

Identification and validation of the angiogenic genes for constructing diagnostic, prognostic, and recurrence models for hepatocellular carcinoma

Jinyu Zhu^{1,2,*}, Bufu Tang^{1,2,*}, Jie Li^{1,2}, Yueli Shi³, Minjiang Chen^{1,4}, Xiuling Lv⁴, Miaomiao Meng⁴, Qiaoyou Weng^{1,4}, Nannan Zhang^{1,2}, Kai Fan^{1,2}, Min Xu^{1,4}, Jiansong Ji^{1,4}

¹Key Laboratory of Imaging Diagnosis and Minimally Invasive Intervention Research, Lishui Hospital, School of Medicine, Zhejiang University, Lishui 323000, China

²Department of Radiology, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310058, China

³Department of Pathology and Pathophysiology, School of Medicine, Zhejiang University, Hangzhou 310058, China

⁴Department of Radiology, The Fifth Affiliated Hospital of Wenzhou Medical University, Lishui 323000, China

*Equal contribution

Correspondence to: Jiansong Ji, Min Xu; **email:** lschrjjs@163.com, lschrxm@163.com

Keywords: angiogenesis, hepatocellular carcinoma (HCC), diagnosis, recurrence, prognosis

Received: December 25, 2019

Accepted: March 9, 2020

Published: May 6, 2020

Copyright: Zhu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Since angiogenesis has an indispensable effect in the development and progression of tumors, in this study we aimed to identify angiogenic genes closely associated with prognosis of HCC to establish diagnostic, prognostic, and recurrence models. We analyzed 132 angiogenic genes and HCC-related RNA sequence data from the TCGA and ICGC databases by Cox and least absolute shrinkage and selection operator (LASSO) regression, and identified four angiogenic genes (ENFA3, EGF, MMP3 and AURKB) to establish prognosis, recurrence and diagnostic models and corresponding nomograms. The prognostic and recurrence models were determined to be independent predictors of prognosis and recurrence ($P < 0.05$). And compared with the low-risk group, patients in the high-risk group had worse overall survival (OS) rates in training cohort ($P < 0.001$) and validation cohort ($P < 0.001$), and higher recurrence rates in training cohort ($P < 0.001$) and validation cohort ($P = 0.01$). The diagnostic models have been validated to correctly distinguish HCC from normal samples and proliferative nodule samples. Through pharmacological analysis we identified piperlongumine as a drug for targeting angiogenesis, and it was validated to inhibit HCC cell proliferation and angiogenesis via the EGF/EGFR axis.

INTRODUCTION

Liver cancer is estimated to be the sixth most common cancer in the world (841,000 cases/year) and the fourth leading cause of cancer (782,000 cases/year), and it poses a serious health threat and economic burden to the world [1]. Hepatocellular carcinoma (HCC) accounts for the vast majority (approximately 75%-

85%) of primary liver cancer, and its increasing worldwide incidence and mortality has generated major clinical interest. There are various treatment strategies for different characteristics of HCC. Surgical resection, transplantation, ablation, and trans-arterial chemoembolization have been shown to produce survival benefits [2, 3], but due to the high recurrence rate after surgery, the prognosis of HCC patients is still

unsatisfactory, and OS and RFS are still very poor [4, 5]. Accurate prognosis assessment is a key step in the effective management and treatment of patients with HCC. In addition to traditional clinical variable, such as age, gender, pathological grade, clinical stage and vascular tumor cell invasion that affect the prognosis of HCC patients, complex molecular pathogenesis is also significantly related to the carcinogenicity and clinical outcome of HCC [6]. To develop more effective treatment management strategies for improving the survival status and clinical outcomes of HCC patients, it is urgent that new, efficient molecular prognostic markers be explored.

Angiogenesis is defined as the process of producing new blood vessels in existing vasculature in a manner known as "germination" [7]. Angiogenesis can occur in physiological and pathological conditions. Physiological angiogenesis is involved in embryonic development, wound healing, and collateral formation to improve organ perfusion, and it is essential in normal human development [8]. Pathological angiogenesis is one of the hallmarks of cancer [9], and it plays a crucial role from the early stages of tissue carcinogenesis to the end of tumor metastasis [10]. Angiogenesis in tumors is driven the overexpression of angiogenic factors and rarely matures or becomes resolved [11]; thus, it is the basis for tumor cell growth, invasion, and metastasis [12]. On the one hand, these new blood vessels provide the necessary oxygen and nutrients to maintain the rapid growth and proliferation of tumor cells, and on the other hand, they provide a channel for tumor cells to enter the circulation and develop distant metastasis [13].

HCC is one of the most common highly vascularized solid tumors [14], and angiogenesis plays a key role in the pathogenesis, growth and progression of HCC [15]. When precancerous lesions develop into clinical HCC, angiogenesis starts, and it continues with the progression of the tumor, inducing rapid growth, proliferation and metastasis of the tumor and causing poor prognosis of HCC patients [16]. Many previous studies have explored and reported the close relationship between angiogenesis and HCC. For example, Park YN [17] analyzed vascular endothelial growth factor (VEGF), which is the most important factor in tumor angiogenesis. Its expression level in tumor tissues increases with the progression of HCC. VEGF-positive HCC patients have a worse prognosis and shorter survival time than their VEGF-negative counterparts. Llovet JM et al. [18] also confirmed that angiogenic markers such as plasma angiopoietin 2 (Ang2) and plasma VEGF are independent predictors of survival prognosis in patients with advanced HCC. However, most of the current research is focused on exploring the effects of individual angiogenesis-related

genes on the development and prognosis of HCC. Few studies have integrated multiple angiogenesis-related genes by high-throughput biomarker sequencing and comprehensively analyzed the relationship between these genes and the prognosis and survival of HCC.

In this study, by analyzing the HCC-related mRNA sequence data in TCGA database and the ICGC database and the angiogenic genes identified in previous publications, we identified four angiogenic genes significantly related to the prognosis of HCC and then constructed and validated diagnostic, prognostic and recurrence models and corresponding nomograms of HCC. The possible anti-tumor mechanisms of piperlongumine, an inhibitor of angiogenic genes, were comprehensively evaluated. Our findings may help further improve the individualized treatment and clinical outcome of HCC.

RESULTS

Differentially expressed HCC-related angiogenic gene identification

First, there is a flowchart that graphically illustrates the analysis process for this study Supplementary Figure 1. We identified 132 angiogenic genes from several studies on angiogenesis, and we matched these genes to HCC-related mRNA sequence data in TCGA and ICGC databases; we compared differentially expressed genes (DEGs) in the mRNA expression profiles between HCC tissues and adjacent noncancerous tissues. Using the threshold of $FDR < 0.05$ and $|\log_2 FC| > 1$, 49 differentially expressed angiogenic genes (40 upregulated and 9 downregulated) were obtained from the database from TCGA (Figure 1A, 1B), and 30 differentially expressed angiogenic genes (24 upregulated and 6 downregulated) were obtained in the ICGC database (Figure 1C, 1D). Then, by matching the angiogenic genes in the two databases, we selected 28 genes for subsequent analysis (23 upregulated and 5 downregulated) (Supplementary Figure 2).

Identification of key angiogenic genes closely associated with prognosis of HCC

To identify angiogenic genes that were significantly associated with HCC survival, we performed LASSO regression and Cox regression analysis. We first performed LASSO to reduce the number of angiogenic genes. A total of 1000 Cox LASSO regression iterations and 10 cross-validations were performed using the R package glmnet. The higher the nonzero coefficient appeared after 1000 iterations of Cox LASSO regression, the stronger the ability of the corresponding gene to predict prognosis. Using the Cox LASSO

regression, we selected 14 angiogenic genes with a strong predictive prognosis (with a recurrence frequency greater than 900 times out of 1000 repetitions) (Supplementary Figure 3). We then used 370 HCC samples from the database from TCGA along with OS time and survival status to investigate the relationship between these 14 angiogenic genes and prognosis using univariate Cox regression analysis. The results confirmed that there were 4 genes (EFNA3, EGF, MMP3 and AURKB) with statistical significance ($P < 0.01$); that is, these four key angiogenic genes were significantly associated with HCC prognosis.

Construction of an angiogenesis-related prognostic signature and validation of its predictive performance

Based on these four angiogenic genes, we established a prognostic signature by multivariate Cox regression as follows: prognostic index (PI) = (β^* expression level of EGF) + (β^* expression level of MMP3) + (β^*

expression level of AURKB) + (β^* expression level of EFNA3). In the HCC cohort from TCGA, 371 HCC patients with survival status were divided into high-risk and low-risk groups using the optimal cut-off value determined by X-tile software. The OS of the high-risk group was significantly worse than that of the low-risk group ($P < 0.001$) (Figure 2A). Figure 2B shows the risk score distribution corresponding to the expression level of the genes. In HCC patients with different survival statuses, we found that the risk score of the dead patients was significantly higher than that of the living patients ($P < 0.01$) (Figure 2C). Time-dependent ROC curve analysis was performed to evaluate the specificity and sensitivity of the prognostic signature. The higher the area under the curve (AUC), the better the predictive performance of the prognostic signature. The AUCs of the prognostic signature reached 0.76 at 0.5 years, 0.74 at 1 year, 0.67 at 3 years and 0.65 at 5 years, indicating a great predictive value of the prognostic signature for OS (Figure 2D). For further validation of the prediction ability, the prognostic signature was applied to the

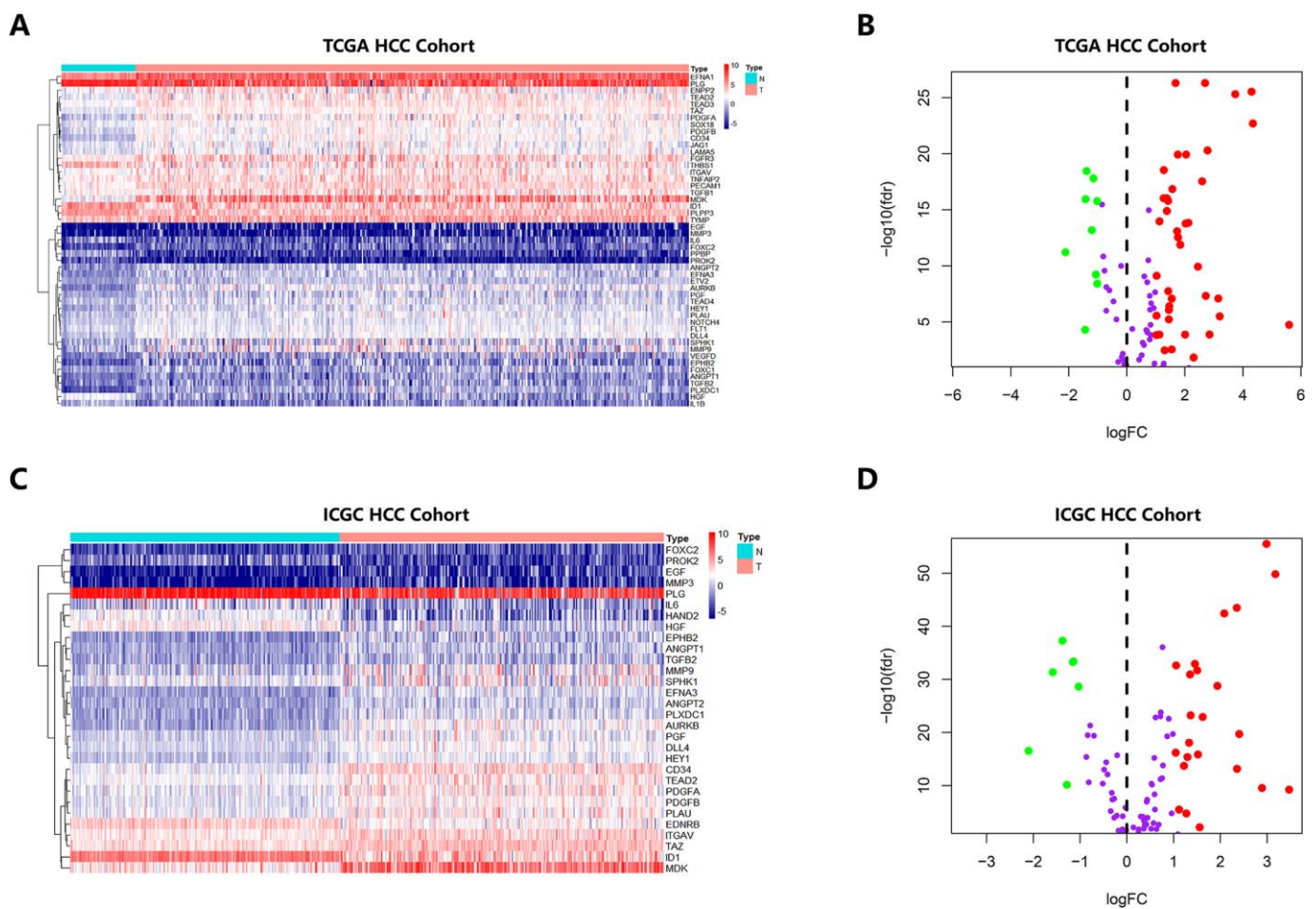


Figure 1. Heatmap and Volcano Plot showed the differentially expressed angiogenic genes of HCC patients in different database. (A, B) Gene expression features in the TCGA database. (C, D) Gene expression levels in the ICGC database.

ICGC database (containing 243 HCC samples) for survival prediction. Utilizing the formula mentioned earlier and the optimal cutoff value identified by X-tile software, 243 HCC patients were classified into high-risk and low-risk groups, and consistent with previous results, patients in the high-risk group had a significantly worse OS than those in the low-risk group ($P < 0.001$) (Figure 2E). Figure 2F shows the risk score distribution corresponding to the expression level of the genes. Consistent with previous findings, the risk score of dead HCC patients was much higher than that of living HCC patients ($P < 0.01$) (Figure 2G). The AUCs of the prognostic signature for OS were 0.73 at 0.5 years, 0.79 at 1 year, 0.64 at 3 years and 0.70 at 4 years (Figure 2H).

A predictive nomogram for OS prediction was constructed and validated in the HCC cohort from TCGA

To evaluate the independent predictive value of this prognostic signature, we performed univariate and multivariate Cox regression analyses on the risk score of the prognostic signature and clinical characteristics (AFP, weight, vascular tumor invasion, sex, pathological grade, and TNM stage) of 370 HCC patients in the cohort from TCGA. The results of the univariable Cox regression analysis showed that age, TNM stage and the risk score of the prognosis signature had predictive value for OS ($HR > 1.00$, $P < 0.05$), while AFP, weight, vascular tumor invasion and sex

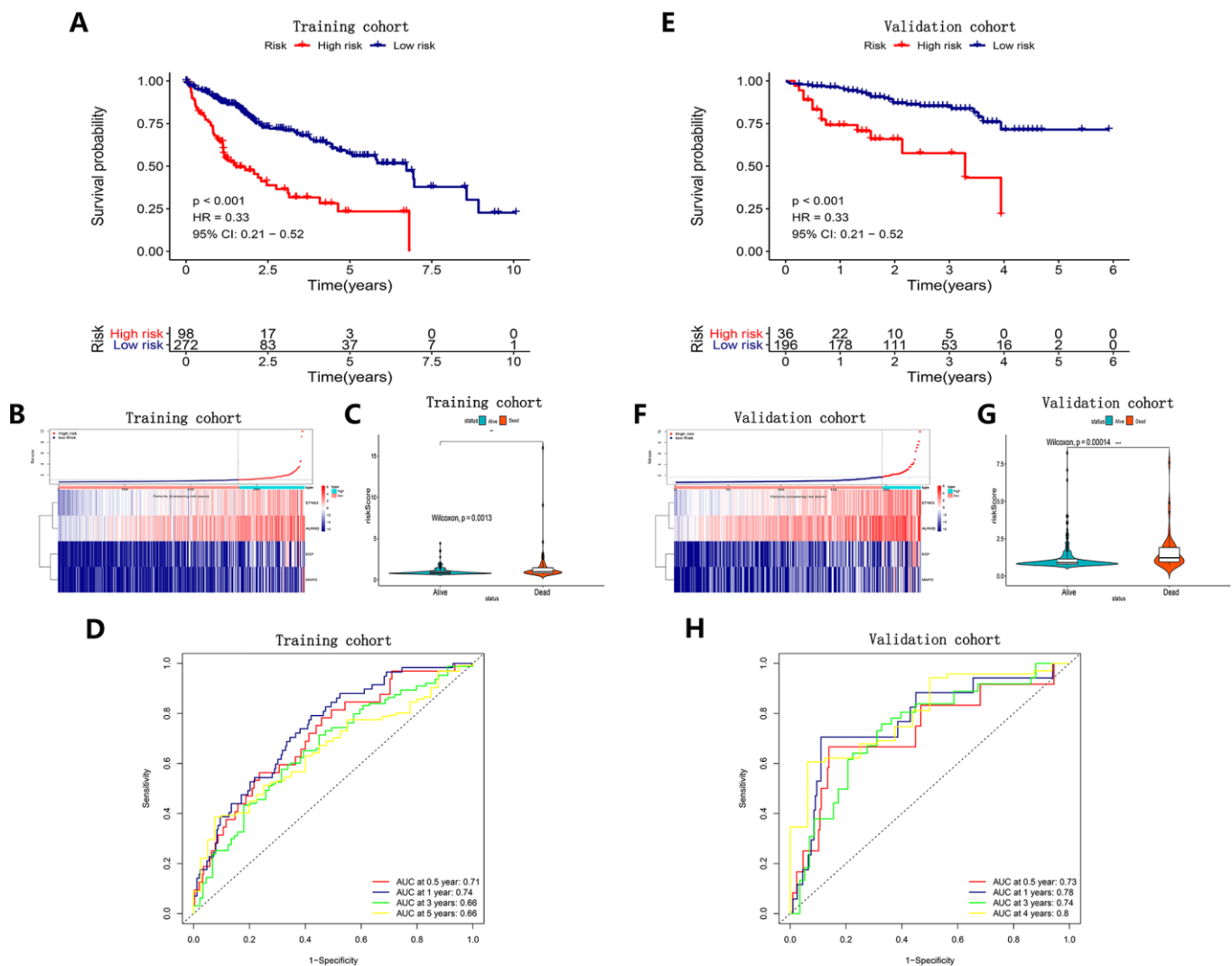


Figure 2. Survival analysis, risk score distribution, Violin plots and time-dependent ROC curves of a prognostic model in the TCGA HCC cohort (A–D) and ICGC HCC cohort (E–H). A and E K–M survival curves showed the OS in the high-risk group was significantly worse than that in the low-risk group ($P < 0.05$). (B, F) Distribution of risk scores of HCC patients with different gene expression levels. (C, G) Risk scores of HCC patients with different survival status. (D, H) ROC curve analysis for OS prediction.

were not associated with OS. Therefore, we incorporated age, TNM stage, and risk score of the prognostic signature in a multivariate Cox regression analysis, and the results showed that TNM stage (HR = 1.961, $P < 0.05$) and the prognostic signature (HR = 1.315, $P < 0.05$) were important independent prognostic factors for OS (Figure 3A). To provide clinicians with a quantitative method to more accurately predict the prognosis and survival status of HCC patients, thereby contributing to more individual clinical decisions and treatment options, we constructed a predictive nomogram, which integrated the two independent prognosis factors based on the multivariable Cox regression analysis (Figure 3B). Each independent prognosis factor was assigned a certain score scale, and the total point value was obtained by adding up the specific points corresponding to the individual's prognosis variable, thereby determining the corresponding OS for the individual at 1, 3, and 5 years. The calibration curves of the nomogram at 1, 3, and 5 years were very close to the best prediction curve, showing a great consistency between the predicted OS rates and the actual observations (Figure 3C–3E).

We then adopted the C-index, time-dependent ROC and DCA to verify the prognostic accuracy of the nomogram in OS prediction. The C index of the nomogram in OS prediction was 0.64, with 1000 cycles of bootstrapping (95% CI: 0.56–0.72), which was superior to the C index for the TNM stage (0.54) and the prognostic signature (0.63); these results were consistent between the nomogram's prediction and actual results. The time-dependent ROC curves showed that the nomogram with the integration of two independent prognosis factors reached 0.641, 0.681 and 0.671 in 1 year, 3 years and 5 years, which was better than a single independent prognosis factor on predictive performance (Figure 3F–3H). The results of the DCA also showed that compared to a single independent prognostic factor, the best net benefit was obtained by prognosis prediction with the nomogram (Figure 3I–3K).

Definition of a recurrence signature based on the key angiogenic genes

We utilized the regression coefficient (β^*) of a multivariate Cox proportional hazards model based on four angiogenic genes and patients with release-free survival (RFS) time and recurrent status in the cohort from TCGA and the GSE14520 cohort. Then, we further constructed a recurrence signature as follows: prognostic index (PI) = (β^* * expression level of EGF) + (β^* * expression level of MMP3) + (β^* * expression level of AURKB) + (β^* * expression level of EFNA3).

Determining the optimal cut-off value by X-tile software, HCC patients were divided into high-risk and low-risk groups. We found that the recurrence rate in HCC patients in the high-risk group was dramatically higher than that of the low-risk group in the same period in the cohort from TCGA ($P < 0.001$) (Figure 4A) and the GSE14520 cohort ($P = 0.01$) (Figure 4D). The distribution of risk scores corresponding to the expression patterns of the angiogenic genes is presented in Figure 4B, 4E. The AUC values for the 0.5-, 1-, 3-, and 5-year recurrence probability with the recurrence signature reached 0.64, 0.68, 0.54 and 0.66 in the cohort from TCGA, respectively (Figure 4C). In addition, the AUCs for 0.5-, 1-, 3-, and 5-year recurrence probability with the recurrence signature reached 0.57, 0.51, 0.6 and 0.57 in the GSE14520 cohort, respectively (Figure 4F). These results showed that the recurrence signature had a great predictive power for recurrence.

Establishment and evaluation of a recurrence nomogram in predicting recurrent probability in the HCC cohort from TCGA

To determine whether the predictive performance of the recurrence signature was independent of traditional clinical features, we performed univariate and multivariate Cox regression analysis on these variables in the cohort from TCGA. The analyses confirmed that age (HR = 2.198, $P < 0.05$), TNM stage (HR = 2.499, $P < 0.05$) and the recurrence signature (HR = 2.883, $P < 0.05$) were independent of other clinical features (Figure 5A). Then, a recurrence nomogram integrated the three significantly independent factors was established (Figure 5B). The calibration curves of the nomogram for 1-, 3-, and 5-year recurrence probabilities, and the best prediction curve (45° line) was nearly coincident, indicating a great agreement between the predicted recurrence probabilities and the actual observations (Figure 5C–5E).

Meanwhile, we compared the prediction accuracy between the nomogram and other single individual prognostic factors via C index, time-dependent ROC curve analysis and DCA. The C index of the nomogram reached 0.71, indicating that the nomogram had an excellent predictive accuracy for recurrence in HCC patients. The AUCs for 1-, 3-, and 5-year recurrence probability with the nomogram reached 0.749, 0.748, 0.54 and 0.722, respectively (Figure 5F–5H). The DCA also demonstrated that the nomogram could help clinicians obtain the best net benefits when making clinical decisions (Figure 5I–5K). It was demonstrated that the nomogram had superior predictive performance and clinical value to that of any single independent recurrence factor.

Identification and validation of a diagnostic model based on the four angiogenic genes

To improve the disease outcome of HCC patients, there is a need for prognostic signatures and recurrence

signatures to help clinicians adjust the treatment strategy for confirmed HCC patients, but in the early diagnosis, there is also a need for sensitive and specific diagnostic criteria to accurately distinguish HCC patients from normal subjects. To this end, we

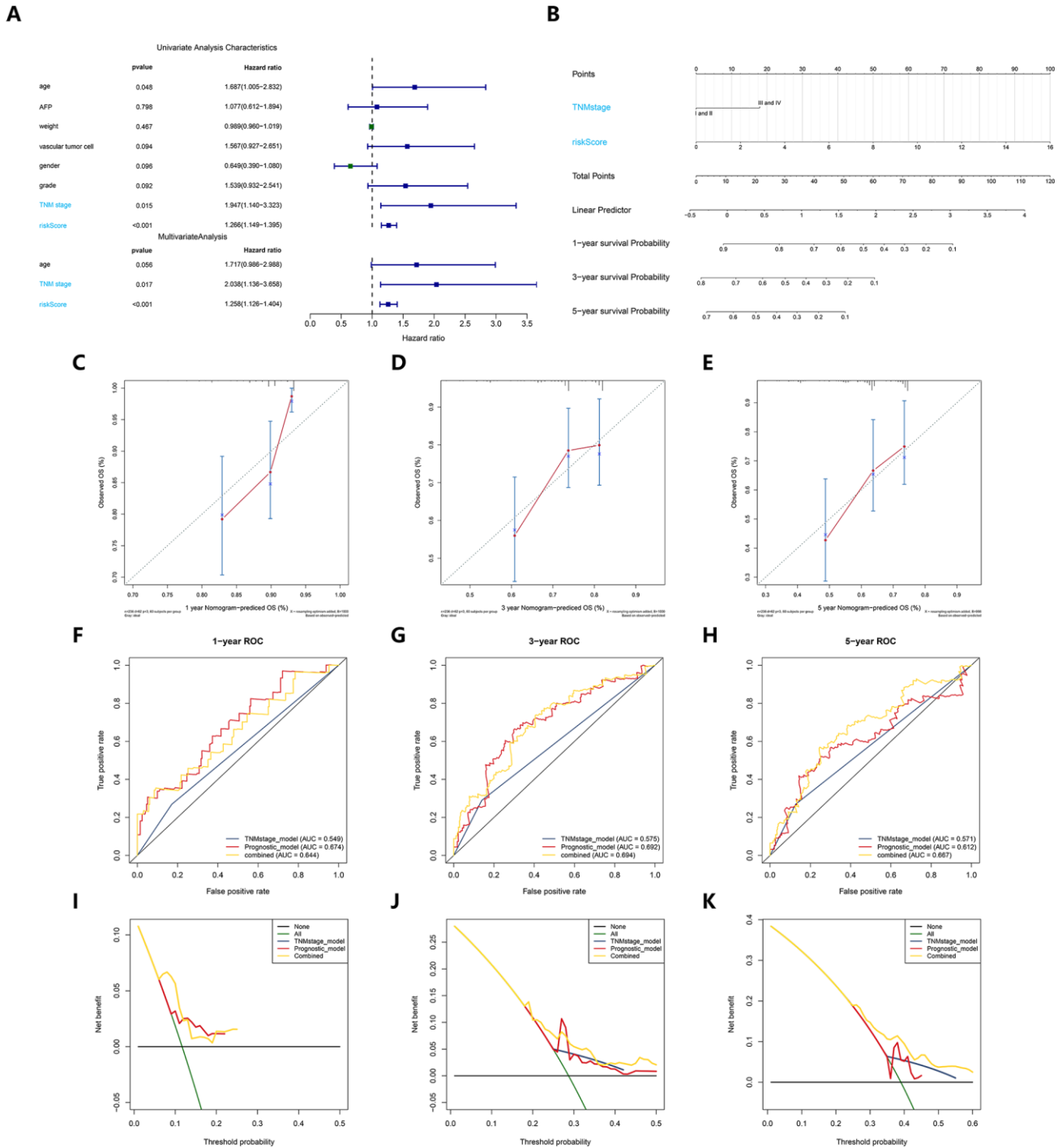


Figure 3. Construction and verification of a predictive nomogram. (A) Univariate and multivariate Cox regression suggested the prognostic model and age were independent prognostic predictors. (B) The nomogram for OS prediction in HCC patients at 1, 3, and 5 years. (C-E) Calibration curves of nomogram for OS prediction at 1, 3, and 5 years. (F-H) ROC curves to evaluate the predictive performance of nomogram. (I-K) DCA curves to evaluate the clinical decision-making benefits of nomogram.

performed a stepwise logistic regression method to establish a diagnostic model that integrated these four angiogenic genes to distinguish between HCC patients and normal subjects. The diagnostic equation was finally identified as follows: $\text{logit}(P = \text{HCC}) = -75.2125977 + (0.85 \times \text{ENFA3 expression level}) + (2.61 \times \text{EGF expression level}) + (3.92 \times \text{MMP3 expression level}) + (9.98 \times \text{AURKB expression level})$. The diagnostic model was confirmed to achieve a specificity of 94.00% and a sensitivity of 90.00% in the HCC cohort from TCGA (containing 50 normal samples and paired 50 HCC samples) (Figure 6A). The AUC for the diagnostic model reached 0.976, which is a diagnostic value that is superior to other methods (Figure 6B). We then validated the diagnostic model using the ICGC HCC cohort (containing 202 normal samples and 243 HCC samples). The diagnostic model diagnosed HCC patients with a specificity of 92.08%, sensitivity of 83.95% (Figure 6D), and AUC of 0.948 (Figure 6E), confirming that the diagnostic model can accurately distinguish HCC samples from normal samples. We show the correspondence between the predicted results of the diagnostic model and the actual status and the corresponding expression characteristics of angiogenic genes in Figure 6C, 6F.

Furthermore, to improve the early screening rate of HCC and reduce the probability of misdiagnosis, we

integrated these four angiogenic genes to establish a diagnostic model in the GSE89377 dataset (containing 22 proliferative nodules samples and 40 HCC samples) to detect whether the diagnostic model could correctly distinguish between HCC and proliferative nodules. We used a stepwise logistic regression method to establish the diagnostic model with a diagnostic formula of $\text{logit}(P = \text{HCC}) = -5.58 + (5.84 \times \text{ENFA3 expression level}) + (1.84 \times \text{EGF expression level}) + (-1.04 \times \text{MMP3 expression level}) + (18.10 \times \text{AURKB expression level})$. The diagnostic model was confirmed to distinguish between HCC and proliferative nodules with high specificity (81.82%) and high sensitivity (90.00%) (Figure 7A). The AUC for the diagnostic model was 0.931, confirming that the diagnosis model had diagnostic capabilities superior to those of other methods (Figure 7B). We then applied this diagnostic model to GSE6764 to further validate its diagnostic performance. The results showed that the diagnostic model had a specificity of 70.59%, a sensitivity of 85.71% (Figure 7D), and an AUC of 0.8848 (Figure 7E), demonstrating that the diagnostic model correctly distinguished HCC and hyperplastic nodules. Figures 7C, 7F show different angiogenic gene expression levels, the correspondence between the predicted results in the diagnostic model and the actual results.

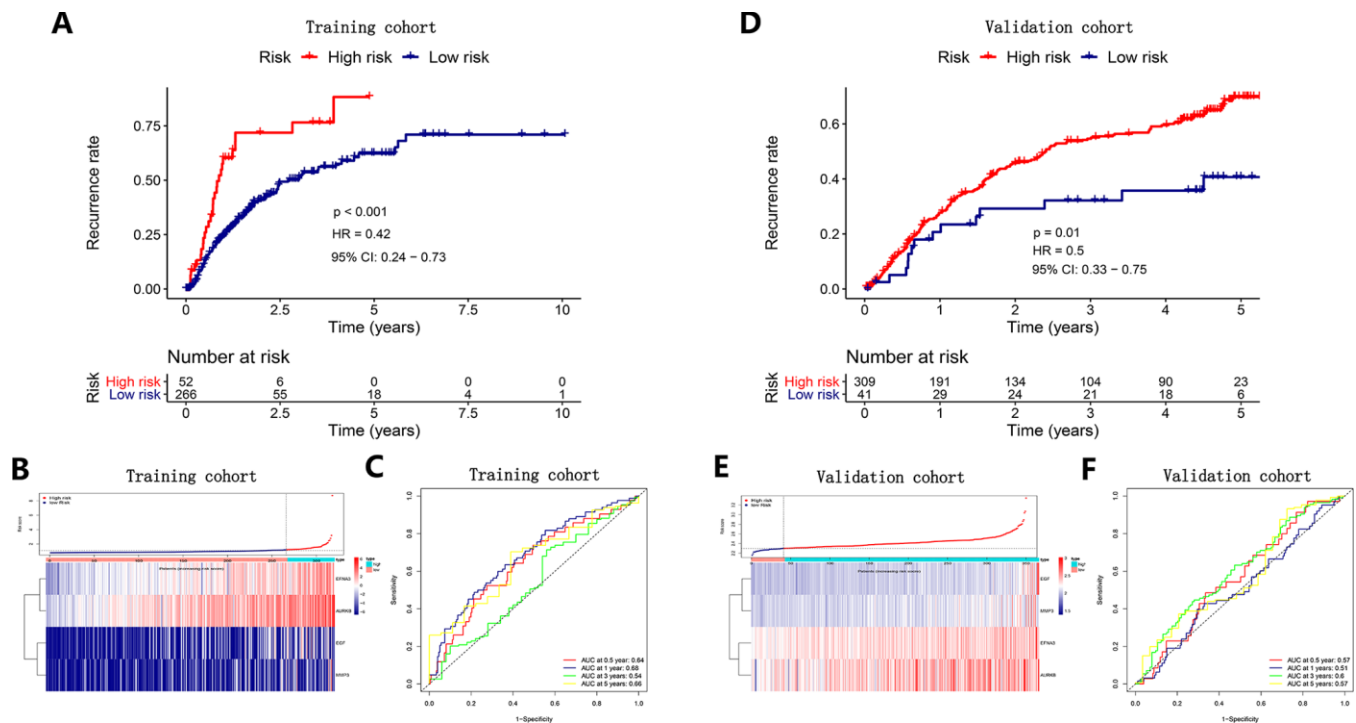


Figure 4. Survival analysis, risk score distribution and time-dependent ROC curves of a recurrence model in the TCGA HCC cohort (A-C) and ICGC HCC cohort (D-F). (A, D) The recurrence rates in the high-risk group was much higher than that in the low-risk group ($P < 0.05$). **(B)** and **E** Gene expression levels and corresponding risk scores in HCC patients. **(C, F)** ROC curve analysis for recurrence prediction.

Further internal and external validation of the expression patterns and prognostic predictive value of the angiogenic genes

To examine the value of these four angiogenic genes (ENFA3, EGF, MMP3, and AURKB) for the

construction of diagnostic, prognostic and recurrence models, we analyzed the expression patterns of the angiogenic genes in the HCC cohorts from ICGC database and different online database, including cBioportal, UALCAN and Human Protein Atlas database. We first performed genetic alteration

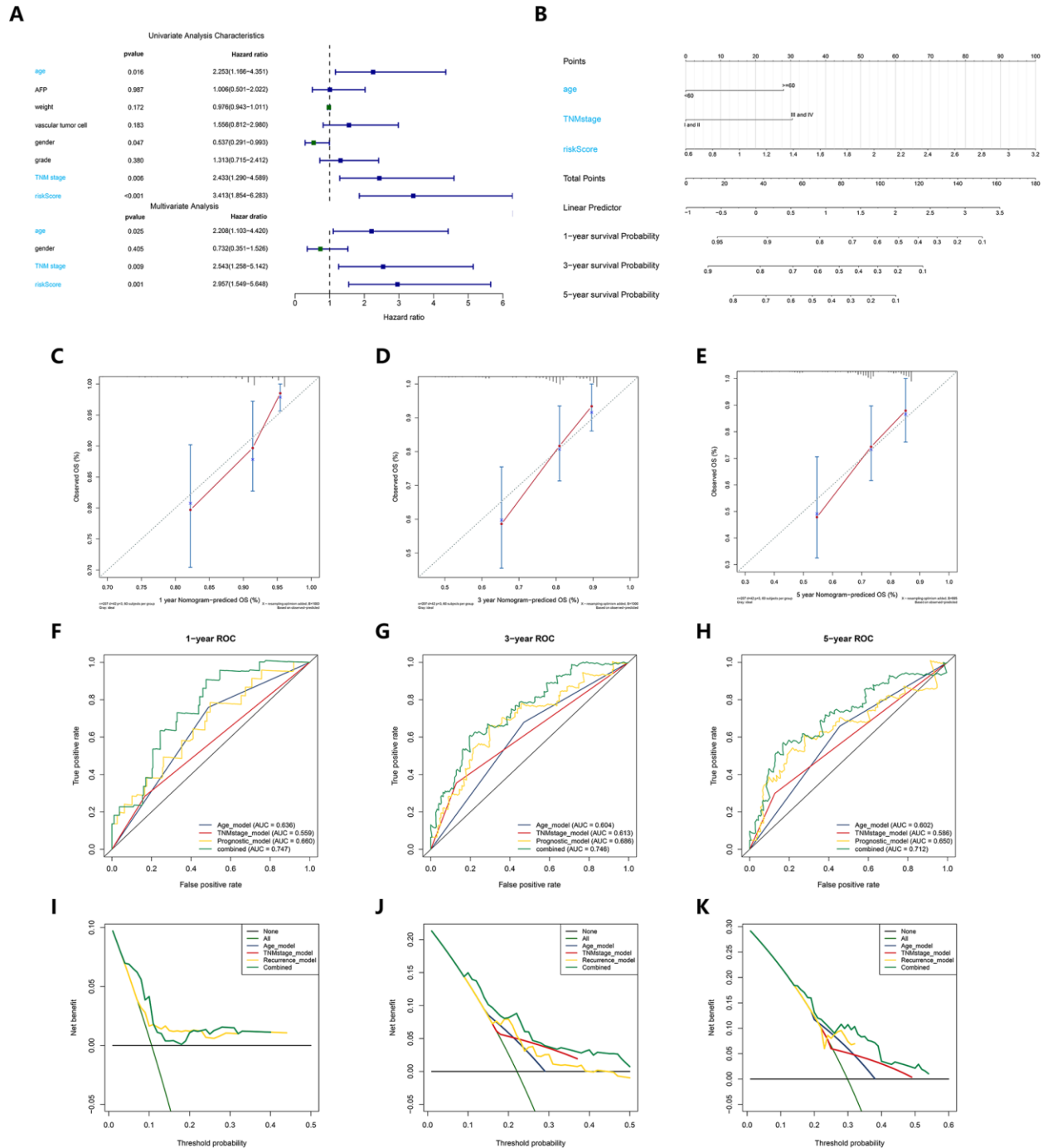


Figure 5. Construction and validation of a recurrence nomogram. (A) Univariate and multivariate Cox regression suggested the recurrence model, TNM stage and age were independent prognostic predictors. (B) The nomogram for recurrence prediction in HCC patients at 1, 3, and 5 years. (C–E) Calibration curves of nomogram for recurrence prediction at 1, 3, and 5 years. (F–H) ROC curves to evaluate the predictive performance of nomogram. (I–K) DCA curves to evaluate the clinical decision-making benefits of nomogram.

detection in the HCC cohort from cBioportal database [19]. It was found that 20.1% (72 HCC samples) showed genetic alterations, 14% of which occurred in ENFA3, where the main alteration was amplification; 5 mutations occurred in EGF (2.2%) (Figure 8A). These genes have the potential to become targeted therapeutic sites and provide new therapeutic targets for the treatment of HCC. Subsequently, in the HCC cohort from UALCAN database [20], we found that the

expression levels of the four angiogenic genes in tumor tissues were significantly higher than it was in normal tissues ($P < 0.001$) (Figure 8B–8E), and this finding was also consistently validated in the ICGC HCC cohort ($P < 0.001$) (Figure 8F–8I). In addition, to verify the clinical relevance of these four angiogenic genes, we analyzed the expression of proteins encoded by the four genes in clinical HCC specimens and normal specimens from the Human Protein Atlas database [21]. The

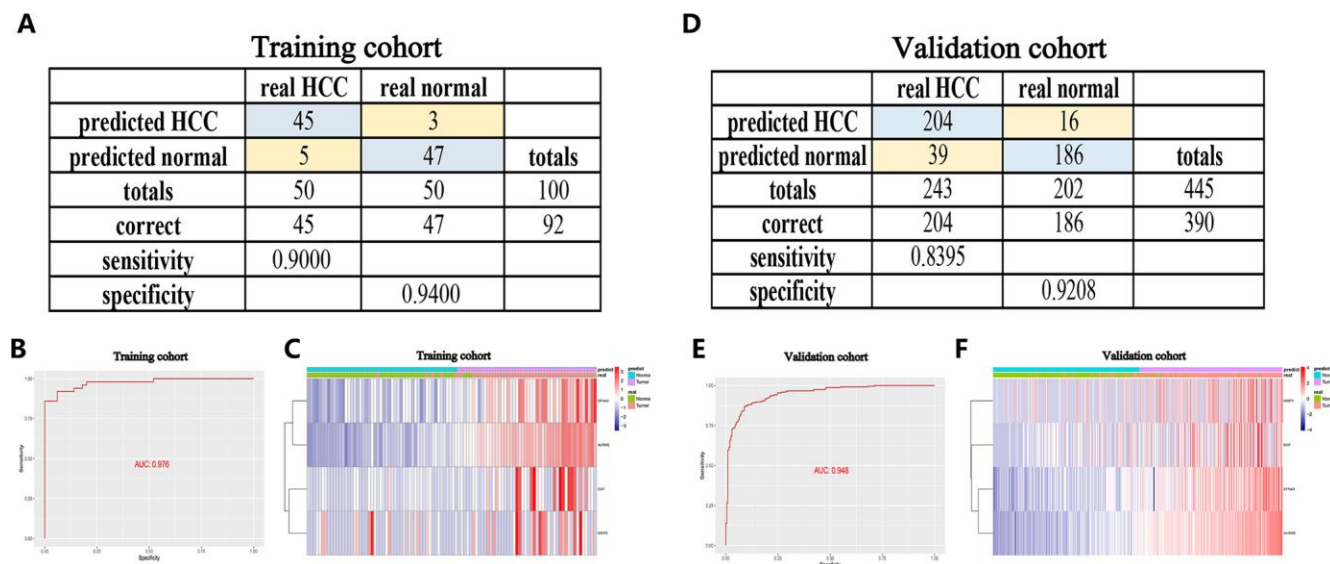


Figure 6. A diagnostic model for distinguishing HCC from normal samples in the TCGA HCC cohort (A–C) and ICGC HCC cohort (D–F). (A, D) Confusion matrix for the binary classification results of the diagnostic model. (B, E) ROC curves for the predictive performance evaluation of the diagnostic model. (C, F) Unsupervised hierarchical clustering of the four angiogenic genes for the diagnostic model.

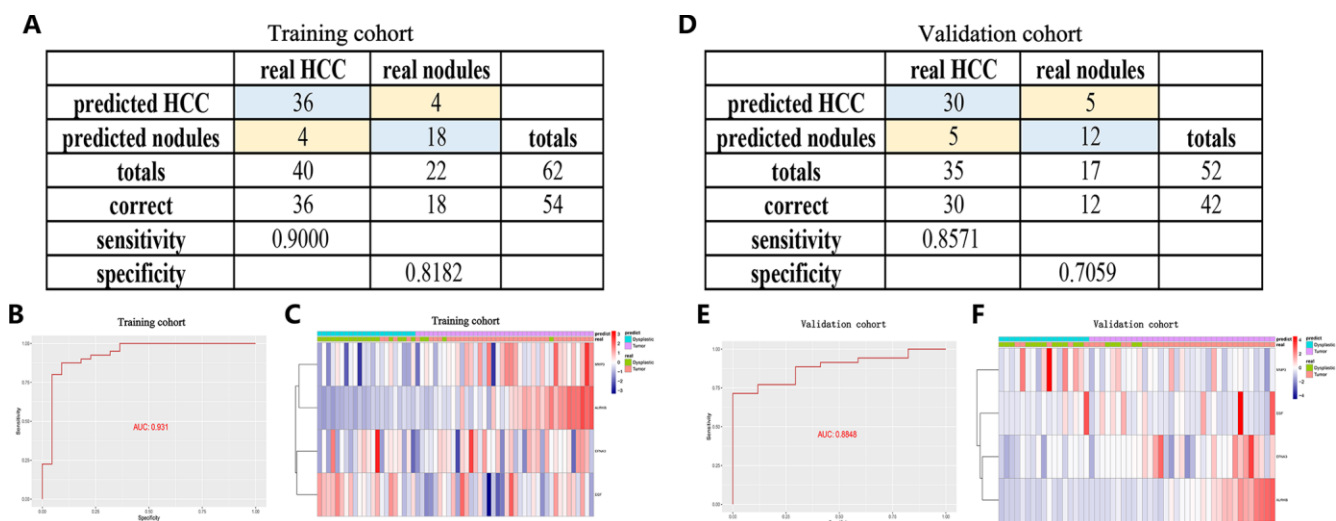


Figure 7. A diagnostic model for distinguishing HCC from proliferative nodules in the training dataset (GSE89377) and validation dataset (GSE6764). (A, D) Confusion matrix for the binary classification results of the diagnostic model. (B, E) ROC curves for the predictive performance evaluation of the diagnostic model. (C, F) Unsupervised hierarchical clustering of the four angiogenic genes for the diagnostic model.

results showed that ENFA3, MMP3 and AURKB were strongly positive in HCC tissues compared to normal liver tissues (Figure 8J), while EGF was not found in the database. The expression characteristics of angiogenic genes were consistent with the prognosis and recurrence analysis, and the finding indicated that the expression characteristics of these four genes are of great value for constructing diagnostic, prognostic and recurrent models. In addition, the expression levels of the four genes in different cell lines are shown in Figure 8K–8N.

Since these four angiogenic genes were highly expressed in HCC patients and were risk factors, we explored the expression correlation of these four genes. We found that ENFA3 and MMP3 have synergistic effects, and their expression levels were positively correlated (Pearson 0.15, $P < 0.05$) (Figure 9A). This result was also verified among other genes (ENFA3 and AURKB, EGF and MMP3, EFNA3 and EGF) ($P < 0.05$) (Figure 9B–9D), while the expression level of AURKB was not significantly correlated with the expression levels of EGF or MMP3 ($P > 0.05$). These results suggested that there was a relationship between these genes and their expression, further contributing to the occurrence of HCC. To verify the association between the expression levels of these four angiogenic genes and the OS of HCC patients, we performed K-M curve analysis of the HCC cohort from TCGA. The analysis showed that the OS of HCC patients with high expression of EFNA3 was significantly lower than that of patients with low expression of ENFA3 ($P < 0.05$). EGF, MMP3 and AURKB also had consistent results; that is, high expression of these genes corresponded to low OS (Figure 9E–9H). The AUCs for 0.5-, 1-, 3- and 5-year OS indicated a great predictive performance based on the expression of the four genes (Figure 9I–9L).

Piperlongumine was found to be an inhibitor of the angiogenic genes

In exploring the clinically targeted therapeutic value of the four angiogenic genes, we identified piperlongumine as an inhibitor of the EGF/EGFR axis through the Cancer Therapeutics Response Portal (CTRP) database (<http://portals.broadinstitute.org/ctrp/>). (Supplementary Figure 4). Piperlongumine is an amide alkaloid of *Piper longum* L. (long piper) that exhibits cytotoxicity against a variety of human cancer cell lines and exhibits antitumor activity in rodents [22]. The 3D structure of piperlongumine is shown in Figure 10A. To investigate the effect of piperlongumine on the angiogenesis signature, we first used Autodock software (Version 3.6.1) to dock piperlongumine and angiogenic genes, and we found that piperlongumine could effectively dock to specific sites of EGF, EGFR and MMP3

(Figure 10B–10D, Supplementary Figure 5). When further validation was performed using PharmMapper (<http://www.lilab-ecust.cn/pharmmapper/>), we found that EGFR and MMP3 were potential targeting regulators of piperlongumine (Figure 10E–10F).

Analysis of the anti-tumor effect of piperlongumine and its possible molecular mechanism

To further verify whether piperlongumine had anti-tumor effects on HCC, we performed in vitro cell proliferation experiments and tube formation experiments. In the Half Maximal Inhibitory Concentration (IC₅₀) assay, the IC₅₀ of piperlongumine was found to be 7.22 μmol in the SK-HEP1 cell line, 13.23 μmol in the SMMC-7721 cell line, 6.67 μmol in the HUVEC cell line, and 29.54 μmol in human LO2 hepatocytes, indicating that HCC cell lines (SK-HEP1 and SMMC-7721) were much more sensitive than human LO2 hepatocytes to piperlongumine (Figure 11A). To assess the effect on tumor cell proliferation in vitro, we treated SK-HEP1 and SMMC-7721 cell lines with different concentrations of piperlongumine and evaluated them by CCK8 assay, and EdU assay. We found by the CCK8 assay that piperlongumine treatment inhibited cell proliferation in a dose-dependent manner (Figure 11B, 11C), and the EdU assay further confirmed this result (Figure 11D–11E). To investigate the effect of piperlongumine on angiogenesis in the EGF/EGFR signaling pathway, we performed an in vitro HUVEC tube formation assay. The results suggested that piperlongumine inhibited angiogenesis via the EGF/EGFR axis signaling pathway (Figure 11F–11G).

Because we have demonstrated that the anti-tumor effect of piperlongumine was achieved by inhibiting proliferation and angiogenesis, we attempted to clarify the specific mechanism of piperlongumine on HCC cells. In the Cancer Therapeutics Response Portal (CTRP) database (<http://portals.broadinstitute.org/ctrp/>), we found that 226 genes could be regulated by piperlongumine, and the coexpression features of the corresponding encoded proteins are shown in Figure 12A. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed on these 226 genes. The results suggested that piperlongumine could cause changes in the EGFR tyrosine kinase inhibitor resistance pathway ($P < 0.01$) (Figure 12B, 12C). To further validate the specific mechanism of piperlongumine, we treated SK-HEP1 and SMMC-7721 cells with different concentrations of piperlongumine and then performed western blot assays. As shown in Figure 12D, 12E, piperlongumine treatment reduced the levels of EGFR, p-Akt and p-Erk in a dose-dependent manner, suggesting that piperlongumine inhibited angiogenesis-induced

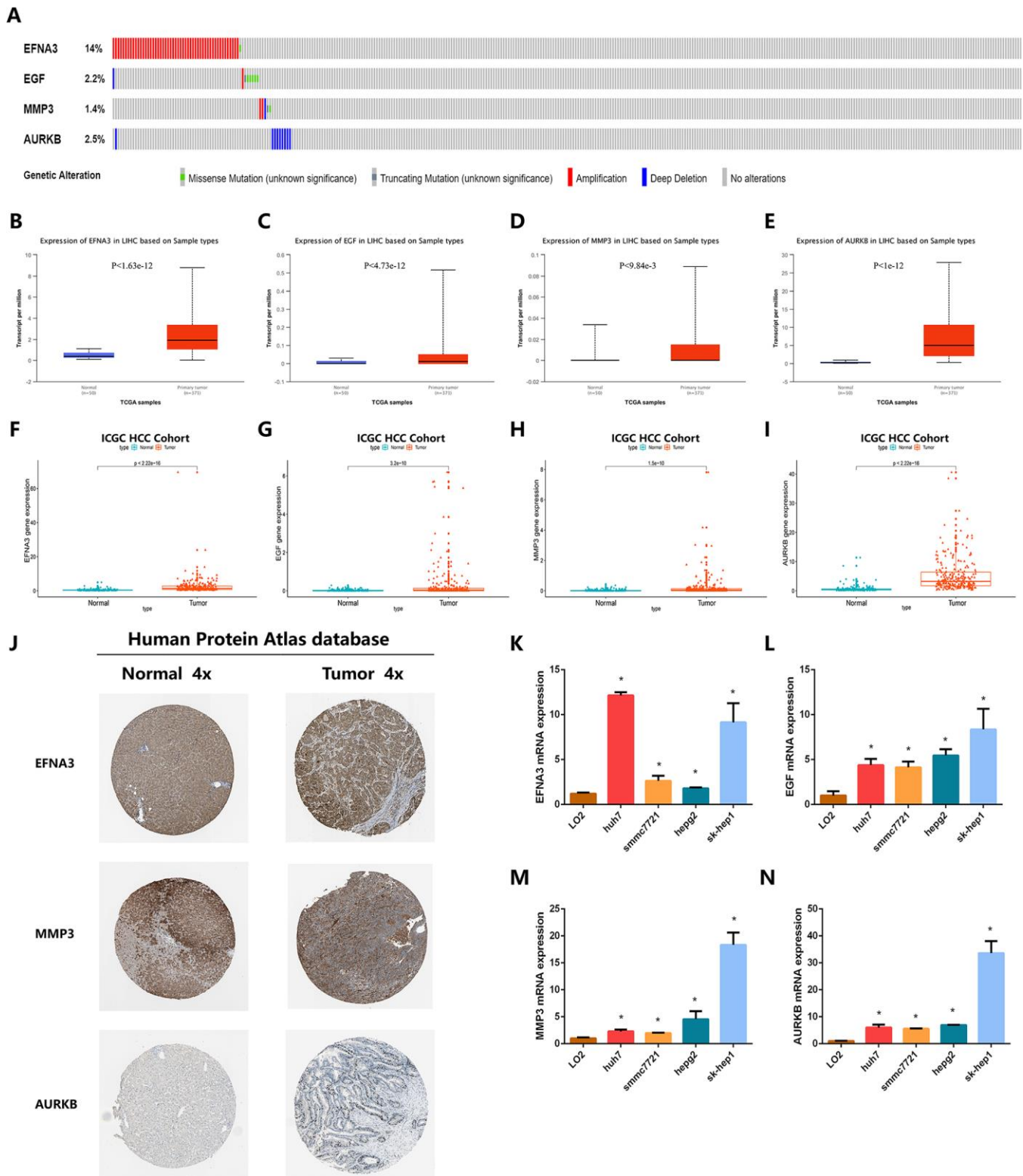


Figure 8. Validation of the expression characteristics of angiogenic genes. (A) Genetic alteration detection of the angiogenic genes in the TCGA HCC cohort. (B–E) Expression levels of angiogenic genes in HCC and normal samples in the TCGA HCC cohort. (F–I) Expression levels of angiogenic genes in HCC and normal samples in the TCGA HCC cohort. (J) The images showed the expression patterns of the angiogenic genes in HCC tissues and normal liver tissues. (K–N) The expression profiles of the angiogenic genes in multiple types of HCC cell lines.

tumorigenesis by targeting EGFR to influence the EGFR/EGF signaling pathway.

DISCUSSION

With the high recurrence rate and low survival rate after surgery, liver cancer is known to be one of the most deadly malignant tumors worldwide [23]. Many methods have been used to evaluate the prognosis of patients with HCC, such as the American Joint Committee on Cancer (AJCC) staging system, the Barcelona Clinic Liver Cancer (BCLC) classification, and the Cancer of the Liver Italian Program (CLIP) staging system [24], but these assessment criteria were based on clinical features and did not consider the critical role of complex molecular pathogenesis in the carcinogenicity and disease outcome of HCC; further, the prognostic predictive performance of these assessment methods was also unsatisfactory. Previous studies have focused on exploring and discovering certain molecular markers and have found that their

combinations have good predictive value in the prognosis and survival of HCC patients [25, 26], but molecular markers that have been identified that have good prognostic value are still rare.

Angiogenesis is required for tissue repair, remodeling in physiological processes such as reproduction, embryo development and wound healing, and it is a hallmark of disease states such as cancer, rheumatoid arthritis and atherosclerosis [27, 28]. As a highly angiogenic tumor, HCC growth and development depends on angiogenesis [29]. Angiogenesis in HCC tissue is primarily triggered by the release of overexpressed angiogenic factors from malignant cells and the tumor microenvironment [14]. The expression patterns of angiogenic factors are closely related to the poor prognosis of HCC patients and can be used as potential therapeutic targets [30]. However, the number of identified angiogenic factors that are associated with the prognosis of HCC patients is still low; further, most of the studies focus on the effects of individual genes on HCC disease outcomes,

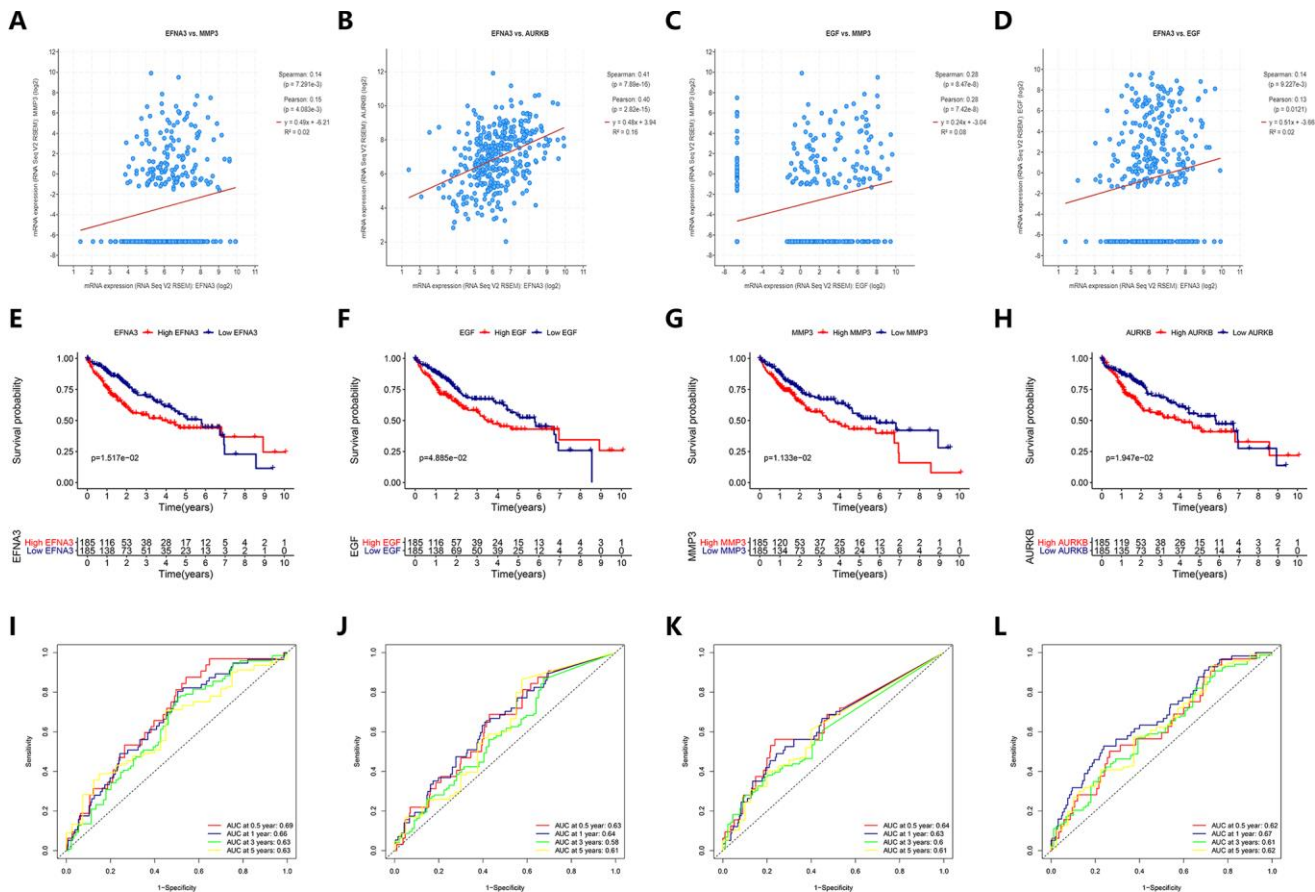


Figure 9. Prediction performance of angiogenic genes for OS and regression analysis between expression levels of different genes. (A–D) The expression of the angiogenic genes was synergistic with each other. **(E–H)** survival curves indicated the OS in the high-expression group was significantly poorer than that in the low-expression group ($P < 0.05$). **(I–L)** ROC curve analysis for OS prediction at 0.5, 1, 3 and 5 years.

and few studies have combined several angiogenic factors to explore the prognosis of HCC.

The widespread use of high-throughput arrays provided an opportunity for us to seek new angiogenic genes involved in the carcinogenicity of HCC and to perform combinatorial analysis [31]. In this study, we identified angiogenic genes that were abnormally expressed in HCC by analyzing HCC-related mRNA sequences in TCGA and ICGC databases, and then we identified expression characteristics of four angiogenic genes (ENFA3, EGF, MMP3 and AURKB) that were significantly associated with HCC prognosis using univariate Cox regression and LASSO Cox regression. Prognostic models, recurrence models, and corresponding nomograms of these four angiogenic genes were constructed by multivariate Cox proportional hazard regression analysis and were determined to have good prognostic/recurrence predictive value. The diagnostic model constructed with these four angiogenic genes was also confirmed to accurately distinguish

between HCC and normal/proliferative nodules. These four angiogenic genes are highly expressed in HCC tissues and work synergistically.

EGF is a growth factor that plays an important role in cell growth, proliferation and migration by binding to its receptor EGFR [32]. Previous immunohistochemistry data have confirmed that EGF is highly expressed in many human malignancies, such as breast cancer and ovarian cancer [33]. Animal model studies have suggested that targeted overexpression of epidermal growth factor (EGF) induces the formation of highly malignant HCC in mice, and its receptor EGFR is also overexpressed in HCC tissues [34]. Given that angiogenesis plays an important role in the development of HCC and that EGFR has been shown to be an angiogenesis promoting factor in previous studies [35], exploring the molecular effects of EGF/EGFR on HCC cells through angiogenesis can better characterize the progression of HCC and identify potential therapeutic targets. Matrix metalloproteinases (MMPs) are a class of

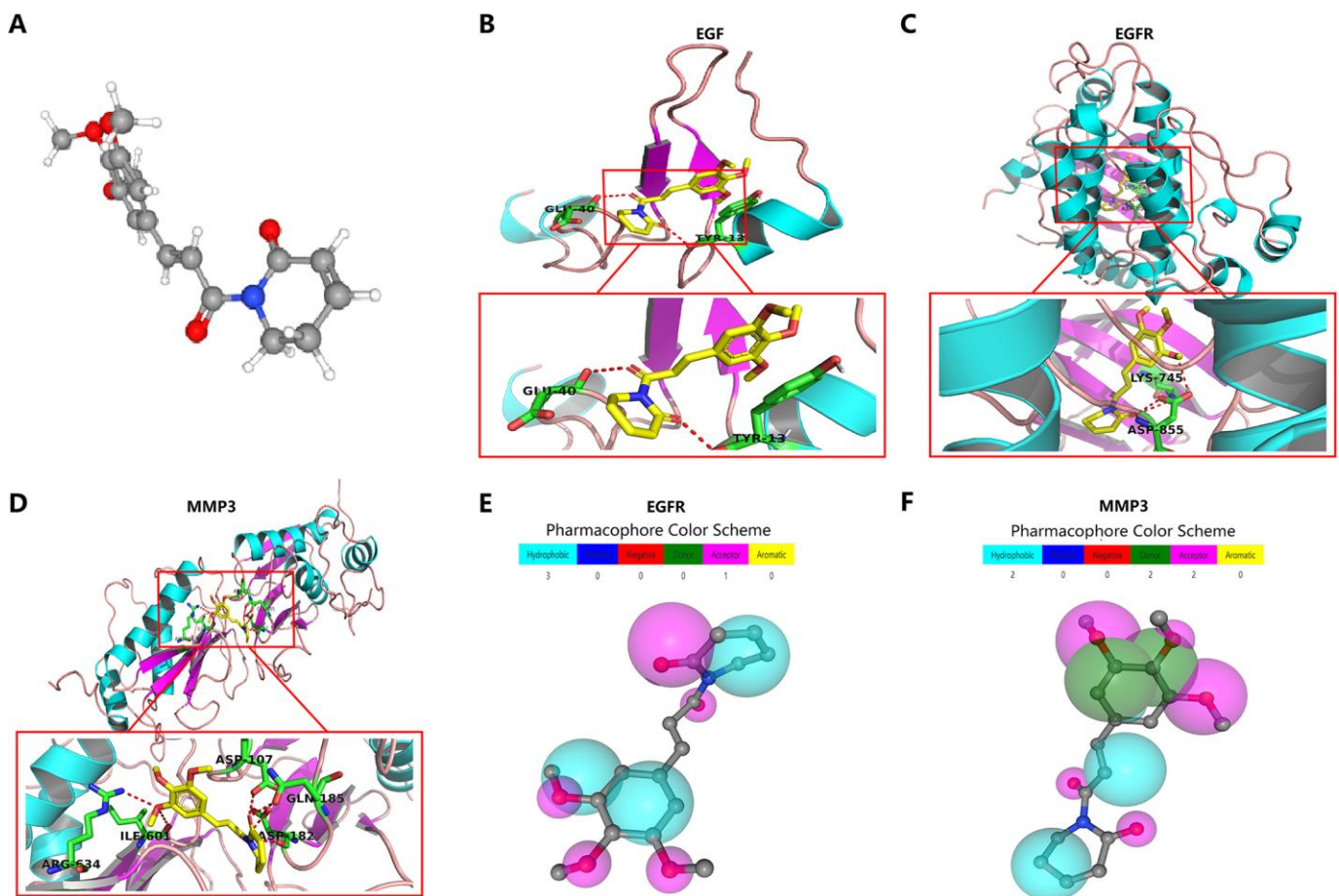


Figure 10. Molecular docking and pharmacophore models for targeted drug identification. (A) The 3D structure of piperlongumine was obtained from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov>). (B–D) Specific binding site of piperlongumine and EGF (B), EGFR (C) and MMP3 (D). (E, F) The pharmacophore model of EGFR (E) and MMP3 (F).

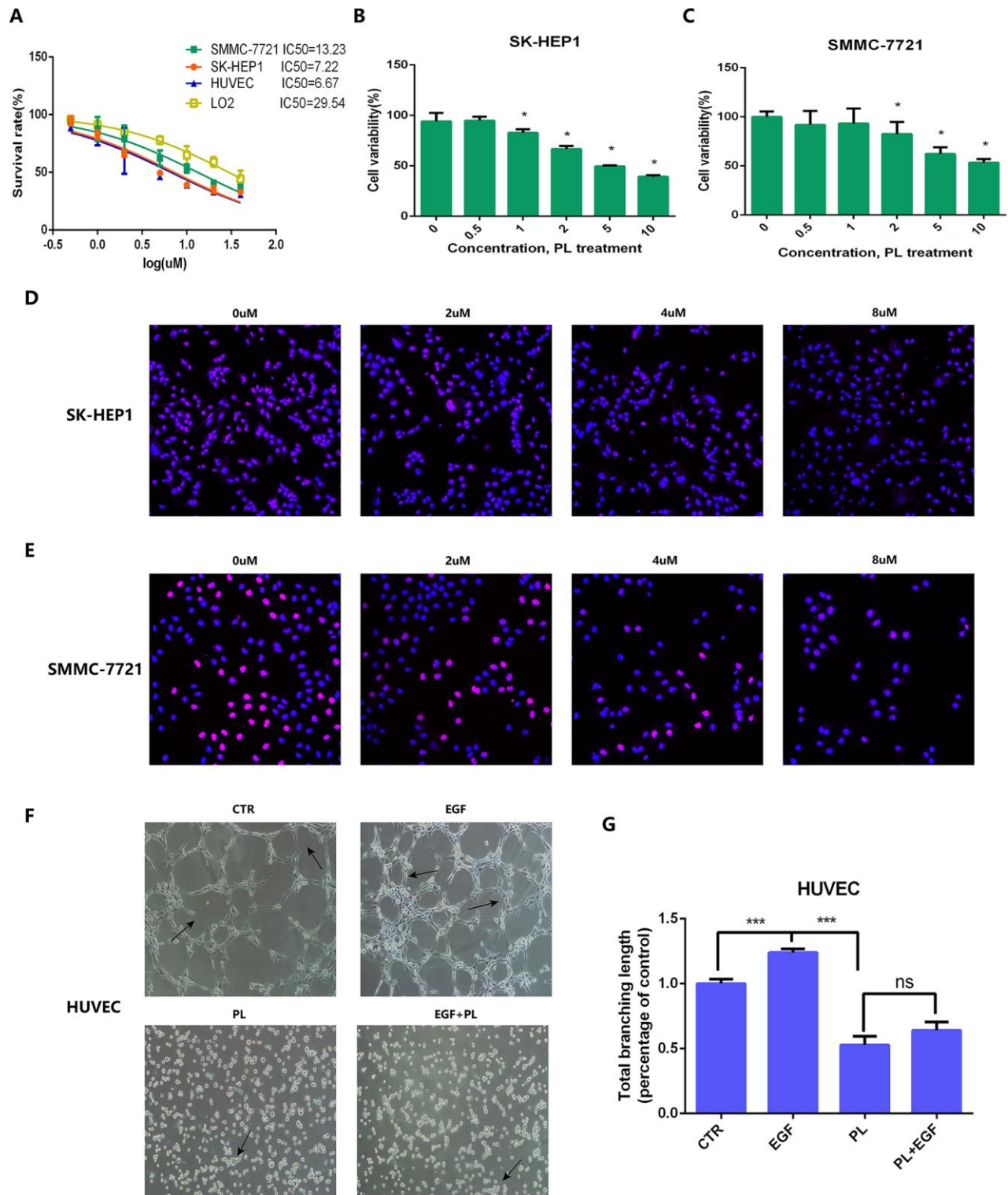


Figure 11. IC50, CCK8 assay, EdU assay and tube formation assay for anti-tumor effect validation of piperlongumine. (A) IC50 of piperlongumine in normal liver cell line (LO2), HUVEC and HCC cell lines. (B, C) CCK8 assay showed piperlongumine inhibited proliferation of SK-HEP1 (B) and SMMC-7721 (C) cell lines in a dose-dependent manner. (D, E) EdU assay showed the inhibition effect of piperlongumine in proliferation of SK-HEP1 (D) and SMMC-7721 (E) cell lines. (F) Tube formation assay suggested that piperlongumine inhibited angiogenesis via EGF/EGFR axis. (G) Statistical analysis to quantify the inhibitory effect of piperlongumine on angiogenesis.

lines, such as gastric cancer, lung cancer, colon cancer and cholangiocarcinoma, and in several animal cancer models [51, 52]. The anti-tumor effect of piperlongumine is achieved through inhibiting cell proliferation [53], promoting apoptosis [54] and inhibiting cell migration [55]. However, the anti-tumor effect of piperlongumine on HCC and its mechanism have not been specifically studied, and the effect of piperlongumine on angiogenesis and its pathways are still unclear. In this study, we performed in vitro cell experiments to confirm that piperlongumine did have an anti-tumor effect on HCC, which was achieved through the inhibition of EGF/EGFR signaling pathway-induced angiogenesis. Thus, our findings also have applicability in clinical practice, which could provide new choices for the treatment and management of HCC patients.

Inevitably, our current research does have some limitations. Firstly, this study used the LASSO method to screen for angiogenic genes related to the prognosis of HCC. In the process of adjusting the regression coefficients to filter the index weights, some important elements that contribute similarly may be ignored. Secondly, the independent verification of the prognosis and recurrence models and the establishment of the nomogram are mainly based on RNA sequencing data and corresponding clinical information from samples in the public data sets TCGA and ICGC. The main considerations are traditionally recognized clinical factors that have a significant impact on the carcinogenicity of HCC, some important clinical factors with similar contributions may be missed, such as endogenous or exogenous chemicals, lifestyle, eating style, and so on, and the difference in sample follow-up time will also affect the scientific prediction of the models. In addition, the study has not been prospectively tested in clinical trials. The predictive performance of prognostic and diagnostic features associated with angiogenesis and corresponding nomograms needs to be further validated in subsequent multicenter clinical trials and prospective studies. While the data used as the validation set of the prognosis model and the recurrence model were obtained from the GEO database, the sample size contained was not large enough, and it also has an impact on the specific verification of predictive performance. The potential mechanism of angiogenesis genes in the prognosis and recurrence of liver cancer remains to be explored.

CONCLUSIONS

In summary, we established and validated diagnostic, prognostic, and recurrence models of four angiogenic genes with good diagnostic and predictive value for HCC. The expression characteristics of these four angiogenic genes were significantly associated with the carcinogenicity and clinical outcome of HCC.

Pharmacological analysis of the inhibitor of EGF/EGFR, piperlongumine, had been shown to have an anti-tumor effect on HCC, which was achieved through inhibiting angiogenesis via the EGF/EGFR signaling pathway. Our findings will provide clinicians with new and effective ideas for the treatment and management of HCC in clinical practice. Additionally, these data can also provide new application ideas for other kinds of tumors.

MATERIALS AND METHODS

Angiogenic genes

A total of 132 angiogenesis-related genes were systematically obtained from relevant studies in the Pubmed, EMBASE, Google Scholar, Web of Science and The Cochrane Library by using the search terms: “angiogenesis” OR “angiogenic”. The relevant public papers were downloaded as of October 2019. The references were referred to additional information.

Acquisition of mRNA sequencing data associated with HCC

The mRNA-sequencing data and corresponding clinical information associated with HCC patients were downloaded from TCGA and ICGC databases. There were 370 HCC tissue samples and 50 normal tissue samples as of July 8, 2019, for a total of 19677 encoding mRNA sequences in TCGA HCC cohort, and 202 normal samples, 243 HCC samples and 22913 mRNA sequences in the ICGC HCC cohort. Since these mRNA-sequencing data were obtained from TCGA and ICGC databases and were freely available to the public, and as this study also strictly followed the publication guidelines and access policies of these databases, this study does not require ethical review and approval from an Ethics Committee.

Identification of differentially expressed genes (DEGs) between HCC and adjacent noncancerous tissues

A total of 132 angiogenic genes were matched with HCC-associated mRNA obtained from the RNA-sequencing platform in TCGA database and the ICGC database to identify angiogenic genes associated with HCC. Then, we determined the differentially expressed genes (DEGs) by limma, an R package, with an absolute log₂-fold change (FC) > 1 and an adjusted P value < 0.05.

Definition of an angiogenesis-related prognostic signature

The expression profiles of differentially expressed angiogenic genes were assessed via LASSO, univariate

and multivariate Cox regression analysis to evaluate the prognostic value of these genes for OS in HCC patients. We first implemented the least absolute shrinkage and selection operator (LASSO) method with an L1 penalty, which reduces the regression coefficient by applying a penalty proportional to its size, so that the number of indicators with a final weight of nonzero is relatively small and so that most potential indicators are reduced to zero [56], thereby reducing the number of angiogenic genes and screening for more relevant angiogenic genes. In the LASSO-penalized Cox regression analysis, we performed 1000 substitution samplings in the data set and selected marker genes with a recurrence frequency greater than 900. Subsequently, we adopted univariate and multivariate Cox regression analysis for these genes to determine key genes with the best prognostic performance. A P value for a gene <0.01 in analyses was considered to represent statistical significance, and the gene was viewed as a crucial prognostic factor in prognostic prediction. Then multivariate Cox regression analysis integrated these statistically significant genes, and the genes univariate Cox regression found no statistical significance but significant clinically relevant, and selected variables by forward selection and backward elimination. The optimal set of genes was obtained according to the test level of the selected variable $>$ the test level of the excluded variable. Finally, we integrated these key genes to establish an optimal apoptosis-related prognostic signature.

Validation analysis of the prognostic signature

Using a linear combination of regression coefficient (β) in a multivariate Cox regression model and angiogenic gene expression levels, we determined a prognostic risk score based on the angiogenesis-related prognostic signature. Prognostic index (PI) = (β^* expression level of EGF) + (β^* expression level of MMP3) + (β^* expression level of AURKB) + (β^* expression level of EFNA3). Using X-tile software, we confirmed the optimal cut-off value and thus divided 370 HCC patients into high-risk and low-risk groups. To assess the predictive performance of this angiogenesis-related prognostic signature, we established a Kaplan-Meier (K-M) survival curve and a time-dependent receptor working characteristic (ROC) curve.

Independence of the prognostic signature from other clinical variables

To assess the independence of this angiogenesis-related prognostic signature in predicting OS in HCC patients compared to other clinical features (age, gender, weight, AFP, vascular tumor cell infiltration, clinical stage and pathological grading), we performed univariate

and multivariate Cox regression analysis on these variables. A bilateral P value <0.05 was considered statistically significant. The hazard ratio (HR) and 95% confidence interval (CI) for each variable were calculated.

Construction and validation of a nomogram

Based on the multivariate Cox regression model, we determined independent prognostic factors associated with OS in HCC patients with HCC ($P < 0.05$) and used rms R software to integrate these independent prognostic factors and construct a predictive nomogram and a corresponding calibration map. The calibration map is validated by calibration and discrimination. In this study, we graphically plotted the probability of collinear prediction to the observed rate to evaluate the alignment curve of the nomogram. The closer the calibration curve was to the reference line, representing the best predictor (at the 45° line), the better the prognostic performance of this nomogram. We also adopted a consistency index (C index), which was calculated using a bootstrap method with 1000 resamples to validate the predictive accuracy between the nomogram and the actual result. The C index ranged from 0.5 to 1.0, which means that the ability to correctly distinguish between results and models ranged from random to perfect. ROC curve analyses were performed to validate and compare the predictive performance between single independent prognostic factors and the nomogram. Meanwhile, the decision curve analysis (DCA) was performed to evaluate the clinical benefit of the nomogram compared to a single prognostic factor. DCA can quantify the clinical applicability of a nomogram by analyzing the clinical outcomes of nomogram-based decisions, and it has great value in determining alternative diagnostic and prognostic strategies [57]. A P value <0.05 was considered to represent statistical significance.

Internal and external validation of gene expression levels of HCC-associated angiogenic genes

Further internal and external validation of the expression patterns of the four angiogenic genes were performed in the cohort from TCGA and the ICGC cohort. Using Wilcoxon signed rank tests in Prism 7.0 (GraphPad, San Diego, CA, USA), we compared the expression profiles of the four angiogenic genes between HCC and normal tissues. A bilateral P value <0.05 was considered to represent statistical significance. In the Human Protein Atlas database, we identified the expression patterns of the four genes using images of HCC and normal tissues. We also explored the extractions among these genes. Finally, the predictive value of each gene for OS in HCC patients was validated using ROC curve analysis.

Cell culture

Human HCC cell lines (SK-HEP1 and SMMC-7721) were purchased from American Type Culture Collection (ATCC) (Manassas, VA, USA). The cell lines were cultured in DMEM (Gibco, Eggenstein, Germany) supplemented with 10% heat-inactivated fetal bovine serum (Gibco, Eggenstein, Germany), and the cell lines were grown in a humidified cell incubator with an atmosphere of 5% CO₂ at 37°C.

Cell counting kit-8 (CCK8) assay

SK-HEP1 and SMMC-7721 cells (3000 cells/well) were plated in 96-well plates (n=5/group) and were incubated in a humidified cell incubator with an atmosphere of 5% CO₂ at 37°C for 24 hours. Cells were treated with piperlongumine (0-10 µmol/ml) for 24, 48 and 72 hours, and CCK8 reagent was added to the wells, and the cells were incubated for an additional 2 hours. The OD was then measured at 450 nm using a microplate reader.

5-ethynyl-2'-deoxyuridine (EdU) cell proliferation assay

SK-HEP1 and SMMC-7721 cells (3000 cells/well) were plated and incubated in complete medium in a 96-well plate (n=5/group) with an atmosphere of 5% CO₂ at 37°C overnight. Then, the cells were treated with piperlongumine (0-8 µg/ml) or EGF (200 ng/ml) and incubated with 50 µM EdU for 6 hours. SK-HEP1 and SMMC-7721 cells were then fixed, permeabilized and EdU stained according to the manufacturer's protocol. Proliferative cells were labeled with a Click-iTTM EdU imaging kit (Ruibo, Guangzhou, China). Cell nuclei were stained with Hoechst 33342 (1 µg/ml) for 10 minutes. The proportion of cells that incorporated EdU was detected by fluorescence microscopy.

In vitro HUVEC tube formation assay

Culture plates (96-well) were coated with ECMatrix™ according to the manufacturer's instructions for 2 hours. HUVECs in endothelial cell growth medium (EGM, CC-3162, Lonza) treated with piperlongumine solutions (2 µmol/ml) were added to the precoated plate at a density of 1×10⁴ cells/ml. The cultures were incubated in a 37°C humidified incubator for 12 hours, and then an inverted light microscope was used to photograph in vitro angiogenesis at the center of the well.

Western blot analysis

SK-HEP1 and SMMC-7721 cells were treated with an inhibitor and lysed with protein lysis buffer. Then, the

debris was removed by centrifugation at 12,000 rpm for 10 min at 4°C. Protein samples were separated using 10% gels via sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and were then transferred to PVDF membranes. Freshly prepared 5% nonfat milk in TBST was used to block the blots for 1.5 h at room temperature on a shaker. The blots were incubated overnight with a specific primary antibody on a shaker at 4°C to probe the blot, and this was followed by incubation with a horseradish peroxidase-conjugated secondary antibody and treatment with an ECL kit (Bio-Rad) for protein detection. The antibodies used in the study were list in supporting information.

Quantitative real-time PCR (qPCR)

Total RNA was extracted from cultured SK-HEP1 cells, SMMC-7721 cells and HUVECs treated with piperlongumine using TRIzol reagent (Invitrogen, Carlsbad, USA) according to the manufacturer's instructions. Then, total RNA was reverse transcribed using a cDNA reverse transcription kit (TransGen, Guangzhou, China), and the obtained cDNA was amplified using a SYBR Green PCR kit (TransGen, Guangzhou, China). Quantitative real-time PCR analysis was performed to detect expression in samples. Each experiment was repeated at least three times. The primers used for qRT-PCR were obtained from RiboBio Co., Ltd. The expression of genes was analyzed using 2- $\Delta\Delta$ CT methodology. The specific content of the relevant primers used in qPCR can be found in the support information.

Statistical analysis

All data are presented as the mean ± standard deviation (SD), and each data point represents at least five samples. The differences between two groups were calculated using Student's two-sided t-tests. P<0.05 was considered to represent a statistical significance.

AUTHOR CONTRIBUTIONS

Jiansong Ji and Min Xu conceived and designed the experiments. Jinyu Zhu, Bufu Tang and Yueli Shi performed the experiments. Jie Li, Minjiang Chen, Xiuling Lv, Miaomiao Meng, Qiaoyou Weng and Nannan Zhang analyzed the data. Jie Li and Kai Fan contributed analysis tools. Jinyu Zhu and Bufu Tang wrote the paper. Jiansong Ji and Min Xu edited the paper.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING

This study was supported by, National Natural Science Foundation of China (Nos. 81803778), and The Key Research and Development Project of Zhejiang Province (No. 2018C0302), and Public Welfare Project of Zhejiang Province (Nos. LGF19H180010, LGD19H160002 and LGF19H180009), and Medical and Health Care Key Project of Zhejiang Province (Nos. 2016146810, 2018KY197 and 2018KY932).

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018; 68:394–424.
<https://doi.org/10.3322/caac.21492>
PMID:[30207593](https://pubmed.ncbi.nlm.nih.gov/30207593/)
2. Llovet JM, Bruix J. Systematic review of randomized trials for unresectable hepatocellular carcinoma: chemoembolization improves survival. *Hepatology.* 2003; 37:429–42.
<https://doi.org/10.1053/jhep.2003.50047>
PMID:[12540794](https://pubmed.ncbi.nlm.nih.gov/12540794/)
3. Wu KT, Wang CC, Lu LG, Zhang WD, Zhang FJ, Shi F, Li CX. Hepatocellular carcinoma: clinical study of long-term survival and choice of treatment modalities. *World J Gastroenterol.* 2013; 19:3649–57.
<https://doi.org/10.3748/wjg.v19.i23.3649>
PMID:[23801868](https://pubmed.ncbi.nlm.nih.gov/23801868/)
4. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet.* 2018; 391:1301–14.
[https://doi.org/10.1016/S0140-6736\(18\)30010-2](https://doi.org/10.1016/S0140-6736(18)30010-2)
PMID:[29307467](https://pubmed.ncbi.nlm.nih.gov/29307467/)
5. Katanoda K, Matsuda T. Five-year relative survival rate of liver cancer in the USA, Europe and Japan. *Jpn J Clin Oncol.* 2014; 44:302–03.
<https://doi.org/10.1093/jjco/hyu025>
PMID:[24596423](https://pubmed.ncbi.nlm.nih.gov/24596423/)
6. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology.* 2007; 132:2557–76.
<https://doi.org/10.1053/j.gastro.2007.04.061>
PMID:[17570226](https://pubmed.ncbi.nlm.nih.gov/17570226/)
7. Staton CA, Reed MW, Brown NJ. A critical analysis of current in vitro and in vivo angiogenesis assays. *Int J Exp Pathol.* 2009; 90:195–221.
<https://doi.org/10.1111/j.1365-2613.2008.00633.x>
PMID:[19563606](https://pubmed.ncbi.nlm.nih.gov/19563606/)
8. Rajabi M, Mousa SA. The Role of Angiogenesis in Cancer Treatment. *Biomedicines.* 2017; 5:5.
<https://doi.org/10.3390/biomedicines5020034>
PMID:[28635679](https://pubmed.ncbi.nlm.nih.gov/28635679/)
9. Carmeliet P, Jain RK. Angiogenesis in cancer and other diseases. *Nature.* 2000; 407:249–57.
<https://doi.org/10.1038/35025220>
PMID:[11001068](https://pubmed.ncbi.nlm.nih.gov/11001068/)
10. Folkman J. What is the evidence that tumors are angiogenesis dependent? *J Natl Cancer Inst.* 1990; 82:4–6.
<https://doi.org/10.1093/jnci/82.1.4>
PMID:[1688381](https://pubmed.ncbi.nlm.nih.gov/1688381/)
11. Ramjiawan RR, Griffioen AW, Duda DG. Anti-angiogenesis for cancer revisited: is there a role for combinations with immunotherapy? *Angiogenesis.* 2017; 20:185–204.
<https://doi.org/10.1007/s10456-017-9552-y>
PMID:[28361267](https://pubmed.ncbi.nlm.nih.gov/28361267/)
12. Ribatti D, Vacca A, Nico B, Sansonno D, Dammacco F. Angiogenesis and anti-angiogenesis in hepatocellular carcinoma. *Cancer Treat Rev.* 2006; 32:437–44.
<https://doi.org/10.1016/j.ctrv.2006.06.002>
PMID:[16870349](https://pubmed.ncbi.nlm.nih.gov/16870349/)
13. Pang R, Poon RT. Angiogenesis and antiangiogenic therapy in hepatocellular carcinoma. *Cancer Lett.* 2006; 242:151–67.
<https://doi.org/10.1016/j.canlet.2006.01.008>
PMID:[16564617](https://pubmed.ncbi.nlm.nih.gov/16564617/)
14. Bishayee A, Darvesh AS. Angiogenesis in hepatocellular carcinoma: a potential target for chemoprevention and therapy. *Curr Cancer Drug Targets.* 2012; 12:1095–118.
PMID:[22873221](https://pubmed.ncbi.nlm.nih.gov/22873221/)
15. Chen JA, Shi M, Li JQ, Qian CN. Angiogenesis: multiple masks in hepatocellular carcinoma and liver regeneration. *Hepato Int.* 2010; 4:537–47.
<https://doi.org/10.1007/s12072-010-9192-4>
PMID:[21063476](https://pubmed.ncbi.nlm.nih.gov/21063476/)
16. Poon RT, Fan ST, Wong J. Clinical implications of circulating angiogenic factors in cancer patients. *J Clin Oncol.* 2001; 19:1207–25.
<https://doi.org/10.1200/JCO.2001.19.4.1207>
PMID:[11181687](https://pubmed.ncbi.nlm.nih.gov/11181687/)
17. Park YN, Kim YB, Yang KM, Park C. Increased expression of vascular endothelial growth factor and angiogenesis in the early stage of multistep hepatocarcinogenesis. *Arch Pathol Lab Med.* 2000; 124:1061–65.
[https://doi.org/10.1043/0003-9985\(2000\)124<1061:IEVEG>2.0.CO;2](https://doi.org/10.1043/0003-9985(2000)124<1061:IEVEG>2.0.CO;2)
PMID:[10888784](https://pubmed.ncbi.nlm.nih.gov/10888784/)
18. Llovet JM, Peña CE, Lathia CD, Shan M, Meinhardt G, Bruix J, SHARP Investigators Study Group. Plasma biomarkers as predictors of outcome in patients with

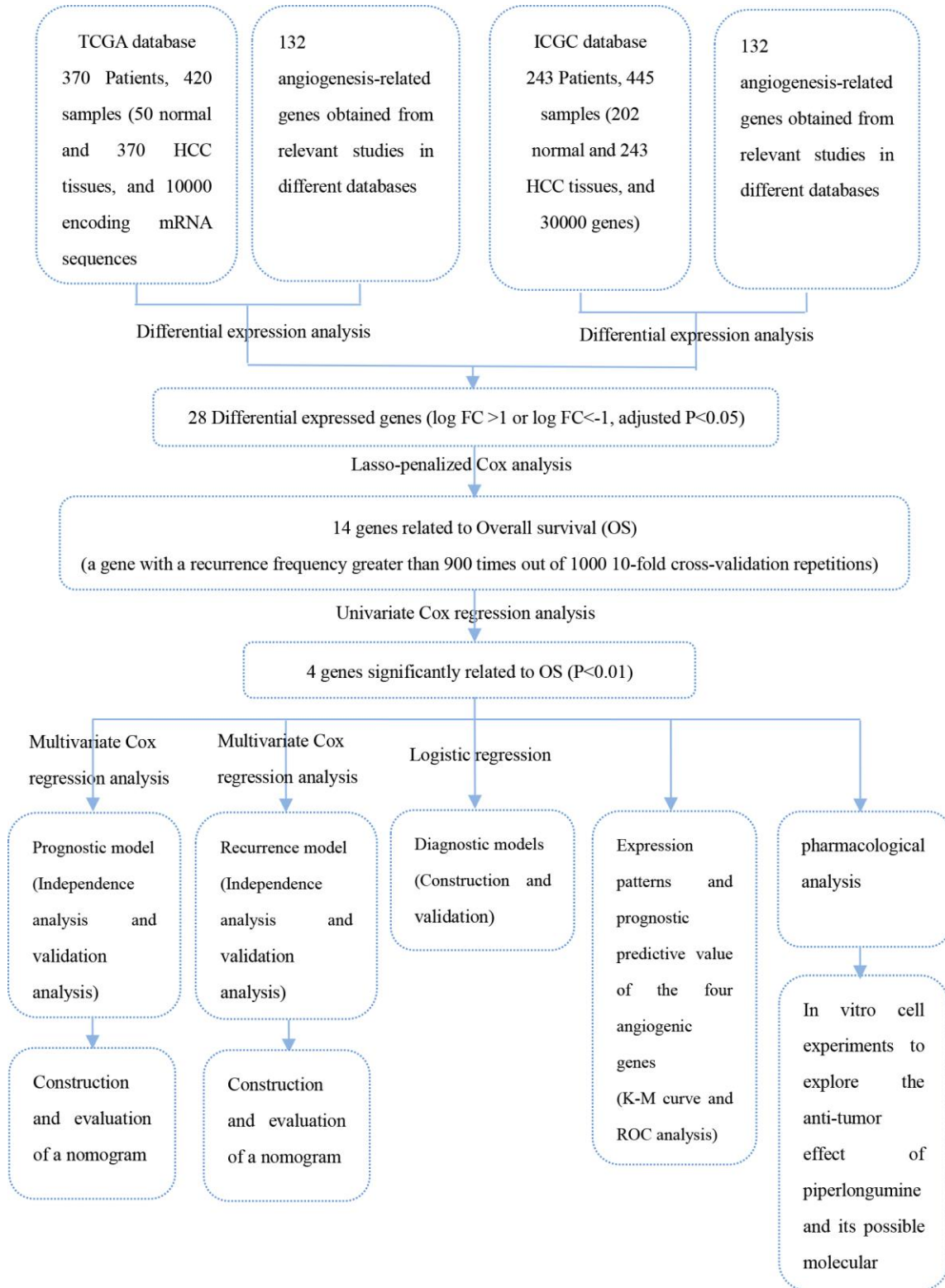
- advanced hepatocellular carcinoma. *Clin Cancer Res*. 2012; 18:2290–300.
<https://doi.org/10.1158/1078-0432.CCR-11-2175>
PMID:[22374331](https://pubmed.ncbi.nlm.nih.gov/22374331/)
19. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012; 2:401–04.
<https://doi.org/10.1158/2159-8290.CD-12-0095>
PMID:[22588877](https://pubmed.ncbi.nlm.nih.gov/22588877/)
20. Chandrashekar DS, Bashel B, Balasubramanya SA, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BV, Varambally S. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia*. 2017; 19:649–58.
<https://doi.org/10.1016/j.neo.2017.05.002>
PMID:[28732212](https://pubmed.ncbi.nlm.nih.gov/28732212/)
21. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015; 347:1260419.
<https://doi.org/10.1126/science.1260419>
PMID:[25613900](https://pubmed.ncbi.nlm.nih.gov/25613900/)
22. Piska K, Gunia-Krzyżak A, Koczurkiewicz P, Wójcik-Pszczola K, Pękala E. Piperlongumine (piplartine) as a lead compound for anticancer agents - Synthesis and properties of analogues: A mini-review. *Eur J Med Chem*. 2018; 156:13–20.
<https://doi.org/10.1016/j.ejmech.2018.06.057>
PMID:[30006159](https://pubmed.ncbi.nlm.nih.gov/30006159/)
23. Gish RG. Hepatocellular carcinoma: overcoming challenges in disease management. *Clin Gastroenterol Hepatol*. 2006; 4:252–61.
<https://doi.org/10.1016/j.cgh.2006.01.001>
PMID:[16527686](https://pubmed.ncbi.nlm.nih.gov/16527686/)
24. Liu PH, Hsu CY, Hsia CY, Lee YH, Su CW, Huang YH, Lee FY, Lin HC, Huo TI. Prognosis of hepatocellular carcinoma: assessment of eleven staging systems. *J Hepatol*. 2016; 64:601–08.
<https://doi.org/10.1016/j.jhep.2015.10.029>
PMID:[26551516](https://pubmed.ncbi.nlm.nih.gov/26551516/)
25. Nault JC, De Reyniès A, Villanueva A, Calderaro J, Rebouissou S, Couchy G, Decaens T, Franco D, Imbeaud S, Rousseau F, Azoulay D, Saric J, Blanc JF, et al. A hepatocellular carcinoma 5-gene score associated with survival of patients after liver resection. *Gastroenterology*. 2013; 145:176–87.
<https://doi.org/10.1053/j.gastro.2013.03.051>
PMID:[23567350](https://pubmed.ncbi.nlm.nih.gov/23567350/)
26. Qin LX, Tang ZY. The prognostic molecular markers in hepatocellular carcinoma. *World J Gastroenterol*. 2002; 8:385–92.
<https://doi.org/10.3748/wjg.v8.i3.385>
PMID:[12046056](https://pubmed.ncbi.nlm.nih.gov/12046056/)
27. Carmeliet P, Jain RK. Principles and mechanisms of vessel normalization for cancer and other angiogenic diseases. *Nat Rev Drug Discov*. 2011; 10:417–27.
<https://doi.org/10.1038/nrd3455> PMID:[21629292](https://pubmed.ncbi.nlm.nih.gov/21629292/)
28. Fox SB, Gasparini G, Harris AL. Angiogenesis: pathological, prognostic, and growth-factor pathways and their link to trial design and anticancer drugs. *Lancet Oncol*. 2001; 2:278–89.
[https://doi.org/10.1016/S1470-2045\(00\)00323-5](https://doi.org/10.1016/S1470-2045(00)00323-5)
PMID:[11905783](https://pubmed.ncbi.nlm.nih.gov/11905783/)
29. Sun HC, Tang ZY. Angiogenesis in hepatocellular carcinoma: the retrospectives and perspectives. *J Cancer Res Clin Oncol*. 2004; 130:307–19.
<https://doi.org/10.1007/s00432-003-0530-y>
PMID:[15034787](https://pubmed.ncbi.nlm.nih.gov/15034787/)
30. Wada H, Nagano H, Yamamoto H, Yang Y, Kondo M, Ota H, Nakamura M, Yoshioka S, Kato H, Damdinsuren B, Tang D, Marubashi S, Miyamoto A, et al. Expression pattern of angiogenic factors and prognosis after hepatic resection in hepatocellular carcinoma: importance of angiopoietin-2 and hypoxia-induced factor-1 alpha. *Liver Int*. 2006; 26:414–23.
<https://doi.org/10.1111/j.1478-3231.2006.01243.x>
PMID:[16629644](https://pubmed.ncbi.nlm.nih.gov/16629644/)
31. Nakamura M, Chiba T, Kanayama K, Kanzaki H, Saito T, Kusakabe Y, Kato N. Epigenetic dysregulation in hepatocellular carcinoma: an up-to-date review. *Hepatol Res*. 2019; 49:3–13.
<https://doi.org/10.1111/hepr.13250>
PMID:[30238570](https://pubmed.ncbi.nlm.nih.gov/30238570/)
32. Herbst RS. Review of epidermal growth factor receptor biology. *Int J Radiat Oncol Biol Phys*. 2004 (Suppl); 59:21–26.
<https://doi.org/10.1016/j.ijrobp.2003.11.041>
PMID:[15142631](https://pubmed.ncbi.nlm.nih.gov/15142631/)
33. Salomon DS, Brandt R, Ciardiello F, Normanno N. Epidermal growth factor-related peptides and their receptors in human malignancies. *Crit Rev Oncol Hematol*. 1995; 19:183–232.
[https://doi.org/10.1016/1040-8428\(94\)00144-1](https://doi.org/10.1016/1040-8428(94)00144-1)
PMID:[7612182](https://pubmed.ncbi.nlm.nih.gov/7612182/)
34. Liu Z, Chen D, Ning F, Du J, Wang H. EGF is highly expressed in hepatocellular carcinoma (HCC) and promotes motility of HCC cells via fibronectin. *J Cell Biochem*. 2018; 119:4170–83.
<https://doi.org/10.1002/jcb.26625>
PMID:[29315755](https://pubmed.ncbi.nlm.nih.gov/29315755/)

35. Keller S, Schmidt MH. EGFR and EGFRvIII Promote Angiogenesis and Cell Invasion in Glioblastoma: Combination Therapies for an Effective Treatment. *Int J Mol Sci.* 2017; 18:18.
<https://doi.org/10.3390/ijms18061295>
PMID:28629170
36. McCawley LJ, Wright J, LaFleur BJ, Crawford HC, Matrisian LM. Keratinocyte expression of MMP3 enhances differentiation and prevents tumor establishment. *Am J Pathol.* 2008; 173:1528–39.
<https://doi.org/10.2353/ajpath.2008.080132>
PMID:18832569
37. Mehner C, Miller E, Khauv D, Nassar A, Oberg AL, Bamlet WR, Zhang L, Waldmann J, Radisky ES, Crawford HC, Radisky DC. Tumor cell-derived MMP3 orchestrates Rac1b and tissue alterations that promote pancreatic adenocarcinoma. *Mol Cancer Res.* 2014; 12:1430–39.
<https://doi.org/10.1158/1541-7786.MCR-13-0557-T>
PMID:24850902
38. Padala C, Tupurani MA, Puranam K, Gantala S, Shyamala N, Kondapalli MS, Gundapaneni KK, Mudigonda S, Galimudi RK, Kupsal K, Nanchari SR, Chavan U, Chinta SK, et al. Synergistic effect of collagenase-1 (MMP1), stromelysin-1 (MMP3) and gelatinase-B (MMP9) gene polymorphisms in breast cancer. *PLoS One.* 2017; 12:e0184448.
<https://doi.org/10.1371/journal.pone.0184448>
PMID:28961241
39. Tsukifuji R, Tagawa K, Hatamochi A, Shinkai H. Expression of matrix metalloproteinase-1, -2 and -3 in squamous cell carcinoma and actinic keratosis. *Br J Cancer.* 1999; 80:1087–91.
<https://doi.org/10.1038/sj.bjc.6690468> PMID:10362121
40. Zheng X, Chopp M, Lu Y, Buller B, Jiang F. MiR-15b and miR-152 reduce glioma cell invasion and angiogenesis via NRP-2 and MMP-3. *Cancer Lett.* 2013; 329:146–54.
<https://doi.org/10.1016/j.canlet.2012.10.026>
PMID:23142217
41. Zhu Q, Ding L, Zi Z, Gao S, Wang C, Wang Y, Zhu C, Yuan Z, Wei F, Cai Q. Viral-Mediated AURKB Cleavage Promotes Cell Segregation and Tumorigenesis. *Cell Rep.* 2019; 26:3657–3671.e5.
<https://doi.org/10.1016/j.celrep.2019.02.106>
PMID:30917319
42. Sorrentino R, Libertini S, Pallante PL, Troncone G, Palombini L, Bavetsias V, Spalletti-Cernia D, Laccetti P, Linardopoulos S, Chieffi P, Fusco A, Portella G. Aurora B overexpression associates with the thyroid carcinoma undifferentiated phenotype and is required for thyroid carcinoma cell proliferation. *J Clin Endocrinol Metab.* 2005; 90:928–35.
<https://doi.org/10.1210/jc.2004-1518>
PMID:15562011
43. Wan B, Huang Y, Liu B, Lu L, Lv C. *AURKB*: a promising biomarker in clear cell renal cell carcinoma. *PeerJ.* 2019; 7:e7718.
<https://doi.org/10.7717/peerj.7718>
PMID:31576249
44. Wang Z, Liu Z, Liu B, Liu G, Wu S. Dissecting the roles of Ephrin-A3 in malignant peripheral nerve sheath tumor by TALENs. *Oncol Rep.* 2015; 34:391–98.
<https://doi.org/10.3892/or.2015.3966>
PMID:25955218
45. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011; 144:646–74.
<https://doi.org/10.1016/j.cell.2011.02.013>
PMID:21376230
46. Klutstein M, Nejman D, Greenfield R, Cedar H. DNA Methylation in Cancer and Aging. *Cancer Res.* 2016; 76:3446–50.
<https://doi.org/10.1158/0008-5472.CAN-15-3278>
PMID:27256564
47. Kou Y, Koag MC, Lee S. N7 methylation alters hydrogen-bonding patterns of guanine in duplex DNA. *J Am Chem Soc.* 2015; 137:14067–70.
<https://doi.org/10.1021/jacs.5b10172>
PMID:26517568
48. Hu C, Li W, Tian F, Jiang K, Liu X, Cen J, He Q, Qiu Z, Kienast Y, Wang Z, Zhang H, Ji Y, Hu J, Hui L. Arid1a regulates response to anti-angiogenic therapy in advanced hepatocellular carcinoma. *J Hepatol.* 2018; 68:465–75.
<https://doi.org/10.1016/j.jhep.2017.10.028>
PMID:29113912
49. Kou Y, Koag MC, Lee S. Promutagenicity of 8-Chloroguanine, A Major Inflammation-Induced Halogenated DNA Lesion. *Molecules.* 2019; 24:3507.
<https://doi.org/10.3390/molecules24193507>
PMID:31569643
50. Koag MC, Kou Y, Ouzon-Shubeita H, Lee S. Transition-state destabilization reveals how human DNA polymerase β proceeds across the chemically unstable lesion N7-methylguanine. *Nucleic Acids Res.* 2014; 42:8755–66.
<https://doi.org/10.1093/nar/gku554>
PMID:24966350
51. Duan C, Zhang B, Deng C, Cao Y, Zhou F, Wu L, Chen M, Shen S, Xu G, Zhang S, Duan G, Yan H, Zou X. Piperlongumine induces gastric cancer cell apoptosis and G2/M cell cycle arrest both in vitro and in vivo. *Tumour Biol.* 2016; 37:10793–804.
<https://doi.org/10.1007/s13277-016-4792-9>
PMID:26874726

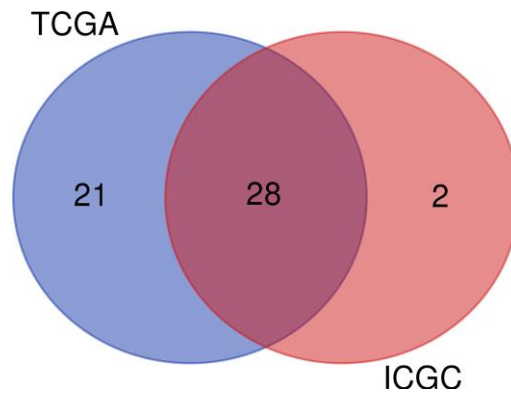
52. Wang F, Mao Y, You Q, Hua D, Cai D. Piperlongumine induces apoptosis and autophagy in human lung cancer cells through inhibition of PI3K/Akt/mTOR pathway. *Int J Immunopathol Pharmacol*. 2015; 28:362–73.
<https://doi.org/10.1177/0394632015598849>
PMID:[26246196](https://pubmed.ncbi.nlm.nih.gov/26246196/)
53. Chen SY, Liu GH, Chao WY, Shi CS, Lin CY, Lim YP, Lu CH, Lai PY, Chen HR, Lee YR. Piperlongumine Suppresses Proliferation of Human Oral Squamous Cell Carcinoma through Cell Cycle Arrest, Apoptosis and Senescence. *Int J Mol Sci*. 2016; 17:17.
<https://doi.org/10.3390/ijms17040616>
PMID:[27120594](https://pubmed.ncbi.nlm.nih.gov/27120594/)
54. Gong LH, Chen XX, Wang H, Jiang QW, Pan SS, Qiu JG, Mei XL, Xue YQ, Qin WM, Zheng FY, Shi Z, Yan XJ. Piperlongumine induces apoptosis and synergizes with cisplatin or paclitaxel in human ovarian cancer cells. *Oxid Med Cell Longev*. 2014; 2014:906804.
<https://doi.org/10.1155/2014/906804>
PMID:[24895529](https://pubmed.ncbi.nlm.nih.gov/24895529/)
55. Zou Y, Zhao D, Yan C, Ji Y, Liu J, Xu J, Lai Y, Tian J, Zhang Y, Huang Z. Novel Ligustrazine-Based Analogs of Piperlongumine Potently Suppress Proliferation and Metastasis of Colorectal Cancer Cells in Vitro and in Vivo. *J Med Chem*. 2018; 61:1821–32.
<https://doi.org/10.1021/acs.jmedchem.7b01096>
PMID:[29424539](https://pubmed.ncbi.nlm.nih.gov/29424539/)
56. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. 1997; 16:385–95.
[https://doi.org/10.1002/\(SICI\)1097-0258\(19970228\)16:4<385::AID-SIM380>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3)
PMID:[9044528](https://pubmed.ncbi.nlm.nih.gov/9044528/)
57. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006; 26:565–74.
<https://doi.org/10.1177/0272989X06295361>
PMID:[17099194](https://pubmed.ncbi.nlm.nih.gov/17099194/)

SUPPLEMENTARY MATERIALS

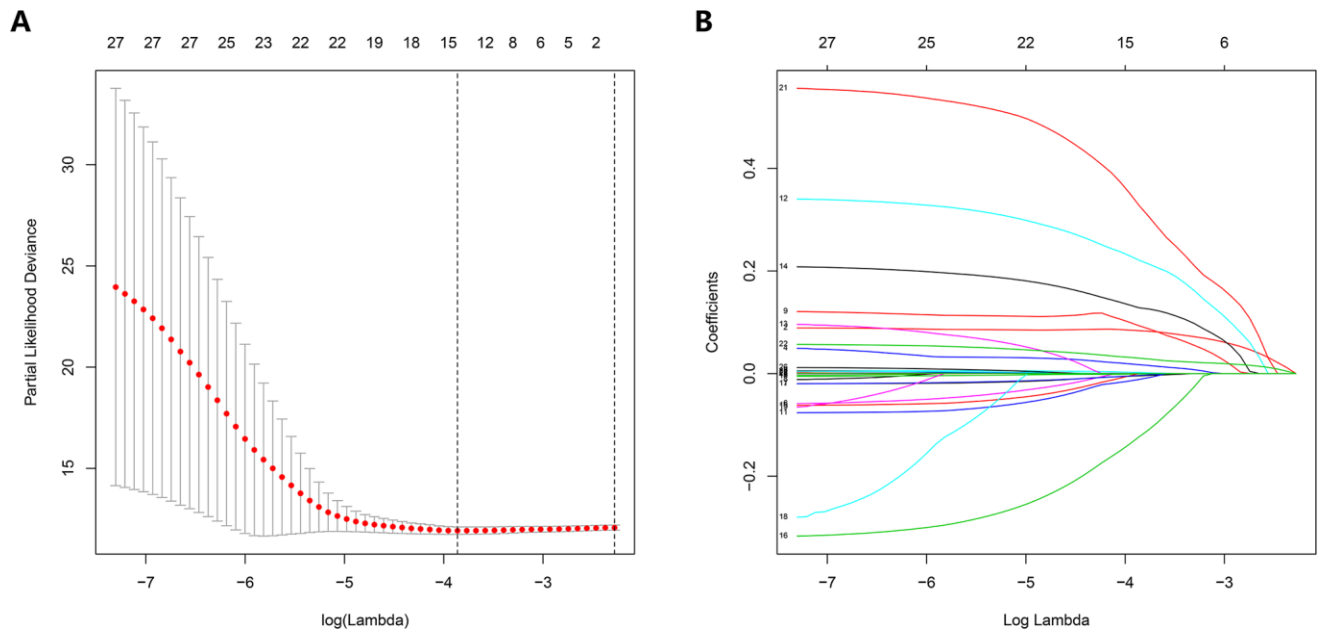
Supplementary Figures



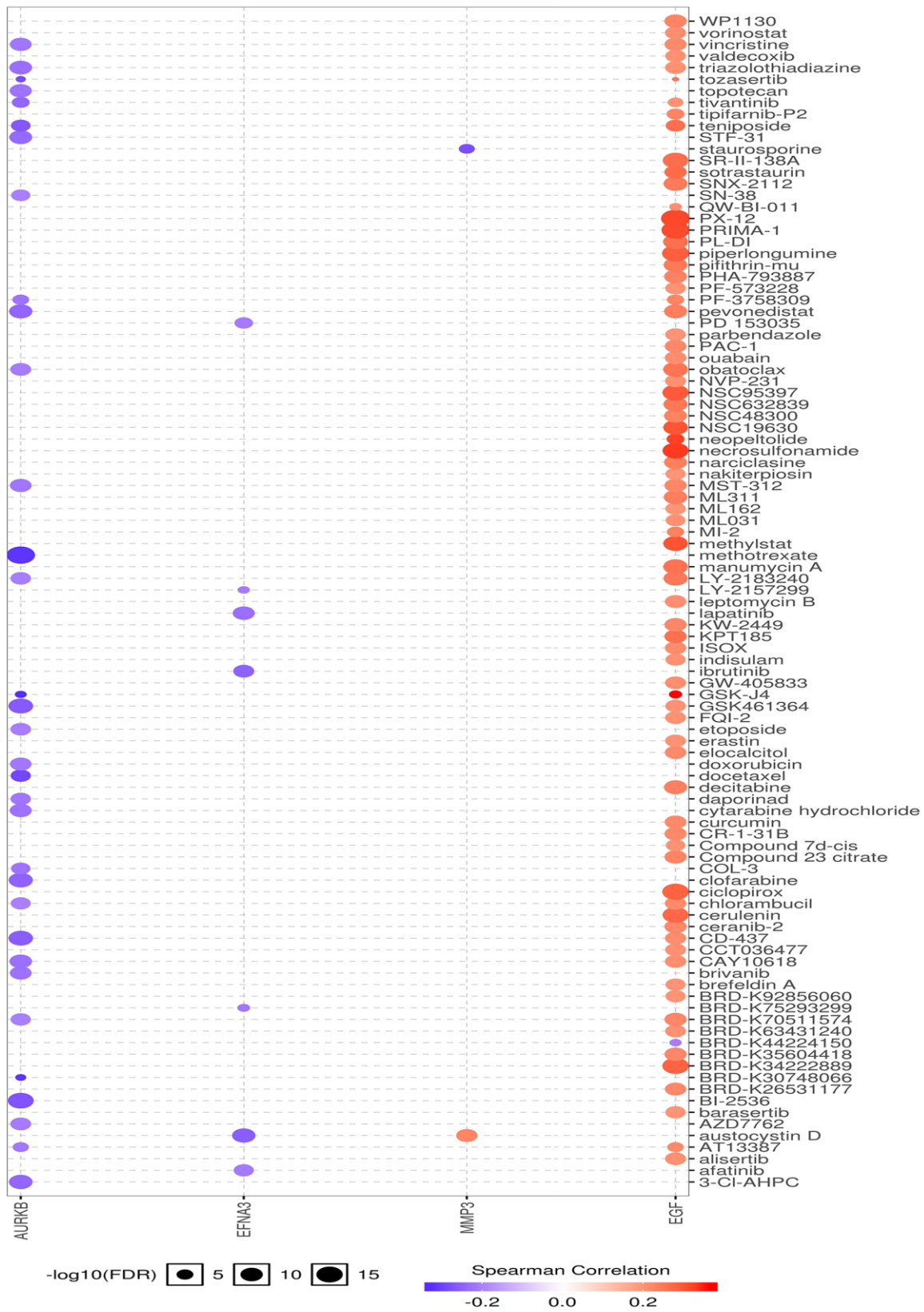
Supplementary Figure 1. An overall flow chart explaining to identify the angiogenic genes for constructing diagnostic, prognostic, and recurrence models in hepatocellular carcinoma.



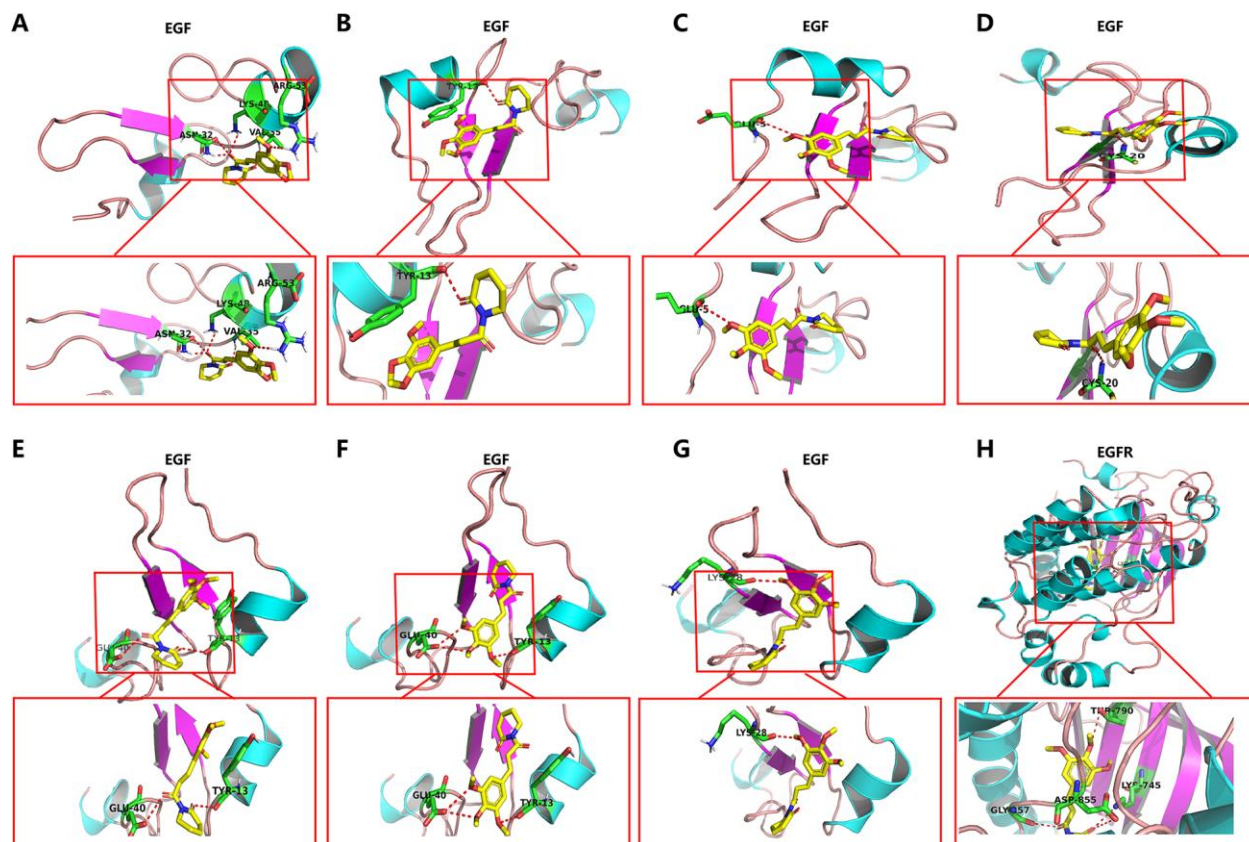
Supplementary Figure 2. Commonly differentially expressed angiogenic genes in TCGA and ICGC databases.



Supplementary Figure 3. Identification of key angiogenic genes closely associated with prognosis of HCC. LASSO Cox regression (A, B) were performed to identify the angiogenic genes closely related to OS of HCC.



Supplementary Figure 4. Inhibitors of the angiogenic genes. Data was obtained from the Cancer Therapeutics Response Portal (CTRP) database (<http://portals.broadinstitute.org/ctrp/>).



Supplementary Figure 5. Molecular docking for targeted drug identification. Specific binding sites of piperlongumine to EGF (A–G) and EGFR (H).