



Original Article

Aminoacyl-tRNA Synthetase-based Prognosis Model and Exploration of Potential Mechanisms in Hepatocellular Carcinoma



Yixun Jin[#], Xinyang Huang[#], Zhuoxin Wang, Hui Xu, Qi Wang* and Lifu Wang*

Department of Gastroenterology, Ruijin Hospital Affiliated to The Shanghai Jiao Tong University School of Medicine, Shanghai, China

Received: 21 June 2022 | Revised: 23 August 2022 | Accepted: 1 November 2022 | Published online: 17 January 2023

Abstract

Background and Aims: Aminoacyl-tRNA synthetases (ARSs) participate in tumor initiation and progression but their involvement in hepatocellular carcinoma (HCC) is not clear. This study aimed to investigate the prognostic value and underlying mechanisms of ARS in HCC. **Methods:** Data were obtained from The Cancer Genome Atlas (TCGA), International Cancer Genome Consortium, Gene Expression Omnibus and Human Protein Atlas databases. The prognostic model was constructed with the use of Cox regression and least absolute shrinkage and selection operator regression. Kaplan-Meier survival analysis, enrichment analysis, single sample gene set enrichment analysis and tumor mutation burden calculation were performed with R to evaluate the model and explore the underlying mechanism. Wilcoxon tests were used for comparisons between groups. **Results:** Aspartyl-tRNA synthetase 2 (DARS2), tyrosyl-tRNA synthetase 1 (YARS1) and cysteinyl-tRNA synthetase 2 (CARS2) were identified as prognostic biomarkers and enrolled in model construction. The area under receiver operating characteristic curve of the model was 0.775. The model was used to assign patients from TCGA into low- and high-risk groups. Those in the high-risk group had a worse prognosis ($p < 0.001$). The clinical significance of the model was tested in different clinical subgroups. Genetic mutation analysis had a higher *TP53* mutation frequency in high-risk group. Enrichment analysis

and study of immune-related cells and molecules found that the high-risk group was characterized by immune-cell infiltration and immunosuppression states. **Conclusions:** A novel ARS family-based model of HCC prognosis was constructed. *TP53* mutation frequency and immune-suppressive status accounted for a worse prognosis in patients included in the high-risk group.

Citation of this article: Jin Y, Huang X, Wang Z, Xu H, Wang Q, Wang L. Aminoacyl-tRNA Synthetase-based Prognosis Model and Exploration of Potential Mechanisms in Hepatocellular Carcinoma. *J Clin Transl Hepatol* 2023;11(4):877–888. doi: 10.14218/JCTH.2022.00301.

Introduction

Hepatocellular carcinoma (HCC) accounts for 90% of liver cancers and is the fourth leading cause of cancer-related mortality worldwide.^{1,2} Liver resection, liver transplant, ablation, transarterial chemoembolization and systemic drugs have dramatically extended patient life expectancy, but the 5-year relative survival rate of liver cancer still hover at around 20%.^{1,3} Improvement in progression in prognosis is needed. Immune therapy has challenged traditional HCC therapy. Immune checkpoint inhibitors (ICIs) are the backbone of current immune therapy. The combination of the PD-L1 inhibitor atezolizumab and the anti-vascular endothelial growth factor (VEGF) agent bevacizumab is accepted as a preferred treatment for advanced HCC.^{4,5} Nevertheless, the overall response rate of immune therapy is less than a third, which limits the survival efficacy.⁶ More specific prognostic models and subclassifications of patients who might benefit from a particular therapy are needed. With the development in multi-omics profiling, various bioinformatic methods are used for HCC-patient subclassification and prognostic assessment. However, many studies enrolled only parameters from the whole transcriptome profile and overlooked their biological processes. Consequently, the results had an inherent bias and did not reflect the intrinsic feature of HCC.

Aminoacyl-tRNA synthetases (ARSs) are a family of enzymes in the cytoplasm and mitochondria, which catalyze the ligation of amino acids to their corresponding tRNAs in protein synthesis. Some combine with ARS-interacting multifunctional proteins to form multi-synthetase complexes.⁷

Keywords: Bioinformatic analysis; Hepatocellular carcinoma; Amino acyl-tRNA synthetases; Prognosis; Immunosuppression.

Abbreviations: AFP, alpha-fetoprotein; ARSs, aminoacyl-tRNA synthetases; CARS2, cysteinyl-tRNA synthetase 2; DARS2, aspartyl-tRNA synthetase 2; DEGs, differentially expressed genes; EMAP11, endothelial monocyte activating polypeptide II; GEO, Gene Expression Omnibus; GO, Gene Ontology; GSEA, gene set enrichment analysis; HCC, hepatocellular carcinoma; HPA, Human Protein Atlas; ICGC, International Cancer Genome Consortium; ICI, immune checkpoint inhibitor; KEGG, Kyoto Encyclopedia of Genes and Genomes; Lasso, least absolute shrinkage and selection operator; LIRI-JP, Liver Cancer-RIKEN Japan; ROC, receiver operating characteristic; ssGSEA, single sample gene set enrichment analysis; TCGA, The Cancer Genome Atlas; TCGA-LIHC, The Cancer Genome Atlas-liver hepatocellular carcinoma; TMB, tumor mutation burden; VEGF, vascular endothelial growth factor; YARS1, tyrosyl-tRNA synthetase 1.

*Contributed equally to this work.

Correspondence to: Lifu Wang and Qi Wang, Department of Gastroenterology, Ruijin Hospital Affiliated to The Shanghai Jiao Tong University School of Medicine, 197 Ruijin Er Road, Shanghai 200001, China. ORCID: <https://orcid.org/0000-0001-5172-9932> (LW) and <https://orcid.org/0000-0002-8437-8224> (QW). Tel: +86-21-63055095, E-mail: lifuwang@sjtu.edu.cn (LW) and wq12147@rjh.com.cn (QW)

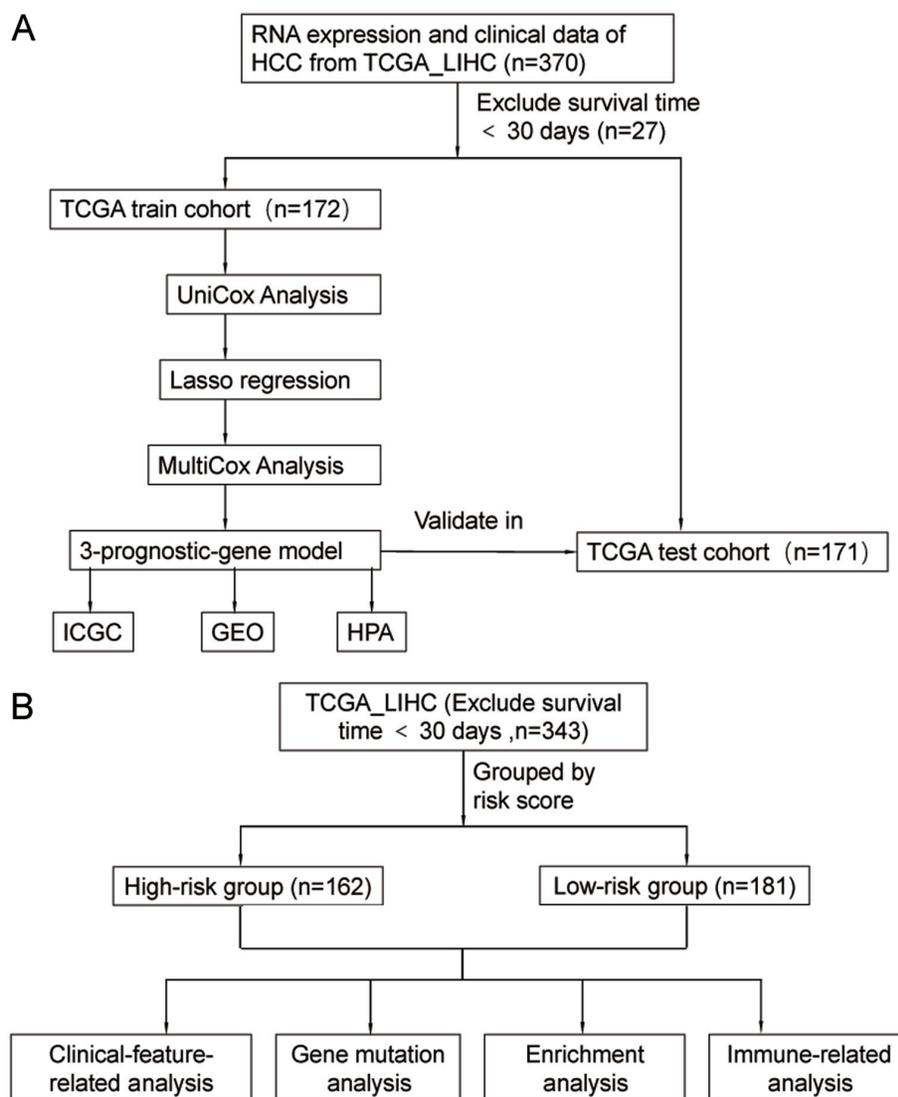


Fig. 1. Study flowchart. (A) ARS prognostic model establishment and (B) comparison of low-risk and high-risk groups. ARS, aminoacyl-tRNA synthetase.

Recent studies have revealed that ARSs are associated with cancer occurrence and progression by their structural metamorphosis and additional motifs and domains including N-terminal helix appendix, leucine zipper, glutathione S-transferase, endothelial monocyte activating polypeptide II (EMAPII) and WHEP domains.^{8,9} Signaling pathways mediated by ARSs also participate in tumorigenesis of multiple organs.^{10–12} Several abnormally expressed ARSs have diagnostic and prognostic value in cancer.^{13–17} Considering their conventional catalytic activity and noncanonical function in tumorigenesis, ARSs may actively participate in initiation and progression of HCC which needs further exploration, and the clinical significance of ARSs in HCC is also unclear.

Herein, we carried out a comprehensive bioinformatic analysis of ARSs in patients from The Cancer Genome Atlas-liver hepatocellular carcinoma (TCGA-LIHC) cohort, constructed a novel prognosis ARS model by Cox regression and least absolute shrinkage and selection operator (Lasso) regression and used it to assess the survival rate of HCC patients. The prognostic value and protein expression level of each enrolled parameter were evaluated in the International Can-

cer Genome Consortium (ICGC), Gene Expression Omnibus (GEO) and Human Protein Atlas (HPA) databases. To understand the underlying mechanism, genetic mutation profiles were analyzed, and enrichment analyses were performed. Results showed differences in immune-related pathways and led us to conduct single sample gene set enrichment analysis (ssGSEA), comparison of the expression of immunosuppressive molecules, and calculation of the tumor mutation burden (TMB). The study design is shown in Figure 1. The study objective was to explore the prognostic value of ARSs and their function in HCC. We established a novel ARS-related prognosis model and preliminarily explored the underlying mechanisms, which will facilitate the design of therapy that improves patient prognosis.

Methods

Public data acquisition and processing

RNA-seq expression data of TCGA-LIHC cohort and corresponding clinical information were obtained from the TCGA

website (<https://portal.gdc.cancer.gov/>). Patients whose survival time was missing or who survived for less than 30 days were excluded, and 343 patients were enrolled in this study. The etiologies of the patients included hepatitis B, hepatitis C, nonalcoholic fatty liver disease, and inherited metabolic diseases such as hemochromatosis and alpha-1 antitrypsin deficiency. We allocated patients randomly to training and test cohorts. Baseline characteristics of patients are shown in Supplementary Table 1. Expression and clinical data of the Liver Cancer-RIKEN Japan (LIRI-JP) cohort and GSE76427 dataset were downloaded from ICGC (<https://dcc.icgc.org/>) and the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) respectively.

Model construction and validation

Univariate Cox regression analysis of 38 ARS genes was conducted in the training cohort (Supplementary Table 2) and 17 prognostic genes were selected. Next, Lasso ten-fold cross-validation was performed by glmnet and survival R packages, and six genes were screened out. Multivariate Cox regression analysis was then used to construct a prognosis model that was confirmed in the test cohort. We used receiver operating characteristic (ROC) curves to evaluate the model performance, which were drawn by the survivalROC R package.

HPA database analysis

DARS2, YARS1 and CARS2 immunohistochemistry profiles were obtained from the HPA (<https://www.proteinatlas.org/>), an open-access database of protein expression of human normal and tumor tissues.

Clinical characteristics analysis

Patients in the TCGA-LIHC cohort were grouped into different subsets by their clinical characteristics including age, sex, alcohol consumption history, hepatitis B history, hepatitis C history, Child-Pugh score, alpha-fetoprotein (AFP) value, tumor grade, tumor stage, residual tumor, and vascular invasion. Differences between the risk scores of the clinical-feature subgroups were compared with Wilcoxon tests. Results were visualized by the ggplot2 package and a nomogram was constructed with the rms package in R.

Acquisition and analysis of mutation profiles

Mutation profiles (VarScan2 format) of TCGA-LIHC cohort were downloaded from the University of Southern California Xena database (<https://xena.ucsc.edu/>) and analyzed with the R maftools package. Waterfall figures were produced with GenVisR.

Screening and enrichment analysis of differentially expressed genes (DEGs)

DEGs between low-risk and high-risk group were found by the R packages DESeq2, edgeR, and limma. Intersection of three DEG sets was obtained and visualized with the nVennR and VennDiagram packages. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of common DEGs was carried out with the clusterProfiler package, which was also used for gene set enrichment analysis (GSEA).

Immune-cell infiltration analysis and TMB calculation

The GSVA package was used to estimate the extent of immune-cell infiltration in different groups,¹⁸ and the strength and direction of the association between risk scores and immune-cell infiltration were evaluated by Spearman's correlation. Differences in the expression of immunosuppressive

molecules between the low-risk group and high-risk group were tested by the Wilcoxon test. The TMB of each sample was calculated with the TCGA mutations package and differences of TMB between the low-risk and high-risk groups were compared with the Wilcoxon test.

Statistical analysis

Statistical analysis was performed with R software (version 4.1.0). We used log-rank test for Kaplan-Meier survival analysis and calculate hazard ratios and 95% confidence intervals. Two-tailed *p*-values <0.05 were considered statistically significant.

Results

Construction of a novel prognostic risk score model based on the ARS

Univariate Cox analysis and Lasso regression identified potential independent prognostic ARSs in the training cohort (Fig. 2A–C). A novel prognostic risk score model containing DARS2, YARS1, and CARS2 was constructed by multivariate Cox analysis. The risk score formula was: Risk Score = 0.089 × DARS2 + 0.126 × YARS1 + 0.102 × CARS2.

DARS2, YARS1, and CARS2 predict prognosis and are highly expressed in HCC

To evaluate the prognostic value of DARS2, YARS1, and CARS2, we performed Kaplan-Meier survival analysis in the ICGC LIRI-JP cohort and GEO GSE76427 dataset. Highly expressed DARS2, YARS1, and CARS2 were linked to shorter survival time in the LIRI-JP cohort (Fig. 3A–C). The survival analysis results in GSE76427 were also statistically significant (Supplementary Fig. 1). We also saw increased levels of DARS2, YARS1, and CARS2 protein expression in the HPA database (Fig. 3D).

Assessment of the ARS prognostic risk score model

The risk score of each patient in the TCGA-LIHC cohort was calculated using the derived model. Using the median score of training cohort, the patients in the TCGA-LIHC cohort were divided into high-risk and low-risk groups. The distribution of risk scores, survival time, and gene expression in the training and test cohorts are shown in Figure 4A and B. The prognostic value of model was estimated by survival analysis of the training and test cohorts. A Kaplan-Meier survival curve of the training cohort (Fig. 4C) showed that HCC patients with high-risk scores had a worse prognosis than those with low-risk scores, which was confirmed in the test cohort (Fig. 4D). Performance of the risk score system in the training and test cohorts was evaluated with ROC curves. The areas under the ROC curves of the training and test cohorts were 0.775 and 0.797, respectively (Fig. 4E, F). High-risk scores were also linked to worse recurrence-free survival (Fig. 4G).

ARS risk score is correlated with clinical features and predicts the prognosis of clinical-feature subgroups

The prognostic significance of the model was evaluated by survival analysis within age, sex, tumor grade, stage, and vascular invasion subsets in the TCGA-LIHC cohort. The results (Fig. 5A–E) showed a high-risk score was significantly associated with worse prognosis in patients who were >65-years of age (*p*<0.001) and ≤65 years of age (*p*=0.003), female (*p*=0.032), male (*p*<0.001), grades 1–2 (*p*<0.001), stages I–II (*p*<0.001), stages III–IV (*p*=0.033),

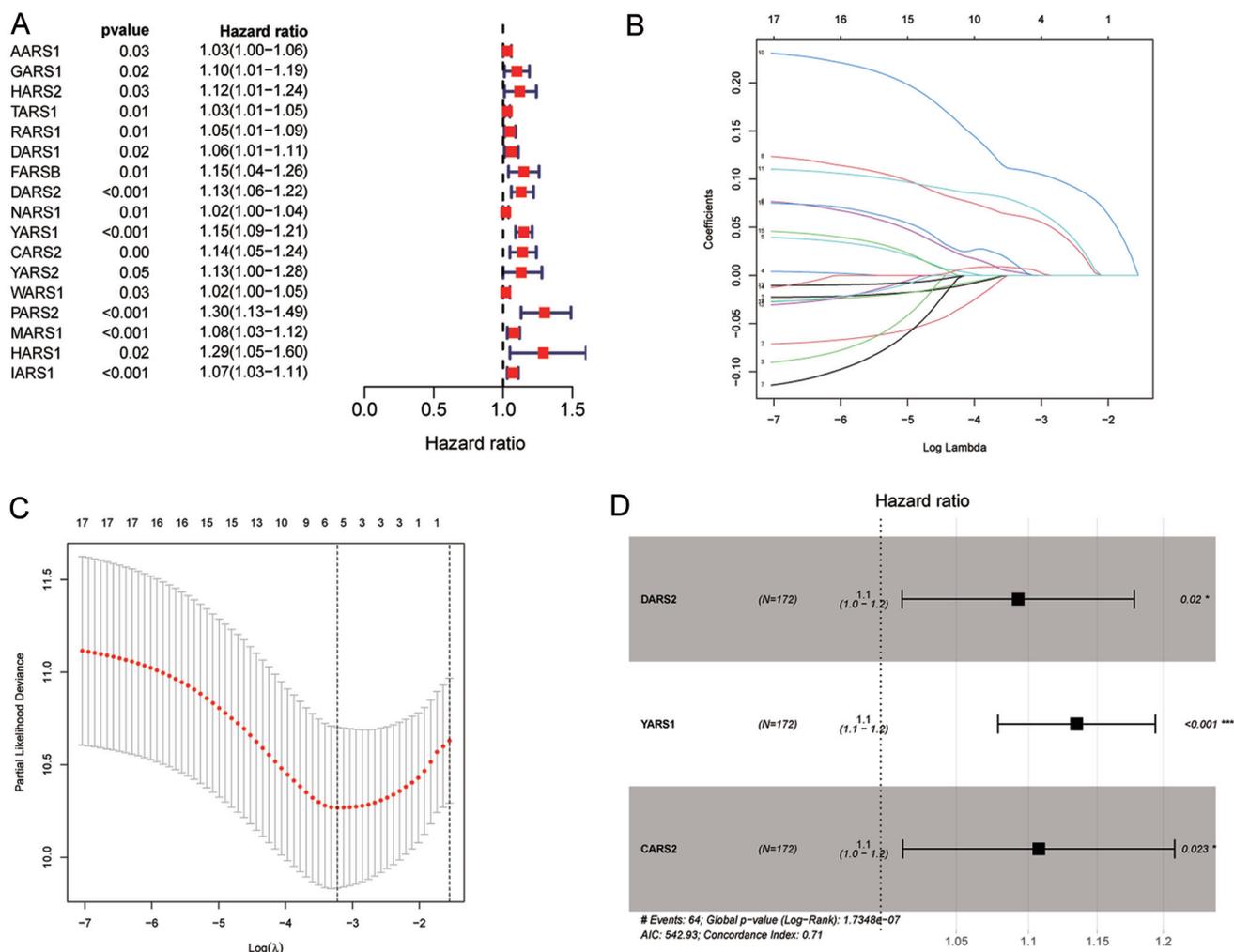


Fig. 2. Prognostic model construction. (A) Forest plot of univariate Cox regression analysis in the training cohort. (B) Lasso coefficient profiles of 17 protein-coding genes. (C) Lasso regression with ten-fold cross-validation obtained six prognostic genes using the minimum lambda value. (D) Forest plot of multivariate Cox regression analysis in the training cohort. Lasso, least absolute shrinkage and selection operator.

and without vascular invasion ($p=0.003$). We also found that patients had higher risk score in grade 3-4, stage III-IV, and vascular invasion subgroup than patients in grade 1-2, stage I-II, and no vascular invasion subgroup respectively. There was no evidence that the risk score was correlated with age, sex, tumor size, AFP, Child-Pugh score, residual tumor, or etiology such as viral hepatitis (Supplementary Fig. 2). A prognostic nomogram (Fig. 5F) was constructed using the risk score and clinical features, which provides an easy-to-use tool to predict the 1-, 2-, and 3-year overall survival of HCC patients.

Analysis of somatic mutations based on risk score model

Because of the importance of gene mutations in carcinogenesis and tumor development, we compared genetic mutational profiles of low-risk and high-risk groups to investigate the underlying mechanism. Information related to gene mutations in the TCGA-LIHC cohort is summarized in Supplementary Figure 3. The frequency of *TTN* mutation was higher in the low-risk than in the high-risk group. *TP53* mutations were more frequent in the high-risk group (Fig. 6A, B).

Functional annotation of DEGs in different risk groups

Differential gene expression analysis was conducted to identify DEGs, and 1590 DEGs were selected for later enrichment analysis (Fig. 7A). GO enrichment results showed that DEGs were enriched in pathways such as organelle fission, nuclear division, mitotic nuclear division, nuclear chromosome and diterpenoid metabolic process (Fig. 7B, Supplementary Fig. 4). KEGG enrichment results indicated that top-ranked enrichment pathways were retinol metabolism, metabolism of xenobiotics by cytochrome P450, cell cycle, drug metabolism-cytochrome P450, chemical carcinogenesis-DNA adducts and neuroactive ligand-receptor interaction (Fig. 7C). GSEA was also applied, and we found genes were enriched in pathways related to activation of immune response, cell cycle and activation of ATR signaling in response to replication stress (Fig. 7D-F, Supplementary Fig. 4).

Correlation between risk score and immune-cell infiltration

Driven by the importance of immune cells in hepatic micro-environment and the result showed in Figure 7D that activation of immune response pathway was enriched in GSEA, we

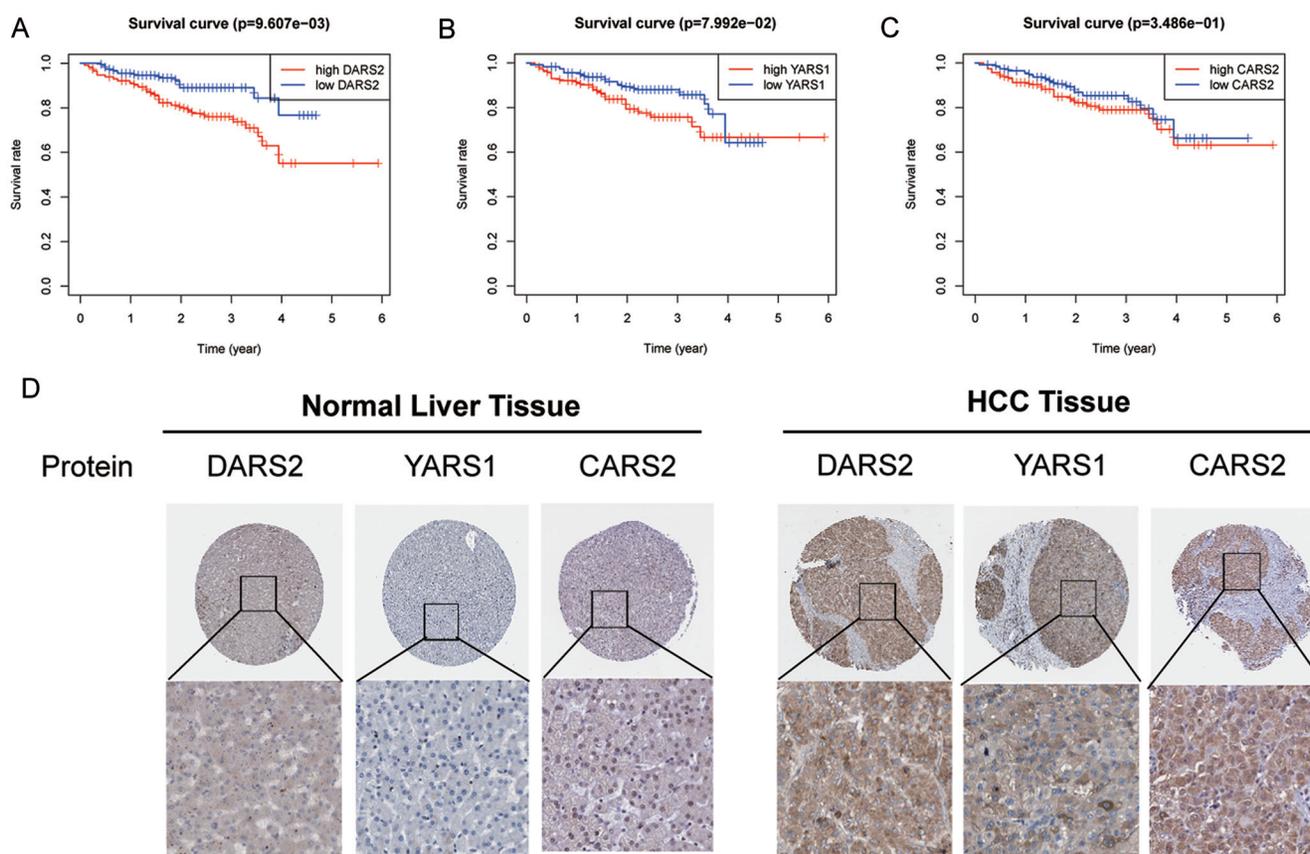


Fig. 3. Survival analysis and Immunohistochemical (IHC) analysis of three prognostic genes determined by other databases. Survival analysis of (A) DARS2, (B) YARS1, and (C) CARS2 in the LIRI-JP cohort of the ICGC database. (D) IHC of DARS2, YARS1, and CARS2 in the HPA database. CARS2, cysteinyl-tRNA synthetase 2; DARS2, aspartyl-tRNA synthetase 2; ICGC, International Cancer Genome Consortium; LIRI-JP, Liver Cancer-RIKEN Japan; HPA, Human Protein Atlas; YARS1, tyrosyl-tRNA synthetase 1.

performed ssGSEA to analyze immune-cell infiltration levels in each tumor sample. Compared low-risk group with high-risk group, we discovered that infiltration levels of activated B cell, activated CD8⁺ T cell, effector memory CD8⁺ T cell, eosinophil, gamma delta T cell, mast cell, memory B cell, and type 1 T helper cell were higher in low-risk group, while infiltration levels of activated CD4⁺ T cell, central memory CD4⁺ T cell and type 2 T helper cell were higher in high-risk group (Fig. 8A). The results of Spearman's correlation analysis showed that risk score was positively correlated with activated CD4⁺ T cell, central memory CD4⁺ T cell but inversely correlated with eosinophil, mast cell and memory B cell (Fig. 8B–F). Higher expression of immunosuppressive molecules such as PD-1, TIM3, CTLA-4 and TIGIT was also discovered in high-risk group (Fig. 8G–J). However, differences in PD-L1, PD-L2, and LAG3 expression in the low- and high-risk groups were not significant. (Fig. 8K, Supplementary Fig. 5). We also calculated TMB of each sample in TCGA-LIHC cohort and found a higher TMB in high-risk group (Fig. 8L).

Discussion

Although various technologies and drugs applied in HCC treatment bring an increase in patient life expectancy, HCC is still a big threat for human health and finding the best treatment remains a conundrum. Therefore, a comprehensive understanding of mechanisms in tumor initiation, progres-

sion, recurrence, metastasis, and resistance is needed. It has been proved that ARSs are involved in tumor progression through their noncanonical functions but their roles in HCC are unspecific.⁸ Herein, we conducted in-silico analyses to explore the role ARSs play and their prognostic value in HCC and presented the first ARS family-based prognostic model.

In this study, univariate Cox regression analysis was conducted in training cohort to find ARSs associated with prognosis and Lasso was carried out considering that multivariate collinearity might exist as ARS genes had similar functions. Multivariate Cox regression was used to select independence prognostic ARS genes and DARS2, YARS1, and CARS2 were screened out, which were then enrolled in the prognostic model. We evaluated the prognostic value of three genes in ICGC and GEO database respectively and tested survival prediction capability of the model in test cohort and subgroups with different clinical features. Although survival time of high-risk patient was shorter in grade 3–4 and vascular invasion subsets, results did not meet the predefined threshold of statistical significance, which might be limited by sample size. In addition, results showed higher risk score was correlated with vascular invasion, higher tumor grade, and stage but irrelevant to hepatitis B or hepatitis C history, which are major causes for HCC.¹⁹ Therefore, the elevation of risk score might not be etiology driven but a universal reaction during hepatocarcinogenesis.

In this model, DARS2, deficiency of which causes leukoencephalopathy, is involved in mitochondrial unfolded

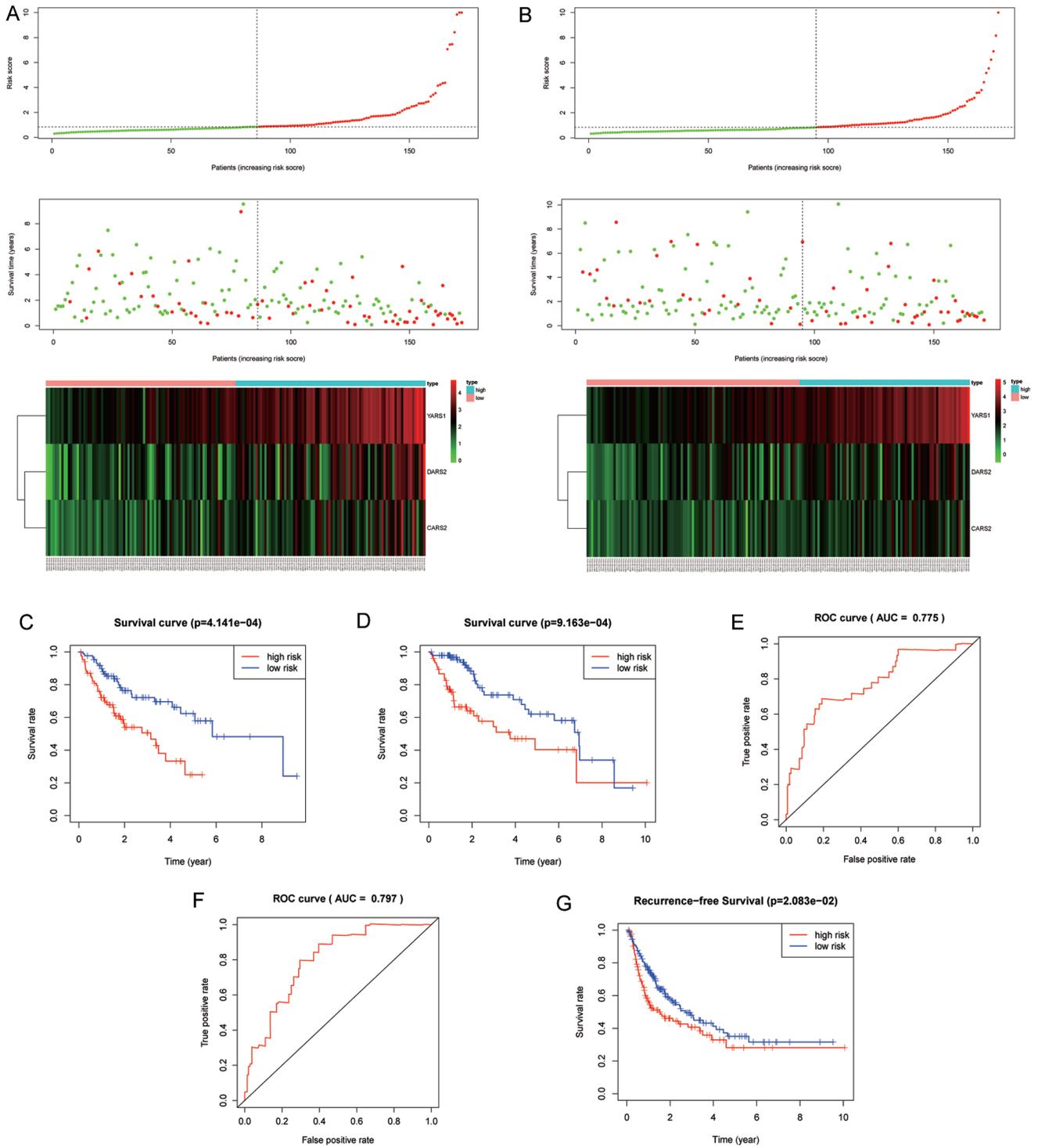


Fig. 4. Prognostic model validation. Survival time, risk score and heat map of the expression of three prognostic genes in (A) the training cohort and (B) the test cohort. Kaplan-Meier survival curves in (C) the training cohort and (D) the test cohort. Receiver operating characteristic curves of the prognostic model in (E) the training cohort and (F) the test cohort. (G) Kaplan-Meier recurrence-free survival curve in the TCGA-LIHC cohort. TCGA-LIHC, The Cancer Genome Atlas-liver hepatocellular carcinoma.

protein response.^{20,21} Liu *et al*,²² have shown that DARS2 is a downstream of NFAT5 and promotes hepatocarcinogenesis through regulating cell cycle and attenuating cell apoptosis, indicating its potential value as a prognosis bio-

marker. YARS1 was recently potentiated to regulate intracellular signaling and associated with chemotherapeutic response.^{23,24} It has been reported that upregulated YARS1 aggravates HCC by mediating PI3K/AKT signaling path-

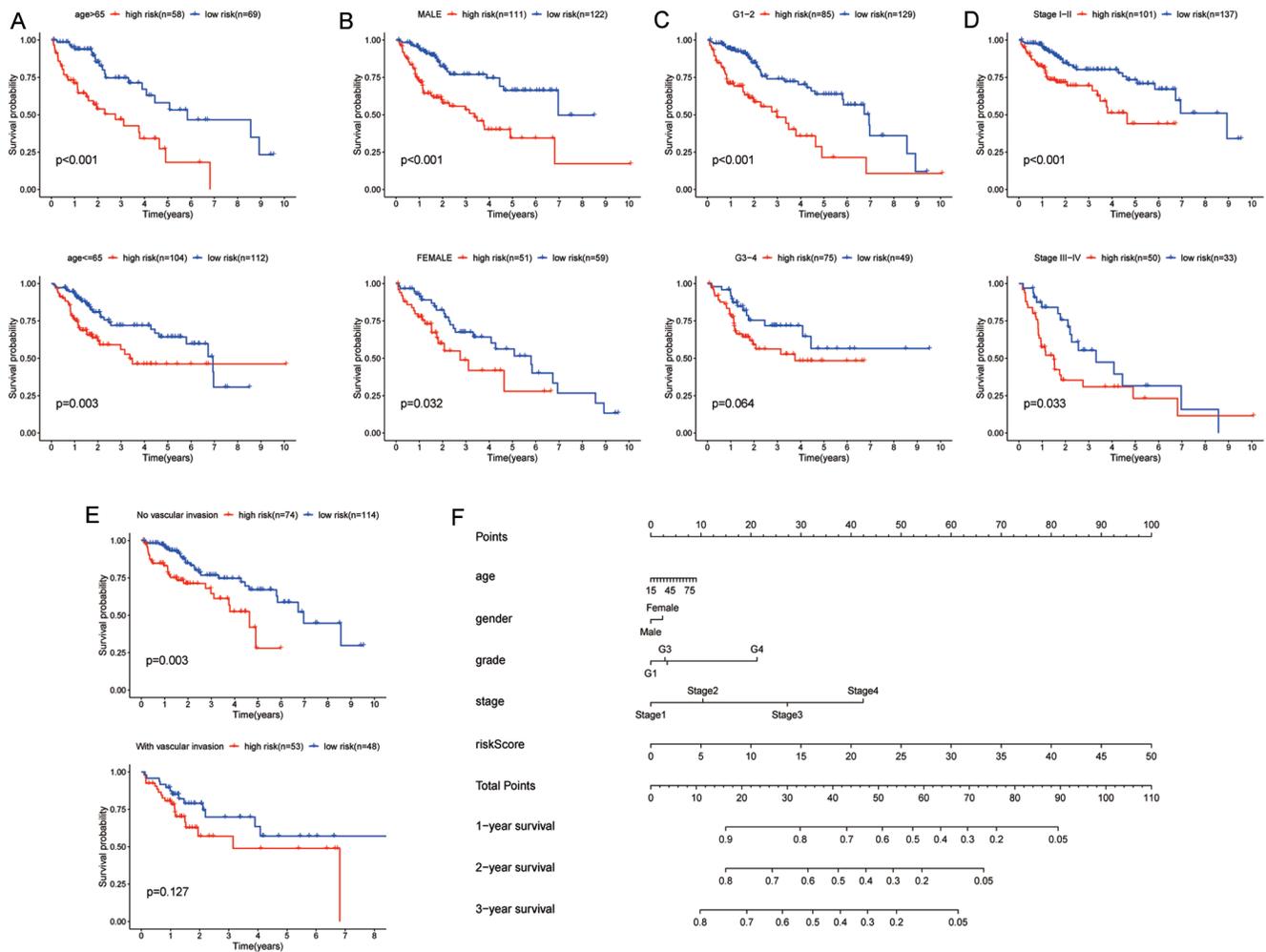


Fig. 5. Prognostic value of the model in clinical-feature subgroups and correlation between clinical features and risk score. Kaplan-Meier survival curves in TCGA-LIHC cohort grouped by (A) age, (B) sex, (C) grade, (D) stage, and (E) vascular invasion. (F) Nomogram of risk score combined with clinical characteristics in the TCGA-LIHC cohort. TCGA-LIHC, The Cancer Genome Atlas-liver hepatocellular carcinoma.

way, which supports our result that YARS1 is related to poor prognosis.^{25,26} Additionally, YARS1 has a C-terminal domain and can be split into two distinct cytokine mimics, one of which is EMAPII.²⁷ EMAPII is a pro-inflammatory cytokine and affects fibronectin disruption and VEGF signaling.²⁸ These results imply that YARS1 could influence prognosis via immune-related mechanism and partially explain why patients with vascular invasion had higher risk score. Although mutation of CARS2 leads to a severe epileptic encephalopathy and complex movement disorder, its role in tumor is still unspecified and further investigations are needed.²⁹ Overall, the three-gene signature successfully predicted the survival of HCC, which drove us to explore the potential underlying mechanisms.

As abnormal and uncontrolled cellular growth in cancer was primarily caused by genetic mutations, we evaluated mutational genes in two distinct groups first and the result showed a higher *TP53* mutation frequency in high-risk group.³⁰ Regulating multiple biologic processes, *TP53* mutational status affect the immune responses and its functional deletion can lead to centrosome amplification and chromosomal instability.^{31,32} It also has been reported that *TP53* mutation frequencies increase significantly with tumor pro-

gression and associate with prognosis of HCC, which can partially account for shorter survival time in high-risk group.³³ Next, we conducted the enrichment analysis to figure out other possible underlying mechanisms and found an enrichment in activation of immune response, showing that immune heterogeneity between low- and high-risk groups is a possible reason for the difference in overall survival.

Immunosuppressive microenvironment is crucial for tumor cells to evade immune surveillance.³⁴ Therefore, we performed ssGSEA to evaluate immune-cell infiltration in tumor microenvironment, which presented a result that infiltration levels of antitumor response related cells, including activated B cell, CD8⁺ T cell, gamma delta T cell, and type 1 T helper cell, were higher in low-risk group, while infiltration levels of cells related to immunosuppression, such as type 2 T helper cell, were higher in high-risk group. We speculated that high-risk score reflected immunosuppression of tumor microenvironment in HCC, which led us to compare expression of immunosuppressive molecules in two distinct groups. Previous studies have shown that PD-1, PD-L1, CTLA-4, TIGIT and TIM-3 are closely related to immunosuppression.^{35,36} Targeting at these immunosuppressive molecules, ICIs herald a new era of systemic treatment. PD-1 inhibitor pembrol-

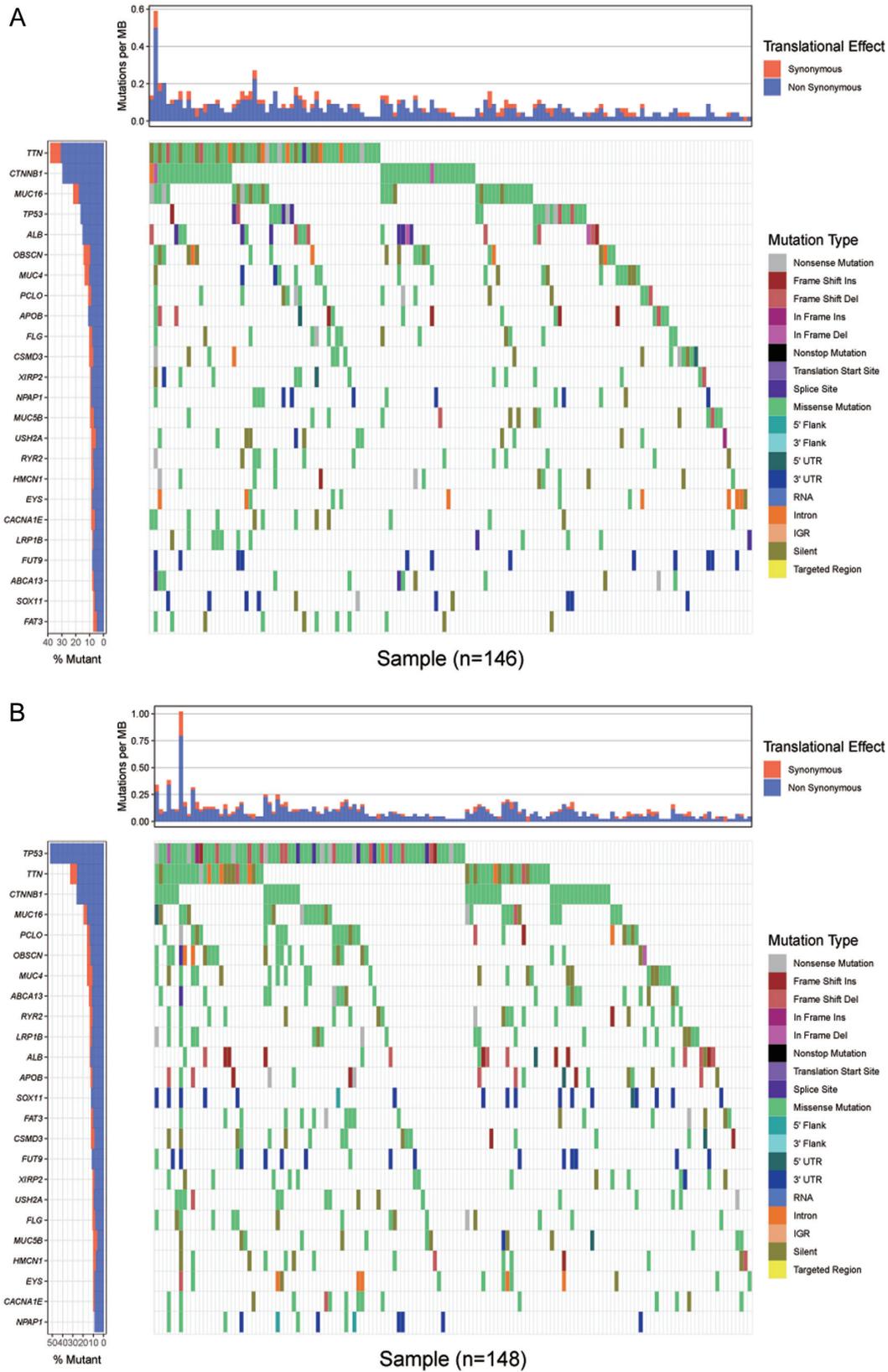


Fig. 6. Mutation profile of the TCGA-LIHC cohort. Waterfall plot of mutation profiles of genes-of-interest in each sample in (A) the low-risk group and (B) the high-risk group. TCGA-LIHC, The Cancer Genome Atlas-liver hepatocellular carcinoma.

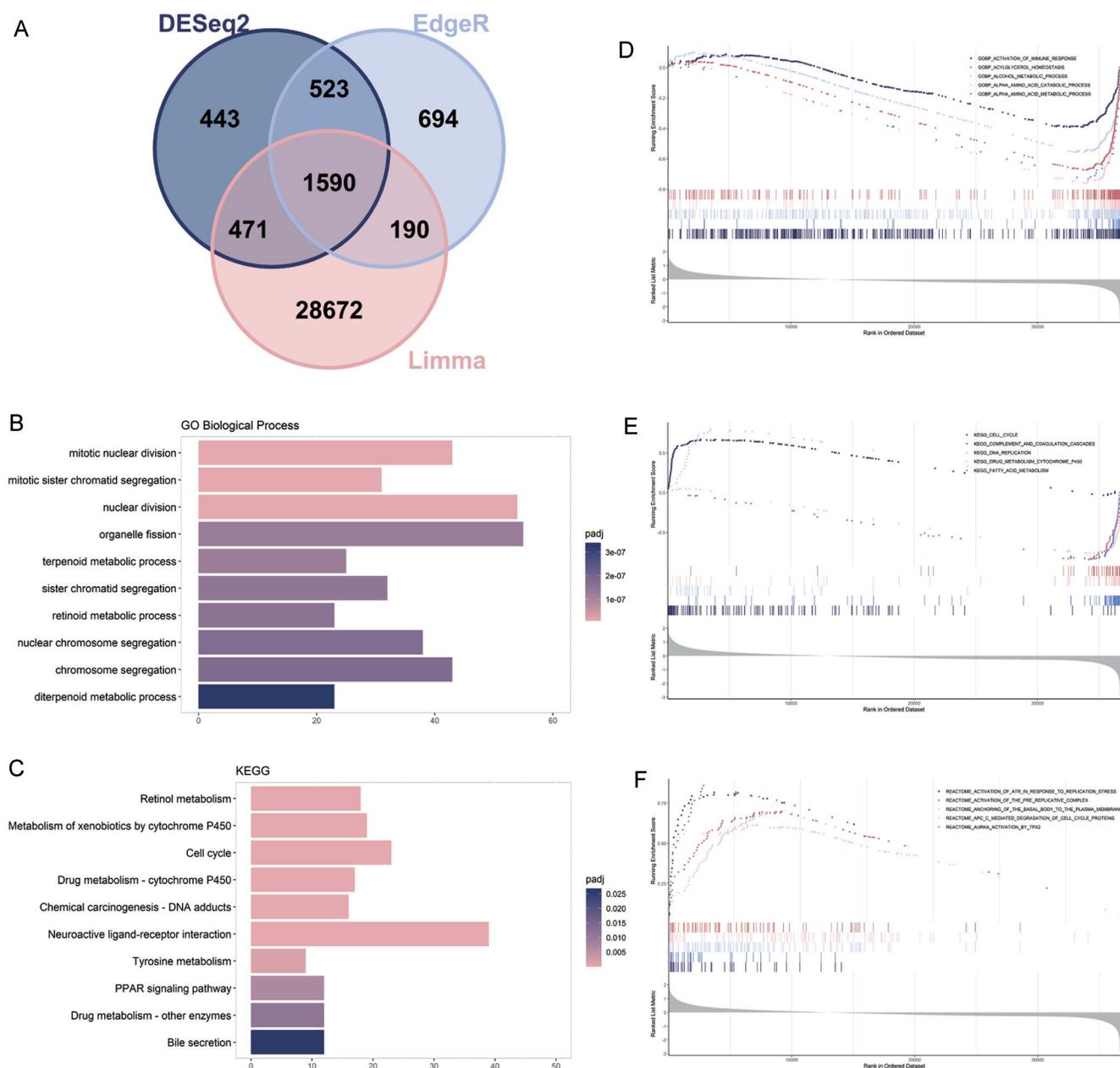


Fig. 7. Enrichment analysis of the low-risk and high-risk groups. (A) Venn diagram of the intersection of differentially expressed genes. (B) Bar plot of Gene Ontology-Biological Process (GO-BP) enrichment analysis results. (C) Bar plot of Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis results. Top five pathway terms of gene set enrichment analysis in (D) GO-BP, (E) KEGG, and (F) REACTOME.

zumab and nivolumab have been accepted as a second-line treatment for patient after sorafenib failure.^{37,38} Combination of ICIs and other agent such as anti-VEGF agent, tyrosine kinase inhibitors will extend patient survival compared with single ICI agent.³⁹ Our results showed that PD-1, CTLA-4, TIGIT, and TIM-3 were highly expressed in the high-risk group, which supported our speculation. Despite the *p*-value not reaching significance, PD-L1 expression was also higher in high-risk group, and further validation is needed. Increasing expression of immunosuppressive molecules in high-risk group hinted that ICIs could be used in high-risk patient treatment and ARS-based prognostic model may act as a predictive biomarker for immunotherapy response, which is

also supported by the result that TMB was elevated in high-risk group.

Although the ARS-based model constructed in this study successfully predicted prognoses of HCC patients and focused on the underlying *TP53* mutation and immune-related mechanism, some limitations remain. First, TCGA cohort did not include all etiological factors for HCC. Although hepatitis B and hepatitis C are not related to risk score, other etiological factors such as nonalcoholic fatty liver disease are not estimated, so whether this model can apply for all etiologies of HCC is remaining for further evaluation. Meanwhile, this model was constructed based on TCGA-LIHC cohort, external validation in prospective cohort and further experimental

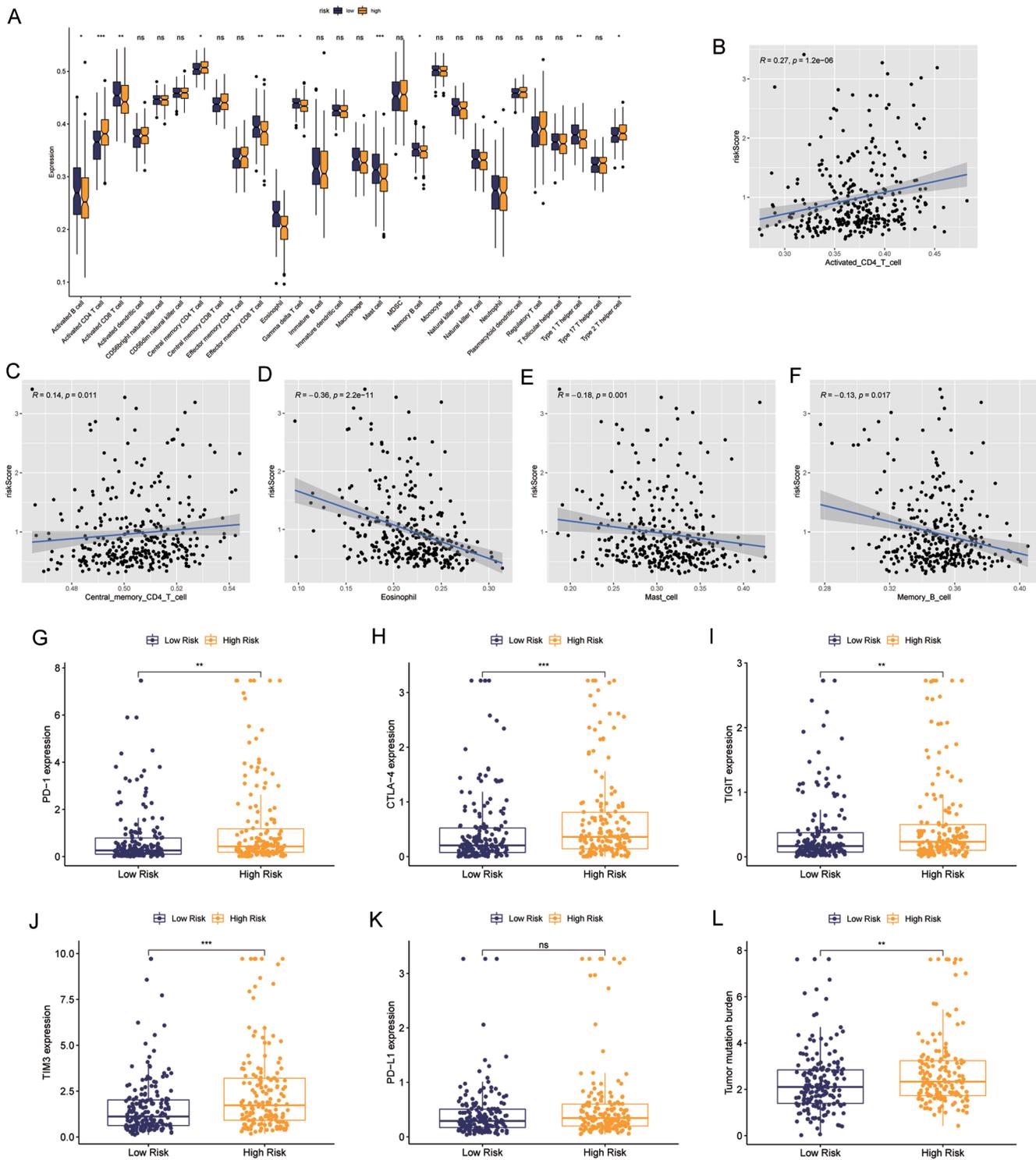


Fig. 8. Analysis of immune-related cells, expression of immunosuppressive molecules, and TMB between low-risk and high-risk groups. (A) Box plot of immune-cell infiltration in the low-risk and high-risk groups. Spearman's correlation analysis of risk score and (B) activated CD4⁺ T cells, (C) central memory CD4⁺ T cells, (D) eosinophils, (E) mast cells, and (F) memory B cells. Box plots of (G) PD-1, (H) CTLA-4, (I) TIGIT, (J) TIM3, and (K) PD-L1 gene expression in low-risk and high-risk groups. Box plot of (L) TMB in the low-risk and high-risk groups. **p*<0.05, ***p*<0.01, ****p*<0.001, ns, not significant. TMB, tumor mutation burden.

verification are required.

This study was the first comprehensive analysis about prognostic value of ARS family members in HCC patients. A

novel ARS-related prognosis model was established and the possible mechanisms behind the model were preliminarily explored. Our findings provided a novel ARS-based prognos-

tic model which might reflect the immune-suppressive state and contribute to developing precision medicine for HCC.

Acknowledgments

We would like to thank individuals and organizations that contributed to TCGA, ICGC, GEO, HPA, GO, KEGG, and REACTOME database and developers of the R packages used in this study. All the websites of these databases and packages have been mentioned above.

Funding

The National Natural Science Foundation of China (Grant Nos. 81870385) supported this study.

Conflict of interest

The authors have no conflict of interests related to this publication.

Author contributions

Conceptualized and designed the study (YJ), collected and interpreted the data (XH), drafted the manuscript (HX, ZW), revised the manuscript for important intellectual content (QW), supervised the study and provided administrative and technical supports (LW). All authors contributed to this study and approved the final manuscript.

Data sharing statement

The cohorts in this study can be found in online repositories. The names and websites of the repositories can be found in the article.

References

[1] Llovet JM, De Baere T, Kulik L, Haber PK, Gretten TF, Meyer T, *et al*. Locoregional therapies in the era of molecular and immune treatments for hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol* 2021;18(5):293–313. doi:10.1038/s41575-020-00395-0, PMID:33510460.

[2] Ringelhan M, Pfister D, O'Connor T, Pikarsky E, Heikenwalder M. The immunology of hepatocellular carcinoma. *Nat Immunol* 2018;19(3):222–232. doi:10.1038/s41590-018-0044-z, PMID:29379119.

[3] Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin* 2022;72(1):7–33. doi:10.3322/caac.21708, PMID:35020204.

[4] Sangro B, Sarobe P, Hervás-Stubbis S, Melero I. Advances in immunotherapy for hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol* 2021;18(8):525–543. doi:10.1038/s41575-021-00438-0, PMID:33850328.

[5] Finn RS, Qin S, Ikeda M, Galle PR, Ducreux M, Kim TY, *et al*. Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma. *N Engl J Med* 2020;382(20):1894–1905. doi:10.1056/NEJMoa1915745, PMID:32402160.

[6] Choi WM, Choi J, Lee D, Shim JH, Lim YS, Lee HC, *et al*. Regorafenib Versus Nivolumab After Sorafenib Failure: Real-World Data in Patients With Hepatocellular Carcinoma. *Hepatol Commun* 2020;4(7):1073–1086. doi:10.1002/hep4.1523, PMID:32626838.

[7] Yao P, Fox PL. Aminoacyl-tRNA synthetases in medicine and disease. *EMBO Mol Med* 2013;5(3):332–343. doi:10.1002/emmm.201100626, PMID:23427196.

[8] Kwon NH, Fox PL, Kim S. Aminoacyl-tRNA synthetases as therapeutic targets. *Nat Rev Drug Discov* 2019;18(8):629–650. doi:10.1038/s41573-019-0026-3, PMID:31073243.

[9] Zhou Z, Sun B, Nie A, Yu D, Bian M. Roles of Aminoacyl-tRNA Synthetases in Cancer. *Front Cell Dev Biol* 2020;8:599765. doi:10.3389/fcell.2020.599765, PMID:33330488.

[10] Han JM, Jeong SJ, Park MC, Kim G, Kwon NH, Kim HK, *et al*. Leucyl-tRNA synthetase is an intracellular leucine sensor for the mTORC1-signaling pathway. *Cell* 2012;149(2):410–424. doi:10.1016/j.cell.2012.02.044, PMID:22424946.

[11] Kim DG, Choi JW, Lee JY, Kim H, Oh YS, Lee JW, *et al*. Interaction of two translational components, Lysyl-tRNA synthetase and p40/37LRP, in plasma membrane promotes laminin-dependent cell migration. *Faseb J* 2012;26(10):4142–4159. doi:10.1096/fj.12-207639, PMID:22751010.

[12] Yum MK, Kang JS, Lee AE, Jo YW, Seo JY, Kim HA, *et al*. AIMP2 Controls Intestinal Stem Cell Compartments and Tumorigenesis by Modu-

lating Wnt/ β -Catenin Signaling. *Cancer Res* 2016;76(15):4559–4568. doi:10.1158/0008-5472.Can-15-3357, PMID:27262173.

[13] Jung JY, Kim EY, Kim A, Chang J, Kwon NH, Moon Y, *et al*. Ratio of Autoantibodies of Tumor Suppressor AIMP2 and Its Oncogenic Variant Is Associated with Clinical Outcome in Lung Cancer. *J Cancer* 2017;8(8):1347–1354. doi:10.7150/jca.18450, PMID:28638448.

[14] Lee JM, Kim T, Kim EY, Kim A, Lee DK, Kwon NH, *et al*. Methionyl-tRNA Synthetase is a Useful Diagnostic Marker for Lymph Node Metastasis in Non-Small Cell Lung Cancer. *Yonsei Med J* 2019;60(11):1005–1012. doi:10.3349/ymj.2019.60.11.1005, PMID:31637881.

[15] Qi L, Zhou B, Chen J, Hu W, Bai R, Ye C, *et al*. Significant prognostic values of differentially expressed-aberrantly methylated hub genes in breast cancer. *J Cancer* 2019;10(26):6618–6634. doi:10.7150/jca.33433, PMID:31777591.

[16] Suh JH, Park MC, Goughnour PC, Min BS, Kim SB, Lee WY, *et al*. Plasma Lysyl-tRNA Synthetase 1 (KARS1) as a Novel Diagnostic and Monitoring Biomarker for Colorectal Cancer. *J Clin Med* 2020;9(2):533. doi:10.3390/jcm9020533, PMID:32075312.

[17] Zhang X, Dong W, Zhang J, Liu W, Yin J, Shi D, *et al*. A Novel Mitochondrial-Related Nuclear Gene Signature Predicts Overall Survival of Lung Adenocarcinoma Patients. *Front Cell Dev Biol* 2021;9:740487. doi:10.3389/fcell.2021.740487, PMID:34760888.

[18] Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013;14:7. doi:10.1186/1471-2105-14-7, PMID:23323831.

[19] Sagnelli E, Macera M, Russo A, Coppola N, Sagnelli C. Epidemiological and etiological variations in hepatocellular carcinoma. *Infection* 2020;48(1):7–17. doi:10.1007/s15010-019-01345-y, PMID:31347138.

[20] Scheper GC, van der Kloot T, van Andel RJ, van Berkel CG, Sissler M, Smet J, *et al*. Mitochondrial aspartyl-tRNA synthetase deficiency causes leukoencephalopathy with brain stem and spinal cord involvement and lactate elevation. *Nat Genet* 2007;39(4):534–539. doi:10.1038/ng2013, PMID:17384640.

[21] Seiferling D, Szczepanowska K, Becker C, Senft K, Hermans S, Maiti P, *et al*. Loss of CLPP alleviates mitochondrial cardiomyopathy without affecting the mammalian UPRmt. *EMBO Rep* 2016;17(7):953–964. doi:10.15252/embr.201642077, PMID:27154400.

[22] Qin X, Li C, Guo T, Chen J, Wang HT, Wang YT, *et al*. Upregulation of DARS2 by HBV promotes hepatocarcinogenesis through the miR-30e-5p/MAPK/NFAT5 pathway. *J Exp Clin Cancer Res* 2017;36(1):148. doi:10.1186/s13046-017-0618-x, PMID:29052520.

[23] Sun J, Lv PC, Zhu HL. Tyrosyl-tRNA synthetase inhibitors: a patent review. *Expert Opin Ther Pat* 2017;27(5):557–564. doi:10.1080/13543776.2017.1273350, PMID:27977303.

[24] Lee KM, Lee H, Han D, Moon WK, Kim K, Oh HJ, *et al*. Combined the SMAC mimetic and BCL2 inhibitor sensitizes neoadjuvant chemotherapy by targeting necrosome complexes in tyrosine aminoacyl-tRNA synthase-positive breast cancer. *Breast Cancer Res* 2020;22(1):130. doi:10.1186/s13058-020-01367-7, PMID:33239070.

[25] Zhou W, Wang J, Zhang J, Wang Y, Jiang L, Guo T, *et al*. LncRNA NCK1-AS1 Aggravates Hepatocellular Carcinoma by the miR-22-3p/YARS Axis to Activate PI3K/AKT Signaling. *J Gastrointest Liver Dis* 2022;31(1):48–59. doi:10.15403/jglid-4077, PMID:35306563.

[26] Zhang C, Lin X, Zhao Q, Wang Y, Jiang F, Ji C, *et al*. YARS as an oncogenic protein that promotes gastric cancer progression through activating PI3K-Akt signaling. *J Cancer Res Clin Oncol* 2020;146(2):329–342. doi:10.1007/s00432-019-03115-7, PMID:31912229.

[27] Wakasugi K, Schimmel P. Two distinct cytokines released from a human aminoacyl-tRNA synthetase. *Science* 1999;284(5411):147–151. doi:10.1126/science.284.5411.147, PMID:10102815.

[28] Awasthi N, Zhang C, Hinz S, Schwarz MA, Schwarz RE. Enhancing sorafenib-mediated sensitization to gemcitabine in experimental pancreatic cancer through EMAP II. *J Exp Clin Cancer Res* 2013;32(1):12. doi:10.1186/1756-9966-32-12, PMID:23497499.

[29] Coughlin CR 2nd, Scharer GH, Friederich MW, Yu HC, Geiger EA, Creadon-Swindell G, *et al*. Mutations in the mitochondrial cysteinyl-tRNA synthase gene, CARS2, lead to a severe epileptic encephalopathy and complex movement disorder. *J Med Genet* 2015;52(8):532–540. doi:10.1136/jmedgenet-2015-103049, PMID:25787132.

[30] Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, *et al*. A compendium of mutational cancer driver genes. *Nat Rev Cancer* 2020;20(10):555–572. doi:10.1038/s41568-020-0290-x, PMID:32778778.

[31] Schwaederlé M, Lazar V, Validire P, Hansson J, Lacroix L, Soria JC, *et al*. VEGF-A Expression Correlates with TP53 Mutations in Non-Small Cell Lung Cancer: Implications for Antiangiogenesis Therapy. *Cancer Res* 2015;75(7):1187–1190. doi:10.1158/0008-5472.Can-14-2305, PMID:25672981.

[32] Long J, Wang A, Bai Y, Lin J, Yang X, Wang D, *et al*. Development and validation of a TP53-associated immune prognostic model for hepatocellular carcinoma. *EBioMedicine* 2019;42:363–374. doi:10.1016/j.ebiom.2019.03.022, PMID:30885723.

[33] Schulze K, Imbeaud S, Letouzé E, Alexandrov LB, Calderaro J, Rebouisson S, *et al*. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet* 2015;47(5):505–511. doi:10.1038/ng.3252, PMID:25822088.

[34] Rabinovich GA, Gabrilovich D, Sotomayor EM. Immunosuppressive strategies that are mediated by tumor cells. *Annu Rev Immunol* 2007;25:267–296. doi:10.1146/annurev.immunol.25.022106.141609, PMID:17134371.

[35] Anderson AC, Joller N, Kuchroo VK. Lag-3, Tim-3, and TIGIT: Co-inhibitory Receptors with Specialized Functions in Immune Regulation. *Immunity* 2016;44(5):989–1004. doi:10.1016/j.immuni.2016.05.001, PMID:2719

- 2565.
- [36] Togashi Y, Shitara K, Nishikawa H. Regulatory T cells in cancer immunosuppression - implications for anticancer therapy. *Nat Rev Clin Oncol* 2019;16(6):356-371. doi:10.1038/s41571-019-0175-7, PMID:30705439.
- [37] El-Khoueiry AB, Sangro B, Yau T, Crocenzi TS, Kudo M, Hsu C, *et al*. Nivolumab in patients with advanced hepatocellular carcinoma (Check-Mate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet* 2017;389(10088):2492-2502. doi:10.1016/s0140-6736(17)31046-2, PMID:28434648.
- [38] Zhu AX, Finn RS, Edeline J, Cattani S, Ogasawara S, Palmer D, *et al*. Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): a non-randomised, open-label phase 2 trial. *Lancet Oncol* 2018;19(7):940-952. doi:10.1016/s1470-2045(18)30351-6, PMID:29875066.
- [39] Ouyang T, Kan X, Zheng C. Immune Checkpoint Inhibitors for Advanced Hepatocellular Carcinoma: Monotherapies and Combined Therapies. *Front Oncol* 2022;12:898964. doi:10.3389/fonc.2022.898964, PMID:35785169.