Journal Pre-proof

A three-gene resistome signature as a prognostic tool in hepatocellular carcinoma

Anabel Sanchez-Martin , Rocio I.R. Macias , Luis Muñoz-Bellvís , Luis M. Gonzalez , Alejandro Forner , Maria Reig , Jose J.G. Marin , Oscar Briz

PII: S1665-2681(25)00404-1
DOI: https://doi.org/10.1016/j.aohep.2025.102179
Reference: AOHEP 102179

To appear in: *Annals of Hepatology*

Received date: 17 July 2025
Accepted date: 15 December 2025

Please cite this article as: Anabel Sanchez-Martin , Rocio I.R. Macias , Luis Muñoz-Bellvís , Luis M. Gonzalez , Alejandro Forner , Maria Reig , Jose J.G. Marin , Oscar Briz , A three-gene resistome signature as a prognostic tool in hepatocellular carcinoma, *Annals of Hepatology* (2025), doi: https://doi.org/10.1016/j.aohep.2025.102179

Original article

A three-gene resistome signature as a prognostic tool in hepatocellular carcinoma

Short Title: Prognostic gene signature in hepatocellular carcinoma

Anabel Sanchez-Martin[a], Rocio I.R. Macias[a,f], Luis Muñoz-Bellvís[b,c], Luis M. Gonzalez[b], Alejandro Forner[d,e,f], Maria Reig[d,e,f], Jose J.G. Marin[a,f‡*], Oscar Briz[a,f‡*]

[a]Experimental Hepatology and Drug Targeting (HEVEPHARM), Institute for Biomedical Research of Salamanca (IBSAL), University of Salamanca, Salamanca, Spain

[b]Service of General and Gastrointestinal Surgery, University Hospital of Salamanca, IBSAL, Salamanca, Spain

[c]Centro de Investigación Biomédica en Red del Cáncer (CIBERONC), Carlos III National Institute of Health, Madrid, Spain

[d]Liver Oncology Unit, Liver Unit, ICMDM, Hospital Clinic of Barcelona. University of Barcelona, Barcelona, Spain

[e]BCLC group, Fundació Recerca Clínic de Barcelona-Institut d'Investigació Biomèdica August Pi i Sunyer (FRCB-IDIBAPS), Barcelona, Spain

[f]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Carlos III National Institute of Health, Madrid, Spain

‡Both authors contributed equally as senior authors to this work.

*Corresponding authors:
Jose J.G. Marin,
Department of Physiology and Pharmacology, University of Salamanca.
Campus Miguel de Unamuno, E.D., Lab-231, 37007.
Salamanca, Spain.
Email: jjgmarin@usal.es

Oscar Briz,
Department of Physiology and Pharmacology, University of Salamanca.
Campus Miguel de Unamuno, E.D., Lab-S05, 37007.
Salamanca, Spain.
Email: obriz@usal.es

Word count: 3642

Number of Figures: 7

Number of Tables: 4

Number of Supplementary Figures: 2

Number of Supplementary Tables: 2

***Introduction and Objectives*:** Outcomes for patients with hepatocellular carcinoma (HCC) are generally poor, partly due to significant multidrug resistance. HCC heterogeneity decreases the accuracy of traditional prediction models. This study aimed to evaluate the prognostic significance of the HCC resistome.

***Materials and Methods*:** Clinical and transcriptomic data from 371 HCC cases available at TCGA were analyzed. An external dataset (604 patients from 4 cohorts) and 40 HCC samples, in which gene expression was determined by RT-qPCR, were used to validate the prognostic model.

Results: The *in silico* analysis revealed two distinct clusters of patients based on the expression of resistome genes. Kaplan-Meier analysis indicated that Cluster 1 (C1) exhibited a better prognosis. Furthermore, the response to sorafenib treatment was better in patients included in C1 than in C2. Cox regression analysis identified the resistome profile as an independent prognostic factor alongside clinicopathological features such as tumor stage and ECOG status. Fifty-eight out of 81 genes examined displayed differential expression between clusters. Thirteen genes demonstrated a correlation between their expression levels and patient survival. In Cox multivariate analysis, *SLC22A1*, *BIRC5*, and *ABCC1* genes emerged as independent prognostic factors, forming the basis for a risk model. *BIRC5* and *ABCC1* upregulation and *SLC22A1* downregulation were associated with worse outcomes. Experimental results confirmed that patients with higher risk scores had a worse prognosis.

Conclusions: A prognostic signature based on the expression levels of three resistome-associated genes has been defined and can serve as a helpful complementary tool in clinical settings to categorize HCC patients.

Keywords: Liver cancer; Multidrug resistance; Prognosis model; Resistome; Tumor heterogeneity.

Abbreviations: FPKM-UQ, fragments per kilobase of transcript per million mapped reads upper quartile; OS, overall survival; qPCR, quantitative PCR; ROC, receiver operating

characteristic; RRS, resistome risk score; TCGA, The Cancer Genome Atlas; TLDA, Taqman Low-Density Array.

1. Introduction

Hepatocellular carcinoma (HCC) is the most common liver cancer in adults, with an estimated annual incidence of more than 900,000 new cases worldwide [1, 2]. It ranks seventh globally among all cancers and third in cancer-related deaths, often due to late detection and its connection to liver cirrhosis [3]. Patient survival is closely tied to the level of liver function. Early-stage treatments such as surgery, ablation, and liver transplantation offer favorable outcomes, although few patients qualify because of asymptomatic early stages and limited early diagnostic methods [4, 5]. For unresectable, intermediate-stage HCC, trans-arterial chemoembolization (TACE) provides a modest chance of prolonging survival [6, 7]. For advanced-stage HCC, treatment options are limited mainly to systemic therapies, predominantly involving immunomodulatory agents and tyrosine kinase inhibitors [6, 8]. While recent therapies used in clinical practice are promising, they typically result in only modest improvements in patient outcomes [9, 10].

The main challenge in treating HCC is its high resistance to anticancer drugs due to the multidrug resistance (MDR) phenotype [11, 12]. This phenotype includes several mechanisms of chemoresistance shared across different cancers [13]. The resistome, a collection of chemoresistance-related genes expressed by the tumor, encodes proteins involved in complex processes also present in healthy cells. In tumor cells, these gene expressions are altered, especially during drug exposure. Our group and others have identified elements of the resistome in HCC [11, 14-16], with significant variability in gene expression between patients linked to chemoresistance and malignancy [17].

The impact of treatment failure on patient outcomes is significant due to the loss of valuable time and an increased risk of tumor progression with enhanced cross-resistance, which can potentially undermine responses to second-line drugs [18]. Therefore, predicting how HCC will respond to treatment is essential for guiding clinical management and boosting survival rates in these patients [19]. While the Barcelona Clinic Liver Cancer (BCLC) system combines tumor

features, liver function, and patient clinical status to offer broadly accepted prognostic information [20], the predictive accuracy of traditional models, including BCLC, remains limited because of the significant molecular diversity inherent in HCC [21, 22]. The addition of new biomarkers is vital for creating more effective predictive tools [23].

Therefore, this study aimed to perform a retrospective analysis to characterize the genetic diversity related to chemoresistance and malignancy in patients with HCC, laying the foundation for creating a prognostic tool for those with this cancer.

2. Materials and Methods

2.1. Data collection for *in silico* prognostic model development

Clinical and transcriptomic data from 371 tumors and 50 paired non-tumor tissue samples available in The Cancer Genome Atlas HCC cohort (TCGA-LIHC) were analyzed. Recurrent tumors were excluded. RNA sequencing (RNA-seq) data, generated on the HiSeq Illumina platform, were aligned to the GRCh38 human genome. Gene expression was measured using Fragments Per Kilobase of transcript per Million mapped reads Upper Quartile (FPKM-UQ) values. Data were retrieved from the NCI Genomic Data Commons (GDC) repository and analyzed using R version 4.2.3 and RStudio, with the "TCGAbiolinks" and "SummarizedExperiment" packages. Additional clinical data were sourced from the cBio Cancer Genomics Portal (http://cbioportal.org) [24].

2.2. Clustering of HCC samples

Expression data for about 100 genes involved in chemoresistance and malignancy (Supplementary Table 1) were analyzed. MDR genes associated with resistance to chemotherapeutic drugs, tyrosine kinase inhibitors, and checkpoint inhibitors were selected based on recent literature reviews [11-13]. Genes with low expression (< 2.0 FPKM-UQ) in 95% of HCC and non-tumor samples were excluded. FPKM-UQ data were log2-transformed before analysis. Clustering of HCC samples, based on resistome-related gene expression profiles,

was conducted using the NG-CHM Builder software (MD Anderson Cancer Center, University of Texas, US) (https://build.ngchm.net/NGCHM-web-builder/Select_Matrix.html?v=2.22.0). Unsupervised hierarchical clustering used Euclidean distance and Ward's minimum variance method.

2.3. Survival analyses

Kaplan-Meier analysis for overall survival (OS) was performed using R with the "TCGAbiolinks" and "Survminer" packages. Survival times between groups were compared using the log-rank test, with risk tables showing patients at risk over time. The optimal gene expression cutoff was determined through the Kaplan-Meier Plotter resource (https://kmplot.com/analysis/) [25]. Cox proportional hazards regression assessed the relationship between OS and resistome or clinicopathological features. Univariate and multivariate Cox regression analyses identified significant variables (p-value < 0.05). Forest plots displaying hazard ratios were generated. From the genes identified in the multivariate Cox analysis, the Resistome Risk Score (RRS) model was developed. The performance of RRS in predicting OS, including sensitivity and specificity, was evaluated using binormal receiver operating characteristic (ROC) curves via online software (http://www.rad.jhmi.edu/jeng/javarad/roc/JROCFITi.html) from Johns Hopkins University School of Medicine, MD, US.

2.4. Validation cohorts

The prognostic model developed from TCGA data was validated using additional cohorts. Data were obtained from the Integrative Molecular Database of HCC (HCCDB), which includes public microarray and RNA-seq datasets from GEO and ICGC repositories. This resource was compiled by Tsinghua University, the National Liver Cancer Center, and Shanghai East Hepatobiliary Surgery Hospital (China) [26] and is freely accessible at http://lifeome.net/database/hccdb.

A second validation cohort included 40 HCC samples from Tumor Biobanks at Salamanca University Hospital and Barcelona Clinic Hospital (Supplementary Table 2). The research protocols were approved by the Ethical Committee for Clinical Research at these institutions and conducted following the principles outlined in the Declaration of Helsinki. All patients provided written consent for the use of tissue samples in biomedical research.

These samples were collected and processed for RNA extraction and gene expression analysis as previously described [27]. Briefly, total RNA was isolated using RNA mini-spin columns treated with RNase-free DNase I, utilizing the Illustra RNAspin Mini Cytiva Kit (Thermo Fisher, Madrid). Subsequently, cDNA was synthesized by reverse transcription (RT) from total RNA using random primers and reverse transcriptase included in the SuperScript VILO cDNA Synthesis Kit (Invitrogen, Thermo Fisher). Quantitative PCR (qPCR) was conducted using AmpliTaq Gold DNA polymerase and Taqman Low-Density Arrays (TLDAs) on an ABI Prism 7900HT Sequence Detection System (Applied Biosystems, Thermo Fisher). The thermal cycling protocol included an initial cycle at 50 °C for 2 min and at 95 °C for 10 min, followed by 40 cycles at 95 °C for 15 seconds and at 60 °C for 60 s. Gene expression levels were normalized twice to GAPDH and ACTB, with 18S rRNA used as a quality control.

The RRS, calculated from the *in silico* study, was applied to samples from both validation cohorts. Patients were categorized into low- and high-risk groups using the risk score formula. To assess the predictive ability of the prognostic model, Kaplan-Meier survival and Cox regression analyses were performed as described above.

2.5. Other statistical analysis

Data processing and statistical analyses were performed using Microsoft Office Excel (version 365) and GraphPad Prism 8. The statistical significance of differences between groups was determined using Pearson's $\chi 2$, Fisher's exact, or Mann-Whitney U tests, as appropriate.

2.6. Ethical Statements

Given the open nature of the data used for the *in silico* study, ethical approval was deemed unnecessary, and all analyses adhered to database access principles. In the case of experiments carried out to validate the model's predictive power, the research protocols were reviewed and approved by the Salamanca Health Area Drug Research Ethics Committee (2019-02-195) and conducted in accordance with the principles outlined in the Declaration of Helsinki. All patients provided written consent for the utilization of tissue samples in biomedical research.

3. Results

3.1. Clustering of HCCs according to the resistome profile

For the present analysis, we selected about one hundred genes based on their significance in the resistome across various cancer types, primarily gastrointestinal, based on data identified by others and our own research group over the past decade [11-13]. The expression of the selected chemoresistance-associated genes had already been validated by RT-QPCR using TaqMan Low-Density Arrays (TLDAs) in HCC samples and in cell lines derived from this cancer [17].

An *in silico* study was then conducted to identify HCC groups based on their expression profiles of these genes, utilizing transcriptomic data from 371 HCC patient tumor samples in the TCGA-LIHC cohort, obtained through RNA-seq. Seventeen genes were excluded due to their consistently low expression (< 2.0 FPKM-UQ) in 95% of both HCC and non-tumor samples. Hierarchical clustering divided the tumors into two main clusters, with significant differences in gene expression profiles: Cluster 1 (C1), comprising 110 patients (29.6%), and Cluster 2 (C2), comprising 261 patients (70.4%) (Figure 1). C2 was further divided into two subgroups, C2a and C2b, though the differences in gene expression between them were less noticeable, as shown by the height of the dendrogram (Figure 1), and had no significant impact on OS (Supplementary Figure 1).

Fifty-eight out of the 81 genes studied showed statistically different median expressions between C1 and C2. Accordingly, they were identified as major contributors to classifying HCC samples into these clusters. Thirty-seven genes were downregulated in C2 compared to C1. *ABCA8*, *SLC22A1*, *CYP1A1*, *CYP1A2*, and *UGT1A4* showed the most significant downregulation (Table 1). Conversely, 21 genes were upregulated in C2 relative to C1. *TYMS*, *BIRC5*, *TOP2A*, *MSH2*, and *ABCC1* exhibited the most notable differences (Table 2). Hierarchical clustering of HCCs using only the 58 top-weighted genes confirmed the presence of two clusters, with 95% of samples remaining in the same cluster as previously described. This demonstrates the persistence of the top-weighted genes alongside the less influential ones (Supplementary Figure 2).

3.2. Relationship between clustering and prognosis

Next, the relationship between the resistome signature and patient outcomes was examined. Using the Kaplan-Meier method, the OS of 370 patients in the previously defined groups was analyzed. At the time of the study, 130 patients (35.0%) had died. C2 showed a poorer prognosis, with a median survival of 45.7 months compared to 70.5 months for C1 (Figure 2A–B). Among patients classified within C1 and C2, only those treated with sorafenib alone or in combination with other drugs were further analyzed as groups C1sor and C2sor, respectively. Patients in the C1sor group demonstrated a better response to treatment, with a longer median survival (52.0 months) than those in C2sor, who had a shorter median survival (21.1 months) (Figure 2C–D).

Beyond the survival analysis, additional clinicopathological parameters were compared between C1 and C2. Statistically significant differences were found in patient age, sex, tumor differentiation grade, performance status according to the Eastern Cooperative Oncology Group (ECOG) scale, tumor stage, and serum alpha-fetoprotein (AFP) levels at diagnosis (Table 3). Given the notable differences in most clinicopathological features between the two clusters, further analysis using Cox regression was conducted to assess their prognostic value for OS. Univariate analysis showed that female sex, advanced tumor stage (III or IV), and reduced patient functional status (ECOG ≥1) were associated with worse survival (Figure 3A). The evaluation of the prognostic ability of selected resistome genes confirmed that C2 was linked to poorer OS, consistent with the previous results obtained using the log-rank test. To determine whether this gene expression signature is an independent prognostic factor, a multivariate analysis was performed, including clinical and molecular variables significantly associated with OS in the univariate analysis. The results indicated that the resistome gene expression profile remained an independent prognostic factor, along with tumor stage and ECOG status (Figure 3B).

To further support the distinct genetic signatures between the two clusters, we evaluated the expression of the classical marker used for HCC monitoring alpha-fetoprotein (*AFP*). Moreover, we also evaluated the differential expression of the epithelial cell adhesion molecule (*EpCAM*) and cytokeratin 19 (*CK19*), two tumor stem cell markers associated with more aggressive behavior and poor prognosis in HCC patients following surgical and adjuvant therapies, which have been proposed as markers of a distinct HCC subtype [28, 29]. Our

results showed that patients in the C2 group had higher AFP, EpCAM, and CK19 expression levels than those in the C1 group (Figure 4).

3.3. Establishment of the prognostic risk model in HCC patients

Among the 58 resistome genes that were differentially expressed in both clusters, we aimed to identify those whose expression levels were individually linked to patient outcomes. The selection was based on the magnitude of the difference in their expression between the two clusters. Values greater than 2-fold (C2 *vs.* C1) were considered as either downregulated (13 genes) or upregulated (6 genes). The log-rank test identified 13 genes with a significant association between their expression levels and patient survival (Table 4). A more stringent univariate Cox regression analysis narrowed the list to 6 genes with prognostic value for OS in HCC, specifically *SLC22A1*, *BIRC5*, *CYP3A4*, *ABCA8*, *ABCC1*, and *CYP1A1* (Figure 5A). In multivariate analysis, three of these genes (*SLC22A1*, *BIRC5*, and *ABCC1*) emerged as independent prognostic factors (Figure 5B). Higher expression of BIRC5 and ABCC1 and lower expression of SLC22A1 were associated with worse outcomes. The estimated regression coefficients, which measure the impact of these three variables on survival and are multiplied by the expression level of each gene, were used to build the "resistome risk-score" (RRS) model (Figure 5C). ROC curve analysis indicated that our prognostic model performed well in predicting OS of HCC patients, with an AUC of 0.787, showing high sensitivity (88.6%) and acceptable specificity (70.4%) (Figure 5D).

3.4. Evaluation of the model's predictive power

The prognostic effectiveness of the RRS model was validated *in silico* across various patient cohorts, considering differences in sample sizes, sample sources, and gene expression measurement techniques (Figure 6A). In all datasets, it was consistently observed that HCC patients with higher RRS values had worse outcomes, confirmed using both the log-rank test (Figure 6A–E) and the Cox regression method (Figure 6F).

Additionally, a validation approach was undertaken using a cohort of 40 patients in which the expression levels of *SLC22A1*, *ABCC1*, and *BIRC5* were experimentally quantified by qPCR. The resistome risk score (RRS) was calculated for each patient and represented as a violin plot (Figure 7A). An optimal cutoff value of –10, determined by minimizing the log-

rank p-value, was used to stratify patients into low-risk (n = 13) and high-risk (n = 27) groups. HCC patients with the highest RRS showed lower *SLC22A1* expression and higher *ABCC1* expression compared with those with the lowest RRS (Figure 7B). The high-RRS group also exhibited a trend toward higher *BIRC5* expression relative to the low-RRS group. Kaplan-Meier log-rank test results highlighted that patients in the high-risk group had significantly shorter OS compared to those in the low-risk group (Figure 7C–D). Cox regression analysis further confirmed the model's reliability in predicting HCC survival, showing that patients with higher RRS values had a worse prognosis (Figure 7E).

4. Discussion

HCC is a neoplasm with considerable variability between individuals, making prognostic predictions difficult. Patients with similar clinical features and tumor stages may have extremely diverse OS outcomes [30]. This heterogeneity also limits the effectiveness of current drug treatments for advanced liver cancer [31]. In a previous study, our group analyzed a small cohort of HCC patients, revealing significant variability in the expression of resistome genes [17]. In the present study, we further explored this question by investigating a potential link between the resistome profile and patient prognosis. The TCGA-LIHC cohort used for the *in silico* study included patients with tumors at various stages and receiving different treatments. Most patients underwent surgical resection without adjuvant pharmacological therapy, while some received ablative treatment or embolization. In the initial analysis, patient prognosis was assessed independently of the treatment each patient received. Median survival times of 45 to 70 months are notably higher than those for patients with intermediate- or advanced-stage HCC. This is because many patients with early-stage disease, for whom surgical specimens were available, are included in this analysis. Although the analysis confirmed significant heterogeneity among patients, it revealed a clear division into two groups based on expression profiles and OS. Patients in the TCGA-LIHC cohort received different antitumor drug therapies, with sorafenib-based treatments being the most common, although some received standard cytotoxic therapies. Among patients receiving sorafenib-based

treatment, those in cluster 2 experienced worse outcomes. Notably, most of these patients had adjuvant drug therapy, suggesting that the tumor's response to the drug might be relatively influenced by other factors, such as tumor malignancy features and the extent of tumor removal.

The comparison of clinicopathological characteristics between the two clusters identified here revealed significant differences in most parameters. One of the most notable differences was related to gender, which may be due to gender-specific risk factors such as alcohol consumption and obesity, impacting hepatocarcinogenesis differently in men and women. Additionally, molecular mechanisms such as the differential expression of androgen and estrogen receptors, which mediate signaling pathways involved in cell proliferation, apoptosis, and DNA repair, may be involved. The serum level of AFP at diagnosis was not predictive of patient prognosis in the overall cohort. However, when examining the identified patient clusters, AFP levels were significantly higher in Cluster 2 compared to Cluster 1. This result is consistent with the widely reported association between AFP levels at diagnosis and outcomes[32, 33].

It is important to note that not all genes showing significant differences between the identified clusters were expected to influence patient prognosis equally. To test this idea, we conducted a study analyzing how each gene affects HCC patient outcomes, which resulted in the creation of a gene signature based on the expression of *SLC22A1*, *BIRC5*, and *ABCC1*, demonstrating independent prognostic ability.

The *SLC22A1* gene encodes organic cation transporter 1 (OCT1), which is highly expressed in hepatocytes and plays a crucial role in transporting various cationic compounds, including both endogenous and xenobiotic compounds, such as the drug sorafenib[34, 35]. Therefore, OCT1 may play a crucial role in the effectiveness of sorafenib because this drug's mechanism of action relies on its ability to inhibit the intracellular domains of tyrosine kinases. Reduced expression of OCT1 has been suggested as a common feature of the multidrug resistance phenotype found in major liver cancer types, such as HCC, cholangiocarcinoma, and hepatoblastoma [17, 35].

The *BIRC5* gene encodes the anti-apoptotic protein survivin, which is overexpressed in many tumors, including liver cancers [17, 36]. Survivin overexpression has been linked to chemoresistance, inhibition of apoptosis, and the development of metastasis, making it a marker of malignant traits in cancer cells [37]. Multidrug resistance-associated protein 1, encoded by the *ABCC1* gene, is expressed by healthy hepatocytes in the basolateral membrane at very low levels under normal conditions. In contrast, significant upregulation of MRP1 has been observed in several liver diseases, including HCC [38, 39]. MRP1 functions as an export pump, mediating the efflux of many drugs, such as several antitumor agents [40], and may be involved in the poor response of HCC cells to sorafenib [41].

The most commonly used models for providing prognostic information, such as the conventional Tumor-Nodule-Metastasis (TNM) staging method [42] and the HCC-specific Barcelona Clinic Liver Cancer staging algorithm [20], incorporate tumor characteristics, liver function, and clinical status but have limited predictive accuracy [22]. Despite advancements in high-throughput genomic techniques, no accurate biomarkers are currently used in routine clinical practice. However, there has been significant effort in searching for novel biomarkers that could enhance the prognostic capacity provided by clinicopathological variables in the future [43]. Several prognostic molecular signatures have been proposed for HCC patients based on genes related to the immune microenvironment [44], hypoxia and methylation [45], or even bile acid metabolism [46, 47], among others. Although multigene signature panels offering a nomogram based on expression profiles using RT-qPCR have been commercialized as prognostic tools for other tumors, such as breast cancer [48], none are currently available for HCC.

Although several studies have proposed gene signatures as prognostic models in cancer, they have had limited impact on routine clinical practice. To date, only a few tests have received approval for clinical use, such as Oncotype DX, MammaPrint, Prolaris, and Decipher. Among these, the Oncotype DX [49] and MammaPrint [50] tests feature multigene expression signatures that include dozens of genes, primarily used in breast cancer to determine the need for adjuvant chemotherapy. Similarly, the Prolaris and Decipher tests [51] are used in prostate cancer to assess the risk of tumor progression and recurrence. Although these tests are approved and utilized in clinical practice, their application remains limited to specific clinical

scenarios. HCC is among the malignancies with the most proposed genetic signatures in the literature; however, their actual clinical use is very low. A significant limitation is the lack of validation in large, diverse patient groups, as most studies rely on small datasets with only tens or a few hundred cases. Additionally, many of these signatures do not significantly improve prognostic accuracy over traditional clinical variables, making their additional cost or technical complexity unjustifiable. Furthermore, transcriptomic assays have not been clinically validated for most tumor types by agencies like the FDA or EMA. Nonetheless, the proven clinical usefulness of existing tests like Oncotype DX and MammaPrint suggests that implementing genetic signatures for prognostic prediction in HCC is feasible.

5. Conclusions

The HCC genetic signature proposed in this study was derived from the resistome, as a crucial trait of malignancy. It was developed with the aim of being used as a complementary tool in clinical settings to categorize patients. This can help identify those with poorer prognoses and those who might benefit from more aggressive treatments. The resistome-based risk score developed here could also be a useful tool for guiding therapy decisions and customizing treatment plans based on each patient's molecular profile, potentially reducing unnecessary interventions and avoiding chemotherapy in low-risk cases. Supporting this, our analysis showed that patients in group C2 of the TCGA-LIHC cohort who received adjuvant sorafenib therapy had worse outcomes compared to those in group C1. Additionally, the resistome genetic signature might be combined with established clinical biomarkers and patient data to improve prognostic accuracy. In our study, tumor stage and ECOG performance status were identified as independent factors affecting prognosis. In summary, we have created and validated a genetic signature based on three resistome-related genes that accurately predicts the prognosis of patients with HCC. This could serve as a valuable tool in clinical practice for tailoring the most appropriate treatment options for each patient.

Funding

This study received funding from various sources, including the CIBERehd (CB06/04/0023) and Fondo de Investigaciones Sanitarias, Instituto de Salud Carlos III, Spain, co-funded by the European Regional Development Fund/European Social Fund, "Investing in your future" (PI22/00526, and PI23/00681); "Junta de Castilla y Leon" (SA113P23, and GRS 2322/A/21), Spain; and AECC Scientific Foundation (2023/2027), Spain. A.S.-M. received support through predoctoral scholarships (FPU) funded by the Ministry of Science, Innovation and Universities, Spain.

Author Contributions

Conceptualization: A.S.-M., J.J.G.M., O.B.; Funding acquisition: J.J.G.M., O.B.; Investigation: A.S.-M., R.I.R.M., L.M.-B., L.M.G., A.F., M.R., O.B., Writing the original draft: A.S.-M., J.J.G.M., O.B.; Review & editing: A.S.-M., R.I.R.M., L.M.-B., L.M.G., A.F., M.R., J.J.G.M., O.B.

Data Availability Statement

Publicly available datasets were used in this study. These can be found in The Cancer Genome Atlas (TCGA) at https://portal.gdc.cancer.gov/, the cBio Cancer Genomics Portal at http://cbioportal.org, and the Integrative Molecular Database of HCC (HCCDB) at http://lifeome.net/database/hccdb. All data generated or analyzed during this study are included in this article and its supplementary material files. Further enquiries can be directed to the corresponding author.

Declaration of interest

None.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024;**74**(3):229-63. https://doi.org/10.3322/caac.21834.
2. Filho AM, Laversanne M, Ferlay J, Colombet M, Pineros M, Znaor A, et al. The GLOBOCAN 2022 cancer estimates: Data sources, methods, and a snapshot of the cancer burden worldwide. *Int J Cancer* 2025;**156**(7):1336-46. https://doi.org/10.1002/ijc.35278.
3. Vogel A, Cervantes A, Chau I, Daniele B, Llovet JM, Meyer T, et al. Hepatocellular carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of oncology : official journal of the European Society for Medical Oncology* 2018;**29**(Suppl 4):iv238-iv55. https://doi.org/10.1093/annonc/mdy308.
4. Reig M, Forner A, Rimola J, Ferrer-Fabrega J, Burrel M, Garcia-Criado A, et al. BCLC strategy for prognosis prediction and treatment recommendation: The 2022 update. *Journal of hepatology* 2022;**76**(3):681-93. https://doi.org/10.1016/j.jhep.2021.11.018.
5. Lehrich BM, Zhang J, Monga SP, Dhanasekaran R. Battle of the biopsies: Role of tissue and liquid biopsy in hepatocellular carcinoma. *Journal of hepatology* 2024;**80**(3):515-30. https://doi.org/10.1016/j.jhep.2023.11.030.
6. Ducreux M, Abou-Alfa GK, Bekaii-Saab T, Berlin J, Cervantes A, de Baere T, et al. The management of hepatocellular carcinoma. Current expert opinion and recommendations derived from the 24th ESMO/World Congress on Gastrointestinal Cancer, Barcelona, 2022. *ESMO open* 2023;**8**(3):101567. https://doi.org/10.1016/j.esmoop.2023.101567.
7. Golfieri R, Giampalma E, Renzulli M, Cioni R, Bargellini I, Bartolozzi C, et al. Randomised controlled trial of doxorubicin-eluting beads vs conventional chemoembolisation for hepatocellular carcinoma. *British journal of cancer* 2014;**111**(2):255-64. https://doi.org/10.1038/bjc.2014.199.
8. Cappuyns S, Corbett V, Yarchoan M, Finn RS, Llovet JM. Critical Appraisal of Guideline Recommendations on Systemic Therapies for Advanced Hepatocellular Carcinoma: A Review. *JAMA oncology* 2024;**10**(3):395-404. https://doi.org/10.1001/jamaoncol.2023.2677.
9. Fortuny M, Sanduzzi-Zamparelli M, Reig M. Systemic therapies in hepatocellular carcinoma: A revolution? *United European gastroenterology journal* 2024;**12**(2):252-60. https://doi.org/10.1002/ueg2.12510.
10. Bicer F, Kure C, Ozluk AA, El-Rayes BF, Akce M. Advances in Immunotherapy for Hepatocellular Carcinoma (HCC). *Curr Oncol* 2023;**30**(11):9789-812. https://doi.org/10.3390/curroncol30110711.
11. Marin JJG, Macias RIR, Monte MJ, Romero MR, Asensio M, Sanchez-Martin A, et al. Molecular Bases of Drug Resistance in Hepatocellular Carcinoma. *Cancers* 2020;**12**(6). https://doi.org/10.3390/cancers12061663.
12. Marin JJG, Romero MR, Herraez E, Asensio M, Ortiz-Rivero S, Sanchez-Martin A, et al. Mechanisms of Pharmacoresistance in Hepatocellular Carcinoma: New Drugs but Old Problems. *Seminars in liver disease* 2022;**42**(1):87-103. https://doi.org/10.1055/s-0041-1735631.

13. Marin JJ, Romero MR, Briz O. Molecular bases of liver cancer refractoriness to pharmacological treatment. *Current medicinal chemistry* 2010;**17**(8):709-40. https://doi.org/10.2174/092986710790514462.
14. Li G, Chen X, Wang Q, Xu Z, Zhang W, Ye L. The roles of four multi-drug resistance proteins in hepatocellular carcinoma multidrug resistance. *Journal of Huazhong University of Science and Technology Medical sciences = Hua zhong ke ji da xue xue bao Yi xue Ying De wen ban = Huazhong keji daxue xuebao Yixue Yingdewen ban* 2007;**27**(2):173-5. https://doi.org/10.1007/s11596-007-0217-8.
15. Gillet JP, Andersen JB, Madigan JP, Varma S, Bagni RK, Powell K, et al. A Gene Expression Signature Associated with Overall Survival in Patients with Hepatocellular Carcinoma Suggests a New Treatment Strategy. *Molecular pharmacology* 2016;**89**(2):263-72. https://doi.org/10.1124/mol.115.101360.
16. Sakurada T, Yoshikawa M, Sunaga M, Kobayashi E, Satoh N, Yokosuka O, et al. Expression of Drug-Resistant Factor Genes in Hepatocellular Carcinoma Patients Undergoing Chemotherapy with Platinum Complex by Arterial Infusion. *Pharmaceutics* 2010;**2**(3):300-12. https://doi.org/10.3390/pharmaceutics2030300.
17. Martinez-Becerra P, Vaquero J, Romero MR, Lozano E, Anadon C, Macias RI, et al. No correlation between the expression of FXR and genes involved in multidrug resistance phenotype of primary liver tumors. *Molecular pharmaceutics* 2012;**9**(6):1693-704. https://doi.org/10.1021/mp300028a.
18. Galun D, Mijac D, Filipovic A, Bogdanovic A, Zivanovic M, Masulovic D. Precision Medicine for Hepatocellular Carcinoma: Clinical Perspective. *Journal of personalized medicine* 2022;**12**(2). https://doi.org/10.3390/jpm12020149.
19. Wang X, Zhao M, Zhang C, Chen H, Liu X, An Y, et al. Establishment and Clinical Application of the Nomogram Related to Risk or Prognosis of Hepatocellular Carcinoma: A Review. *Journal of hepatocellular carcinoma* 2023;**10**:1389-98. https://doi.org/10.2147/JHC.S417123.
20. Llovet JM, Bru C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. *Seminars in liver disease* 1999;**19**(3):329-38. https://doi.org/10.1055/s-2007-1007122.
21. Chan LK, Tsui YM, Ho DW, Ng IO. Cellular heterogeneity and plasticity in liver cancer. *Seminars in cancer biology* 2022;**82**:134-49. https://doi.org/10.1016/j.semcancer.2021.02.015.
22. Chavez-Villa M, Dominguez-Rosado I. Overview of Current Hepatocellular Carcinoma Staging Systems: Is There an Optimal System? *Surgical oncology clinics of North America* 2024;**33**(1):29-41. https://doi.org/10.1016/j.soc.2023.06.010.
23. Pourbagheri-Sigaroodi A, Fallah F, Bashash D, Karimi A. Unleashing the potential of gene signatures as prognostic and predictive tools: A step closer to personalized medicine in hepatocellular carcinoma (HCC). *Cell biochemistry and function* 2024;**42**(1):e3913. https://doi.org/10.1002/cbf.3913.
24. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* 2012;**2**(5):401-4. https://doi.org/10.1158/2159-8290.CD-12-0095.
25. Gyorffy B. Integrated analysis of public datasets for the discovery and validation of survival-associated genes in solid tumors. *Innovation (Camb)* 2024;**5**(3):100625. https://doi.org/10.1016/j.xinn.2024.100625.

26. Lian Q, Wang S, Zhang G, Wang D, Luo G, Tang J, et al. HCCDB: A Database of Hepatocellular Carcinoma Expression Atlas. *Genomics, proteomics & bioinformatics* 2018;**16**(4):269-75. https://doi.org/10.1016/j.gpb.2018.07.003.
27. Sanchez-Martin A, Sanchon-Sanchez P, Romero MR, Marin JJG, Briz O. Impact of tumor suppressor genes inactivation on the multidrug resistance phenotype of hepatocellular carcinoma cells. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* 2023;**165**:115209. https://doi.org/10.1016/j.biopha.2023.115209.
28. Zhuo JY, Lu D, Tan WY, Zheng SS, Shen YQ, Xu X. CK19-positive Hepatocellular Carcinoma is a Characteristic Subtype. *J Cancer* 2020;**11**(17):5069-77. https://doi.org/10.7150/jca.44697.
29. Terris B, Cavard C, Perret C. EpCAM, a new marker for cancer stem cells in hepatocellular carcinoma. *Journal of hepatology* 2010;**52**(2):280-1. https://doi.org/10.1016/j.jhep.2009.10.026.
30. Zhang Q, Lou Y, Yang J, Wang J, Feng J, Zhao Y, et al. Integrated multiomic analysis reveals comprehensive tumour heterogeneity and novel immunophenotypic classification in hepatocellular carcinomas. *Gut* 2019;**68**(11):2019-31. https://doi.org/10.1136/gutjnl-2019-318912.
31. Suresh A, Dhanasekaran R. Implications of genetic heterogeneity in hepatocellular cancer. *Advances in cancer research* 2022;**156**:103-35. https://doi.org/10.1016/bs.acr.2022.01.007.
32. Chan MY, She WH, Dai WC, Tsang SHY, Chok KSH, Chan ACY, et al. Prognostic value of preoperative alpha-fetoprotein (AFP) level in patients receiving curative hepatectomy-an analysis of 1,182 patients in Hong Kong. *Translational gastroenterology and hepatology* 2019;**4**:52. https://doi.org/10.21037/tgh.2019.06.07.
33. Ma WJ, Wang HY, Teng LS. Correlation analysis of preoperative serum alpha-fetoprotein (AFP) level and prognosis of hepatocellular carcinoma (HCC) after hepatectomy. *World journal of surgical oncology* 2013;**11**:212. https://doi.org/10.1186/1477-7819-11-212.
34. Swift B, Nebot N, Lee JK, Han T, Proctor WR, Thakker DR, et al. Sorafenib hepatobiliary disposition: mechanisms of hepatic uptake and disposition of generated metabolites. *Drug metabolism and disposition: the biological fate of chemicals* 2013;**41**(6):1179-86. https://doi.org/10.1124/dmd.112.048181.
35. Herraez E, Lozano E, Macias RI, Vaquero J, Bujanda L, Banales JM, et al. Expression of SLC22A1 variants may affect the response of hepatocellular carcinoma and cholangiocarcinoma to sorafenib. *Hepatology* 2013;**58**(3):1065-73. https://doi.org/10.1002/hep.26425.
36. Yu J, Wang Z, Zhang H, Wang Y, Li DQ. Survivin-positive circulating tumor cells as a marker for metastasis of hepatocellular carcinoma. *World journal of gastroenterology* 2021;**27**(43):7546-62. https://doi.org/10.3748/wjg.v27.i43.7546.
37. Dai D, Liang Y, Xie Z, Fu J, Zhang Y, Zhang Z. Survivin deficiency induces apoptosis and cell cycle arrest in HepG2 hepatocellular carcinoma cells. *Oncology reports* 2012;**27**(3):621-7. https://doi.org/10.3892/or.2011.1544.
38. Hoffmann K, Shibo L, Xiao Z, Longerich T, Buchler MW, Schemmer P. Correlation of gene expression of ATP-binding cassette protein and tyrosine kinase signaling pathway in patients with hepatocellular carcinoma. *Anticancer research* 2011;**31**(11):3883-90, http://www.ncbi.nlm.nih.gov/pubmed/22110214.

39. Bonin S, Pascolo L, Croce LS, Stanta G, Tiribelli C. Gene expression of ABC proteins in hepatocellular carcinoma, perineoplastic tissue, and liver diseases. *Mol Med* 2002;**8**(6):318-25, http://www.ncbi.nlm.nih.gov/pubmed/12428063.
40. He SM, Li R, Kanwar JR, Zhou SF. Structural and functional properties of human multidrug resistance protein 1 (MRP1/ABCC1). *Current medicinal chemistry* 2011;**18**(3):439-81. https://doi.org/10.2174/092986711794839197.
41. Chang YS, Su CW, Chen SC, Chen YY, Liang YJ, Wu JC. Upregulation of USP22 and ABCC1 during Sorafenib Treatment of Hepatocellular Carcinoma Contribute to Development of Resistance. *Cells* 2022;**11**(4). https://doi.org/10.3390/cells11040634.
42. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA: a cancer journal for clinicians* 2017;**67**(2):93-9. https://doi.org/10.3322/caac.21388.
43. Hoshida Y, Toffanin S, Lachenmayer A, Villanueva A, Minguez B, Llovet JM. Molecular classification and novel targets in hepatocellular carcinoma: recent advancements. *Seminars in liver disease* 2010;**30**(1):35-51. https://doi.org/10.1055/s-0030-1247131.
44. Su L, Zhang G, Kong X. A Novel Five-Gene Signature for Prognosis Prediction in Hepatocellular Carcinoma. *Frontiers in oncology* 2021;**11**:642563. https://doi.org/10.3389/fonc.2021.642563.
45. Ren M, Fan B, Cao G, Zong R, Feng L, Sun H. Exploration and validation of a combined Hypoxia and m6A/m5C/m1A regulated gene signature for prognosis prediction of liver cancer. *BMC genomics* 2023;**24**(1):776. https://doi.org/10.1186/s12864-023-09876-3.
46. El-Mir MY, Badia MD, Luengo N, Monte MJ, Marin JJ. Increased levels of typically fetal bile acid species in patients with hepatocellular carcinoma. *Clin Sci (Lond)* 2001;**100**(5):499-508, http://www.ncbi.nlm.nih.gov/pubmed/11294690.
47. Qu Y, Gong X, Zhao Z, Zhang Z, Zhang Q, Huang Y, et al. Establishment and Validation of Novel Prognostic Subtypes in Hepatocellular Carcinoma Based on Bile Acid Metabolism Gene Signatures Using Bulk and Single-Cell RNA-Seq Data. *International journal of molecular sciences* 2024;**25**(2). https://doi.org/10.3390/ijms25020919.
48. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. *The New England journal of medicine* 2018;**379**(2):111-21. https://doi.org/10.1056/NEJMoa1804710.
49. Verrill M, Lux MP, Gligorov J, Geisler J, Duchnowska R, Elsberger B, et al. Genomic tests to guide management of breast cancer in Europe: regulation, reimbursement, adoption, and challenges. *Future Oncol* 2025:1-11. https://doi.org/10.1080/14796694.2025.2577335.
50. Glas AM, Floore A, Delahaye LJ, Witteveen AT, Pover RC, Bakx N, et al. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC genomics* 2006;**7**:278. https://doi.org/10.1186/1471-2164-7-278.
51. Nguyen HG, Welty CJ, Cooperberg MR. Diagnostic associations of gene expression signatures in prostate cancer tissue. *Curr Opin Urol* 2015;**25**(1):65-70. https://doi.org/10.1097/MOU.0000000000000131.

Figure Legends

Figure 1. Classification of the HCC patients from the TCGA-LIHC cohort into two main clusters (C1 and C2) and two subgroups (C2a and C2b) based on the expression profiles of genes linked to the resistome. The heatmap shows unsupervised hierarchical clustering of 371 HCC samples and 81 genes using RNA-seq data (log_2 FPKM-UQ). Clustering was carried out using Euclidean distance and Ward's minimum variance method. Columns represent tumor samples, and rows correspond to genes.

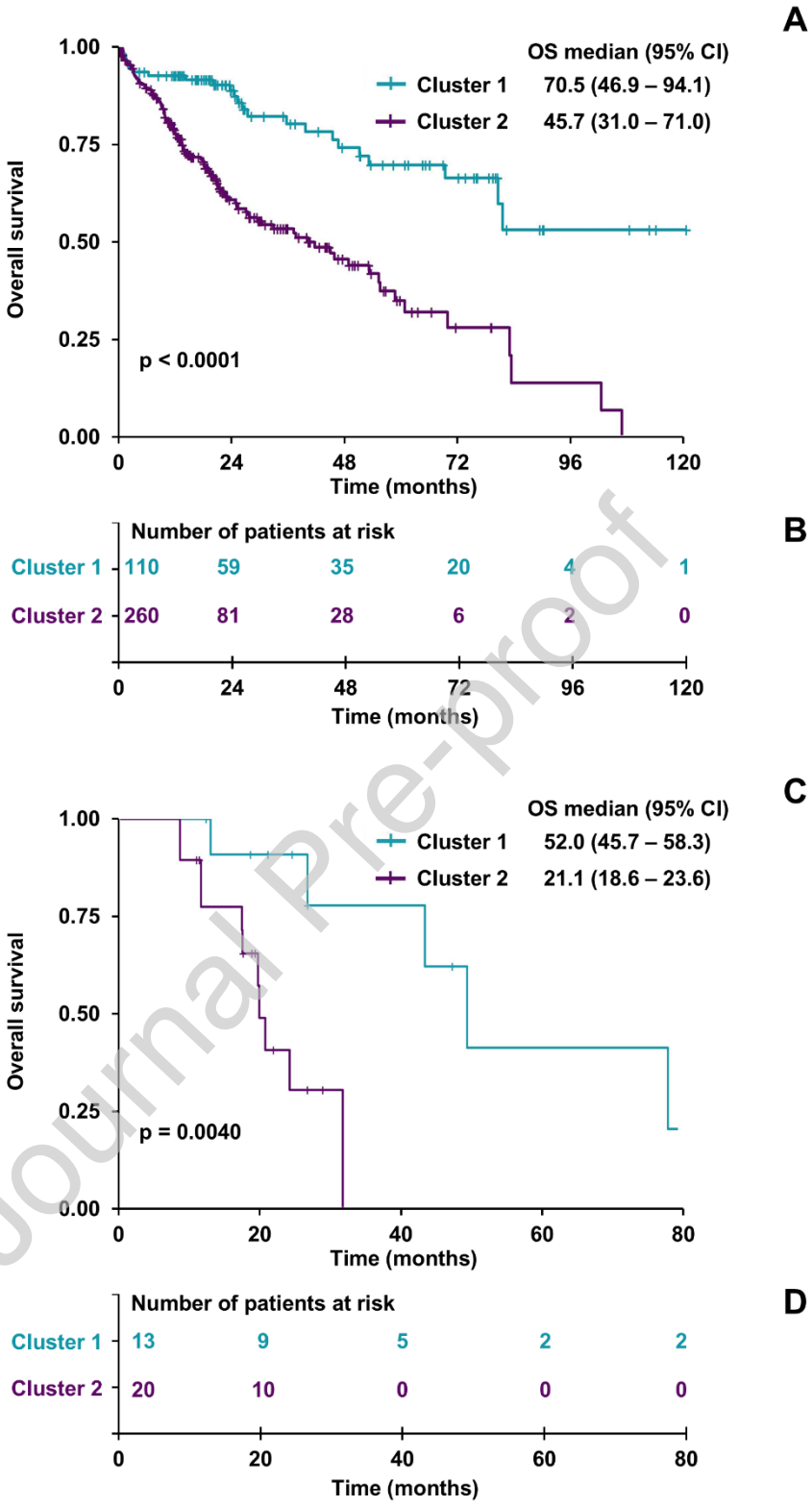


Figure 2

Figure 2. Impact of the expression profile of resistome-associated genes on the overall survival (OS) of HCC patients from the TCGA-LIHC cohort. Kaplan-Meier curves compare OS of 370 HCC patients classified into Cluster 1 or Cluster 2 based on the resistome (A). A similar analysis was conducted with only 33 members of both clusters treated with sorafenib (C1sor and C2sor, respectively) (C). The log-rank test was used to compare survival times between groups. (B, D) Risk tables show the number of patients at risk for each group at different time points. Impact of clinicopathological and molecular characteristics on the prognosis of HCC patients from the TCGA-LIHC cohort.

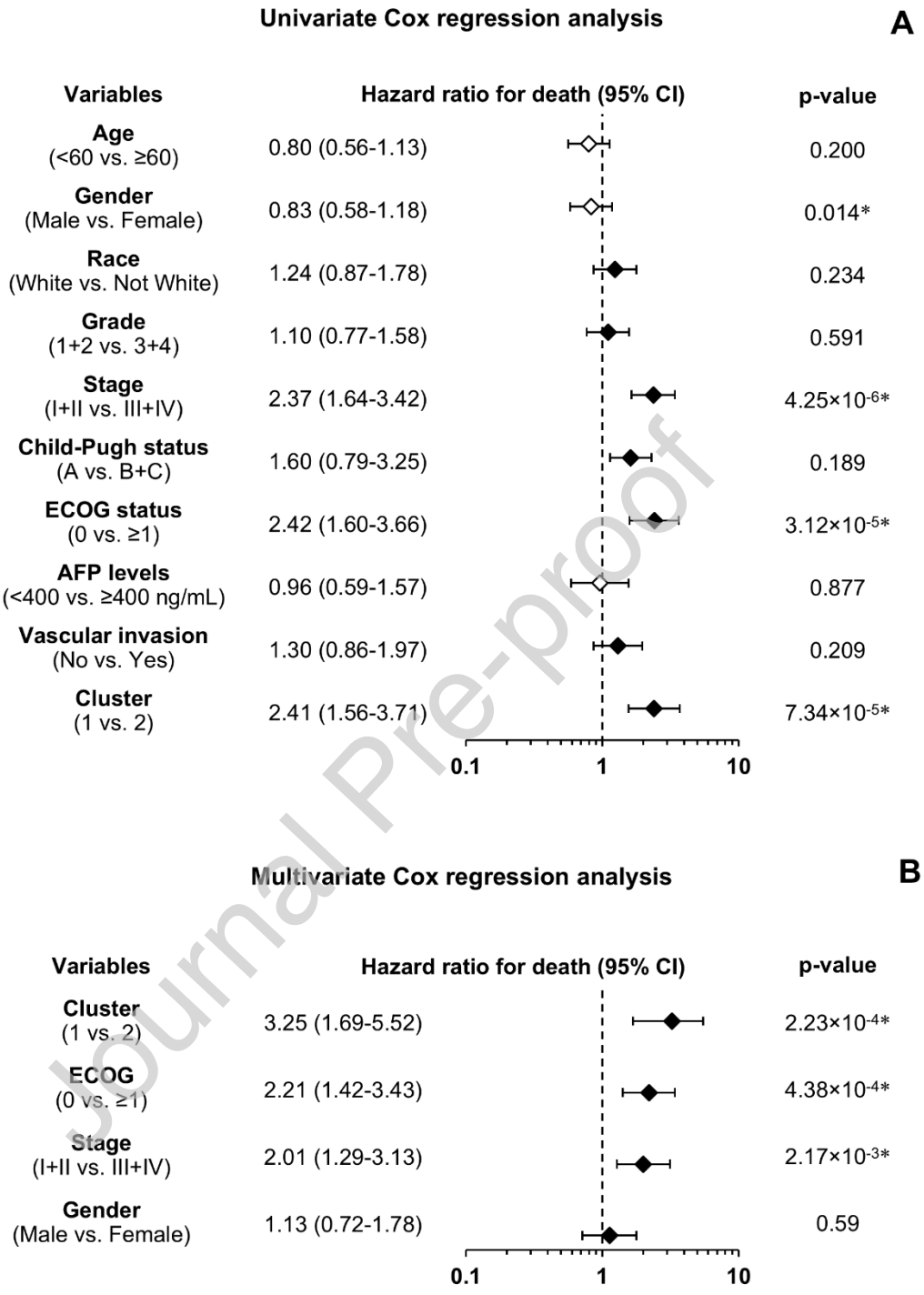


Figure 3

Figure 3. Impact of resistome-associated gene expression on the prognosis of HCC patients from the TCGA-LIHC cohort. Univariate analysis using the Cox regression model for overall survival (A) and multivariate analysis (B) with variables found significantly different in the univariate analysis. The number of patients ranged from 246 to 370, depending on the variable. The forest plot displays hazard ratios for death (diamonds) and 95% confidence intervals (CI) (error bars). AFP, α-fetoprotein; ECOG, Eastern Cooperative Oncology Group.

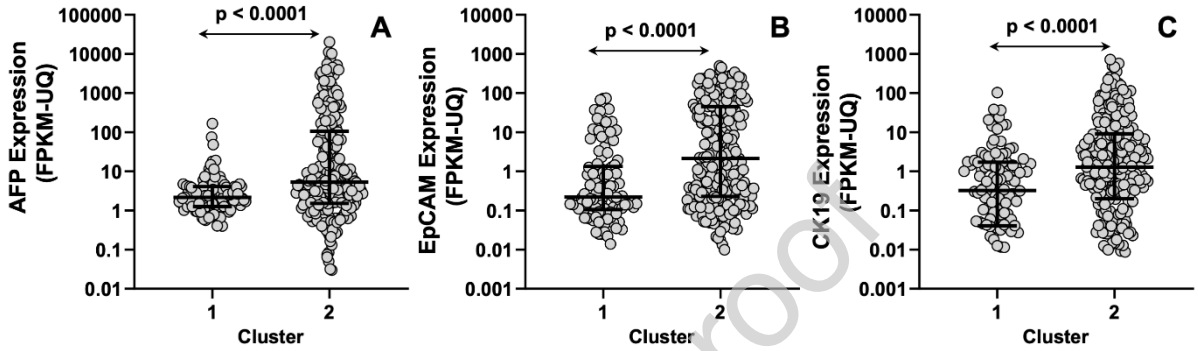


Figure 4. *In silico* analysis of AFP (A), EpCAM (B), and CK19 (C) expression in patients with hepatocellular carcinoma (HCC) classified into Cluster 1 or Cluster 2 based on their resistome profile. The analysis includes mRNA expression levels measured by RNA sequencing (RNA-seq) in tumor samples (n = 370), obtained from The Cancer Genome Atlas (TCGA) database. Data are shown as the median with interquartile range (IQR) (lines) and individual data points (circles). The comparison between C1 and C2 was conducted using the Mann–Whitney U test for unpaired samples ($p < 0.0001$).

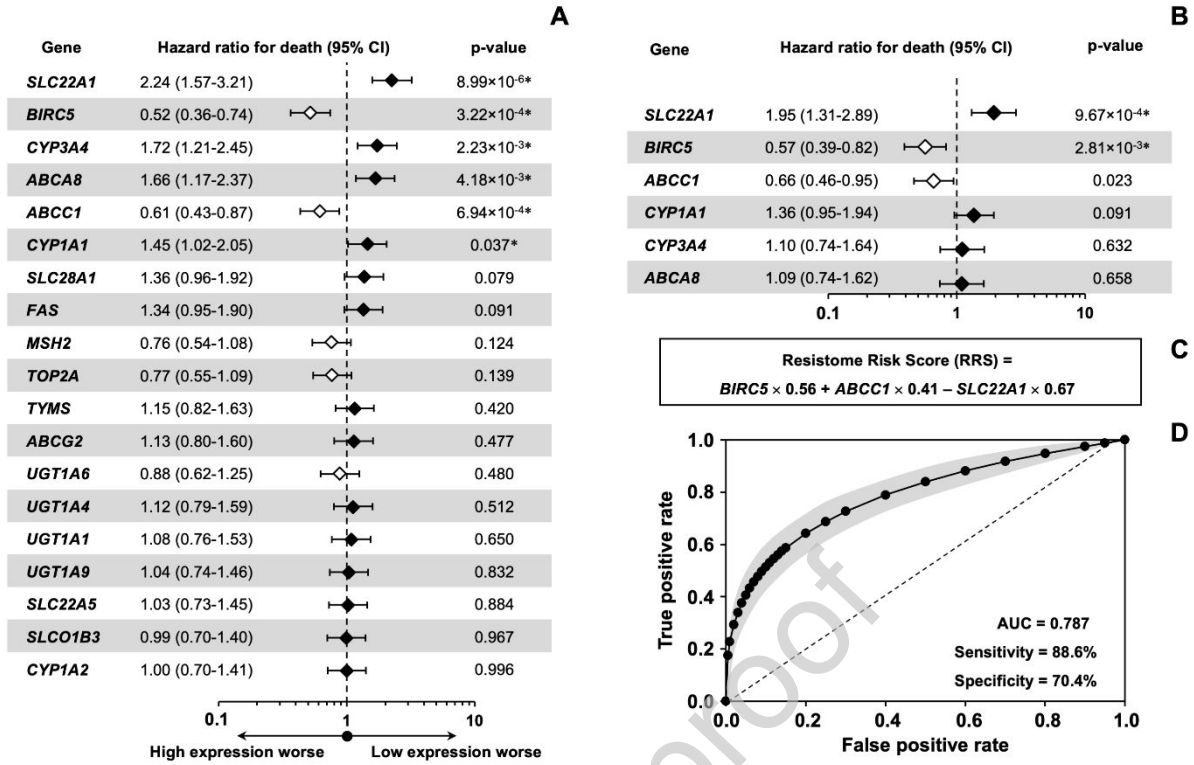
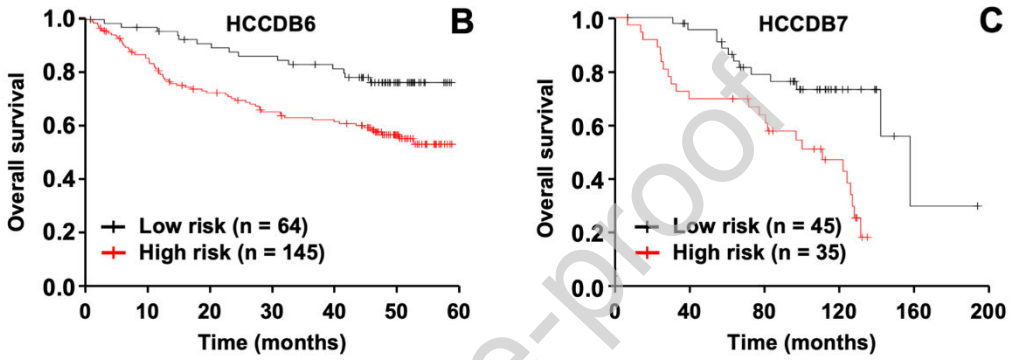
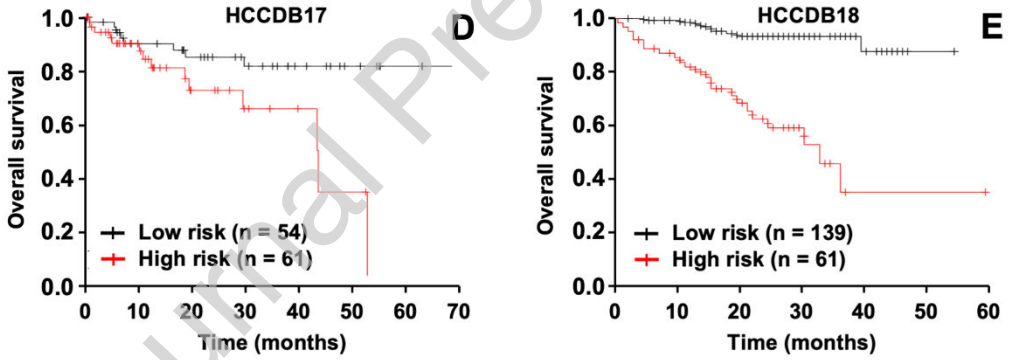


Figure 5. Impact of resistome-associated gene expression on the prognosis of HCC patients from the TCGA-LIHC cohort. Univariate analysis using the Cox regression model for overall survival (A) and multivariate analysis (B) with variables found significantly different in the univariate analysis. The forest plot displays hazard ratios for death (diamonds) and 95% confidence intervals (CI) (error bars). (C) The formula for calculating the resistome risk score (RRS) is derived from the coefficients obtained by the multivariate Cox regression model, which measure the impact of covariates and are multiplied by the expression level of each gene. (D) Maximum likelihood estimation of binormal receiver operating characteristic (ROC) curve to validate the risk score's performance in predicting OS. Displayed in gray are the lower and upper limits of the asymmetric 95% CI for the fraction of true positives across a range of false-positive fractions on the estimated ROC curve. AUC, area under the ROC curve.

A

Dataset	Source	Measurement technique	Number of HCC	Median OS (months)		
				Low-risk group	High-Risk group	Log-rank p-value
HCCDB6	GSE14520	Affymetrix v2.0 microarray	209	52.7	16.5	2.7×10^{-3}
HCCDB7	GSE10143	Illumina DASL assay	80	152.7	96.8	8.5×10^{-4}
HCCDB17	GSE76427	Illumina v4.0 microarray	115	75.5	47.5	0.015
HCCDB18	ICGC-LIRI-JP	RNA sequencing	200	72.0	39.5	1.8×10^{-10}





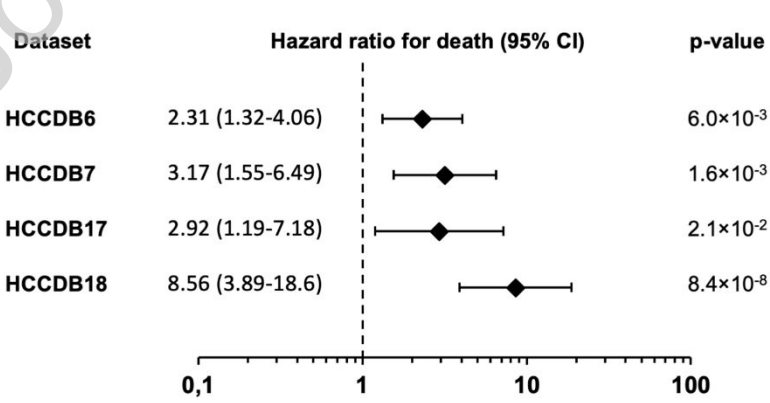


Figure 6. *In silico* validation of the HCC prognostic model using various publicly available databases. (A) Analysis of the relationship between the resistome risk score (RRS) and survival of HCC patients across different public databases. The log-rank test was used to compare survival times between two patient groups classified according to the RRS. The optimal cutoff point for gene expression, which yielded the maximum survival difference between the groups with the lowest log-rank p-value, was used. OS stands for overall survival. (B-E) Kaplan-Meier curves showing OS of HCC patients from different datasets, divided into low or high RRS groups. (F) Univariate analysis with the Cox regression model to evaluate the effect of RRS on OS in HCC patients. The forest plot displays hazard ratios for death (diamonds) and 95% confidence intervals (CI) (error bars).

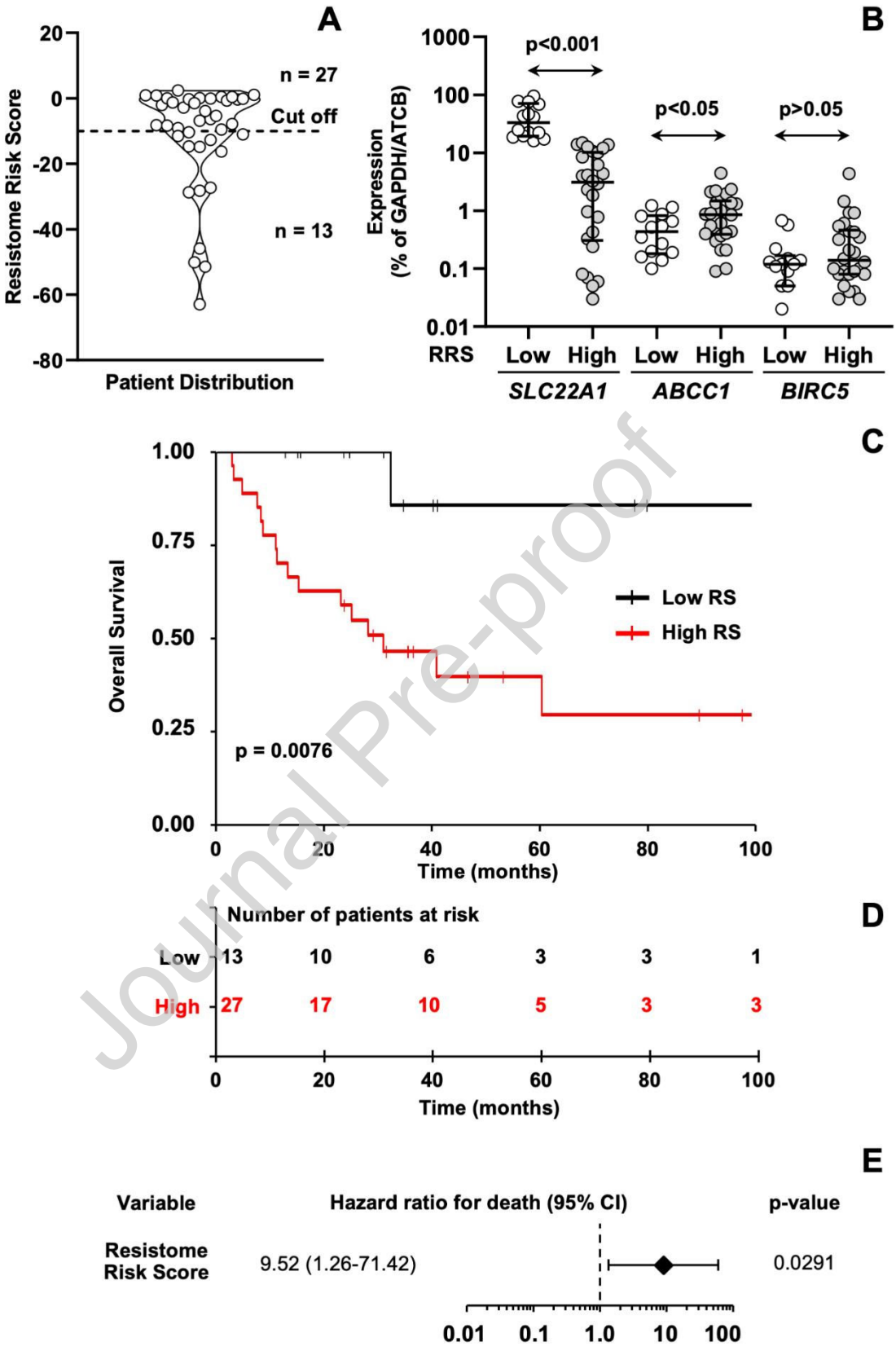


Figure 7. (A) Violin plot of resistome risk scores (RRS) for each sample from the validation cohort. The best expression cutoff point that resulted in the maximum survival difference between the two groups with the lowest log-rank p-value was used. (B) Expression of *SLC22A1*, *ABCC1*, and *BIRC5* genes in HCC patients from the validation cohort (n = 40) classified into low or high RRS groups. mRNA levels were determined by RT-qPCR using TLDAs and normalized based on GAPDH and ACTB expression levels. Individual data, median, and interquartile range (IQR) are represented. (C) Kaplan-Meier curves comparing overall survival (OS) of 40 HCC patients from the validation cohort classified into low or high RRS groups. Log-rank test was used to compare survival time between two groups of patients. (D) Risk tables showing the number of patients at risk for each group at different times. (E) Univariate analysis using the Cox regression model to evaluate the impact of the RRS on the OS of HCC patients. The forest plot includes hazard ratios for death (diamond) and the 95 % confidence interval (CI) (error bars).

Table 1. Comparison of the expression of genes involved in resistome between the two clusters of tumor samples from HCC patients identified in the TCGA-LIHC cohort.

Gene	Median in cluster 1 (FPKM-UQ)	Median in cluster 2 (FPKM-UQ)	Decrease fold-change (cluster 2 vs. 1)	p-value
ABCA8	5.6	0.11	51.3	$<10^{-15}$
SLC22A1	68.1	1.5	46.4	$<10^{-15}$
CYP1A1	24.3	1.2	19.6	$<10^{-15}$
UGT1A4	36.5	2.0	18.3	$<10^{-15}$
CYP1A2	30.6	3.1	10.0	$<10^{-15}$
CYP3A4	91.4	9.2	9.9	$<10^{-15}$
ABCG2	10.8	1.3	8.2	3.48×10^{-13}
UGT1A6	6.3	0.92	6.8	8.00×10^{-14}
FAS	4.6	1.1	4.1	6.17×10^{-12}
UGT1A9	18.9	5.0	3.8	$<10^{-15}$
UGT1A1	24.2	6.8	3.6	4.02×10^{-7}
SLC28A1	17.3	6.2	2.8	6.58×10^{-8}
SLCO1B3	11.2	4.4	2.6	$<10^{-15}$
KDR	8.1	4.9	1.7	4.73×10^{-7}
SLC22A3	16.5	10.3	1.6	1.48×10^{-10}
ABCB1	16.4	10.4	1.6	6.12×10^{-8}
CES2	56.0	36.1	1.5	$<10^{-15}$
PMS2	0.31	0.20	1.5	9.96×10^{-3}
ABCC4	1.1	0.72	1.5	1.20×10^{-3}
SLC47A1	24.7	16.3	1.5	8.40×10^{-14}
CES1	113.8	80.1	1.4	$<10^{-15}$
MYC	15.0	10.6	1.4	2.20×10^{-3}
SLCO1B1	43.5	32.5	1.3	2.73×10^{-13}
SLCO2B1	24.4	18.4	1.3	5.97×10^{-7}
ABCC2	25.9	20.0	1.3	1.69×10^{-3}
ALDH1A1	95.2	74.8	1.3	$<10^{-15}$
SLC31A1	27.4	22.2	1.2	1.06×10^{-9}
NFE2L2	21.6	17.6	1.2	5.70×10^{-14}
YAP1	13.2	11.1	1.2	1.42×10^{-4}
MET	22.6	19.5	1.2	1.58×10^{-4}
BCL2L1	27.7	24.1	1.1	2.51×10^{-6}
PTEN	11.8	10.3	1.1	6.24×10^{-5}
NFKB1	9.1	8.1	1.1	2.41×10^{-2}
XIAP	9.0	8.0	1.1	4.00×10^{-4}
MTDH	26.3	24.1	1.1	1.39×10^{-2}
CFLAR	4.6	4.3	1.1	3.50×10^{-2}
GSTA1	87.1	80.1	1.1	2.37×10^{-2}

The Mann-Whitney U test was used to compare gene expression between the two groups. Only the expression of genes for which statistically significant differences were found are shown. HCC, hepatocellular carcinoma. FPKM-UQ, fragments per kilobase of transcript per million mapped reads upper quartile.

Table 2. Comparison of the expression of genes involved in resistome between the two clusters of tumor samples from HCC patients identified in the TCGA-LIHC cohort.

Gene	Median in cluster 1 (FPKM-UQ)	Median in cluster 2 (FPKM-UQ)	Increase fold-change (cluster 2 vs. 1)	p-value
TYMS	0.11	1.2	10.4	1.44×10^{-10}
BIRC5	1.8	10.4	5.6	$<10^{-15}$
TOP2A	2.8	11.8	4.2	$<10^{-15}$
MSH2	0.22	0.87	3.9	2.88×10^{-11}
ABCC1	1.7	4.2	2.5	3.00×10^{-4}
SLC22A5	0.14	0.32	2.3	6.00×10^{-4}
UPP1	3.7	5.7	1.5	2.46×10^{-6}
SLC29A2	6.1	9.0	1.5	2.05×10^{-5}
TGFB1	8.9	12.9	1.5	1.17×10^{-5}
PMS1	1.0	1.4	1.4	8.17×10^{-4}
SOX9	8.0	10.7	1.3	2.30×10^{-3}
GSTP1	12.7	16.5	1.3	5.55×10^{-4}
MVP	21.9	28.3	1.3	1.95×10^{-5}
DPYD	8.2	10.2	1.2	2.02×10^{-2}
DHFR	4.6	5.6	1.2	1.64×10^{-2}
BAX	20.5	24.9	1.2	4.90×10^{-8}
MSH6	4.6	5.6	1.2	4.97×10^{-4}
ERCC1	8.4	10.0	1.2	3.87×10^{-6}
TP53	8.3	9.7	1.2	1.43×10^{-2}
UNG	18.9	20.5	1.1	1.24×10^{-3}
RPL6	49.5	53.5	1.1	3.18×10^{-7}

The Mann-Whitney U test was used to compare gene expression between the two groups. Only the expression of genes for which statistically significant differences were found are shown. HCC, hepatocellular carcinoma. FPKM-UQ, fragments per kilobase of transcript per million mapped reads upper quartile.

Table 3. Comparison of clinical and pathologic characteristics between the two clusters of tumor samples from HCC patients identified in the TCGA-LIHC cohort.

	Cluster		
Characteristics	**1 (n=110)**	**2 (n=261)**	**p-value**
Age, median (95% CI)	65 (62-67)	59 (58-61)	0.0003
Sex, n (%)			0.0002
Male	90 (81.8)	165 (61.8)	
Female	20 (18.2)	102 (38.2)	
Race, n (%)			0.061
White	62 (59.0)	125 (47.7)	
Asian	36 (34.3)	125 (47.7)	
Black, American Indian, or Alaskan Native	7 (6.7)	12 (4.6)	
NA	5	5	
Underlying disease, n (%)			0.031[a]
Alcoholic	45 (34.1)	73 (25.6)	
Hepatitis B	30 (22.7)	78 (27.5)	
Hepatitis C	23 (17.4)	33 (11.6)	
NAFLD	8 (6.1)	12 (4.2)	
Hemochromatosis	3 (2.3)	5 (1.8)	
Others	2 (1.5)	11 (3.9)	
No history of disease	21 (15.9)	72 (25.4)	
Vascular invasion, n (%)			0.70
No	66 (67.3)	144 (64.9)	
Yes	32 (32.7)	78 (35.1)	
NA	11	45	
Tumor grade, n (%)			0.043[b]
G1	62 (57.4)	123 (46.0)	
G2	27 (25.0)	68 (25.5)	
G3	17 (15.7)	64 (24.0)	
G4	1 (0.9)	12 (4.5)	
NA	3	0	
Child-Pugh status, n (%)			0.78[b]
A	71 (91.0)	153 (91.1)	
B	7 (9.0)	14 (8.3)	
C	0 (0.0)	1 (0.6)	
NA	32	101	
ECOG performance status, n (%)			0.011
0	52 (59.1)	113 (55.4)	
1	33 (37.5)	53 (26.0)	
2	2 (2.3)	24 (11.8)	
3	1 (1.1)	11 (5.4)	
4	0 (0.0)	3 (1.4)	
NA	21	63	
Stage (TNM classification), n (%)			0.035
I	59 (59.0)	116 (45.8)	
II	24 (24.0)	63 (24.9)	
III	17 (17.0)	69 (27.3)	
IV	0 (0)	5 (2.0)	
NA	10	14	
AFP (ng/mL), n (%)			0.0011
<400	77 (89.5)	143 (71.8)	

>400	9 (10.5)	56 (28.2)
NA	24	70

AFP, alpha-fetoprotein; ECOG, Eastern Cooperative Oncology Group; NA, not available. To calculate the statistical significance of the differences between the groups Pearson's χ^2, Fisher's exact, or Mann-Whitney U tests were used as appropriate. [a] Less frequent underlying diseases (hemochromatosis and NAFLD) were grouped in the "Others" category to apply Pearson's χ^2 test. [b] Tumor grades G3 and G4, Child-Pugh B and C, ECOG 3 and 4 were grouped to apply the Pearson's χ^2 test.

Table 4. Analysis of the association between the expression of genes involved in resistome and survival of HCC patients from the TCGA-LIHC cohort (n=371).

		Median OS (months)	
Gene	**Log-rank p-value**	**Low expression cohort**	**High expression cohort**
BIRC5	7.4×10^{-7}	71.0	22.9
SLC22A1	9.2×10^{-7}	30.0	84.4
MSH2	7.0×10^{-5}	70.5	25.6
TYMS	8.4×10^{-5}	71.0	30.0
ABCC1	8.9×10^{-5}	84.4	42.4
TOP2A	1.00×10^{-4}	71.0	30.0
CYP3A4	0.0017	46.6	81.9
FAS	0.0058	47.4	84.7
ABCA8	0.0061	49.7	81.9
SLCO1B3	0.0065	17.8	26.7
SLC28A1	0.0077	30.0	59.7
UGT1A4	0.016	45.7	81.9
UGT1A6	0.018	81.9	46.2
ABCG2	0.068	46.6	81.9
CYP1A2	0.074	45.7	59.7
SLC22A5	0.081	81.9	49.7
UGT1A9	0.179	46.6	70.5
UGT1A1	0.242	54.1	70.5
CYP1A1	0.313	59.7	46.6

Based on the expression of each gene, expressed as FPKM-UQ (Fragments Per Kilobase of transcript per Million mapped reads Upper Quartile), patients were classified into two groups and the association between survival and gene expression was determined. The Log-rank test was used to compare the survival time between the two groups. The best expression cutoff that produced the maximum difference in survival between the two groups with the lowest Log-rank p-value was used. OS, overall survival.

Graphical Abstract