Heliyon 10 (2024) e33637

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Integrative nomogram model based on anoikis-related genes enhances prognostic evaluation in colorectal cancer

Yuexiao Zhang [a], Xia Xue [a], Fazhan Li [a], Bo Zhang [a], Pengyuan Zheng [a,b,*], Yang Mi [a,b,**]

[a] Henan Key Laboratory of Helicobacter Pylori & Microbiota and Gastrointestinal Cancer, Marshall B. J. Medical Research Center, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, 450052, PR China
[b] Department of Gastroenterology, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, 450052, PR China

ARTICLE INFO

Keywords:
Anoikis
LASSO machine learning algorithm
Nomogram
Clinical prediction model
Generalizability

ABSTRACT

Background: Revealing the role of anoikis resistance plays in CRC is significant for CRC diagnosis and treatment. This study integrated the CRC anoikis-related key genes (CRC-AKGs) and established a novel model for improving the efficiency and accuracy of the prognostic evaluation of CRC.

Methods: CRC-ARGs were screened out by performing differential expression and univariate Cox analysis. CRC-AKGs were obtained through the LASSO machine learning algorithm and the LASSO Risk-Score was constructed to build a nomogram clinical prediction model combined with the clinical predictors. In parallel, this work developed a web-based dynamic nomogram to facilitate the generalization and practical application of our model.

Results: We identified 10 CRC-AKGs and a risk-related prognostic Risk-Score was calculated. Multivariate COX regression analysis indicated that the Risk-Score, TNM stage, and age were independent risk factors that significantly associated with the CRC prognosis($p < 0.05$). A prognostic model was built to predict the outcome with satisfied accuracy (3-year AUC = 0.815) for CRC individuals. The web interactive nomogram (https://yuexiaozhang.shinyapps.io/anoikisCRC/) showed strong generalizability of our model. In parallel, a substantial correlation between tumor microenvironment and Risk-Score was discovered in the present work.

Conclusion: This study reveals the potential role of anoikis in CRC and sets new insights into clinical decision-making in colorectal cancer based on both clinical and sequencing data. Also, the interactive tool provides researchers with a user-friendly interface to input relevant clinical variables and obtain personalized risk predictions or prognostic assessments based on our established model.

* Corresponding author. Department of Gastroenterology, The Fifth Affiliated Hospital of Zhengzhou University; Marshall B. J. Medical Research Center of Zhengzhou University, Kangfu Street No. 3, Zhengzhou, 450052, PR China.

** Corresponding author. Department of Gastroenterology, The Fifth Affiliated Hospital of Zhengzhou University; Marshall B. J. Medical Research Center of Zhengzhou University, Kangfu Street No. 3, Zhengzhou, 450052, PR China.

E-mail addresses: medp7123@126.com (P. Zheng), yangmi198@zzu.edu.cn (Y. Mi).

https://doi.org/10.1016/j.heliyon.2024.e33637
Received 27 February 2024; Received in revised form 24 June 2024; Accepted 25 June 2024
Available online 26 June 2024

1. Introduction

Colorectal cancer (CRC) poses a significant risk to global health with substantial morbidity and mortality [1]. The primary therapeutics nowadays for CRC are surgical resection, adjuvant radiotherapy, and chemotherapy [2,3]. However, the high heterogeneity among CRC patients, along with limitations in current prognostic assessment protocols, can result in the oversight of critical factors during the early stages and ultimately lead to prognostic evaluation failures [3,4]. In parallel, the selection of appropriate adjuvant radiotherapy and chemotherapy is challenging and brings a diverse survival prognosis [5]. Thereby, adopting an efficient and accurate assessment model is significant in predicting and evaluating the prognosis for CRC individuals.

Anoikis is a programmed death of normal cells that occurs under a prolonged suspension [6]. This cellular suicide is induced by cell detachment from the extracellular matrix (ECM) [7], which is an adaptation strategy to maintain cellular homeostasis and prevent tumor metastasis [8]. Instead of anoikis, cancer cells lose their attachment to surrounding tissues and enter the circulation system during tumor progression [9]. Cancer cells have the ability to evade clearance by anoikis and invade other tissues, leading to the formation of metastases. This is achieved through the acquisition of resistance to anoikis [10]. The 14-3-3σ protein makes liver cancer cells acquire resistance to anoikis by activating the ERK1/2 pathway, thus promoting the metastasis of liver cancer [11]. Multiple anoikis-related genes (ARGs) have been found to regulate the occurrence and development of cancer. For instance, CEMIP increases prostate cancer invasion and metastasis by promoting anoikis resistance in prostate cancer [12]. Increased CLDN1 enhances gastric cancer metastasis [13]. However, the role of anoikis-related genes in CRC remains unclear.

Prediction models that incorporate diverse types of data and algorithms are increasingly recognized as more reliable and robust strategies in precision medicine [14]. These models integrate various predictors, including clinical data, pathology results, and sequencing data, to enhance their predictive capabilities [15,16]. By employing multiple calculating methods such as logistic regression, COX regression, nomograms, and machine learning, these models ensure the robustness and accuracy of the predictions [17].

Nomogram clinical prediction models have the advantage of transforming complex regression equations into visual graphics. This visual representation makes the results of the prediction model highly understandable and facilitates their interpretation during clinical evaluation [18,19], including ferroptosis-related prognostic model [20], cuproptosis-related prognostic model [21], and disulfidptosis-related prognostic model [22]. However, a lack of prediction models based on anoikis genes limits the explanation of CRC prognosis evaluation. We herein aim to develop a model for assessing CRC prognosis using ARGs. In parallel, we try to provide an available model into an interactive website for wide and open-assess use.

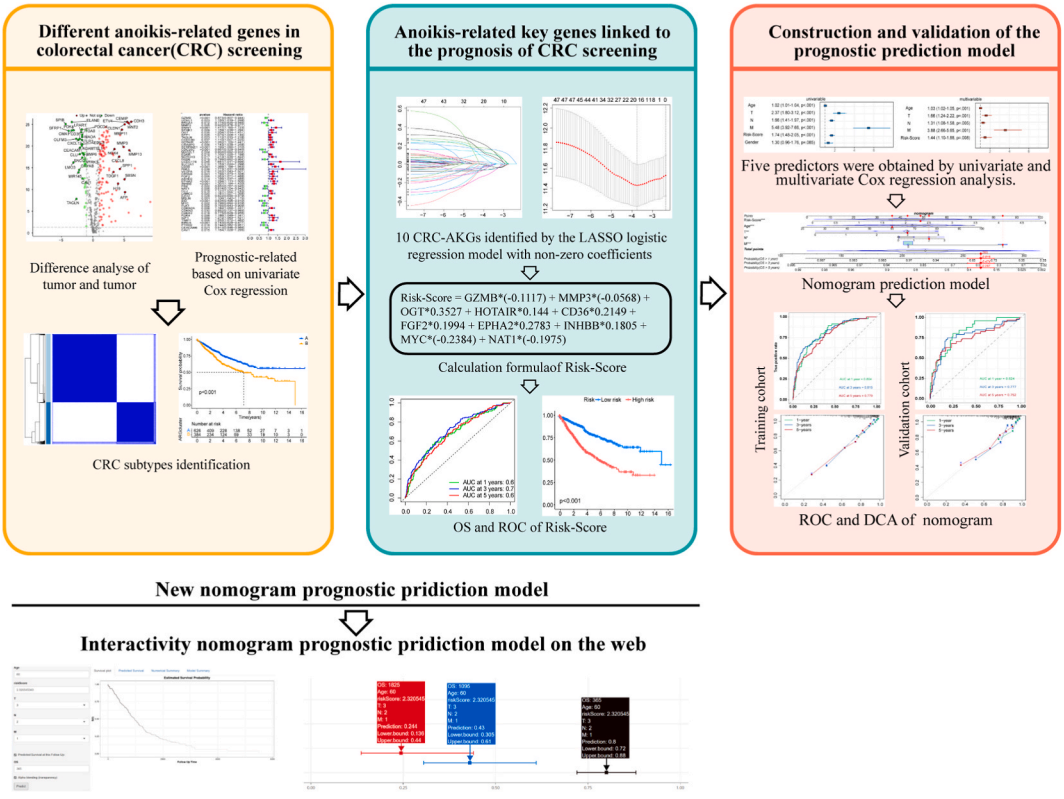

Fig. 1. The flow chart of this study.

2. Methods and materials

In this study, we investigated the role of anoikis-related key genes (AKGs) in CRC and evaluated the rational integration of CRC-AKGs, the Lasso machine learning algorithm, and nomogram in clinical prediction models. The primary research concepts and procedures of this study were described in Fig. 1.

2.1. Data collection

The CRC patient data from TCGA (https://www.cancer.gov/ccg/research/genome-sequencing/tcga) and Gene Expression Omnibus databases(GSE39582, https://www.ncbi.nlm.nih.gov/geo/) were obtained and filtered. We obtained a total of 1010 samples, of which 918 had complete clinical information. The relevant clinical factors were presented in Supplementary Table S1.

2.2. ARGs associated with prognosis in CRC screening

RNA sequencing data from TCGA was subjected to differential gene expression analysis. ARGs were collected from the anoikis database (https://www.genecards.org/, https://maayanlab.cloud/Harmonizome/). The differentially expressed ARGs between tumor and normal cases were identified with the limma package (v3.52.4) and applied the following criteria: $|\log FC|>1$ and $p.adj<0.05$. By integrating clinical prognostic data, CRC-ARGs were calculated using univariate Cox regression hazard analysis [23].

2.3. Tumor mutational burden

We analyzed the CRC-ARGs mutations from the tumor tissues using the maftools package (v2.16.0) and visualized it by complexheatmap package (v2.12.1). In parallel, the genetic locus and copy number variation (CNV) of CRC-ARGs in RCircos plot was generated using RCircos package (v1.2.2).

2.4. CRC clusters identification

We employed sva package (v3.44.0) [24] to eliminate batch effects of expression profiles from TCGA and GEO. The data were normalized and combined for further analysis. The hierarchical clustering analysis was carried out using the package ConsensusClusterPlus (v1.60.0) [25]. The number of clusters(K) with the lowest proportion of ambiguous clustering (PAC) score was considered the optimal cluster number. The Principal Component Analysis (PCA) was performed using the "prcomp" R function and plotted using the program ggplot2 (v3.4.4). By integrating clinical data and prognosis, the prognostic significance of two clusters was investigated using Kaplan-Meier survival curve analysis and receiver operating characteristic (ROC) curve analysis [26].

2.5. Evaluation of LASSO Risk-Score and its prognostic significance

We randomly divided the combined dataset into training (70 %) and validation (30 %) sets [27]. The LASSO was performed on training set using 10-fold cross validation and identified the variables with a non-zero β-coefficient [28,29]. The CRC-AKGs were identified for the construction and calculation of the Risk-Score. The calculation method and formula for the Risk-Score, denoted as Risk-Score = coefficientα*geneA + coefficientβ*geneB + coefficientγ*geneC, were recorded and demonstrated [28]. The median LASSO Risk-Score was set as the cut-off value [30], patients were classified into "High risk group" and "Low risk group". In addition, Kaplan-Meier survival analysis and ROC curve [26] between the two groups were performed. We used the Student's t-test/Kruskal-Wallis test to evaluate LASSO Risk-Score and clinical factors. All differences were considered significant as $p < 0.05$.

2.6. CRC-AKGs expression and prognostic significance in CRC

We investigated the DNA and protein expression levels of these CRC-AKGs using the HPA database (https://www.proteinatlas.org/). Moreover, we conducted survival analysis and ROC curve to evaluate their prognostic significance in CRC [26].

2.7. Predictive nomogram construction, utility verification and promotion

To screen out appropriate predictors, the LASSO Risk-Score and clinical factors were subjected to univariate and multivariate COX analysis. The threshold for statistical significance was set as $p < 0.05$. We calculated the area under the ROC curve (AUC) to quantify the predictive value [31]. The diagnostic power of the test was found to be good, moderate, and poor when the area under the curve (AUC) was greater than 0.8, between 0.7 and 0.8, and less than 0.7, respectively [31,32]. Calibration curves was used to evaluate the calibration of the model [33]. The closer the prediction line is to the reference line (diagonal), the better the model is fitted [34]. Decision curves analysis (DCA) and area of decision curves (AUDC) were used to test clinical utility of the model [35]. In DCA, the ordinate axis is the net benefit. Thus, "Nomogram" line in the DCA plot was above the "None" and "All" lines, indicating that the predictive model had strong clinical utility [36]. If the AUDC of the nomogram line is significantly higher than the AUDC of all lines, it means that the higher the overall position of the nomogram line in the axes, the greater its clinical utility [37].

Model validation was performed on the validation dataset. The R packages were dependent on the following support package:

glmnet v4.1, rms v6.7-1, dcurves v0.4.0, and timeROC v0.4.0. In the shinyapp platform (https://www.shinyapps.io/), a dynamic nomogram has been developed and is now available online (https://yuexiaozhang.shinyapps.io/anoikisCRC/). Dynamic nomogram can be accessed and used totally freely through network devices, which greatly promotes the generalizability and utility of our model.

2.8. Evaluation of immune environment

To estimate the relative abundance of immune cells within the tumor microenvironment (TME), we used the CIBERSORT algorithm [38]. The differential analysis between "High risk group" and the "Low risk group" was carried out using the limma package (v3.52.4). A non-parametric Spearman's rank correlation was used to find out the association between CRC-AKGs and immune cells. The ESTIMATE algorithm was used to evaluate the relationship between Risk-Score and ImmuneScore, StromalScore, and ESTIMATEScore.

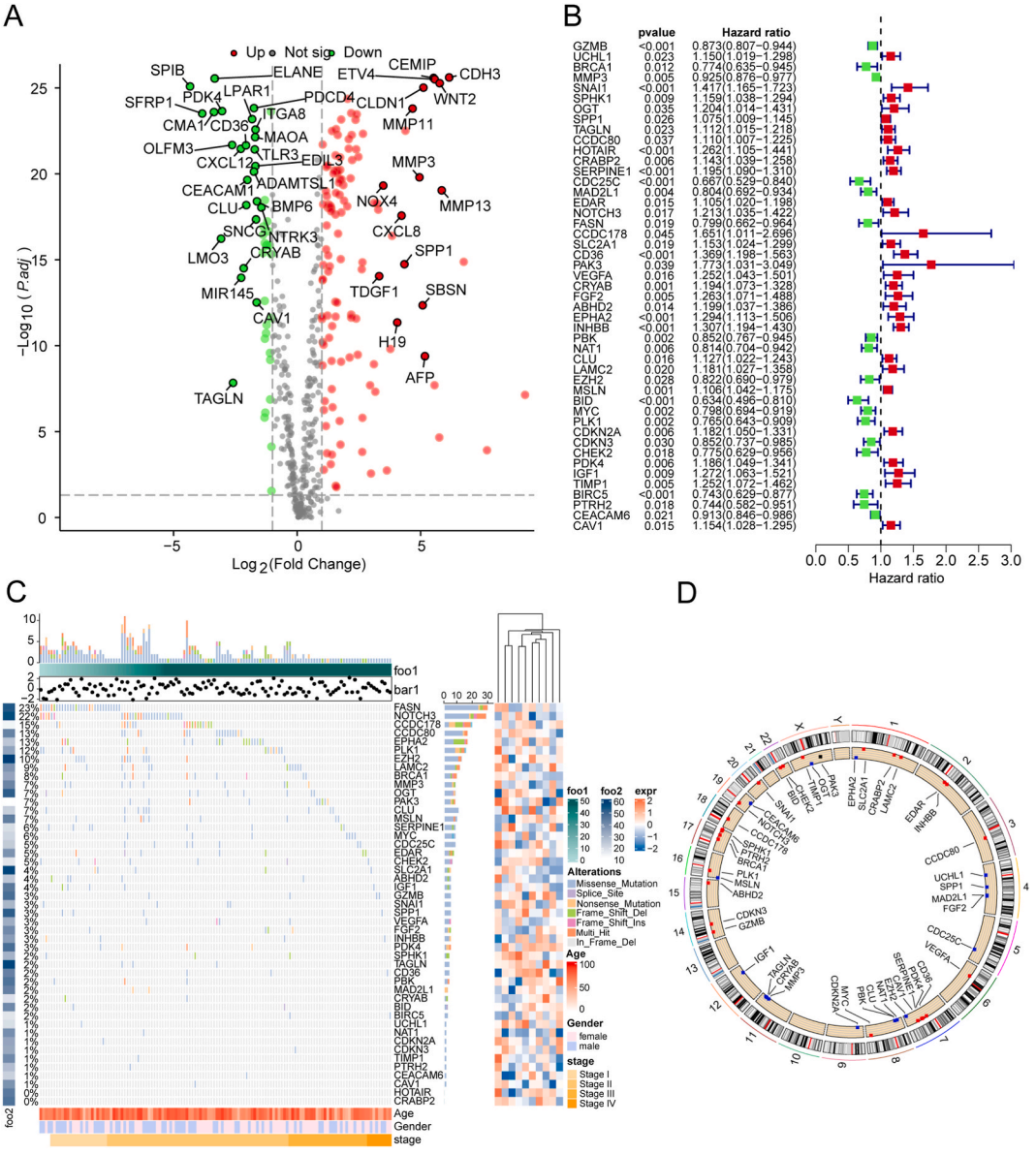

Fig. 2. Screening the CRC-ARGs in CRC. (A). Volcano plot of DEGs in CRC. (B). Univariate Cox risk regression analysis: forest plot of the association between CRC-ARGs and CRC survival. (C). Somatic mutation status of CRC-ARGs in CRC visualized via waterfall plot. (D). The localization of the CRC-ARGs on TCGA-GBM 23 chromosomes. Red represents high levels of CNV, and blue represents low levels of CNV.

2.9. Single cell sequencing analysis

Based on our prior study, this work investigated the distribution of various immune cells in CRC by using public data (GSE108989) from the Tumor Immunity Single Cell Hub (TISCH, http://tisch.comp-genomics.org/). We investigated the expression and distribution of CRC-AKGs in CRC cells and various immune cells as well.

2.10. Statistical analysis

The statistics in this study were conducted using R software (v4.0.1), with the corresponding algorithms and codes documented in Supplementary File 1. The primary data were expressed as mean ± standard deviation. Statistical comparisons between two subgroups were conducted by using Student's t-test, while multiple comparisons were conducted by using Kruskal-Wallis test. The signifcant differences were considered as $p < 0.05$.

3. Results

3.1. Screening and clustering of CRC-ARGs

From the TCGA-CRC dataset, we extracted 168 differentially expressed ARGs (adjusted $p < 0.05$), with 120 upregulated and 48 downregulated genes (Fig. 2A). Univariable Cox regression analysis showed 47 genes associated with the CRC prognosis, which were referred to be CRC-ARGs. The 30 genes were correlated with poor clinical prognosis (Fig. 2B). The waterfall chart showed the mutation of CRC-ARGs in the CRC (Fig. 2C). We found a total of 45 CRC-ARGs were mutated in CRC, with *FASH* and *NOTCH3* showed above 20 % mutation rates. In addition, the CNV of CRC-ARGs in CRC showed common, and their location on the chromosome was shown in Fig. 2D.

After screening CRC-ARGs, we performed a hierarchical clustering analysis on the combined dataset. According to the PAC score, we determined that the optimal number of clusters was 2. We classified the CRC patients into cluster A and B based on CRC-ARGs expression (Fig. 3A and B). Kaplan-Meier survival analysis indicated that the differences in survival between the clusters were significant ($p < 0.001$, Fig. 3C). ROC analysis showed poor predictive performance for 1,3,5-year survival with AUC values of 0.561, 0.566, 0.562, respectively (Fig. 3D).

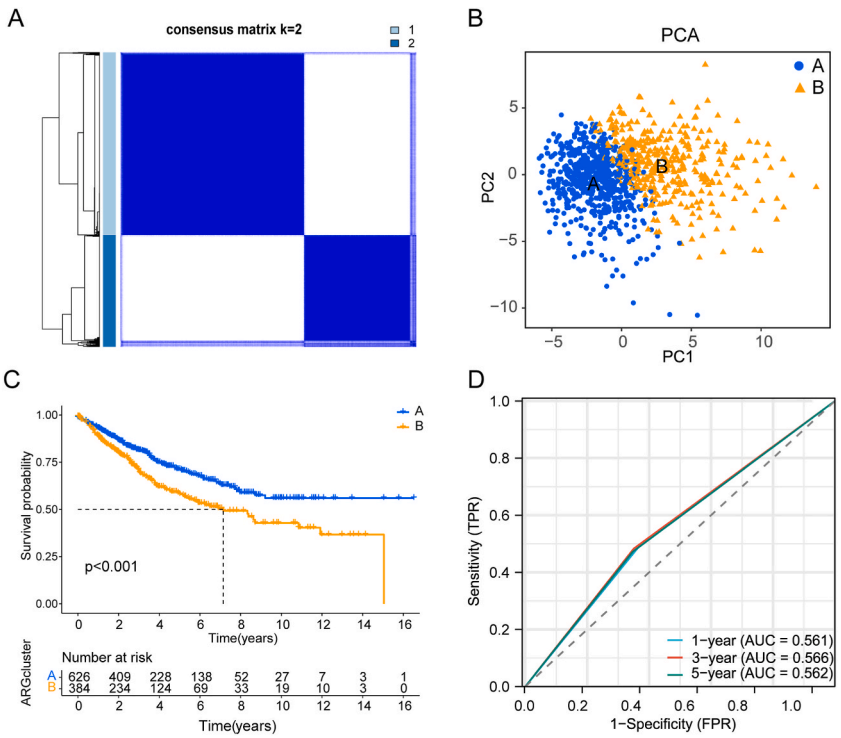

Fig. 3. Identification of intrinsic clusters of CRC by CRC-ARGs. (A). The patients with CRC were divided into two distinct gene clusters. (B). PCA for different clusters. (C). Kaplan-Meier survival curve analysis for the association of the different clusters and overall survival times. (D). Time-dependent ROC curve analysis at 1,3 and 5 years showing the AUC for overall survival.

3.2. Construction and validation of a prognostic signature related to anoikis

Due to the large number of genes in CRC-ARGs (over 40), their generalised application was challenging. The prediction accuracy of CRC prognosis based on clusters calculated by the expression of CRC-ARG showed poor (all AUCs were less than 0.6, as shown in Fig. 3D). Therefore, we constructed a more accurate prediction model to identify CRC-AKGs from CRC-ARGs.

Based on LASSO regression analysis, we selected *NAT1*, *MYC*, *MMP3*, *GZMB*, *INHBB*, *EPHA2*, *FGF2*, *CD36*, *HOTAIR*, and *OGT* as CRC-AKGs for prognostic model construction (Fig. 4A and B). We used model coefficients obtained from LASSO regression analysis to calculate individual risk scores for each CRC sample, and the formula for the Risk-Score was:

Risk-Score = *GZMB**(−0.1117) + *MMP3**(−0.0568) + *OGT**0.3527 + *HOTAIR**0.144 + *CD36**0.2149 + *FGF2**0.1994 + *EPHA2**0.2783 + *INHBB**0.1805 + *MYC**(−0.2384) + *NAT1**(−0.1975).

Based on Risk-Scores, we classified CRC samples into high-risk and low-risk groups. We found that the survival time of the patients was shorter as the scores increased and the levels of various CRC-AKGs differed among groups at high and low risk (Fig. 4C). Genes *NAT1*, *MYC*, and *MMP3* showed higher expression in the high-risk group than it in the low-risk group, while *INHBB*, *EPHA2*, *FGF2*, *CD36*, *HOTAIR*, and *OGT* showed the opposite trend. Furthermore, we employed Kaplan-Meier survival and ROC analysis to evaluate the prognostic values of Risk-Score. The results showed that patients classified as low risk performed significantly higher overall survival rates in CRC compared with it from the high-risk group ($p < 0.001$, Fig. 4E). The AUCs of 1-, 3-, and 5-year were 0.681, 0.709, 0.678, respectively (Fig. 4D).

We explored the correlation between the Risk-Score and various clinical factors (Fig. 5A). The distribution of CRC-AKGs, high-risk group, low-risk group, and TNM staging was investigated. The heat map showed that the high-risk group had more patients with high TNM stage compared with the low-risk group. Fig. 5B–D indicated a significant difference in Risk-Score across T, N, and M stage ($p < 0.001$). However, no significant statistical difference was found between age and gender strata ($p > 0.05$, Fig. S1). We also validated CRC-AKGs with public databases and conducted a survival analysis of single gene, revealing differential expression of OGT proteins in tumor and normal tissues (Fig. S2).

3.3. Construction and validation of predictive nomogram

We identified age, T, N, M, and Risk-Score as independent risk factors affecting the prognosis of CRC patients using univariate and multivariate cox regression analyses in the training set ($p < 0.001$, Fig. 6A and B). We then constructed a nomogram model to predict the 1, 3, 5-year time survival rate of CRC patients with age, T, N, M, and Risk-Score as predictors (Fig. 6C). The nomogram included the calculation of the composite score of these 5 variables, enabling the estimation of survival rate at 1, 3, 5-year time from CRC patients.

We evaluated the prediction accuracy using ROC. The results showed that the 1-, 3-, and 5-year AUCs for nomogram (0.804, 0.815,

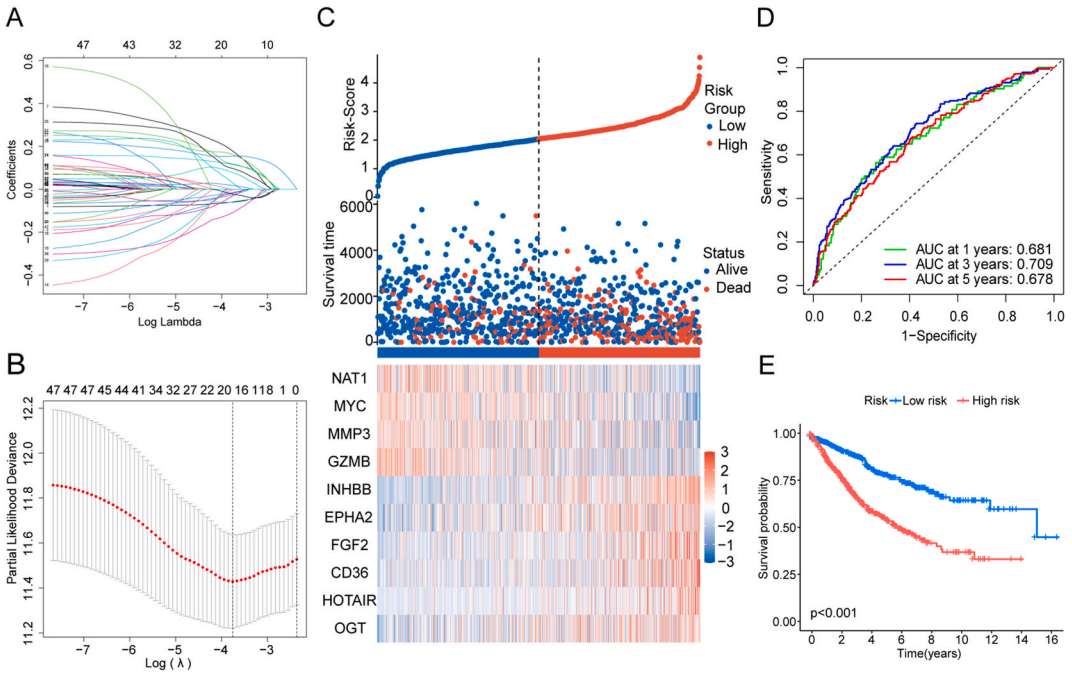

Fig. 4. CRC-AKGs were selected by the LASSO logistic regression model. (A and B). The LASSO method of CRC-ARGs with CRC prognosis. (C). The distribution of risk score, survival status, and expression levels of the 10 CRC-AKGs among CRC patients. (D). Time-dependent ROC curve analysis between low and high group at 1,3 and 5 years showing the AUC for overall survival. (E). Kaplan-Meier survival curve analysis for the association of the different risk groups and overall survival times.

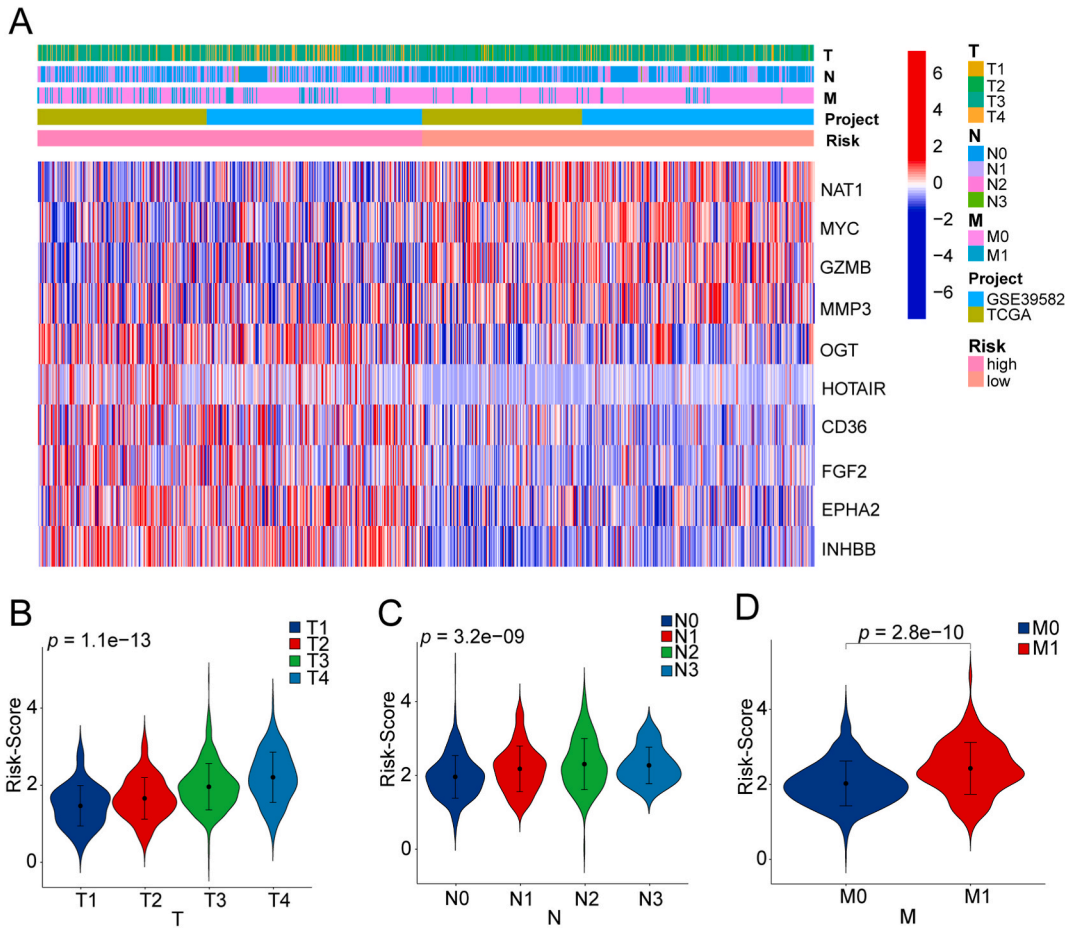

Fig. 5. Relationship between clinical factors and Risk-Score based on CRC-AKGs. (A). Heatmap of TNM stage, expression of CRC-AKGs and Risk-Score, showing their distribution and differences across groups (high-risk group and low-risk group). (B–D). Risk score differences among subgroups for TNM stage.

and 0.779, Fig. 6D) were significantly ($p < 0.01$) higher than the 1-, 3-, and 5-year AUCs for Risk-Score (0.681, 0.709, 0.678, Fig. 4D). The calibration curves showed a strong concurrence between the estimations conducted using the nomogram and the actual observations (Fig. 6E). The DCA demonstrated that the AUDC for the model was much higher than "ALL", which indicated that the nomogram was feasible for making valuable and informed judgements of the prognosis (Fig. 6F). Furthermore, we performed the evaluations in the validation set, and the results were consistent with the training set (Fig. 6G–I).

3.4. Human-computer interaction nomogram construction and promotion

Since the predictive model visualized in Fig. 6C is not user-friendly, we constructed a dynamic column line diagram (Fig. 7). This diagram is accessible through the Open Platform (https://www.shinyapps.io/) and provides a more convenient use. By inputting the patient's age, Risk-Score, T, N, and M in required fields and selecting the desired OS for prediction, the patient's customized survival curve can be displayed (Fig. 7B). The probability of a patient in Fig. 7A predicting 1, 3, 5-year time of OS by the model was shown in Fig. 7C.

3.5. Comprehensive analysis of the immune microenvironment of LASSO Risk-Score and CRC-AKGs

To identify differences in immune cell infiltration between high-risk and low-risk groups based on Risk-score, we quantified the level of immune cell infiltration in CRC patients. We found elevated levels of immune cells such as plasma cells, $CD8^+$ T cells and activated $CD4^+$ memory T cells in the low-risk group, while higher levels of neutrophils, macrophages M0 and macrophages M2 were found in the high-risk group (Fig. 8A). We investigated the correlation between CRC-AKGs and Risk-Score and the level of immune cell infiltration (Fig. 8B), it showed that Risk-Score was positively correlated with M2 macrophages and neutrophils ($p < 0.001$, $r = 0.13$; $p < 0.001$, $r = 0.12$, Fig. 8C), while the Risk-score was negatively correlated with activated $CD4^+$ T cells and resting NK cells ($p < 0.001$, $r = 0.25$; $p < 0.001$, $r = 0.13$, Fig. 8C).

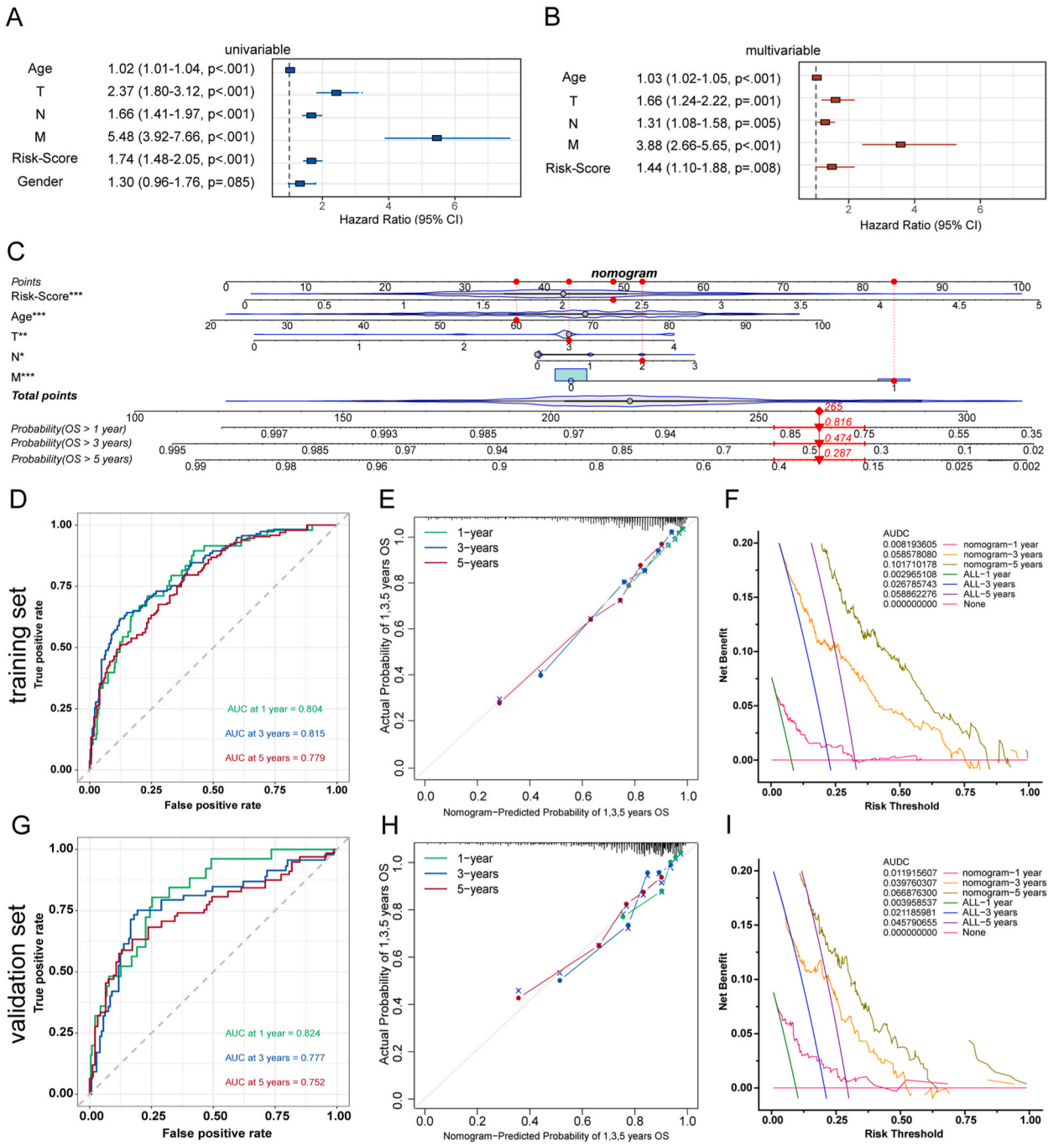

Fig. 6. Construction and validation of the prognostic prediction model. (A and B). Univariate and multivariate analyses based on Cox revealed the association between each predictor with OS. (C). Nomogram integrated the age, T stage, N stage, M stage and Risk-Score. (D and G). Time-dependent ROC of the training set and the validation set. (E and H) Calibration curves for the nomogram model in the training set and the validation set. (F and I). DCA curves of the training set and the validation set.

The data set from TISCH showed significant differences in the number of different cells among immune cells at CRC sites, with $CD4^+$ T cells and plasma cells showed the highest numbers (Figs. S3A and B). In addition, the expression of CRC-AKGs varied in different immune cells, which the gene *GZMB* showed a higher expression in $CD8^+$ T cells, DC, T prolif cells; the expression of CD36 was higher in DC, macro cells (Fig. S3).

To explore the potential impact of the Risk-Score in the complex TME, we investigated the correlation between risk score and TME indices. The results revealed that stromalscore and estimatescore were significantly elevated in high-risk group compared with it in the

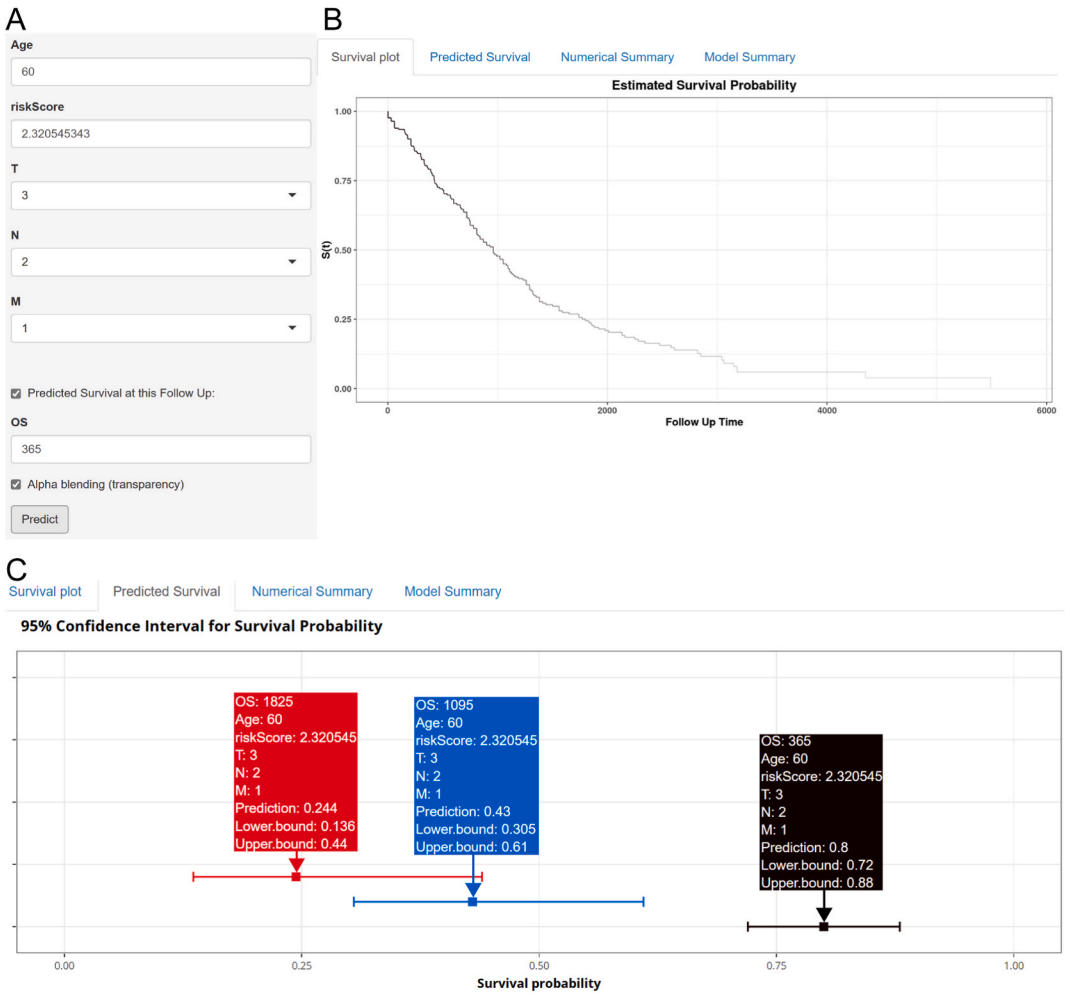

Fig. 7. An interactive nomogram clinical prediction model for predicting the prognosis of CRC patients. Interactive nomogram could be applied on networked devices. The corresponding URL was: https://yuexiaozhang.shinyapps.io/anoikisCRC/. Here, the slider on the left allows you to enter 5 specific characteristics. Enter the OS you wish to predict and click the "Predict" button. Here, this figure showed an example of CRC patients.

low-risk group ($p < 0.001$, Fig. 8D), indicating the presence of an augmented immune response in the high-risk group.

4. Discussion

CRC is a globally prevalent malignancy associated with high mortality rates, particularly in advanced stages [39]. After analyzing the RNAseq data from TCGA-CRC, we identified 10 genes as CRC-AKGs and calculated their weights and Risk-Score. Furthermore, our models are now available online and open-assessed, thus ensuring user-friendly interactions. This approach proves more convenient for medical workers and patients, in addition, effectively enhances the assessment of CRC prognosis in the real world.

We identified 47 genes that are linked to the prognosis of CRC (Figs. 2), 45 of them shown mutations in colon cancer, indicating the potential use of CRC-ARGs as identifying markers for CRC. Despite the observed gene expression differences in the gene cluster (Fig. 3 A and B), further enhancement is required in its prognostic evaluation value (Fig. 3C and D). To focus on more CRC-related genes, we identified 10 CRC-AKGs by LASSO. Notably, the prognostic evaluation value of the LASSO Risk-Score still requires further enhancement, as the AUC value was only approximately 0.7. The omission of key clinical factors may contribute to the lack of prognostic value of the LASSO Risk-Score. Both calibration and decision curves indicated that the model performed a good prediction effect. However, the apparent performance of a model, as indicated by its estimated performance using the data utilized for model development, often results in an overly optimistic estimate due to overfitting of the model to that specific dataset [40]. We thus performed the validation set to further evaluate the model. The AUC, calibration curves and DCA of validation set were similar to those of the training set, which support the robustness of our mode for clinical application.

Although there has been extensive research on predictive modelling, they are generally static and not user-friendly in clinical settings [41,42]. Nomograms make complex regression equations easier to read [43]. In the clinical work, by substituting the patient's

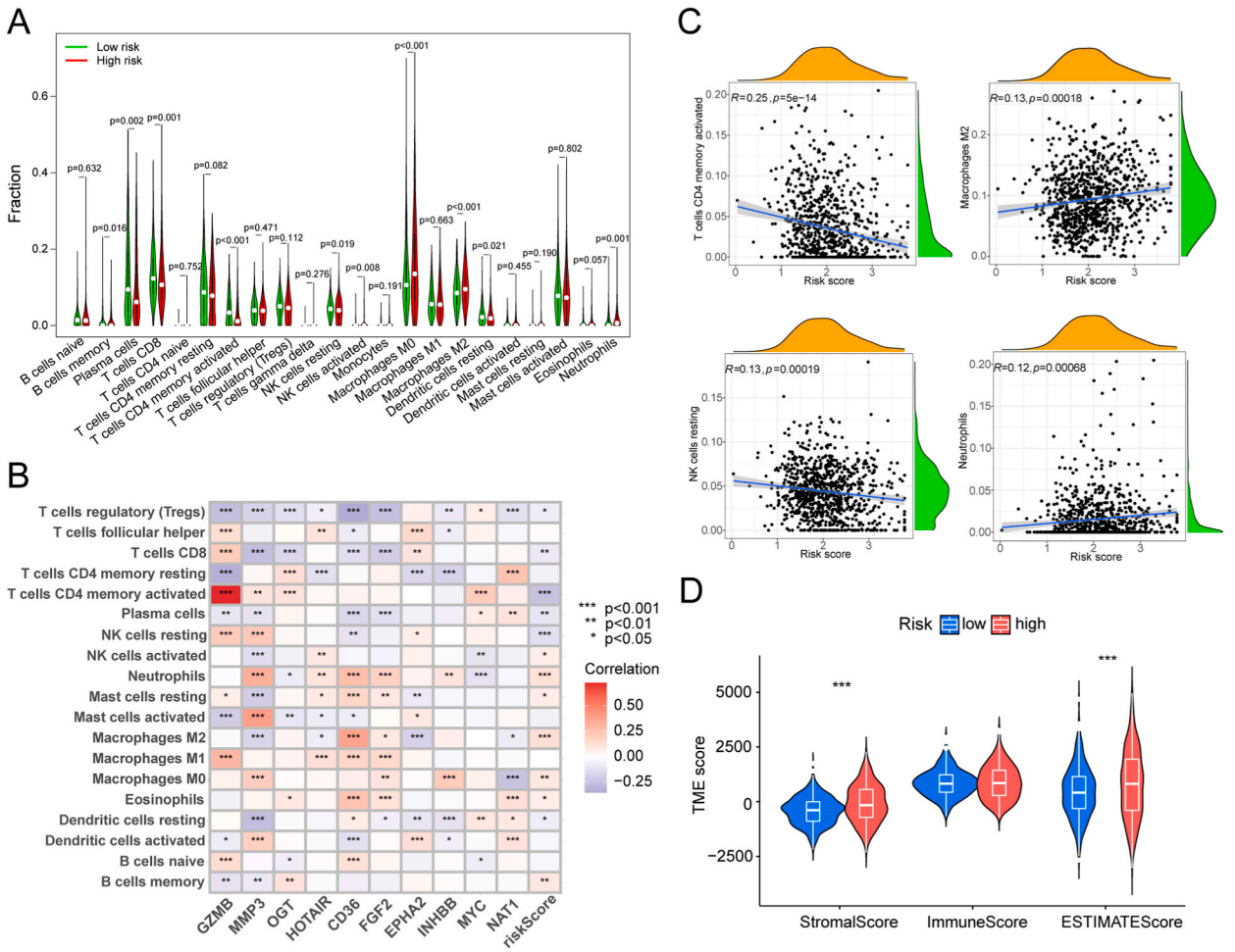

Fig. 8. Relationship between CRC-AKGs and CRC Tumor Microenvironment. (A). Violin plot of differences in immune cell infiltration in high-risk and low-risk group. (B). Heatmap visualizes the correlation of each CRC-AKG and Risk-Score with immune cell. (C). Scatterplots showing the correlation between Risk-Score and immune cell. (D). Violin plot of differences in immune-related indicators in samples from high-risk and low-risk groups.

age, TNM classification and risk score into the present model, we provide an accurate digital survival or risk probability for each CRC patient, which can assist the clinician in decision-making, reflecting the idea of individualized medicine. A prognostic prediction model for lung adenocarcinoma and a prognostic prediction model for low-grade endometrial stromal sarcoma both implemented through nomogram models [44,45]. To increase the interactivity of our models, we developed an innovative open platform online. This enables users from around the globe to conveniently access our model and obtain the 1, 3, and 5-year time survival rate along with the corresponding confidence level through a few straightforward selections and inputs. The interactive nature of our models enhances the user experience, thereby significantly enhancing the clinical utility of the model.

We found the screened CRC-AKGs show good predictive power. Understanding these gene interactions is vital for the development of diagnostic tools and targeted therapies in CRC research. Studies on genes such as *GZMB*, *MMP3*, *OGT*, *HOTAIR*, *CD36*, *FGF2*, *EPHA2*, *INHBB*, *MYC*, and *NAT1* in the context of colon cancer has provided valuable insights into the molecular mechanisms underlying the disease. For instance, *GZMB* was found in natural killer cells (NK cells) and cytotoxic T lymphocytes (CTLs) [46]. It cleaves and activates various apoptosis-associated proteins [47]. A study has shown that infiltration of $GZMB^{+}$cells can enhance the antitumor effects of TME [48]. *GZMB* was also identified as a protective factor in a model predicting immunotherapy response in gastric cancer [49]. Genes *MMP3*, *OGT*, *HOTAIR*, *CD36*, *FGF2*, *EPHA2*, *INHBB*, *MYC*, and *NAT1* contribute to various facets of colon cancer development and progression, such as signaling pathways, metastasis, and cell growth regulation. *MMP3* is associated with immune function and can be used as a therapeutic target for the treatment of UC (ulcerative colitis) patients, as well as a biomarker of immune cell infiltration in CRC [50]. *FGF2* is associated with cell proliferation and angiogenesis [51]. It shows upregulated in various human cancers and are correlated with poor prognosis [52]. Besides, *OGT*, *CD36*, *EPHA2*, and *INHBB* have strong associations with CRC prognosis. *OGT* is a potential tumor marker [53], while *EPHA2* is associated with the invasion, migration, and metastasis of cancer cells, making it a marker of poor prognosis in CRC [54]. In a study on CRC, *INHBB* was identified as a key gene for predicting the efficacy of

radiotherapy [55].

After calculating the median risk score, we divided the patients into high- and low-risk groups. We observed distinct patterns of immune cell infiltration between the two groups. In the low-risk group, we found elevated levels of Macrophages M2, T cells $CD4^+$ and $CD8^+$, plasma cells, and NK cells resting. In contrast, the high-risk group exhibited elevated levels of Macrophages M2 and M0, B cell memory, Neutrophils, Eosinophils, and NK cells activated. Macrophages M2 cells inhibit T cells and NK cells activity, suppressing immune function and promoting immune evasion and tumor progression, which fits with data from other studies [56]. In the low-risk group, $CD4^+$ and $CD8^+$ T cells were found to be elevated. $CD4^+$ T cells could activate $CD8^+$ T cells, which encourage them to proliferate and differentiate into effector T cells, and $CD8^+$ T cells can directly kill tumor cells [57,58]. Synergism between $CD4^+$ and $CD8^+$ T cells is essential for immune function and tumor resistance. Furthermore, our analysis indicated that stromal cells were significantly more abundant in the high-risk group compared with the low-risk group. Tumor stromal cells have been widely recognized for their significant role in tumor growth, metastasis, progression, and treatment. These cells contribute to tumor development and spread by promoting tumor growth and enhancing metastatic potential. Additionally, they suppress immune function through various mechanisms, thereby creating an immunosuppressive microenvironment that aids in tumor evasion from the immune system [59,60]. The observed difference between the high- and low-risk groups suggests that reduced infiltration of tumor-killing cells, coupled with a surge of immune cells promoting immune evasion and an increase in tumor stromal cells, can potentially contribute to the relatively lower survival rates in the high-risk group. These findings provide compelling evidence for the strong predictive power of our model. However, it is important to note that further experimental validation is necessary to elucidate the precise mechanisms underlying these observations. Such validation will not only enhance our understanding of the underlying processes but also pave the way for potential clinical translation of our model.

5. Conclusion

The present study constructed a robust prognosis prediction model in CRC based on CRC-AKGs, which also combined with clinical data to perform better in precision medicine. However, our study does have limitations. Validation with clinical samples and experimental data is necessary, and further exploration would be valuable for deciphering the mechanisms of the gene signature in CRC progression.

Funding information

National Key Research and development program of China (No. 2020YFC2006100); Zhengzhou Major Collaborative Innovation Project (No. 18XTZX12003); Key projects of discipline construction in Zhengzhou University (No. XKZDJC202001).

Ethics approval

Not applicable.

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author (medp7123@126.com) on reasonable request.

CRediT authorship contribution statement

Yuexiao Zhang: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Xia Xue:** Writing – review & editing. **Fazhan Li:** Software, Methodology, Investigation. **Bo Zhang:** Software, Methodology. **Pengyuan Zheng:** Conceptualization. **Yang Mi:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Thanks for Academy of Medical Science, Zhengzhou University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e33637.

References

[1] Y. Hao, et al., Polymeric nanoparticles with ROS-responsive prodrug and platinum nanozyme for enhanced chemophotodynamic therapy of colon cancer, Adv. Sci. 7 (2020) 2001853, https://doi.org/10.1002/advs.202001853.
[2] L.H. Biller, D. Schrag, *JAMA*. Diagnosis and treatment of metastatic colorectal cancer, A Review 325 (2021) 669–685, https://doi.org/10.1001/jama.2021.0106.
[3] B. Cheng, A. Rong, Q. Zhou, W. Li, LncRNA LINC00662 promotes colon cancer tumor growth and metastasis by competitively binding with miR-340-5p to regulate CLDN8/IL22 co-expression and activating ERK signaling pathway, J. Exp. Clin. Cancer Res. 39 (2020) 5, https://doi.org/10.1186/s13046-019-1510-7.
[4] B. Pardini, et al., Int J Cancer, DNA Repair and Cancer in Colon and Rectum: Novel Players in Genetic Susceptibility, vol.146, 2020, pp. 363–372, https://doi.org/10.1002/ijc.32516.
[5] R. Dienstmann, R. Salazar, J.J. Tabernero, Personalizing colon cancer adjuvant therapy: selecting optimal treatments for individual patients, Clin. Oncol. 33 (2015) 1787–1796, https://doi.org/10.1200/JCO.2014.60.0213.
[6] A.P. Gilmore, Cell Death Differ. Anoikis 12 (Suppl 2) (2005) 1473–1477, https://doi.org/10.1038/sj.cdd.4401723.
[7] L. Jin, et al., *Mol cell*. The PLAG1-GDH1 Axis promotes anoikis resistance and tumor metastasis through CamKK2-AMPK signaling in LKB1-deficient lung, Cancer 69 (2018) 87–99 e87, https://doi.org/10.1016/j.molcel.2017.11.025.
[8] B. Peltanova, M. Raudenska, M. Masarik, Effect of tumor microenvironment on pathogenesis of the head and neck squamous cell carcinoma: a systematic review, Mol. Cancer 18 (2019) 63, https://doi.org/10.1186/s12943-019-0983-5.
[9] B. El Hassouni, et al., The dichotomous role of the glycolytic metabolism pathway in cancer metastasis: interplay with the complex tumor microenvironment and novel therapeutic strategies, Semin. Cancer Biol. 60 (2020) 238–248, https://doi.org/10.1016/j.semcancer.2019.08.025.
[10] Y. Dai, et al., Cell commun signal, Anoikis resistance–protagonists of breast cancer cells survive and metastasize after ECM detachment 21 (2023) 190, https://doi.org/10.1186/s12964-023-01183-4.
[11] J. Song, et al., The 14-3-3sigma protein promotes HCC anoikis resistance by inhibiting EGFR degradation and thereby activating the EGFR-dependent ERK1/2 signaling pathway, Theranostics 11 (2021) 996–1015, https://doi.org/10.7150/thno.51646.
[12] G. Ye, et al., Theranostics, Nuclear MYH9-induced CTNNB1 transcription, targeted by staurosporin, promotes gastric cancer cell anoikis resistance and metastasis 10 (2020) 7545–7560, https://doi.org/10.7150/thno.46001.
[13] Y. Yu, et al., *J Exp Clin Cancer Res*. CircCEMIP Promotes Anoikis-Resistance by Enhancing Protective Autophagy in Prostate Cancer Cells, vol.41, 2022, p. 188, https://doi.org/10.1186/s13046-022-02381-7.
[14] M.C. Lenert, M.E. Matheny, C.G. Walsh, *J Am Med Inform Assoc*. Prognostic models will be victims of their own success, unless 26 (2019) 1645–1650, https://doi.org/10.1093/jamia/ocz145.
[15] A.Y. Abuhelwa, et al., Machine learning for prediction of survival outcomes with immune-checkpoint inhibitors in urothelial cancer, Cancers (2021) 13, https://doi.org/10.3390/cancers13092001.
[16] F. Shamout, T. Zhu, D.A. Clifton, IEEE Rev Biomed Eng. Machine Learning for Clinical Outcome Prediction 14 (2021) 116–126, https://doi.org/10.1109/RBME.2020.3007816.
[17] R.C. Deo, Circulation, Machine Learning in Medicine 132 (2015) 1920–1930, https://doi.org/10.1161/CIRCULATIONAHA.115.001593.
[18] S.Y. Park, Nomogram: an analogue tool to deliver digital knowledge, J. Thorac. Cardiovasc. Surg. 155 (2018) 1793, https://doi.org/10.1016/j.jtcvs.2017.12.107.
[19] V.P. Balachandran, M. Gonen, J.J. Smith, R.P. DeMatteo, Nomograms in oncology: more than meets the eye, Lancet Oncol. 16 (2015) e173–e180, https://doi.org/10.1016/S1470-2045(14)71116-7.
[20] Z. Xu, et al., Construction of a ferroptosis-related nine-lncRNA signature for predicting prognosis and immune response in hepatocellular carcinoma, Front. Immunol. 12 (2021) 719175, https://doi.org/10.3389/fimmu.2021.719175.
[21] Z. Sun, et al., Identification and validation of an anoikis-associated gene signature to predict clinical character, stemness, IDH mutation, and immune filtration in glioblastoma, Front. Immunol. 13 (2022) 939523, https://doi.org/10.3389/fimmu.2022.939523.
[22] S. Zhao, et al., Crosstalk of disulfidptosis-related subtypes, establishment of a prognostic signature and immune infiltration characteristics in bladder cancer based on a machine learning survival framework, Front. Endocrinol. 14 (2023) 1180404, https://doi.org/10.3389/fendo.2023.1180404.
[23] T.M. Therneau, P.M. Grambsch, T.M. Therneau, P.M. Grambsch, The Cox Model, Springer, 2000.
[24] J.T. Leek, W.E. Johnson, H.S. Parker, A.E. Jaffe, J.D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, Bioinformatics 28 (2012) 882–883, https://doi.org/10.1093/bioinformatics/bts034.
[25] M.D. Wilkerson, D.N. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking, Bioinformatics 26 (2010) 1572–1573, https://doi.org/10.1093/bioinformatics/btq170.
[26] E.W. Steyerberg, Y. Vergouwe, Towards better clinical prediction models: seven steps for development and an ABCD for validation, Eur. Heart J. 35 (2014) 1925–1931, https://doi.org/10.1093/eurheartj/ehu207.
[27] K. Yang, et al., Letter to the Editor: clinical utility of urine DNA for noninvasive detection and minimal residual disease monitoring in urothelial carcinoma, Mol. Cancer 22 (2023) 25, https://doi.org/10.1186/s12943-023-01729-7.
[28] S.W. Ng, et al., *Nature*. A 17-gene Stemness Score for Rapid Determination of Risk in Acute Leukaemia, vol.540, 2016, pp. 433–437, https://doi.org/10.1038/nature20598.
[29] B. Engelhard, et al., Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons, Nature 570 (2019) 509–513, https://doi.org/10.1038/s41586-019-1261-9.
[30] M. Riester, et al., Risk prediction for late-stage ovarian cancer by meta-analysis of 1525 patient samples, J. Natl. Cancer Inst. (2014) 106, https://doi.org/10.1093/jnci/dju048.
[31] M.N. DeSalvo, N. Tanaka, L. Douw, A.J. Cole, S. M. Radiology Stufflebeam, Contralateral Preoperative Resting-State Functional MRI Network Integration Is Associated with Surgical Outcome in Temporal Lobe Epilepsy 294 (2020) 622–627, https://doi.org/10.1148/radiol.2020191008.
[32] J.N. Mandrekar, Receiver operating characteristic curve in diagnostic test assessment, J. Thorac. Oncol. 5 (2010) 1315–1316, https://doi.org/10.1097/JTO.0b013e3181ec173d.
[33] C. Jin, et al., *Nat Commun*. Predicting Treatment Response from Longitudinal Images Using Multi-Task Deep Learning, vol.12, 2021, p. 1851, https://doi.org/10.1038/s41467-021-22188-y.
[34] A.C. Alba, et al., Discrimination and calibration of clinical prediction models: users' guides to the medical literature, JAMA 318 (2017) 1377–1384, https://doi.org/10.1001/jama.2017.12126.
[35] K.F. Kerr, M.D. Brown, K. Zhu, H.J. Janes, Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use, Clin. Oncol. 34 (2016) 2534–2540, https://doi.org/10.1200/JCO.2015.65.5654.
[36] B. Van Calster, et al., Reporting and interpreting decision curve analysis: a guide for investigators, Eur. Urol. 74 (2018) 796–804, https://doi.org/10.1016/j.eururo.2018.08.038.
[37] S. Chen, et al., Prognostic dynamic nomogram integrated with inflammation-based factors for non-small cell lung cancer patients with chronic hepatitis B viral infection, Int. J. Biol. Sci. 14 (2018) 1813–1821, https://doi.org/10.7150/ijbs.27260.
[38] A.M. Newman, et al., Robust enumeration of cell subsets from tissue expression profiles, Nat. Methods 12 (2015) 453–457, https://doi.org/10.1038/nmeth.3337.
[39] S.C. Shah, S.H. Itzkowitz, *Gastroenterology*. Colorectal cancer in inflammatory bowel disease, Mechanisms and Management 162 (2022) 715–730 e713, https://doi.org/10.1053/j.gastro.2021.10.035.
[40] N. Kreuzberger, et al., Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis, Cochrane Database Syst. Rev. 7 (2020) CD012022, https://doi.org/10.1002/14651858.CD012022.pub2.

[41] Z. Di, et al., Single-cell and WGCNA uncover a prognostic model and potential oncogenes in colorectal cancer, Biol. Proced. Online 24 (2022) 13, https://doi.org/10.1186/s12575-022-00175-x.
[42] J.H. Lee, et al., Prognostic nomogram of hypoxia-related genes predicting overall survival of colorectal cancer-Analysis of TCGA database, Sci. Rep. 9 (2019) 1803, https://doi.org/10.1038/s41598-018-38116-y.
[43] A. Iasonos, D. Schrag, G.V. Raj, K.S. Panageas, How to build and interpret a nomogram for cancer prognosis, J. Clin. Oncol. 26 (2008) 1364–1370, 0.1200/JCO.2007.12.9791.
[44] F. Xu, X. Huang, Y. Li, Y. Chen, L. Lin, Mol ther nucleic acids. m, A-related lncRNAs are potential biomarkers for predicting prognoses and immune responses in patients with LUAD 24 (6) (2021) 780–791, https://doi.org/10.1016/j.omtn.2021.04.003.
[45] J. Wu, et al., A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: a population-based analysis, Cancer Commun. 40 (2020) 301–312, https://doi.org/10.1002/cac2.12067.
[46] Y. Wang, et al., SMAD4 promotes TGF-beta-independent NK cell homeostasis and maturation and antitumor immunity, J. Clin. Invest. 128 (2018) 5123–5136, https://doi.org/10.1172/JCI121227.
[47] G. Sharma, C.M. Rive, R.A. Holt, Rapid selection and identification of functional CD8(+) T cell epitopes from large peptide-coding libraries, Nat. Commun. 10 (2019) 4553, https://doi.org/10.1038/s41467-019-12444-7.
[48] J.B. Wang, et al., Tumor immunophenotyping-derived signature identifies prognosis and neoadjuvant immunotherapeutic responsiveness in gastric cancer, Adv. Sci. 10 (2023) e2207417, https://doi.org/10.1002/advs.202207417.
[49] J.H. Cheong, et al., Predictive test for chemotherapy response in resectable gastric cancer: a multi-cohort, retrospective analysis, Lancet Oncol. 19 (2018) 629–638, https://doi.org/10.1016/S1470-2045(18)30108-6.
[50] Z. Pan, et al., Identification of gene signatures associated with ulcerative colitis and the association with immune infiltrates in colon cancer, Front. Immunol. 14 (2023) 1086898, https://doi.org/10.3389/fimmu.2023.1086898.
[51] H. Caiado, et al., Data on the evaluation of FGF2 gene expression in Colorectal Cancer, Data Brief 31 (2020) 105765, https://doi.org/10.1016/j.dib.2020.105765.
[52] S.M. Bang, et al., Regulation of mRNA export through API5 and nuclear FGF2 interaction, Nucleic Acids Res. 48 (2020) 6340–6352, https://doi.org/10.1093/nar/gkaa335.
[53] F. Zhou, et al., The role and potential mechanism of O-Glycosylation in gastrointestinal tumors, Pharmacol. Res. 184 (2022) 106420, https://doi.org/10.1016/j.phrs.2022.106420.
[54] M. De Robertis, et al., Dysregulation of EGFR pathway in EphA2 cell subpopulation significantly associates with poor prognosis in colorectal cancer, Clin. Cancer Res. 23 (2017) 159–170, https://doi.org/10.1158/1078-0432.CCR-16-0709.
[55] L. Qian, X. Lai, B. Gu, X. Sun, An immune-related gene signature for predicting neoadjuvant chemoradiotherapy efficacy in rectal carcinoma, Front. Immunol. 13 (2022) 784479, https://doi.org/10.3389/fimmu.2022.784479.
[56] L.M.E. Janssen, E.E. Ramsay, C.D. Logsdon, W.W. Overwijk, The immune system in cancer metastasis: friend or foe? J Immunother Cancer 5 (2017) 79, https://doi.org/10.1186/s40425-017-0283-9.
[57] M. Saxena, S.H. van der Burg, C.J.M. Melief, N. Bhardwaj, Therapeutic cancer vaccines, Nat. Rev. Cancer 21 (2021) 360–378, https://doi.org/10.1038/s41568-021-00346-0.
[58] Y. Liu, et al., Tumour heterogeneity and intercellular networks of nasopharyngeal carcinoma at single cell resolution, Nat. Commun. 12 (2021) 741, https://doi.org/10.1038/s41467-021-21043-4.
[59] Y. Mao, E.T. Keller, D.H. Garfield, K. Shen, J. Wang, Cancer metastasis rev, Stromal cells in tumor microenvironment and breast cancer 32 (2013) 303–315, https://doi.org/10.1007/s10555-012-9415-3.
[60] F. Li, M.C. Simon, Dev cell, Cancer Cells Don't Live Alone: Metabolic Communication within Tumor Microenvironments 54 (2020) 183–195, https://doi.org/10.1016/j.devcel.2020.06.018.