

Predicting the Impact of ARHGAP33 Gene on Liver Cancer Prognosis Based on Multi-Algorithm Model

Yaohui Zhao^{1-3,*}, Chenjie Wang^{1-3,*}, Xiaotong Shen¹⁻³, Xinhui Cao¹⁻³, Zi Wang¹⁻³, Huijiao Jiang¹⁻³, Xueling Chen¹⁻³, Xiangwei Wu¹⁻⁴

¹School of Medicine, Shihezi University, Shihezi, 832000, People's Republic of China; ²NHC Key Laboratory of Prevention and Treatment of Central Asia High Incidence Diseases, Shihezi University, Shihezi, 832000, People's Republic of China; ³The Clinical Research Center for Infectious Diseases of Xinjiang Production and Construction Corps, Shihezi University, Shihezi, 832000, People's Republic of China; ⁴Department of Hepatobiliary Surgery, The First Affiliated Hospital of Shihezi University, Shihezi, 832000, People's Republic of China

*These authors contributed equally to this work

Correspondence: Xiangwei Wu; Xueling Chen, School of Medicine, Shihezi University, North Fourth Road, Shihezi City, Xinjiang, 832000, People's Republic of China, Email Wu_xiangweii@outlook.com; XXuelingchen@outlook.com

Objective: To investigate the impact of Rho GTPase-activating protein 33 (ARHGAP33) and its synergistic interaction with SFPQ on the prognosis of hepatocellular carcinoma (HCC) through bioinformatics and experimental research.

Materials and Methods: RNA sequencing data from The Cancer Genome Atlas (TCGA) were analyzed to assess ARHGAP33 expression in hepatocellular carcinoma (HCC). Co-expressed genes were identified using WGCNA and GSVA, and integrated into a multi-algorithm consensus prognostic model.

Results: The analysis of the TCGA database indicated a marked overexpression of ARHGAP33 mRNA in tissues from hepatocellular carcinoma (LIHC), with a statistically significant finding ($P < 0.001$). WGCNA revealed that SFPQ is a gene associated with ARHGAP33. In the developed consensus prognostic model, survival analysis using Kaplan-Meier (K-M) alongside the CoxBoost model demonstrated that the overall survival time for patients classified as high-risk was significantly less than that of those classified as low-risk ($P < 0.05$). The Institutional Review Board at Shihezi University granted ethical approval for this research (Ethics Application No.: KJ2025-290-01).

Conclusion: The expression level of ARHGAP33 affects HCC prognosis, and its synergistic overexpression with SFPQ impairs the prognosis of HCC patients. ARHGAP33 could potentially be used as a biomarker for evaluating prognosis in hepatocellular carcinoma (HCC), offering a new theoretical foundation for enhancing HCC outcomes.

Keywords: ARHGAP33 gene, hepatocellular carcinoma, multi-algorithm, prognostic model, bioinformatics

Introduction

Worldwide, liver cancer is a harmful neoplasm that poses significant threats to human health, being the 7th most prevalent and the 3rd most deadly among various tumor types.¹ In China, liver cancer incidence is ranked 5th, while its mortality rate is 2nd.² Hepatocellular carcinoma (HCC) represents the most common type of primary liver cancer, making up about 75% to 85% of all cases.³ The disease frequently manifests in a subtle manner, resulting in the majority of patients receiving a diagnosis only at a later stage, which often means that the best chance for surgical intervention is lost. In addition, hepatocellular carcinoma (HCC) shows a lack of response to both radiotherapy and chemotherapy, leading to a grim overall prognosis, with a 5-year survival rate approximately at 12.1%. Consequently, a thorough investigation into the underlying mechanisms of HCC, coupled with the discovery of reliable prognostic indicators and treatment targets, is essential for increasing the survival rates of those affected by liver cancer. Currently, AFP and PIVKA-II, although widely used as clinical biomarkers, remain insufficient for the diagnosis and management of hepatocellular carcinoma, particularly in the context of elevated tumor-derived factors. Research

on alternative biomarkers and targeted therapies for liver cancer has made notable progress, including targets such as NAT10,⁴ MET, FGFR4, and DDR1, as well as several promising therapeutic biomarkers like Claudin 18.2, BMI1, and MDK. Nevertheless, these markers do not sufficiently fulfill the requirements of clinical diagnostics. Therefore, a continued search for more effective indicators is vital for advancing early detection, timely treatment, and prognostic assessment of HCC.

The gene ARHGAP33, which is also identified by other names such as SNX26, TCGAP, and NOMA-GAP (hereafter referred to as ARHGAP33), can be found in the 19q13.12 area of human chromosome 19. Its complete designation is Rho GTPase-activating protein 33, and it plays a crucial role as a part of the Rho GTPase-activating protein (RhoGAPs) family.^{5–9} The RhoGAPs family plays a pivotal role in cellular activities by primarily regulating the activity of Rho GTPases, thereby achieving precise modulation of intracellular signal transduction and cellular morphological changes. It plays a role in multiple biological activities, such as the reorganization of the cytoskeleton, cell migration, cell proliferation, differentiation, and programmed cell death. Within cells, Rho GTPases can be found in two states: one that is active and associated with guanosine triphosphate (GTP), and another that is inactive and linked to guanosine diphosphate (GDP). ARHGAP33 enhances the GTP hydrolysis activity of Rho GTPases, promoting their conversion from active Rho-GTP to inactive Rho-GDP, thus terminating the signaling pathway mediated by Rho GTPases. During cell migration, ARHGAP33 influences dynamic cytoskeletal reorganization by regulating Rho GTPase activity: when cells receive migration signals, Rho GTPases are activated, promoting the polymerization and cross-linking of actin filaments to form structures such as pseudopodia for cell movement; ARHGAP33 timely reduces Rho GTPase activity, leading to cytoskeletal depolymerization and thus coordinating the speed and direction of cell migration. During cell division, ARHGAP33 participates in regulating spindle formation and chromosome separation to ensure the normal progression of cell division. Abnormal function of ARHGAP33 may lead to errors in chromosome separation, resulting in instability of cellular genetic material and an increased risk of cell carcinogenesis.^{10–14} Research has shown that unusual levels of expression of the ARHGAP33 gene are linked to the onset and progression of different human disorders, encompassing neurological and cardiovascular conditions. In the field of oncology, the role of ARHGAP33 has gradually attracted attention, with studies suggesting that it may be a risk factor for prostate cancer. However, its specific function and molecular mechanism in liver cancer remain unclear.

The SFPQ gene encodes a multifunctional protein located in the nucleus that is essential for numerous biological functions, such as regulating gene transcription, processing and splicing RNA, and repairing DNA damage.¹⁵ SFPQ establishes a complex regulatory network through its interactions with other proteins and nucleic acid molecules, influencing cell proliferation, differentiation, and apoptosis. Throughout the processes of tumorigenesis and development, irregular expression or malfunction of SFPQ can interfere with standard cellular regulatory functions, consequently facilitating the proliferation and longevity of cancer cells. Recent research has progressively clarified the atypical expression and significance of SFPQ across different types of tumors;^{16,17} however, research specifically focused on liver cancer remains in its nascent stages. Some investigations suggest that SFPQ may function as a necroptosis factor, impacting the prognosis of liver cancer patients via the mTOR signaling pathway.¹⁸

This study intends to explore the influence of ARHGAP33 on the prognosis of individuals diagnosed with hepatocellular carcinoma (HCC) through the application of bioinformatics methods and the use of resources and tools like The Cancer Genome Atlas (TCGA). Moreover, we plan to develop a prognostic risk model for HCC that incorporates genes associated with ARHGAP33. This study aims to deepen our comprehension of ARHGAP33's role in the onset and advancement of liver cancer, paving the way for new biomarkers and possible therapeutic targets that facilitate early diagnosis, prognosis assessment, and customized treatment approaches for liver cancer. Furthermore, exploring ARHGAP33 may reveal new mechanisms involved in the progression and development of liver cancer. This exploration could serve as a theoretical foundation for the development of new cancer treatments, holding significant clinical importance and wider societal consequences.

Materials and Methods

Bioinformatics Analysis

TCGA Database Analysis

The Cancer Genome Atlas (TCGA) database integrates gene expression data with clinical follow-up information across multiple cancer types, serving as an indispensable resource for investigating tumorigenesis and cancer progression. In the

present study, ARHGAP33 expression data and corresponding clinical information (n=371) were extracted from the TCGA database. Cox regression analysis and Kaplan-Meier survival analysis were performed using the survival and survminer packages in R language to evaluate the associations between ARHGAP33 expression levels and overall survival (OS). The Log rank test was employed to determine the statistical significance of differences in survival curves ($P < 0.05$). RNA-seq data (transcript-level, FPKM-normalized format) and relevant clinical information of 371 hepatocellular carcinoma (HCC) patients were retrieved based on the following inclusion criteria: 1. clinical follow-up duration ≥ 1 month; 2. complete documentation of survival status (alive/deceased); 3. no missing values in gene expression data (samples with a missing rate $< 5\%$ were imputed via the k-nearest neighbor method, while those with a missing rate $\geq 5\%$ were directly excluded); 4. exclusion of samples from patients with other concurrent malignant tumors or incomplete clinical data.

Cox regression analysis was performed using the survival package (version 3.5–7) in R language (version 4.3.1), and the Kaplan-Meier (K-M) survival curves were plotted with the survminer package (version 0.4.9). The association between ARHGAP33 expression levels and overall survival (OS) of patients was analyzed via the Log rank test, with the significance level set at $P < 0.05$.

Gene Expression Omnibus Database and International Cancer Genome Consortium

Gene expression datasets related to hepatocellular carcinoma were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>), including GSE109211 (n=140), GSE144269 (n=10), GSE113996 (n=30), GSE76427 (n=115), and GSE54236 (n=20). All datasets have undergone raw data normalization (log2 transformation was used to unify the expression level scale). Data from the ICGC-LIRI cohort (n=232) was retrieved from the International Cancer Genome Consortium (ICGC) database (<https://dcc.icgc.org/>), and the gene expression matrix along with corresponding clinical survival information were extracted for external validation of the model.

CancerSEA Database

Using 14 tumor cell functional states annotated by the CancerSEA database (<http://bioacc.hrbmu.edu.cn/CancerSEA/>) (including DNA damage, DNA repair, hypoxia, invasion, metastasis, etc), the gene set z-scores corresponding to each functional state were calculated with the GSVA package (version 1.50.1) in R language. The z-scores were standardized using the scale function to obtain gene set scores. Pearson correlation analysis (two-tailed test) was performed to evaluate the correlation between ARHGAP33 expression levels and the gene set scores of each functional state, with the significance level set at $P < 0.05$.

Weighted Correlation Network Analysis

A gene co-expression network was constructed using the WGCNA package (version 1.72–1) in R language, following the steps below: 1. Data preprocessing: The gene expression matrix of the TCGA-LIHC dataset was filtered to retain genes with an expression level > 0.5 FPKM in at least 50% of samples; a total of 12,483 genes were finally included for network construction.; 2. Soft threshold screening: A scale-free network model was adopted, with the soft threshold range set at 1–30. The optimal soft threshold was determined as 6 by analyzing the network topological structure ($R^2 > 0.85$, average connectivity approaching 0); 3. Module partitioning: The dynamic tree cut algorithm was used, with the minimum number of module genes set at 30 and the module merging threshold at 0.25, to cluster genes with similar expression patterns into different modules; 4. Module-phenotype association: The module eigengene (ME) of each module was calculated, and the correlation between ME and the ARHGAP33 high/low expression phenotypes (classified by the median of ARHGAP33 expression levels) was analyzed to screen out core modules significantly correlated with the phenotypes ($|r| > 0.7$, $P < 1e-200$); 5. Hub gene screening: Within the core modules, genes with gene significance (GS) > 0.5 and module membership (MM) > 0.8 were screened as potential prognostic genes associated with ARHGAP33.

Multi-Algorithm Modeling

Algorithm parameter settings: 1. Lasso regression: The glmnet package (version 4.1–8) in R language was used, with alpha set to 1; the optimal lambda value (lambda.min) was determined via 10-fold cross-validation; 2. Elastic net: alpha was set to 0.5, and the remaining parameters were consistent with those of Lasso regression; 3. Ridge regression: alpha

was set to 0, and the remaining parameters were consistent with those of Lasso regression.; 3. Stepwise Cox regression: The stepAIC function of the survival package in R language was adopted; genes were screened using the forward stepwise selection method with the Akaike Information Criterion (AIC) as the criterion; 4. CoxBoost model: The CoxBoost package (version 1.0–11) in R language was used, with the number of iterations set to 1000 and the learning rate set to 0.01; the optimal model was determined via 10-fold cross-validation.

Risk score calculation: For the model constructed by each algorithm, the risk score of each sample was calculated using the formula $\text{Risk Score} = \sum(x_i \times \beta_i)$ based on gene expression levels (x_i) and their corresponding regression coefficients (β_i).

Model evaluation metrics: The timeROC package (version 0.4–9) in R language was used to plot the 1-year, 3-year, and 5-year receiver operating characteristic (ROC) curves, and the area under the curve (AUC) was calculated. The average AUC value (arithmetic mean of the 1-year, 3-year, and 5-year AUCs) was used to evaluate the predictive performance of the model. The surv_cutpoint function (method="median") of the survminer package was adopted to determine the optimal cut-off value of the risk score, and patients were divided into a high-risk group and a low-risk group (the proportion of samples in both groups was ≥ 0.3). The Log rank test was performed to compare the survival differences between the two groups of patients.

A Meta-analysis was performed on the results of univariate Cox survival analysis for each dataset (TCGA-LIHC, ICGC-LIRI, GSE109211, etc). The inverse variance method in the metafor package (version 3.8–1) of R language was adopted, with the logarithmic hazard ratio (log HR) and its 95% confidence interval as effect size indicators, to test the prognostic significance of the model across different datasets. For heterogeneity testing, the I^2 statistic was used ($I^2 < 50\%$ indicates low heterogeneity, 50%-75% indicates moderate heterogeneity, and $> 75\%$ indicates high heterogeneity).

Immunohistochemical Analysis

Immunohistochemical (IHC) images depicting hepatocellular carcinoma (LIHC) alongside normal tissues were obtained from The Human Protein Atlas (HPA) database, and an analysis of the expression variation of ARHGAP33 between these tissue types was conducted.

Results

Expression and Functional Analysis of ARHGAP33 in Liver Cancer

Data from The Cancer Genome Atlas (TCGA) regarding liver cancer indicate that the expression levels of ARHGAP33 in tumor samples were significantly higher than those found in nearby non-tumor samples. (Figure 1A). To assess the connection between ARHGAP33 expression levels and the overall survival (OS) of patients with liver cancer, Kaplan-Meier curve analysis was conducted (Figure 1B). When ARHGAP33 was utilized as a biomarker for hepatocellular carcinoma (HCC) patients, its area under the receiver operating characteristic (ROC) curve (AUC) for predicting overall survival was found to be 0.898 (with a 95% confidence interval^{12,19} of 0.857–0.936) (Figure 1C). Additionally, we observed a statistically significant difference in ARHGAP33 expression levels across clinical stages I–IV of HCC (Figure 1D). This suggests that ARHGAP33 expression may exhibit a consistent trend with HCC progression (ie, advancing clinical stage), thereby providing robust statistical support for its potential utility as a stage-associated molecular biomarker in HCC. The findings indicated that patients with hepatocellular carcinoma (HCC) who exhibited elevated levels of ARHGAP33 had notably reduced overall survival (OS) compared to those with lower ARHGAP33 levels. In addition, images from immunohistochemical (IHC) staining of both normal liver tissues and hepatocellular carcinoma tissues (refer to Figure 1E) were obtained from the Human Protein Atlas (HPA) database. The examination of these images revealed that the expression levels of ARHGAP33 were higher in hepatocellular carcinoma tissues compared to those found in normal liver tissues. Employing the gene set enrichment analysis (GSEA) feature, we conducted an analysis using the Hallmark gene set alongside the KEGG (Kyoto Encyclopedia of Genes and Genomes) metabolic gene set (see Figure 1F). Significant differences were observed across different gene sets when contrasting the high-expression group with the low-expression group, particularly in regards to pathways related to the cell cycle, such as the G2M checkpoint and E2F target pathways. The use of hierarchical visualization facilitated a clearer understanding of

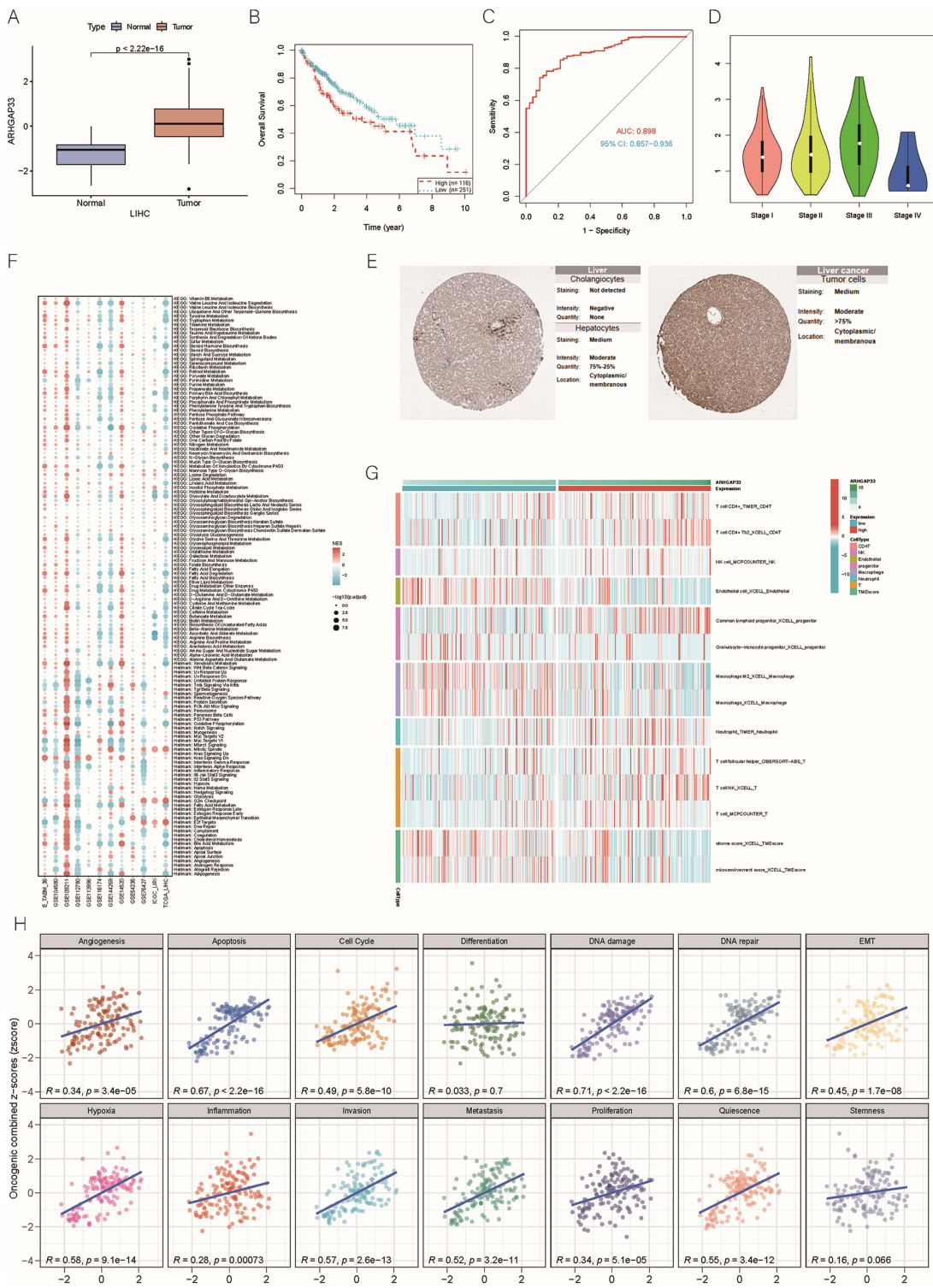


Figure 1 Analysis of the Expression, Prognostic Value, and Tumor Immune Microenvironment Association of ARHGAP33 in Liver Cancer. **(A)** The expression of ARHGAP33 in tumor tissues is higher than that in adjacent non-tumor tissues; **(B)** The overall survival (OS) rate of the ARHGAP33 high-expression group is lower than that of the low-expression group, ($p=0.025$); **(C)** The area beneath the receiver operating characteristic (ROC) curve for ARHGAP33 regarding the overall survival of patients with hepatocellular carcinoma (HCC) is notably high; **(D)** Analysis of variance (ANOVA) revealed a statistically significant difference in ARHGAP33 expression among HCC Stage I–IV ($F=5.17$, $Pr(>F)=0.00166$); notably, the expression levels in Stage III/IV were significantly higher than those in Stage I/II; **(E)** Immunohistochemical staining shows that ARHGAP33 is highly expressed in tumor tissues; **(F)** Enrichment analysis of ARHGAP33 gene-related pathways was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database; **(G)** Analysis via the Tumor Immune Estimation Resource (TIMER) database reveals that the expression of ARHGAP33 is correlated with immune cell infiltration; **(H)** Gene Set Variation Analysis (GSVA) shows that different functional phenotypes of tumor cells are positively correlated with the high expression of ARHGAP33.

the enrichment scores and levels of significance among the different gene sets. These findings offer valuable insights into the biological roles of ARHGAP33 at varying expression levels and establish a groundwork for additional functional studies and possible clinical applications. Importantly, within the GSE109211 gene set, the extent of enrichment for ARHGAP33 displayed notable differences. Various software tools were utilized to assess the differences in immune cell infiltration between the categories of high and low expression of ARHGAP33. A heatmap was generated to visualize immune cells with significant differences, where samples were ordered from left to right according to the increasing expression level of ARHGAP33. It was found that the infiltration levels of CD4+ T cells and natural killer T (NKT) cells were higher in the high ARHGAP33 expression group (Figure 1G). Using the CancerSEA database, 14 distinct functional states of tumor cells in the GSE109211 gene set were organized. The analysis revealed that high ARHGAP33 expression in liver hepatocellular carcinoma (LIHC) was strongly positively correlated with DNA damage, DNA repair, hypoxia, invasion, and metastasis. This indicates that the level of expression of ARHGAP33 is strongly linked to pro-tumor biological activities, including tumor invasion and metastasis (Figure 1H). Collectively, these results indicate a significant relationship between the expression levels of ARHGAP33 and key biological processes that promote the progression of malignant tumors. Additionally, ARHGAP33 could act as a standalone predictor for unfavorable outcomes in patients and may also hold promise as a clinical biomarker.

Construction of a Liver Cancer Prognostic Model Based on ARHGAP33-Associated Genes Screened via WGCNA

Utilizing weighted gene co-expression network analysis (WGCNA) along with subsequent integration analysis (Figure 2A), we determined that the Turquoise module serves as the crucial functional gene set influencing prognosis. This module exhibited a positive correlation within the high-expression group (Q1) while demonstrating a negative correlation in the low-expression group (Q4) (Figure 2B and C). Utilizing various algorithms, we developed prognostic models with the relevant gene sets and carried out a thorough assessment based on the average AUC values over 1, 3, and 5 years. Our analysis revealed that the CoxBoost model exhibited outstanding performance concerning average AUC values at these three intervals (Figure 2D). A univariate Cox regression analysis was conducted by us, complemented by its meta-analysis, utilizing risk scores obtained from the CoxBoost model (Figure 2E). The outcomes indicated that the prognostic model we devised was validated as a risk factor in the evaluation of different survival periods across several datasets, suggesting that the CoxBoost model possesses excellent predictive accuracy and generalizability. Consequently, the CoxBoost model was ultimately identified as the most effective algorithm (Figure 2D). Among the gene sets contained in the Turquoise module, multiple prognostic models were used with visualization methods to identify that the core prognostic gene related to ARHGAP33 is SFPQ (Figure 2F). Ultimately, the KM analysis illustrated in Figure 2G indicates that among four cohorts differentiated by varying survival durations, individuals in the high-risk category exhibited a significantly worse prognosis, whereas those in the low-risk category fared comparatively better. Similarly, poorer prognosis was observed in the high-risk groups of ICGC LIRI, GSE144269, and GSE54236 (Figure S1). The results emphasize the strong predictive precision and widespread applicability of the algorithmic model we created, indicating its promise as an innovative instrument for predicting liver cancer outcomes. Furthermore, our WGCNA analysis resulted in the discovery of the SFPQ gene, which may be linked to ARHGAP33.

Coordinated Expression of ARHGAP33 and SFPQ Promotes Liver Cancer Progression

Utilizing liver cancer-related information sourced from The Cancer Genome Atlas (TCGA) database, researchers discovered that tumor tissue exhibited a markedly elevated expression of SFPQ in comparison to adjacent non-tumor tissues (Figure 3A). To explore the correlation between SFPQ expression levels and overall survival (OS) among hepatocellular carcinoma (HCC) patients, a Kaplan-Meier curve analysis was conducted (Figure 3B). The findings indicated that HCC patients with elevated SFPQ expression experienced significantly reduced OS compared to those with lower SFPQ levels, with consistent expression patterns observed for ARHGAP33 and SFPQ in liver cancer tissues (Figure 3C). Analyses of the data from the TCGA database utilized Spearman rank correlation, Pearson correlation (P. cor) analysis, and Fisher's exact test. The findings indicated that increased expression of SFPQ frequently correlated with

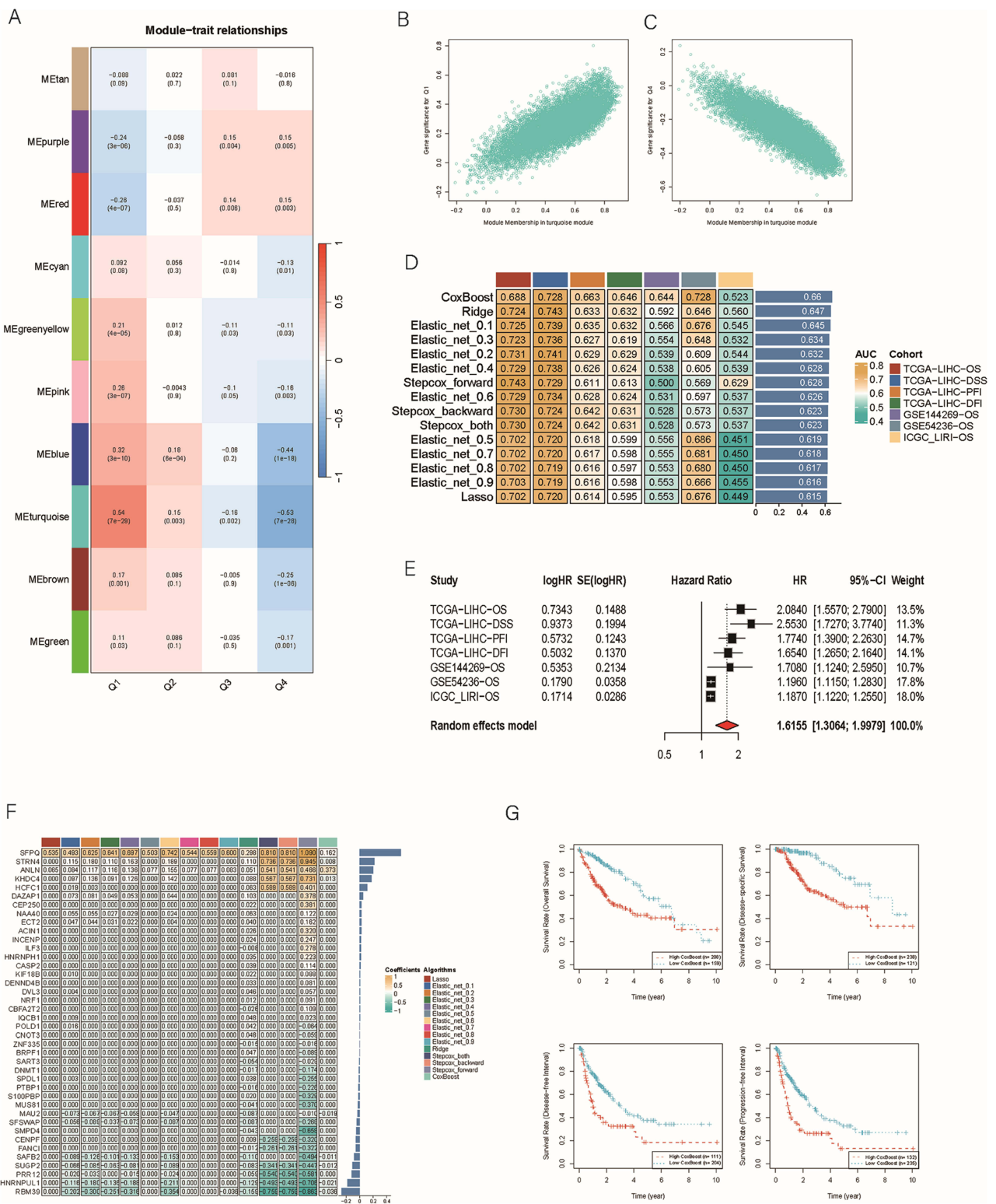


Figure 2 Integration of High-Correlation Gene Sets and Validation of High-Performance Models via WGCNA Analysis. (A) Through Weighted Gene Co-expression Network Analysis (WGCNA), it was found that the Turquoise module showed a significant correlation among Q1-Q4; (B) The ME Turquoise module exhibited a positive correlation in Q1, ($cor > 0.77, p < 1e-200$); (C) The ME Turquoise module exhibited a negative correlation in Q4, ($cor = -0.87, p < 1e-200$); (D) Comparison of the average 1-year, 3-year, and 5-year Area Under the ROC Curve (AUC) values of prognostic models constructed by different algorithms revealed that CoxBoost was the prognostic model algorithm with the most stable performance over the time span; (E) In the analysis of various survival durations across multiple datasets, the CoxBoost model emerged as a significant risk factor, demonstrating both high predictive accuracy and strong generalization capabilities; Heterogeneity: $I^2 = 86\%, \tau^2 = 0.0646, p < 0.01$; (F) A heatmap was used to calculate the coefficient of each gene in different algorithms, suggesting that the SFPQ gene had relatively high coefficients across multiple algorithms; (G) The outlook for the high-risk group was considerably worse, whereas the outlook for the low-risk group was comparatively more favorable, The Log rank test P-values for overall survival (OS), disease-specific survival (DSS), disease-free interval (DFI), and progression-free interval (PFI) in the TCGA-LIHC cohort were all less than 0.001.

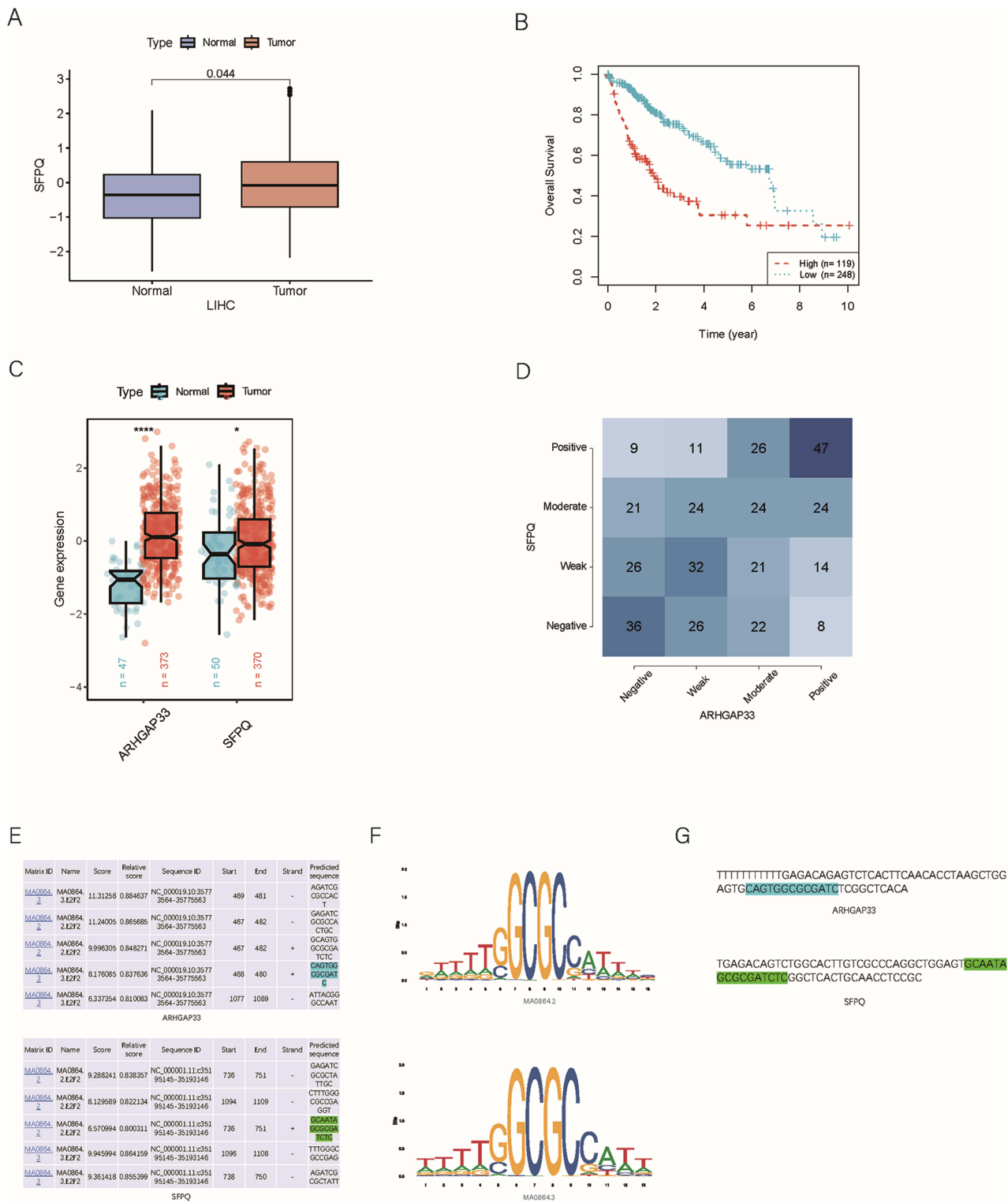


Figure 3 Impact of combined expression of ARHGAP33 and SFPQ on hepatocellular carcinoma (LIHC). **(A)** In hepatocellular carcinoma (HCC) tissues, the level of SFPQ expression was notably greater compared to that found in adjacent non-tumor tissues ($P=0.004$); **(B)** Individuals exhibiting elevated levels of SFPQ expression experienced a notably reduced overall survival (OS) ($P<0.001$); **(C)** Compared with adjacent normal tissues, ARHGAP33 was significantly highly expressed in hepatocellular carcinoma (HCC) tissues ($p \leq 0.0001$); SFPQ was also highly expressed in cancer tissues with a statistically significant difference ($p \leq 0.05$). In addition, ARHGAP33 and SFPQ exhibited consistent expression patterns in HCC tissues. (Among them, $p \leq 0.0001$ corresponds to **** and $p \leq 0.05$ corresponds to *); **(D)** Correlation analysis of ARHGAP33 and SFPQ in liver hepatocellular carcinoma (LIHC) showed that their correlation coefficient rho was 0.44; the P-value of Pearson correlation test was 3.01×10^{-19} , and the P-value of Fisher's test was 5×10^{-4} , indicating a significant positive correlation between the expression levels of ARHGAP33 and SFPQ; **(E)** Following the retrieval of the transcriptional regulatory sequences for ARHGAP33 and SFPQ from the National Center for Biotechnology Information (NCBI), sequences associated with E2F2 were obtained through JASPAR; **(F)** The key regulatory motifs were screened out using JASPAR; **(G)** The key regulatory motifs of E2F2 were present in both ARHGAP33 and SFPQ.

higher levels of ARHGAP33 expression (Figure 3D), indicating a favorable relationship between the levels of SFPQ and ARHGAP33 expressions, alongside a possible co-expression trend for these two genes. Consequently, we compiled RNA-seq data sourced from the TCGA database and recognized several shared potential transcriptional regulators for

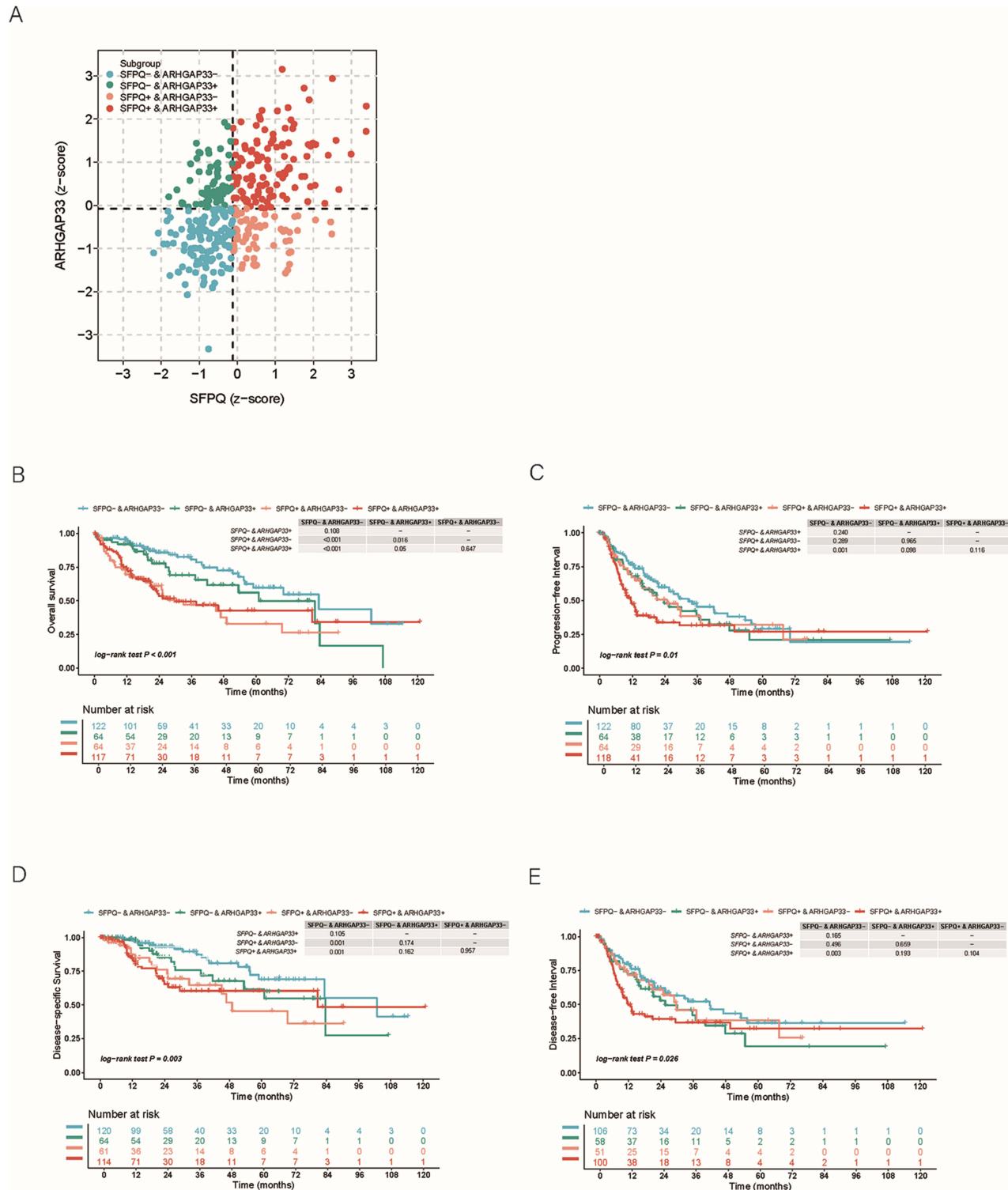


Figure 4 Impact of Coordinated Expression of ARHGAP33 and SFPQ on the Prognosis of Liver Cancer Patients. (A) There were more samples with the coordinated high expression of ARHGAP33 and SFPQ; (B–E) In different survival cohorts, patients with the coordinated high expression of ARHGAP33 and SFPQ had the worst prognostic outcomes.

both ARHGAP33 and SFPQ, with E2F2 exhibiting a notably strong correlation with each gene. The JASPAR database facilitated the analysis of motif sequence characteristics pertaining to the E2F2 transcription factor (Figure 3E). An analysis of motif enrichment was performed to examine the sequences linked to ARHGAP33 and SFPQ (refer to Figure 3F), resulting in the identification of significant regulatory motifs. Furthermore, the locations of these motif segments within ARHGAP33 and SFPQ were illustrated (see Figure 3G). These findings suggest a potential regulatory role of the E2F2 transcription factor on the expression levels of ARHGAP33 and SFPQ, which may, in turn, influence the prognosis for patients with liver cancer.

Coordinated Expression of ARHGAP33 and SFPQ Promotes Liver Cancer Progression

Samples from The Cancer Genome Atlas (TCGA) database were categorized into four subgroups according to gene expression profiles: (1) high levels of both ARHGAP33 and SFPQ (ARHGAP33+SFPQ+); (2) low ARHGAP33 and high SFPQ levels (ARHGAP33⁻SFPQ+); (3) low expression of both ARHGAP33 and SFPQ (ARHGAP33⁻SFPQ⁻); and (4) high ARHGAP33 expression with low SFPQ levels (ARHGAP33+SFPQ⁻) (Figure 4A). Similarly, in other cohorts, we observed co-expression patterns between ARHGAP33 and SFPQ (Figure S2). A Kaplan-Meier survival curve analysis was conducted to create four survival cohorts based on the co-expression levels of ARHGAP33 and SFPQ. The findings indicated that liver cancer patients who displayed “coordinated high expression” of both ARHGAP33 and SFPQ experienced the worst prognostic outcomes (Figure 4B–E). Therefore, it is hypothesized that ARHGAP33 and SFPQ may have a functional association, forming an oncogenic molecular axis that collectively promotes tumor progression in liver cancer patients.

Similarly, in other cohorts, we observed co-expression patterns between ARHGAP33 and SFPQ.

Conclusions

Liver cancer represents a highly aggressive tumor known for its considerable metastasis rate and unfavorable prognosis. In China, the incidence and mortality rates for liver cancer make up more than 50% of the worldwide totals and have shown a steady increase in recent years. At present, the main treatment options for liver cancer involve surgery, interventional radiotherapy, and chemotherapy. Nonetheless, many patients receive their diagnosis at later stages, which frequently leads to disappointing treatment results. Therefore, early detection, accurate diagnosis, and timely intervention are essential and effective approaches to improve the outcomes in liver cancer treatment.

In the past few years, advancements in bioinformatics, genomics, proteomics, transcriptomics, and multi-omics technologies have helped to find numerous types of signaling transduction pathways and cell death pathways involved in the occurrence and development of prostate cancer. This signalling includes the receptor tyrosine kinase pathway, the RAF/MEK/ERK signalling cascade, PI3K/AKT/mTOR pathway Wnt/ β -catenin signalling pathway, ubiquitin/proteasome degradation process, Hedgehog signalling pathway PANoptosis, Apoptosis Pyroptosis,¹⁹ and Necroptosis. Such findings provide valuable references for the early diagnosis and treatment of liver cancer.²⁰ Additionally, studies have indicated that genes encoding RhoGAPs (ARHGAP) can inactivate Rho-like GTPases. Among these, ARHGAP9, 15, 18, 19, 25, and 30 have been linked to breast cancer,²¹ while ARHGAP33 is associated with prostate cancer; however, its function and molecular mechanisms in liver cancer remain unclear.

This study explores the combined impact of the Rho GTPase ARHGAP33 gene alongside the SFPQ gene on hepatocellular carcinoma (HCC) prognosis through bioinformatics analysis. The results show that the expression levels of both ARHGAP33 and SFPQ are heightened in HCC tissues in comparison to normal liver tissues. Additionally, liver cancer patients with high independent expression levels of these genes tend to have a worse prognosis, potentially affecting overall survival by altering the tumor microenvironment or immune response instead of directly hindering disease progression. Further investigations are essential to clarify the molecular mechanisms associated with ARHGAP33 and to assess its possible applications in personalized treatment. These results indicate that the ARHGAP33 gene may function as an encouraging biomarker for forecasting the prognosis of early-stage HCC.

To investigate further the relationship between ARHGAP33, SFPQ, and the prognosis of hepatocellular carcinoma (HCC), we examined pertinent data obtained from databases including NCBI and JASPAR. Our results suggest that E2F2, a transcription factor, may have a significant role in regulating the transcription of ARHGAP33 and SFPQ, thus affecting the expression of related genes and influencing the prognosis for patients with liver cancer. By synthesizing

information from the TCGA database, WGCNA, GEO algorithms, and additional resources, we developed a prognostic model for HCC rooted in ARHGAP33 using multiple algorithms. Comprehensive analyses indicated the possible roles of ARHGAP33 and SFPQ in the prognosis of HCC. After a thorough assessment of average AUC values at 1, 3, and 5 years, the CoxBoost model emerged as the most effective algorithm.

Univariate Cox regression analysis of risk scores derived from the optimal algorithm (CoxBoost model) revealed that the constructed prognostic model serves as a significant risk factor across various survival periods in multiple datasets. This finding indicates that the model demonstrates high predictive accuracy and generalizability. Among the four cohorts representing different survival durations, the prognosis for the high-risk group was markedly poorer, whereas the low-risk group exhibited relatively better outcomes. This observation further substantiates the reliability and practicality of the CoxBoost model in clinical prognosis evaluation, offering a crucial reference for future research and clinical applications.

This research marks the initial effort to demonstrate a notable relationship between ARHGAP33 and SFPQ within a prognostic framework, employing a multi-algorithm analysis that identifies the CoxBoost model as the most suitable approach. Our findings suggest that ARHGAP33 and SFPQ are regulated by a shared transcription factor, E2F2, which could be involved in the initiation and advancement of liver cancer through the Rho/ROCK pathway and RNA splicing; nonetheless, the precise mechanisms require additional investigation.

Despite its contributions, this study has certain limitations. To begin with, although the model's robustness has been confirmed through analyses involving public databases, the intrinsic biases tied to retrospective studies require additional correction through prospective cohort or experimental investigations. Furthermore, the relationship between ARHGAP33 and prognostic genes is presently backed only by correlative findings. Conducting *in vitro* and *in vivo* research, which includes gene knockout techniques, overexpression studies, and chromatin immunoprecipitation, is vital to clarify their direct mechanisms of action. The strategies for pharmacotherapy of tumors have become increasingly diverse. In addition to targeted therapeutic agents, plant-derived traditional medicines have also demonstrated emerging application prospects in cancer treatment.²² However, this study has not explored the regulatory effects of the aforementioned drugs on the expression and function of ARHGAP33 and SFPQ. While this study focuses on transcriptomic signatures, future integration of physiological and imaging-based biomarkers—such as hepatic perfusion dynamics and tumor-induced vascular compression—could enhance patient stratification and help guide personalized combination therapies in distinct HCC subtypes.

In summary, this research not only validates the clinical predictive significance of the prognostic model involving ARHGAP33 and SFPQ for patients with hepatocellular carcinoma (HCC), but it also highlights the dual function that ARHGAP33 might serve as both a potential prognostic biomarker and a therapeutic target. Additionally, the collaborative interactions among various genes may provide an essential foundation for upcoming investigations. As a result, ARHGAP33 holds promise as a notable prognostic biomarker for HCC, and an in-depth examination of its regulatory functions within this context may lead to innovative therapeutic approaches and renewed optimism in HCC treatment.

Data Sharing Statement

The datasets used and/or analysed during the current study were publicly available from the The Cancer Genome Atlas (TCGA) database.

Ethics Approval and Consent to Participate

This study was conducted in accordance with the declaration of Helsinki. The Institutional Review Board at Shihezi University granted ethical approval for this research (Ethics Application No.: KJ2025-290-01). The written informed consent was obtained from the participants or their guardians.

Acknowledgments

Thanks for the support from the Tianshan Young Talent Scientific and Technological Innovation Team (2023TSYCTD0020) and the Corps Guidance Science and Technology Plan Project (2022ZD041).

Author Contributions

Yaohui Zhao and Chenjie Wang contributed equally to this paper, both as the first authors. All authors made a significant, contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This work was funded by the Tianshan Young Talent Scientific and Technological Innovation Team (2023TSYCTD0020) and the Corps Guidance Science and Technology Plan Project (2022ZD041).

Disclosure

All authors declare that they have no conflict of interests.

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA*. 2024;74(3):229–263. doi:10.3322/caac.21834
2. Wypkl W. Interpreting the 2022 global cancer statistics report. *J Multidisciplin Cancer Manage*. 2024;10(3):1–16. doi:10.12151/JMCM.2024.03-01
3. Li Q, Xia C, Li H, et al. Disparities in 36 cancers across 185 countries: secondary analysis of global cancer statistics. *Front Med*. 2024;18(5):911–920. doi:10.1007/s11684-024-1058-6
4. Zhang Y, Dong Y, Chen S, et al. Targeting NAT10 inhibits hepatocarcinogenesis via ac4C-mediated SMAD3 mRNA stability. *Exploration*. 2025:20250075. doi:10.1002/EXP.20250075
5. Chiang SH, Hwang J, Legendre M, Zhang M, Kimura A, Saliel AR. TCGAP, a multidomain Rho GTPase-activating protein involved in insulin-stimulated glucose transport. *EMBO J*. 2003;22(11):2679–2691. doi:10.1093/emboj/cdg262
6. Kim Y, Ha CM, Chang S. SNX26, a GTPase-activating protein for Cdc42, interacts with PSD-95 protein and is involved in activity-dependent dendritic spine formation in mature neurons. *J Biol Chem*. 2013;288(41):29453–29466. doi:10.1074/jbc.M113.468801
7. Liu H, Nakazawa T, Tezuka T, Yamamoto T. Physical and functional interaction of Fyn tyrosine kinase with a brain-enriched Rho GTPase-activating protein TCGAP. *J Biol Chem*. 2006;281(33):23611–23619. doi:10.1074/jbc.M511205200
8. Rosario M, Schuster S, Juttner R, Parthasarathy S, Tarabykin V, Birchmeier W. Neocortical dendritic complexity is controlled during development by NOMA-GAP-dependent inhibition of Cdc42 and activation of cofilin. *Genes Dev*. 2012;26(15):1743–1757. doi:10.1101/gad.191593.112
9. Shen PC, Xu DF, Liu JW, et al. TC10 β /CDC42 GTPase activating protein is required for the growth of cortical neuron dendrites. *Neuroscience*. 2011;199:589–597. doi:10.1016/j.neuroscience.2011.08.053
10. Hall A. Rho GTPases and the actin cytoskeleton. *Science*. 1998;279(5350):509–514. doi:10.1126/science.279.5350.509
11. Jaffe AB, Hall A. Rho GTPases: biochemistry and biology. *Annu Rev Cell Dev Biol*. 2005;21:247–269. doi:10.1146/annurev.cellbio.21.020604.150721
12. Bos JL, Rehmann H, Wittinghofer A. GEFs and GAPs: critical elements in the control of small G proteins. *Cell*. 2007;129(5):865–877. doi:10.1016/j.cell.2007.05.018
13. Teherkejian J, Lamarche-Vane N. Current knowledge of the large RhoGAP family of proteins. *Biol Cell*. 2007;99(2):67–86. doi:10.1042/BC20060086
14. Heasman SJ, Ridley AJ. Mammalian Rho GTPases: new insights into their functions from in vivo studies. *Nat Rev Mol Cell Biol*. 2008;9(9):690–701. doi:10.1038/nrm2476
15. Thivierge C, Bellefeuille M, Diwan SS, et al. Paraspeckle-independent co-transcriptional regulation of nuclear microRNA biogenesis by SFPQ. *Cell Rep*. 2024;43(9):114695. doi:10.1016/j.celrep.2024.114695
16. Yang L, Gilbertsen A, Jacobson B, et al. SFPQ and its isoform as potential biomarker for non-small-cell lung cancer. *Int J Mol Sci*. 2023;24(15). doi:10.3390/ijms241512500
17. Klotz-Noack K, Klinger B, Rivera M, et al. SFPQ depletion is synthetically lethal with BRAF(V600E) in colorectal cancer cells. *Cell Rep*. 2020;32(12):108184. doi:10.1016/j.celrep.2020.108184
18. Song H, Ge Y, Xu J, et al. Identification and validation of novel signature associated with hepatocellular carcinoma prognosis using single-cell and WGCNA analysis. *Int J Med Sci*. 2023;20(7):870–887. doi:10.7150/ijms.79274
19. Xiang J, Li Y, Mei S, et al. Novel diagnostic and therapeutic strategies based on PANoptosis for hepatocellular carcinoma. *Cancer Biol Med*. 2025:20250150. doi:10.20892/j.issn.2095-3941.2025.0150
20. Sia D, Villanueva A, Friedman SL, Llovet JM. Liver cancer cell of origin, molecular class, and effects on patient prognosis. *Gastroenterology*. 2017;152(4):745–761. doi:10.1053/j.gastro.2016.11.048
21. Chen WX, Lou M, Cheng L, et al. Bioinformatics analysis of potential therapeutic targets among ARHGAP genes in breast cancer. *Oncol Lett*. 2019;18(6):6017–6025. doi:10.3892/ol.2019.10949
22. Nawaz S, Wajid A, Nawaz A, et al. Calotropis procera: a review of molecular mechanisms, bioavailability, and potential anticancer property. *Biomed Eng Commun*. 2024;3:17–22. doi:10.53388/BMEC2024021

International Journal of General Medicine

Publish your work in this journal

The International Journal of General Medicine is an international, peer-reviewed open-access journal that focuses on general and internal medicine, pathogenesis, epidemiology, diagnosis, monitoring and treatment protocols. The journal is characterized by the rapid reporting of reviews, original research and clinical studies across all disease areas. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-general-medicine-journal>

Dovepress
Taylor & Francis Group