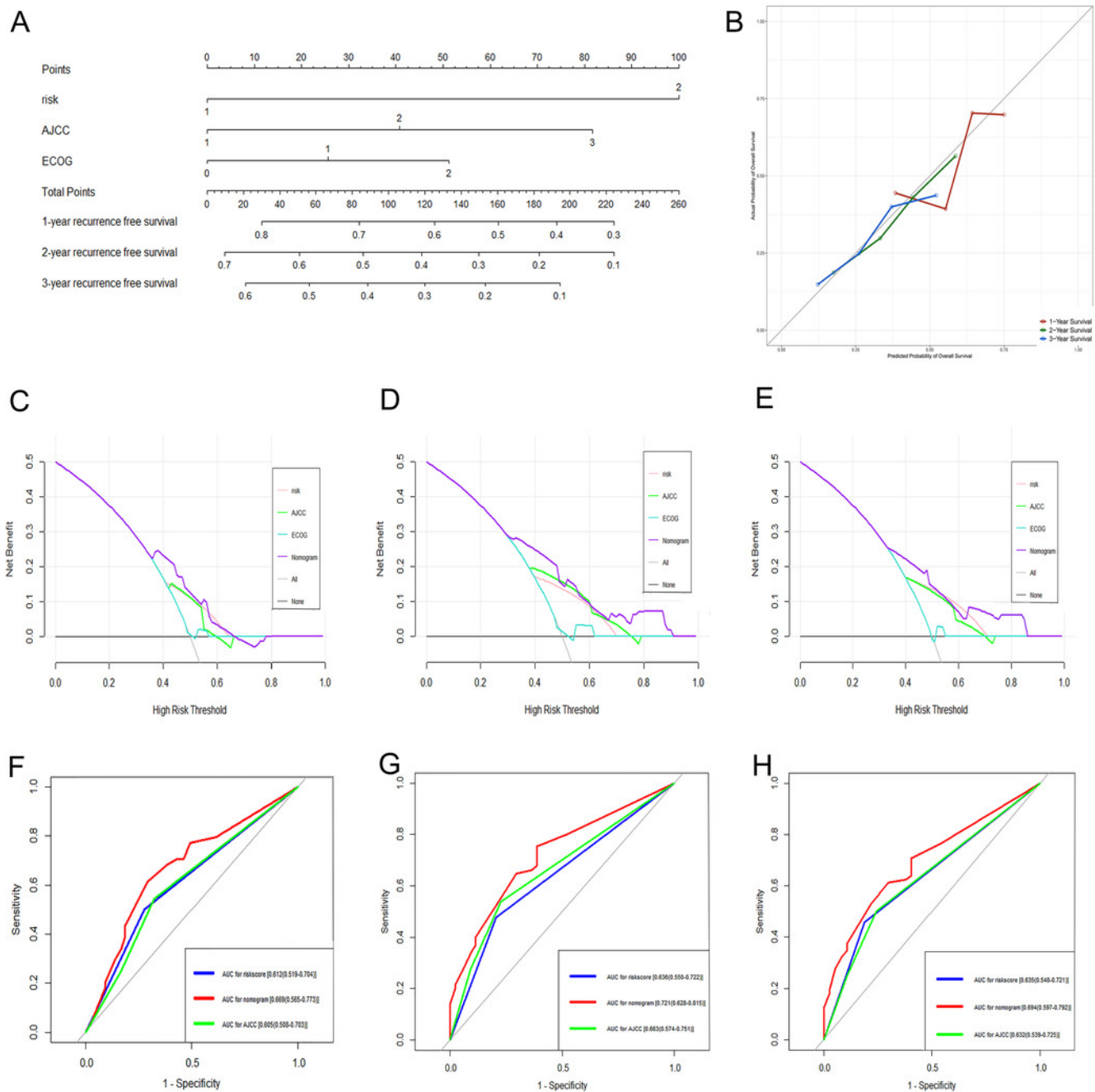
Figure 8

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.