Identification of amino acid metabolic signature to predict prognosis and guide clinical therapy in patients with Hepatocellular carcinoma

Dalin Chen¹, HuiZhong Jiang¹, Bing Yang¹, Bin Guo¹, Dan Li¹, Guangze Ye¹, Zhu Yang¹, Dongxin Tang¹, and FengXi Long¹

¹Guizhou University of Traditional Chinese Medicine

September 8, 2023

Abstract

Background The high heterogeneity of Hepatocellular carcinoma (HCC) has led to poor clinical outcomes. The critical role of amino acid metabolic reprogramming in tumor growth is gradually emerging. However, amino acid metabolism in HCC is less studied, and the mechanisms still need to be clarified. Methods We acquired transcriptome information on HCC patients from public databases. Amino acid metabolism-related genes (ACRGs) were used as prognostic markers. We built the prognosis-related ACRG_score model using Univariate/Multivariate COX and Lasso regression analyses following stratification by consensus clustering. Furthermore, we assigned HCC patients with high ACRG expression to the high-risk category and others to the low-risk category. We compared clinical characteristics, immune cell infiltration, somatic mutations, and immune checkpoint (IC) genes among the groups. Finally, drug sensitivity and molecular docking analyses were used to identify therapeutic candidates targeting the essential ACRG target proteins. Result We constructed a six-gene ACRG_score model that better predicted the survival prognosis for liver cancer patients, and we validated it using internal and external datasets. In HCC patients, ACRG_score are associated with clinicopathological characteristics and have proven to be an independent prediction factor. Nomogram and calibration curves illustrated the model could correctly forecast patient prognosis. In addition, immune infiltration, Tumor Mutational Burden (TMB), and ACRG_score were revealed to be significantly correlated. IC genes were also present. According to immunohistochemical analysis, HCC tissues had higher EZH2, SLC2A1, and SPP1 expression levels than normal tissues. Additionally, we identified seven candidate small-molecule medications that may bind four of the ACRG essential target proteins. Conclusions: In this study, the ACRG_score model was created to analyze the prognosis, TMB, IC, and therapy responsiveness for HCC patients. This model can predict patient prognosis, guide immunotherapy, provide clinical dosing suggestions, and supply valuable tools for individualized patient treatment.

Hosted file


Hosted file

Identification of amino acid metabolic signature to predict prognosis and guide clinical therapy in patients with Hepatocellular carcinoma

Da-lin Chen¹, Hui-Zhong Jiang¹, Bing Yang², Bin Guo³, Dan Li², Guang-Ze Ye², Zhu Yang³, Feng-Xi Long³, Dong-Xin Tang²

¹School of Basic Medicine, Guizhou University of Traditional Chinese Medicine, GuiYang, China.
²The First Affiliated Hospital of Guizhou University of Traditional Chinese Medicine, Guiyang, China.
³Guizhou University of Traditional Chinese Medicine, GuiYang, China.

da-lin Chen and Hui-Zhong Jiang should be considered joint first authors.
Bing Yang, Bin Guo, Dan Li, Guang-Ze Ye, Zhu Yang, Feng-Xi Long, Dong-Xin Tang should be considered joint senior authors

Correspondence
Feng-Xi Long, Guizhou University of Traditional Chinese Medicine, GuiYang, China. Email:longfx0321@163.com
Dong-Xin Tang, The First Affiliated Hospital of Guizhou University of Traditional Chinese Medicine, Guiyang, China. Email:tangdongxin@sina.com

Abstract:

Background The high heterogeneity of Hepatocellular carcinoma (HCC) has led to poor clinical outcomes. The critical role of amino acid metabolic reprogramming in tumor growth is gradually emerging. However, amino acid metabolism in HCC is less studied, and the mechanisms still need to be clarified.

Methods We acquired transcriptome information on HCC patients from public databases. Amino acid metabolism-related genes (ACRGs) were used as prognostic markers. We built the prognosis-related ACRG_score model using Univariate/Multivariate COX and Lasso regression analyses following stratification by consensus clustering. Furthermore, we assigned HCC patients with high ACRG expression to the high-risk category and others to the low-risk category. We compared clinical characteristics, immune cell infiltration, somatic mutations, and immune checkpoint (IC) genes among the groups. Finally, drug sensitivity and molecular docking analyses were used to identify therapeutic candidates targeting the essential ACRG target proteins.

Result We constructed a six-gene ACRG_score model that better predicted the survival prognosis for liver cancer patients, and we validated it using internal and
external datasets. In HCC patients, ACRG_score are associated with clinicopathological characteristics and have proven to be an independent prediction factor. Nomogram and calibration curves illustrated the model could correctly forecast patient prognosis. In addition, immune infiltration, Tumor Mutational Burden (TMB), and ACRG_score were revealed to be significantly correlated. IC genes were also present. According to immunohistochemical analysis, HCC tissues had higher EZH2, SLC2A1, and SPP1 expression levels than normal tissues. Additionally, we identified seven candidate small-molecule medications that may bind four of the ACRG essential target proteins.

**Conclusions:** In this study, the ACRG_score model was created to analyze the prognosis, TMB, IC, and therapy responsiveness for HCC patients. This model can predict patient prognosis, guide immunotherapy, provide clinical dosing suggestions, and supply valuable tools for individualized patient treatment.

**Keywords**

Hepatocellular carcinoma, Amino acid metabolic, Prognosis model, Immune infiltration, Tumor microenvironment.

1 Introduction

HCC is the third most common cancer-related cause of death and the fifth most prevalent illness on the worldwide scale[1]. If the disease growth is not adequately controlled, the number of new cases and fatalities from HCC will rise by more than 50% over the next 20 years[2-3]. For the past few years, multi-kinase inhibitors like sorafenib and lenvatinib are employed as first-line therapy in treating advanced liver cancer. Yet, toxicity and drug resistance frequently result in little change in the survival outlook of patients[4]. There is a high incidence of immunological adverse events despite the benefits of immunotherapy in treating HCC having been clinically demonstrated and gradually playing a significant role[5]. Given the heterogeneity in liver cancer, Finding individualized tools in evaluating prognosis and predicting treatment effect is a crucial step toward improving clinical diagnosis and therapy.

Amino acids are one of the important substances for maintaining cellular functions and are closely related to the process of tumor development[6]. The process of altering amino acid metabolic pathways to adapt survival is an vital feature of tumor growth[7-8]. According to several studies, reprogramming of amino acid metabolism can boost biosynthesis, reduce reactive oxygen species produced by fast development, and increase the energy supply for malignancies[9-11]. In addition, amino acid metabolism reprogramming can change the tumor microenvironment (TME) and, as a result, modify immune cell function, enabling tumor cells to evade immune system identification[12]. In recent years, amino acid metabolism has been a hotspot in cancer research, and the development of related targeted drugs has been very successful[13].
Considering the role of tumor-related amino acid metabolism in therapy, the promise of discovering valuable markers of amino acid metabolism to build clinical prognostic models is particularly relevant. Therefore, we created a 6-gene ACRG_score and forecast immune cell infiltration, Tumor Mutational Burden (TMB), and immunotherapy effectiveness based on the ACRG_score. Furthermore, we compared the ACRG level variation between HCC and healthy liver tissues using immunohistochemistry. Meanwhile, molecular docking and drug sensitivity analyses were used to screen potential medications. Finally, we successfully constructed a prognostic model capable of predicting the multidimensional characteristics of HCC patients.

2 Materials and Methods

2.1 Data acquisition


2.2 Identification of ACRG-DEGs in tumor and paracancer samples

We obtained a gene set encompassing 460 genes relevant to amino acid metabolism from the MSigDB database. When comparing tumor and para-cancer samples, we set the "limma" R package's criteria of $|\log FC| > 1$ and FDR < 0.05 to look for ACRG-DEGs. Besides, we used the "clusterProfiler" R package to understand further the potential molecular functions which perform GO and KEGG functional enrichment analysis in ACRG-DEGs.

2.3 Consensus Clustering Analysis of ACRG-DEGs

We analyzed the expression of ACRG-DEGs in the sample tissue and divided the patients into distinct molecular subtypes using the "ConsensusClusterPlus" R package. Meanwhile, we repeated the process a thousand times to confirm classification stability. Effective downscaling and model identification for each cluster subtype of data and grouping visualization was accomplished using the principal component analysis (PCA) method. Additionally, we evaluated the overall survival (OS) rate difference between subgroups employing the "survival" R package. We visualized the relationship between various molecular subtypes and clinicopathological features by applying the "pheatmap" R package. Utilizing the "GSVA" R algorithm, we conducted Gene Set Variation Analysis (GSVA) to analyze the two subgroups' biological features[15].
2.4 Construction of amino acid metabolism-related prognostic ACRG_Score

We conducted differential genetic analysis in both patient groupings and discovered 172 differential genes. By using univariate Cox regression analysis, we identified 66 genes that are related. Six candidate genes were selected following the multivariate Cox analysis and Lasso regression analysis screening of 66 genes, and the ACRG_score was subsequently created. ACRG_score=gene expression (1) × Corresponding coefficient (1) + gene expression (2) × Corresponding coefficient (2) +⋯⋯+ gene expression [n] × Corresponding coefficient [n]. Randomly divide all patients into a training set (n=299), testing set (n=299), and entire set (n=598), and include separate TCGA set and GEO set as external validation cohorts. Using ACRG_score median as dividing line, we separated each cohort's low- and high-risk categories. We contrasted the survival difference of categories on various cohorts utilizing the "survminer" R package and evaluated the ACRG_score model's prediction power using Receiver Operating Characteristic (ROC) curves.

2.5 ACRG_score clinical evaluation

The survival disparities among two risk categories patients were examined in several clinical subgroups to assess the clinical efficacy of ACRG_score. In addition, to determine if the ACRG_score may serve as an independent prognostic factor, we conducted uniCox and multiCox analyses in several cohorts.

2.6 Construction and validation of nomogram

We utilized the "rms" R package\[16\] to build the Nomogram by adding the ACRG_score and clinicopathological parameters such as gender, age, and tumor stage into the reference to forecast HCC patient OS at 1-, 3-, and 5 years. Furthermore, we evaluated the nomogram model accurateness utilizing the ROC and calibration plot.

2.7 Immunolandscape analysis

We synthetically assessed the relationship between ACRG and the immunological landscape with various tools. First, the ESTIMATE algorithm\[17\] evaluates ESTIMATEScore, stromal score, and immunescore to determine the purity of the tumor. Second, using the CIBERSORT algorithm, we evaluated 22 different immune cell types\[18\]. Third, compare immune cell infiltration relative levels among the two risk groups using single sample gene set enrichment analysis (ssGSEA)\[19\]. In addition, We applied the tools TIMER, XCELL, QUANTISEQ, MCPCOUNTER, EPIC, and CIBERSORT-ABS to measure the relative level of immune cell infiltration. We constructed a heat map to display the outcomes. Furthermore, we examined the disparity of 47 immune checkpoint (IC) genes among two risk categories to assess the potential effectiveness of ICs-related treatment. Last but not least, we obtained Immune Cell Proportion Score (IPS) data from The Cancer Imaging Archive (TCIA) database, and predictive assessment of anti-PD-1 and anti-CTLA-4 curative effect was carried out for patients in each risk group.
2.8 TMB and drug sensitivity analysis

We analyzed the gene mutation of dissimilar risk groupings using the "maftools" R package and portrayed the 20 highest mutation frequency genes via oncoplots.

2.9 Analysis of ACRG expression on several dimensions

We compared the 6 ACRGs expression level among healthy and tumor samples and two risk groupings, adopting the "limma" R algorithm. Besides, we verified the expression of ACRG protein using immunohistochemistry data from healthy liver and HCC tissues in the Human Protein Atlas (HPA) database[20].

2.10 Analysis of drug sensitivity and forecasting of candidate medicines

We evaluated the half maximal inhibitory concentration (IC50) values of two subgroups for chemotherapy and targeted medication therapy using the "pRRophetic" R package[21]. In addition, we utilized the Cancer Therapeutics Response Portal (CTRP) and Genomics of Drug Sensitivity in Cancer (GDSC) database[22] to identify commonly prescribed chemotherapeutic medications whether significantly linked to ACRG. To validate molecular docking, we also acquired 2D structures of medicinal compounds based on the Pubchem platform(http://pubchem.ncbi.nlm.nih.gov)[23] and protein structures based on the PDB database(http://www.rcsb.org)[24]. Finally, SYBYL-X 2.0 software was applied to the molecular docking of drugs and protein targets, and the docking results were simulated and screened employing its scoring function to obtain a total score, with a total score $\geq 5$ indicating a greater affinity between the receptor and the ligand.

2.11 Statistical Analysis

In virtue of Kaplan Meier(K-M), we plotted prognostic survival curves. Furthermore, With the help of the Pearson test, we carried out a correlation analysis. Besides, we compared the two groups' data by Wilcoxon test. Finally, We used R software for drug sensitivity analyses and other data processing, modification, and charting (version 4.1.2).

3 Results

3.1 Acquisition of ACRG-DEGs

By comparing the expression of 460 ACRGs in 374 HCC samples with 50 normal samples in the TCGA database, we were able to identify 137 ACRG-DEGs (Figure 1A, B). GO enrichment algorithm on ACRG-DEGs displayed that they predominantly connected with the cellular amino acid metabolic, alpha-amino acid catabolic process and tRNA aminoacylation for protein translation-related processes (Figure 1C). Furthermore, we discovered through KEGG enrichment analysis that these genes were primarily linked to the metabolism of Alanine, aspartate, and glutamate, the synthesis of aminoacyl-tRNA, and the metabolism of glutathione, the
metabolism of cysteine and methionine, arginine and proline, and other related amino acid metabolic pathways are shown in Figure 1D.

**FIGURE 1.** Obtaining ACRG-DEGs. (A) Heatmap of the ACRGs between the normal and the tumor tissues. (B) Volcano plot of the ACRG-DEGs in HCC. (C) Chord diagram of GO enrichment analysis in ACRG-DEGs. (D) KEGG enrichment analysis.

### 3.2 Generation of ACRG-DEGs clusters, TME characteristics and biological features between two subgroups

In this study, we combined the TCGA and GEO data cohorts (total sample size, n=647), set the cluster variable (k) from 2 to 9 to conduct the consensus clustering algorithm, and found that k=2 was the best choice, naming the two subgroups ACRGclusterA (sample size n=213) and ACRGclusterB (sample size n=434), respectively (Figure 2A). According to Kaplan-Meier survival curves, ACRGcluster...
B had an OS considerably greater than ACRGcluster A (P=0.039; Figure 2B). Principal component analysis (PCA) also demonstrated intense discrimination, showing the precision of the clustering approach (Figure 2C). Additionally, we compared the clinical data and gene expression of the two ACRG clusters, finding that ACRGcluster B expressed more ACRGs than ACRG cluster A (Figure 2E). Meanwhile, we compared the immune cell abundances in two ACRGclusters using ssGSVA because immune cell abundance can influence TME and control tumor progression. The findings showed significant differences in 16 immune cell abundances between the two ACRGclusters, with 12 being greater in ACRGcluster A (Figure 2D). Finally, we conducted GSVA enrichment analysis further to understand the variations in biological function across the ACRGclusters. The findings revealed that ACRG cluster A is enriching in metabolic-related processes (Figure 2F).

**FIGURE 2.** Gene clusters were constructed based on ACRG-DEGs. (A) Consensus matrix of all samples with applicable k values(k=2). (B) Kaplan-Meier curve reveals different OS between two ACRGclusters. (C) PCA verified the accuracy of consensus clustering. (D) Boxplot shows the discrepancy of immune cell infiltration levels among the two categories. (E) Heatmap shows the difference in clinical information and ACRGs expression among the two categories. (F) Heatmap displays GSVA of biological pathways between two clusters. *P<0.05, **P<0.01, ***P<0.001.

3.3 Building gene clusters based on ACRG-DEGs

We used the R Program "limma" to find 172 differentially expressed genes (DEGs) among two ACRGclusters. We conducted GO and KEGG analysis on DEGs
to investigate the underlying biological properties. Additionally, we screened 66 DEGs utilizing univariate Cox regression analysis on 172 DEGs (Figure 3C). We repeated a consensus clustering analysis in light of the 66 DEGs expression level, and all patients were split into two subtypes: gene clusters A and B (Figure 3D). According to PCA (Figure 3E), there were substantial differences in the distribution of the two gene subtypes. Survival analysis revealed OS of gene cluster B was noticeably superior to gene cluster A (Figure 3F).

3.4 Construction and validation of ACRG_score model

We performed Lasso and MultiCox analysis on 66 DEGs and established a Risk score (ACRG_score) consisting of 6 genes, including SLC2A1, EZH2, GZMA, IGFBP3, SPP1, CXCL9. The following formula calculates the Risk score: 

$$\text{ACRG\_score} = \text{expression of SLC2A1} \times 0.2659 + \text{expression of EZH2} \times 0.2788 + \text{expression of GZMA} \times (-0.2200) + \text{expression of IGFBP3} \times 0.2015 + \text{expression of SPP1} \times 0.0547 + \text{expression of CXCL9} \times (-0.1256)$$

We split all HCC patients into training (n = 300) and test (n = 300) groups at random in a 1:1 ratio. Moreover, using the median ACRG_score as a dividing line, we separated training group patients into low- and high-risk groups. Figure 4A depicted the prognosis, Risk score distribution and the six ACRGs expression level in training group patients. Our results showed that as the ACRG_score increases, the patient's survival time shortens, and mortality rises. Meanwhile, compared to GZMA and CXCL9, the expression levels of SLC2A1, EZH2, IGFBP3, and SPP1 were greater in the high-risk category. We compared the OS variations between two risk groups using survival analysis. ROC curves were used to illustrate the model's prediction performance, and the areas under the curves for the 1-, 3-, and 5-year periods were 0.767, 0.773, and 0.764, respectively. The K-M plot revealed that the low-risk category had remarkably longer survival times than another category (P < 0.001; Figure 4D). We verified the model using the independent TCGA with GEO (GSE76427) cohort, and the findings aligned with the training set (Figure 4B, C, E, and F). Similarly, the testing set and the entire set exhibit parallel results (Supplementary Figure 1A, B, C, D, F, G). Data from multiple cohorts illustrate how accurate the model is. Furthermore, We compared Progression-free survival (PFS) among two categories and the PFS of the low-risk category was significantly longer than another to illustrate further the model's excellent predictive potential (Supplementary Figure 1E, H, I, G). Finally, we created a Sankey plot to depict the distribution between two ACRG clusters, two gene clusters, ACRG_score, and survival status (Figure 4G). When analyzing the ACRG_score differences, we discovered that ACRG cluster A was substantially greater than ACRG cluster B (p < 0.001; Figure 4H) and gene cluster A was considerably greater than gene cluster B (p < 0.001; Figure 4I).
FIGURE 3. Building gene clusters based on ACRG-DEGs. (A, B) GO and KEGG analysis of DEGs between two ACRG clusters. (C) Univariate Cox analysis of DEGs. (D) Heatmap illustrating the association of two gene clusters with clinicopathological traits. (E) The consensus clustering matrix divides patients into two types of gene
clusters. (F) The effectiveness of consensus clustering is demonstrated by PCA. (G) Kaplan-Meier curve reveals different OS between two gene clusters.

3.5 ACRG_score stratified analysis in clinicopathological variables

We discovered a significant difference between the ACRG_score and Stage, but not between age, gender, or ACRG_score when analyzing the correlation between ACRG_score and clinicopathological variables (Figure 5A, B, C, D). The prediction power of the ACRG_score was then assessed in patients with various clinicopathological variables (Age, Gender, Stage) using survival analysis. We discovered that the high-risk category OS was often lower than another, however, individuals older than 65 were an exception (Figure 5E, F, G).

3.6 Independent prognostic analysis of ACRG_score

Age, Gender, Stage, and ACRG_score were incorporated into Univariate/Multivariate Cox analysis to evaluate whether these pathological factors have independent predictive value. We uncovered stage and ACRG_scores were substantially related with OS in the entire cohort (p < 0.001; Figure 6A, B). Meanwhile, the ROC curves combining ACRG_score and clinicopathological variables were able to depict the exactitude of each component to forecast patients' OS in 1-, 3-, and 5-year, with the ACRG_score having the highest Area Under ROC Curve (AUC) value, suggesting its great predictive ability (Figure 6C, D, E). Furthermore, in the TCGA, Train, and Test cohorts, the final independent prognostic analysis was consistent with the entire cohort (Figure 6F-J; Supplementary Figure 2A-J). Consequently, ACRG_score can be utilized as a standalone indicator of the prognosis of liver cancer.

3.7 Construction and validation of a nomogram

Based on the ACRG_score's ability to predict outcomes, we integrated it with clinicopathological variables to create a nomogram that could predict patients' OS at 1, 3, and 5 years (Figure 7A). We plotted Calibration and ROC curves to evaluate the line plot model's effectiveness. The findings revealed that the nomogram model could accurately forecast patients' OS. (Figure 7 B-E).
FIGURE 4. Development and validation of ACRG_score model. (A, B, C) ACRG_score risk distribution, survival status, and associated gene expression in the training set, TCGA set, and GEO set. (D, E, F) In the training set, TCGA set, and GEO set respectively, the K-M curve showing the OS discrepancy among two risk groupings, and the ROC curve assessing the efficacy that the model predicts 1-, 3-,
and 5-year OS. (G) Sankey diagram demonstrating the connection between the ACRGclusters, Geneclusters, ACRG_score, and survival status. (H) Analysing the discrepancy in ACRG_score between two ACRGclusters. (I) Comparing the difference in ACRG_score between two Geneclusters.

3.8 Immune landscape analysis based on ACRG_score

Because tumor growth necessitates evasion of immune recognition\textsuperscript{25}, and changes in tumor metabolism can influence immune cell function by changing the TME\textsuperscript{26}, we explored the link between ACRG_score, TME, and immunological infiltration further. The ESTIMATE program assessed both risk groups' TME scores and found that the low-risk category was considerably higher stromalscore, immunescore, and ESTIMATEscore than another (Figure 8A). We examined the correlation between immune infiltration and ACRG_score using the CIBERSORT algorithm and discovered that M0 and M2 macrophages showed significantly positive pertinency with ACRG_score. In contrast, CD8 T cells and plasma cells exhibited remarkably negative correlations with ACRG_score (Figure 8C). Meanwhile, immune cell number among different risk categories was assessed using ssGSEA. The findings revealed that CD56dim natural killer cells, immature dendritic cells, natural killer T cells, and multiple other immune cells had higher levels in the high-risk group than another (Figure 8B). We develop a heatmap to contrast the outcomes of the TIMER, CIBERSORT, CIBERSORT-ABS, QUANTISEQ, MCPCOUNTER, XCELL, and EPIC programs among two risk groups (Figure 8E).

When assessing the pertinency between the ACRG_score and 47 IC genes, we discovered a substantial difference between the two risk groups and a strong correlation between the ACRG_score and 23 IC genes (Figure 8D, F). We derived IPS for patients with liver cancer from the TCIA database to comprehend the response to immunotherapy. According to the findings, the low-risk group responded better to PD-1 and CTLA-4 inhibitors. (Figure 8G-J).

3.9 Relationship between Somatic cell mutation and ACRG_score

TMB is a reliable immunotherapy molecular indicator and can increase prognostication precision. Therefore, We analyzed the variations in genomic mutations between risk groupings to explore the connection between ACRG_score and somatic mutations. The five highest mutation rates genes in both risk groups included TP53, CTNNB1, TTN, MUC16, and PCLO, with TP53 mutation rates remarkably greater in the high-risk grouping than the other and CTNNB1 mutation frequencies more significant in the low-risk grouping (Figure 9A, B). According to the mutation frequency, We separated HCC patients into Low and High TMB groupings. A survival study for these two groups demonstrated that the Low TMB group OS was distinctly better than the other (P=0.025; Figure 9C). Additionally, on
the basis of TMB and ACRG_score, a subgroup survival analysis was conducted, and the high risk patients with High TMB had a worse OS (P<0.0001; Figure 9D).

**Figure 5.** ACRG_score stratifying analysis in clinicopathological variables. (A) Heatmap of the connection between clinicopathological factors and ACRG_score. (B-D) Proportion of clinicopathological features in the two risk groupings, and association between Risk score and clinical characteristics. (E-G) Kaplan-Meier curves displaying the discrepancy in OS between the two risk groups for each clinical categories. ***P<0.001.
Figure 6. Independent prognostic analysis of ACRG_score. (A) Univariate Cox regression analysis in the entire set. (B) Multivariate Cox regression analysis in the entire set. (C-E) ROC curves assessing the forecast accuracy of multiple factors at 1-, 3-, and 5- years in the entire cohort. (F) Univariate Cox regression analysis in the training set. (G) Multivariate Cox regression analysis in the training set. (H-I) ROC curves assessing the forecast accuracy of multiple factors at 1-, 3-, and 5- years in the training set.
3.10 Comprehensive analysis of ACRGs expression

We compared the expression differences of six ACRGs in two risk groups, and the results revealed that CXCL9 and GZMA were significantly highly expressed in the low-risk grouping. However, EZH2, SPP1, SLC2A1, and IGFBP3 were the opposite (p<0.001; Figure 10A). We further demonstrated that based on differential gene expression analysis between tumor and para cancer samples from TCGA HCC cases, EZH2, SPP1, SLC2A1, and CXCL9 were strongly expressed in tumor samples, yet IGFBP3 substantially expressed in normal samples(p<0.001; Figure 10B). In addition, EZH2, SPP1, and SLC2A1 are highly expressed in tumor tissues, according to an immunohistochemical analysis of the HPA database (Figure 10C, D, E). Using tissue staining and quantity scores, we analyzed the protein expression level in healthy and tumor samples, which revealed that EZH2, SPP1, and SLC2A1 had remarkably higher expression in tumor tissues (P<0.01; Figure 10F-H).

Figure 7. Development and validation of a nomogram. (A) Nomogram for predicting 1-, 3-, 5-years OS of HCC patients. (B) Calibration curve to evaluate how well the nomogram predicts. (C-E) ROC curves assessing the forecast accuracy of the nomogram at 1-, 3-, and 5- years.
Figure 8. Immune landscape analysis based on ACRG_score. (A) The ESTIMATE algorithm analyses the discrepancy in TME score between different risk groups. (B) Two risk categories immunoassay based on ssGSEA. (C) Evaluating the discrepancy in TME immune cell levels between the two risk groupings. (D) Analysis of ICs gene expression discrepancy among the two risk categories. (E) Heatmap displaying the difference in immune cell levels in the two risk groups based on multiple comprehensive analysis. (F) Correlation analysis between ICs gene expression and
Risk score. (G-J) IPS assessed the response to immunotherapy in both low- and high-risk populations. *P<0.05, **P<0.01, ***P<0.001.

3.11 Drug sensitivity analysis and candidate drug prediction

We researched the IC50 of common chemotherapy agents versus targeted medications in the two risk groups to anticipate the connection between ACRGs and anticancer treatments. The findings revealed that the majority of medicines IC50 considerably varied between the two risk categories (P<0.001; Figure 11A-I). For example, Cisplatin and Doxorubicin, two drugs often used to treat HCC, exhibited lower IC50 values in the high-risk category. As a result, we anticipated that Cisplatin and Doxorubicin would be more effective in treating the high-risk patient population (Figure 11B, H). Additionally, we obtained correlation data between ACRGs and alternative medications from the GDSC and CTRP databases (Supplementary Table 1). After selecting ACRGs-drugs with FDR<0.05 and high correlation, we performed molecular docking verification and demonstrated its structure and efficiency score (Figure 12A-H). According to the scoring data, Nilotinib, Dasatinib, and Doxorubicin each had docking scores of more than five for IGFBP3, SLC2A1, and SLC2A1. (Supplementary Table 1).

4 Discussion

Since Warburg discovered aberrant changes in tumor cell glucose metabolism, the effect of metabolic reprogramming on the growth of tumors has been a significant focus of oncology research\textsuperscript{[27]}. Amino acids play a crucial role in tumor cells as one of the three metabolites. It is anticipated to be a potential approach for treating malignancies by identifying the amino acid metabolic-related targets in tumor progression and then blocking them to inhibit the amino acid metabolism reprogramming to regulate tumor progression and enhance prognosis\textsuperscript{[28]}. Studying the mechanism of hepatocellular cancer from the standpoint of amino acid metabolism is particularly important because the liver serves as the central hub of amino acid metabolism. For instance, the growth and proliferation of alanine-mediated HCC may be stopped by inhibiting the GPT1 protein\textsuperscript{[29]}, whereas OGDHL knockdown promotes the growth and survival of HCC by controlling glutamine metabolic pathways\textsuperscript{[30]}. Because prior research had certain relative limitations, we combined multiple clustering analyses with clinical analysis to create and validate the model and performed a multidimensional immunoassay before screening small molecule therapies with clinical promise.
Figure 9. Association between ACRG_score and TMB. (A, B) Waterfall diagrams of somatic mutations between different risk groups. (C) Survival analysis of TMB. (D) Survival analysis with ACRG_score and TMB.

In this study, we make use of consensus cluster analysis to categorize the patients into two clusters on the basis of the 117 ACRG-DEGs expression level. We discovered that immune cells were more widely dispersed in cluster A. GO and KEGG algorithm analyzed the DEGs in both subgroups, and results showed these genes were highly enriched in pathways associated with metabolism. After that, we carried out yet another clustering analysis based on DEGs. The results demonstrated...
significant differences in the two groupings' survival prognoses, suggesting that DEGs may be employed as a predictor of survival prognosis for clinical patients. Furthermore, We constructed an ACRG_score model including SLC2A1, EZH2, GZMA, IGFBP3, SPP1, and CXCL9. We graphed survival and ROC curves to evaluate the model accuracy in the multiple data sets. Additionally, the ACRG Risk score demonstrated reliable overall survival prediction across numerous clinicopathological strata, including age, gender, and stage. Univariate versus multifactorial independent prognostic analysis supported the ACRG_score's status as an independent prognostic factor. Finally, We combined the ACRG_score, age, gender, and stage to develop a nomogram model. The ROC and calibration curves confirmed the model's excellent applicability in clinical prediction. According to the results above, the ACRG_score model could accurately predict the clinical prognosis of individuals with HCC.

We further explored the potential functions and research progress of 6 genes of ACRGs (EZH2, GZMA, IGFBP3, SPP1, CXCL9, SLC2A1). Enhancer of zeste homolog 2 (EZH2) is a polycomb group genes (PcGs) family member that controls epigenetics through transcriptional repression\(^{[31]}\), influencing cell cycle progression, autophagy, apoptosis, and preventing cellular senescence\(^{[32]}\). Studies have shown that disruption of the role of EZH2 as a transcriptional master regulator can promote cancer progression\(^{[33]}\). EZH2 expression is upregulated in HCC\(^{[34]}\), while EZH2 inhibitors can exert anti-HCC and other cancers by activating tumor suppressor miRNAs\(^{[35]}\).

Granzyme A (GzmA) is a significant serine protease family member\(^{[36]}\) that is primarily expressed by a group of cytotoxic cells and promotes apoptosis via the death receptor route or the granule secretion pathway\(^{[37-38]}\). It was shown that GZMA can bind to F2R in HCC patients and trigger apoptosis by activating the JAK2/STAT1 signaling pathway\(^{[39]}\). IGFBP-3 is a p53 tumor suppressor regulatory protein involved in various diseases such as cancer, diabetes, and fatty liver\(^{[40]}\). P53 activates IGFBP-3 expression and mediates cell cycle arrest and apoptosis\(^{[41-42]}\). Highly phosphorylated glycoprotein Osteopontin (OPN) is encoded by Secreted Phosphoprotein 1 (SPP1). Numerous pathological processes, including cancer, heart disease, diabetes, and inflammation, are impacted by SPP1\(^{[43-44]}\). Studies have confirmed that SPP1 is one of the significant genes involved in HCC metastasis\(^{[45]}\). In the meantime, as an essential independent prognostic biomarker, plasma OPN can be utilized to assess the effectiveness of HCC therapy and the recurrence rate\(^{[46-47]}\). Using single-cell transcriptomics, Lichun Ma showed that SPP1 expression is directly linked to developing HCC cells and reprogramming the TME\(^{[48]}\). The results of our investigation are consistent with this research.
Figure 10. Comprehensive analysis of the ACRGs expression. (A) ACRG expression differences between two risk groupings. (B) Differential expression of ACRG in
tumor and paracancerous tissues. (C-E) Immunohistochemical detection of ACRG protein expression in HCC and normal liver tissue. (F-H) Differences in immunohistochemical scores of Normal and Tumour samples. *P<0.05, **P<0.01, ***P<0.001.

C-X-C motif chemokine ligand 9 (CXCL9) is known as Monokine induced by gamma interferon (MIG)[49]. CXCL9 was discovered to boost CD8+ T cell infiltration in tumor cells and induce an immunological response to destroy tumor cells. High expression of CXCL9 significantly increased tumor cell sensitivity to immune checkpoint blockade therapy[50]. Solute carrier family 2 member 1 (SLC2A1) encodes Glucose transporters 1 (GLUT1), the body's most widely distributed glucose transporter, which is involved in the transport of glucose to erythrocytes to maintain blood glucose concentration[51-52]. Research has shown that numerous tumor tissues exhibit aberrant GLUT1 expression, which is directly correlated with tumor aggressiveness[53-54]. We compared Six genes' expression levels in samples from the normal/tumor and high/low-risk groups and discovered that EZH2, SPP1, and SLC2A1 had the same expression difference, were significantly highly expressed in HCC tissues, and might be employed as possible indicators to foretell the onset of HCC patients.

TME is thought to be a complex, multicellular environment that supports the growth of tumor cells[55]. Immune cells are the most prevalent component of TME, which include both tumor-promoting and tumor-antagonizing immune cells and have a bidirectional impact on tumor formation[56-57]. The importance of the immunological landscape in TME cannot be overstated because immunotherapy for TME is now opening up fresh perspectives in oncology treatment. After comparing the immune infiltration patterns among the two risk groupings, we found that the High-risk category had a considerably higher abundance of Macrophages M0 and M2 than the other group. Tumor-associated macrophages (TAM) are the most abundant adaptive immune population in the TME and regulate tumor progression[58-59]. According to the consensus, M0 macrophages are at rest and polarize in response to varied stimuli into the classically activated M1 macrophage phenotype and the alternatively activated M2 macrophage phenotype[60]. M1 macrophage phenotype triggers inflammation against tumors by releasing TNF-α, IL-1β, IL-12, and IL-23, while M2 macrophage phenotype suppresses inflammatory response by secreting IL-4, IL-10, and TGF-β, thus promoting tumor proliferation and invasion[61-62]. M0 macrophages were previously thought to have no immune effects, but by building a predictive model of M0 macrophage-related HCC, it has been shown that M0 macrophages may be a factor in promoting HCC. However, the conclusion still needs to be further validated[63]. In contrast, several lymphocytes, including T cells CD8, were a lot lower in the group at high risk than those at low risk. CD8+ T cells are the primary immune population targeting cancer cells, induced by the interaction of dendritic cells (DC), natural killer cells (NK), and CD4+ T cells to form toxic T lymphocytes (CTL), which in turn recognize and remove cancer cells with MHC-I-like molecules[64]. Such results were consistently demonstrated in the ssGSEA analysis.
Figure 11. Drug sensitivity analysis. Differential responses of two risk populations to the frequently administered chemotherapeutic drug (A) Dasatinib; (B) Cisplatin; (C) Temsirolimus; (D) Nilotinib; (E) Imatinib; (F) Gefitinib; (G) Etoposide; (H) Doxorubicin; (I) Docetaxel.

In addition to evaluating immune infiltration, we discovered changes in the somatic mutation features between the two risk groupings. In the high-risk category, TP53 mutations were more common than in the other group, while the frequency of CTNNB1 mutations was the opposite. Such variation may lead to a worse survival prognosis for individuals in the high-risk category. It is evident from the survival curves of the L-TMB and H-TMB subgroups that the forecast for the H-TMB group is worse. It has been shown that TMB can generate new antigens presented by major histocompatibility complex (MHC) proteins, increasing the chance of recognition of tumor cells by T cells and targeting tumor cells for destruction[65]. Therefore, compared to the L-TMB group, the H-TMB group ought to have had a better
Nevertheless, this is not in conflict with the current study, as more and more findings show that the complexity of immune processes in the TME highlights the limitations of TMB as a biomarker, making it more important to combine with other predictive factors for clinical application\textsuperscript{[66-67]}.  

**Figure 12.** Molecular docking result. Docking site of the EZH2 active region with Dasatinib (A), Etoposide (B). Docking site of the GZMA active region with Etoposide (C), Temsirolimus (D). Docking site of the IGFBP3 active region with Imatinib (E), Nilotinib (F). Docking site of the SLC2A1 active region with Dasatinib (G), Doxorubicin (H).  

ICs are an essential class of receptor-ligand molecules\textsuperscript{[68]} that regulate adaptive immune processes and promote tumor development mainly by mediating immune
escape\textsuperscript{69}. For cancer patients, immune checkpoint inhibitors in tumor immunotherapy offer novel possibilities. For instance, PD-1 and PD-L1 inhibitors can terminate the dephosphorylation of the T-cell activating enzymes which results in T-cell activation, returning T cells' ability to kill\textsuperscript{70}. The combination VEGF/PD-(L)1 blocker treatment approach, now the first-line medicine for treating HCC in clinical trials, can significantly increase OS in patients with advanced HCC\textsuperscript{71}. Furthermore, we analyzed the relationships between 46 immunological checkpoints with ACRG\_score and found that the majority of ICs highly expressed in the high-risk group. Therefore, this means that immune checkpoint treatment may have better benefits in high-risk groups.

Finally, we compared the IC\textsubscript{50} discrepancy in the two risk categories for common chemotherapeutic agents to explore new ways for ACRG to guide clinical drug use. The results demonstrated that those in the high-risk category were more susceptible to Gefitinib and Temsirolimus. The low-risk patients, however, were more sensitive to Dasatinib, cisplatin, Nilotinib, Imatinib, Etoposide, Doxorubicin, and Docetaxel. Additionally, We screened drug-target combinations using the GDSC and CTRP databases, looking for combinations with FDR<0.05 for molecular docking validation and docking the targets consistently related to chemotherapeutic medicines. We looked for stable binding docking outcomes and discovered that doxorubicin, dasatinib, and nilotinib would be suitable therapeutic alternative drugs. Our work shows the potential benefit of these drugs in the HCC therapy based on amino acids metabolic, even if dose predictions from current data alone have not yet been clinically validated. Although our study made novel findings, it still has some restrictions. This is a retrospective analysis using public databases. Even after internal and external validation with multiple datasets, prospective studies with large multicenter samples and adequate verification with in vivo and in vitro experiments are still needed.

5 Conclusion

In conclusion, our effort was effective in developing the ACRG\_score and ACRG-based clinical risk models, which are capable of reliably predicting the clinical prognosis of HCC patients. At the same time, ACRG has shown unique advantages in revealing differences in TME and immunotherapy effects among patients and provides compelling clues for clinical drug use. The study explores effective biomarkers for HCC treatment from the amino acid metabolic reprogramming perspective and guides clinical prognosis and individualized comprehensive treatment of HCC patients. To give a more robust theoretical foundation for targeted therapy of HCC, we will further in-depth research the mechanisms of the six ACRGs in developing HCC.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS


CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

DATA AVAILABILITY STATEMENT

The unique contributions made in the work are described in the article as well as the Supplementary Material. Any further questions can be directed to the authors.

ETHICS STATEMENT

All the data in this research were accessed from publicly database or publications with proper citation. It is therefore not subject to ethical approval and informed consent is not required.

ORCID

Dalin Chen: https://orcid.org/0009-0002-8733-9199
Reference


Supplementary Table 1: Molecular correlation and docking score

<table>
<thead>
<tr>
<th>ACRGs</th>
<th>Drug</th>
<th>GDSC</th>
<th>CTRP</th>
<th>Molecular docking binding force</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cor</td>
<td>fdr</td>
<td>cor</td>
<td>fdr</td>
</tr>
<tr>
<td>EZH2</td>
<td>Dasatinib</td>
<td>0.22378872</td>
<td>0.00015701</td>
<td>0.136808123</td>
</tr>
<tr>
<td>EZH2</td>
<td>Etoposide</td>
<td>-0.135218119</td>
<td>0.000940689</td>
<td>-0.371153243</td>
</tr>
<tr>
<td>GZMA</td>
<td>Etoposide</td>
<td>-0.141872631</td>
<td>0.000496604</td>
<td>-0.247340226</td>
</tr>
<tr>
<td>GZMA</td>
<td>Temsirolimus</td>
<td>-0.134273478</td>
<td>0.001093997</td>
<td>-0.106619501</td>
</tr>
<tr>
<td>IGFBP3</td>
<td>Imatinib</td>
<td>-0.290348991</td>
<td>5.33648E-07</td>
<td>-0.179786546</td>
</tr>
<tr>
<td>IGFBP3</td>
<td>Nilotinib</td>
<td>0.106533949</td>
<td>0.011331079</td>
<td>0.335299126</td>
</tr>
<tr>
<td>SLC2A1</td>
<td>Dasatinib</td>
<td>-0.266975722</td>
<td>4.42396E-06</td>
<td>-0.167790222</td>
</tr>
<tr>
<td>SLC2A1</td>
<td>Doxorubicin</td>
<td>0.128646219</td>
<td>0.001720648</td>
<td>0.411883654</td>
</tr>
</tbody>
</table>