BRITISH JOURNAL OF HOSPITAL MEDICINE

ARTICLE

Construction and Value Analysis of a Prognostic Assessment Model Based on Radiomics and Genetic Data for Colorectal Cancer

Yongna Cheng[1], Ziming Feng[2], Xiangming Wang[1,*]

[1]Department of Radiology, Yiwu Central Hospital, Yiwu, Zhejiang, China
[2]Department of Cardiovascular Medicine, Yiwu Central Hospital, Yiwu, Zhejiang, China
*Correspondence: 18846451824@163.com (Xiangming Wang)

Abstract

Aims/Background Colorectal cancer (CRC) is one of the major global health problems, with high morbidity and mortality, underscoring the need for new diagnostic and prognostic tools. Therefore, this study aims to evaluate the significance of integrating radiomics with genetic data in CRC prognostic assessment and improve the accuracy of prognosis prediction.

Methods This study included computed tomography (CT) images from 225 CRC patients and RNA-seq information from 654 patients, obtained from the TICA database. Key radiomics features and genes were identified through radiomics feature extraction, least absolute shrinkage and selection operator (LASSO) regression analysis, and Kaplan-Meier survival analysis. Furthermore, a CRC prognostic model was constructed using these key genes and radiomics features.

Results This study identified 170 key radiomics features. Out of them, five were significantly associated with CRC prognosis. Transcriptome data analysis identified 8 key genes, among which the high expressions of Inhibin Subunit Beta B (*INHBB*), Potassium Voltage-Gated Channel Subfamily Q Member 2 (*KCNQ2*), and Ubiquilin Like (*UBQLNL*) were significantly correlated with poor prognosis. Age, tumor stage, pathological T stage, and pathological N stage were determined as independent prognostic factors. Moreover, immune infiltration analysis demonstrated that the immune score of the low-risk group was higher than that of the high-risk group, with significant differences in some immune cells, and key genes were correlated with immune cells. Additionally, the constructed CRC prognostic model incorporating three genes, *INHBB*, *KCNQ2*, and *UBQLNL*, exhibited high prediction accuracy in the validation set, with area under the curve (AUC) values of 0.80, 0.87, and 0.84 at 1-year, 3-year, and 5-year, respectively, indicating good prediction performance and reliability of the model.

Conclusion The multimodal data combining radiomics features and gene expression data can improve the accuracy of CRC prognostic assessment, providing a valuable prognostic prediction tool for clinical practice and helping to guide the selection and optimization of treatment regimens.

Key words: colorectal cancer; radiomics; genomics; prognostic; bioinformatics

Submitted: 6 September 2024 Revised: 12 December 2024 Accepted: 18 December 2024

How to cite this article:
Cheng Y, Feng Z, Wang X. Construction and Value Analysis of a Prognostic Assessment Model Based on Radiomics and Genetic Data for Colorectal Cancer. Br J Hosp Med. 2025. https://doi.org/10.12968/hmed.2024.0620

Introduction

Colorectal cancer (CRC), a widespread gastrointestinal tumor worldwide, ranks among the primary causes of cancer-induced fatalities. Its prevalence presents substantial challenges to public health, casting a long shadow over the well-being of populations worldwide (Chen et al, 2020). In 2020, CRC ranked third globally in

incidence, accounting for approximately 11% of all cancer cases, and second in mortality, accounting for about 9%, with the emerging trend of CRC onset in younger individuals under 50 (Klimeck et al, 2023; Sung et al, 2021). According to projections, by 2040, China is likely to witness the peak incidence and mortality rates for cancer (Zhang et al, 2023). Despite advancements in therapeutic approaches such as radiotherapy, chemotherapy, and surgical interventions, the current screening and treatment options for CRC remain inadequate to address its high morbidity and mortality rates, particularly in developing countries with large populations (Cao et al, 2024). Therefore, it highlights the need for comprehensive research focused on improving the diagnosis, treatment, and prognosis of CRC to achieve a significant reduction in its morbidity and mortality.

In recent years, advancements in surgical techniques, the introduction of new adjuvant therapies, and the rapid development in diagnostic imaging have significantly improved the treatment of CRC (Li et al, 2024). Among these, radiomics has emerged as a crucial approach to treating rectal cancer. By extracting a wide array of quantitative features from medical imaging data, which is rich in numerical information but invisible to the naked eye, radiomics utilizes machine learning and artificial intelligence algorithms for comprehensive analysis. This analytical technique captures the biological characteristics and heterogeneity of tumors, providing a low-risk and efficient method for cancer diagnosis, classification, and prognosis (Aerts et al, 2014; Fan et al, 2022; Limkin et al, 2017).

Since its inception in 2012, radiomics has been widely applied in oncology, demonstrating its potential as a non-invasive imaging biomarker and promoting the shift of imaging technology from qualitative to quantitative (Lambin et al, 2017). It is crucial in cancer diagnosis, staging, prognosis assessment, and patient monitoring (Caruso et al, 2021). Particularly in the diagnosis and management of CRC, radiomics holds promise in assisting clinicians in identifying high-risk diseases. An investigation has indicated that radiomics features are associated with the pathological response of rectal cancer and can predict pathological complete response (pCR) (Wen et al, 2023). Furthermore, radiomics is increasingly applied to the prognosis prediction of rectal cancer. By combining clinical information with radiomics features, predictive models such as nomograms can be constructed, providing more personalized treatment plans for patients (Inchingolo et al, 2023; Xue et al, 2022).

Genetic data is increasingly crucial in the prognosis assessment of CRC. By using whole-genome and transcriptome sequencing analyses, researchers can identify driver genes associated with different stages of CRC and uncover key prognostic factors (Xu et al, 2023; Zhang et al, 2021). For instance, the largest global multi-omics study on CRC has revealed numerous disease-related genes, including 24 newly discovered potential driver genes associated with the wingless/integrated (WNT), epidermal growth factor receptor (EGFR), and transforming growth factor-beta (TGF-β) pathways, which are significantly correlated with the survival of CRC (Nunes et al, 2024). The integration of genetic data is transforming the prognosis assessment of CRC, providing more precise molecular typing and prognostic markers. This technological improvement helps develop personalized and effective treatment strategies (Mortezapour et al, 2023; Wei et al, 2023). Despite these

advancements, there are limited studies combining radiomics with genomic data for the prognosis assessment of CRC. Our study retrieved radiomics and transcriptome sequencing data from CRC liver metastasis, as well as from the Cancer Genome Atlas Colon Adenocarcinoma (TCGA-COAD) and Rectal Adenocarcinoma (TCGA-READ) databases. Subsequently, by integrating radiomics features with key genetic markers, we constructed a predictive model for CRC prognosis, providing a basis for more precise and personalized treatment strategies.

Methods

Data Collection and Preprocessing

The CRC liver metastasis, TCGA-COAD, and TCGA-READ datasets were obtained from the TICA database (https://public.cancerimagingarchive.net/). These datasets include computed tomography (CT) tumor images from 225 CRC patients and RNA-seq information from 654 patients. The original CT images were segmented using 3D Slicer software (version 5.1.0, National Institutes of Health (NIH); available at https://www.slicer.org/, Cities of Boston and Cambridge, MA, USA). Manual segmentation was performed to outline the region of interest (ROI) and a 3D image with the segmented ROI was developed for subsequent analysis.

Data Feature Extraction

Feature extraction was performed using the "PyRadiomics" plugin within 3D Slicer v5.1.0 software. This tool effortlessly extracted 851 radiomic features from each ROI. These features included 162 initial statistical parameters (e.g., power, disorder, average, standard deviation, and maximum value), 107 shape and size-based features (e.g., maximum 3D diameter, volume, and surface area), 675 texture attributes, and 744 wavelet transform-based features. The extracted features were standardized via Z-scores. Feature selection was performed using least absolute shrinkage and selection operator (LASSO) regression analysis, implemented with the 'glmnet' package in R (version 4.0.1; R Core Team; https://www.r-project.org). Finally, radiomics features significantly associated with CRC were selected.

Differential gene expression analysis of RNA sequencing data from TCGA-COAD and TCGA-READ datasets was performed using the 'DESeq2' package in R software. Principal component analysis (PCA) was subsequently conducted on the datasets using R software. Subsequently, the results were visualized using the "ggplot" and "pheatmap" packages, and then volcano plots and heatmaps were generated to obtain key differentially expressed genes (DEGs).

Risk Grouping and Prognostic Analysis of Radiomics Features

Survival analysis was performed on selected key features to calculate predictive risk scores using the "survival" package in the R program. Determine the optimal risk score threshold based on the risk score and the patient's survival status using the default parameters of the "survminer" package in R. Then, individuals were divided into high-risk and low-risk groups according to the cut-off value. The significance of the survival curves between the two groups was evaluated using

Log-Rank test, with the optimal cut-off point defined as the value indicating the largest difference in survival rate. The core formula for this calculation is:

$$Z(c) = \frac{O_1 - E_1}{\sqrt{V_1}}$$

Where: O_1 indicates the number of deaths observed in the high-risk group. E_1 is the expected deaths in the low-risk group, and V_1 is the variance.

Employing this formula, the risk score was derived for each sample. The 225 imaging samples were categorized into high-risk and low-risk groups based on the optimal threshold. Kaplan-Meier curves were generated to compare survival probabilities between these groups. Statistical significance was achieved at $p < 0.05$.

Risk Grouping and Survival Analysis of RNA Sequencing Information

From the TCGA-COAD and TCGA-READ datasets, 654 eligible samples were randomly allocated to training (n = 455) and validation (n = 199) sets (7:3 ratio). Risk scores were then determined for the validation set based on the expression of key DEGs. LASSO regression analysis was utilized to investigate the DEGs, with key genes selected through 10-fold cross-validation. The LASSO model calculated linear predictive values (i.e., risk scores) using the features with non-zero regression coefficients and their corresponding values. The prediction function from the "survminer" package in R was used to compute the risk scores for each patient by taking the dot product of each patient's feature matrix and the vector of LASSO-selected coefficients. The equation for the risk score is given below:

$$\text{Risk Score}_{\ i} = \sum_{j=1}^{p} \beta_j X_{ij}$$

where β_j is the coefficient of regression for the feature selected by LASSO, and X_{ij} is the value of the jth feature for the ith patient. A higher risk score indicates a greater risk of death for the patient. Patients were stratified into high- and low-risk groups based on their risk scores. Kaplan-Meier curves were plotted to compare the survival rates between the two groups. Furthermore, multivariate COX regression analysis was performed on the selected core genes using the "survival" and "survminer" packages in R. The findings were visualized with a forest plot generated using the "forestplot" package.

Analysis of Independent Prognostic Factors

Differences in risk scores between the two groups were assessed by comparing the high-risk and low-risk groups using *t*-tests in terms of age, sex, tumor grade, pathological T-stage, and pathological N-stage. Univariate COX regression analyses then evaluated the independent effect of each factor on survival. Following this, all factors were incorporated into multivariate COX regression analyses, controlling for the effects of other variables and identifying independent prognostic factors.

Assessment of Infiltrating Immune Cells

The relative composition of infiltrating immune cells and their association with gene expression levels were assessed using the Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) method. Variations in tumor-infiltrating immune cell populations between high- and low-risk groups were subsequently examined using the Wilcoxon test. The correlation between immune cell types and selected genes was visually displayed using a heatmap created with the R package "ggplot2".

Gene-Radiomics Combined Analysis and Evaluation

The TCGA-COAD and TCGA-READ datasets were randomly partitioned into training and validation sets at a 7:3 ratio. Within the training set, multivariable COX regression analysis was conducted on selected genes and imaging features to identify key indicators associated with CRC prognosis. Forward-backward stepwise regression was used for feature selection, and findings were visualized with a forest plot.

Time-dependent receiver operating characteristic (ROC) curves and calibration plots were employed to evaluate model performance in the validation set. A nomogram predicting CRC prognosis based on key gene expression levels and risk scores, was then generated using the "rms" package in R. The final calibration plot visually displayed the calibration curve and its performance.

Model performance, as determined by the LASSO model, was assessed through ROC curves, which plotted the true positive rate (TPR) as the vertical coordinate and the false positive rate (FPR) as the horizontal coordinate. A larger area under the curve (AUC) indicates better model performance. Calibration curves were then utilized to evaluate the model's predictive accuracy, and a p value less than 0.05 was considered statistically significant. The predicted probabilities were plotted on the horizontal axis, and the actual observed proportion was plotted on the vertical axis, at various probability levels. These curves offered valuable insights into the accuracy of the model in forecasting outcomes.

Statistical Analysis

Statistical analyses were conducted using the R program. Manual segmentation of the ROIs was performed using 3D Slicer v5.1.0 software. Imaging feature extraction from each ROI was conducted using the "PyRadiomics" plugin in 3D Slicer v5.1.0 software.

Continuous variables were compared using Student's t-tests after assessing normality with the Kolmogorov-Smirnov test. However, categorical variables underwent comparison were compared using the chi-square tests. Overall survival (OS), a critical measure for evaluating treatment efficacy outcomes and prognosis in oncology, was defined as the time from randomization or treatment initiation to death from any cause. OS serves as the gold standard for assessing patient survival benefits. Survival curves were developed using the Kaplan-Meier method, and group differences were evaluated using the log-rank test. To identify independent prog-

nostic factors, univariate and multivariate COX proportional hazards regression analyses were performed with a significance level set at $p < 0.05$.

Results

Analysis of Imaging Histological Characteristics and Prognostic Evaluation of CRC Patients

Initially, 170 key features were identified through the quasi-quantification of radiomic data from 225 patients extracted from the TICA database before performing LASSO regression analysis (Fig. 1A). Subsequently, to validate the reliability of the LASSO regression results, 10-fold cross-validation was conducted, resulting in the identification of five crucial features (Fig. 1B). These five key elements were then evaluated for their correlation with survival. Using the risk calculation formula based on imaging histological features, the optimal threshold for risk classification was found to be 0.02774729. However, during our analysis, we discovered that out of 225 radiomic samples, only 220 patient samples had corresponding survival data. Consequently, we were only able to use these 220 patient samples for further analysis. Based on the optimal threshold, 44.55% (n = 98) of these 220 patients were categorized as high-risk, while 55.46% (n = 122) were classified as low-risk. It should be noted that this high-risk and low-risk classification is determined by comprehensively considering the five crucial features according to the optimal threshold, reflecting the overall situation of these features in risk stratification. Kaplan-Meier survival plots were constructed based on the calculated risk scores. The results demonstrated that patients in the high-risk group had a significantly lower survival rate compared to those in the low-risk group during the same period ($p < 0.05$, Fig. 1C–G).

This indicates that all five significant characteristics are valuable for prognostic evaluation in CRC patients and can assist in differentiating between patient groups with disparate prognoses.

Differential Analysis and Prognosis Assessment of RNA Sequencing Information in CRC

The results of the transcriptome data demonstrated 7155 down-regulated genes and 1370 up-regulated genes (Fig. 2A). Further LASSO regression analysis identified 319 DEGs. Meanwhile, to verify the robustness of the LASSO results, a 10-fold cross-validation was implemented, leading to the selection of eight key genes (Inhibin Subunit Beta B (*INHBB*), Procollagen C-Endopeptidase Enhancer 2 (*PCOLCE2*), Parathyroid Hormone 1 Receptor (*PTH1R*), Homeobox D4 (*HOXD4*), Potassium Voltage-Gated Channel Subfamily Q Member 2 (*KCNQ2*), Ubiquilin Like (*UBQLNL*), Malignant T Cell Invasion and Dissemination 2 (*MID2*), Homeobox C4 (*HOXC4*)) that were closely associated with the CRC prognosis (Fig. 2B,C).

Furthermore, to evaluate the clinical significance of these 8 genes, we evaluated their expression levels in the validation set. Using the risk score formula, a threshold value of 0.29 was calculated, with patients scoring higher categorized as high-risk and those below classified as low-risk (Fig. 3B). The results of the Kaplan-

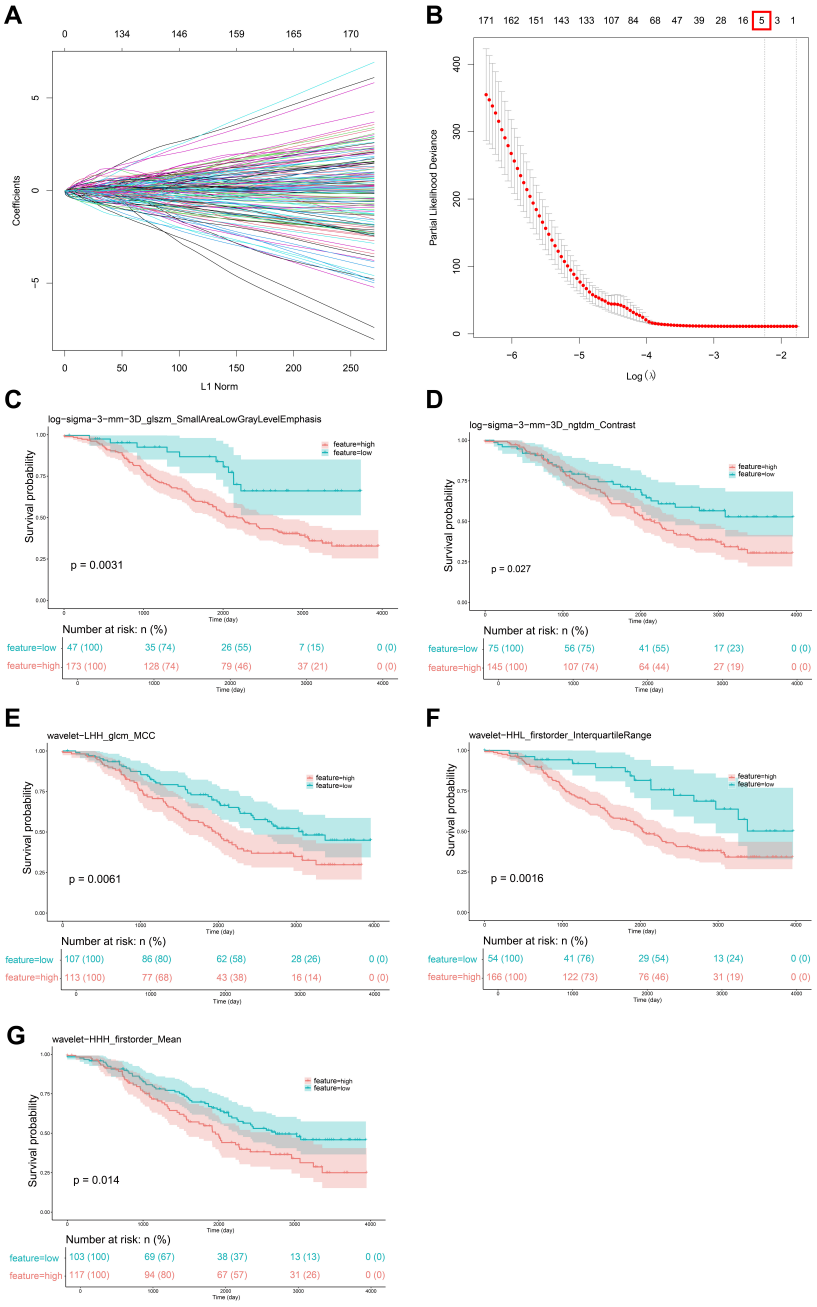


Fig. 1. Feature selection and prognosis analysis of 225 imaging histology data. (A) This figure depicts the distribution of gene coefficients derived from the LASSO regression model. The horizontal axis illustrates the penalty parameter (λ) applied within the LASSO model, whereas the vertical axis denotes the magnitude of the corresponding feature coefficients. Coefficients approaching zero signify a reduced influence of the respective features on predictive outcomes. (B) Cross-validation of the LASSO regression model by a factor of 10. Note: The cross axis indicates the range of values of the penalty parameter λ and the longitudinal axis indicates the prediction error of the model. The two dashed lines on the graph generally select the parameter corresponding to the dashed line on the left as the coefficients of the important features. The red-framed "5" represents five key features. (C–G) Kaplan-Meier survival of the risk scores of the 5 important features assessment. "LHH", "HHL", and "HHH" refer to Wavelet Transform Subbands. LASSO, least absolute shrinkage and selection operator.

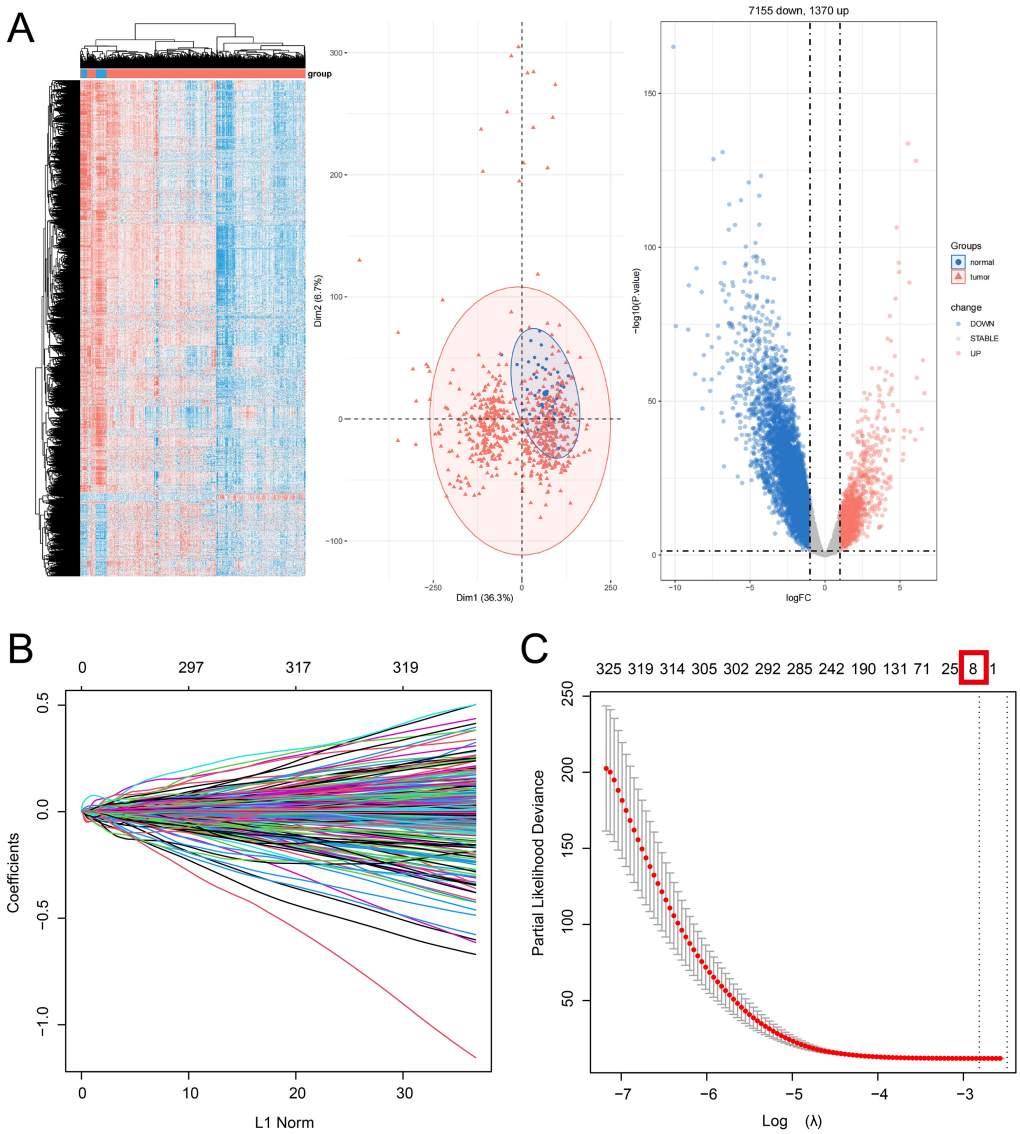


Fig. 2. Transcriptomic data of 654 CRC patients were analyzed for DEGs. (A) Heatmap and Volcano Plot. Heatmap shows gene expression differences between samples and Volcano Plot shows DEGs. (B) Distribution of gene coefficients of LASSO regression model. The penalty parameter λ, inherent to the LASSO model, is displayed on the horizontal axis, while the magnitude of coefficients for corresponding features is presented on the vertical axis. Smaller coefficients indicate a diminished impact of the features on the predictive outcome. (C) Cross-validation of the LASSO regression model by a factor of 10. The cross axis indicates the range of values of the penalty parameter λ and the longitudinal axis indicates the prediction error of the model. The two dashed lines on the graph generally select the parameter corresponding to the dashed line on the left as the coefficient of the important feature. The red-framed "8" represents 8 key genes. CRC, colorectal cancer; DEGs, differentially expressed genes.

Meier curves (Fig. 3A) showed that low expression of these 8 key genes was linked to improved survival of CRC patients ($p < 0.05$).

Moreover, COX regression and forest plot (Fig. 3B,C) revealed that only the high expression of three genes, *INHBB* ($p < 0.001$), *KCNQ2* ($p = 0.002$), and *UBQLNL* ($p = 0.004$), were significantly linked to a higher risk of CRC-related

death. Conversely, the *p*-values of *PCOLCE2*, *PTH1R*, *HOXD4*, *MID2*, and *HOXC4* were greater than 0.05, suggesting that these genes did not show significant differences in their influence on CRC prognosis.

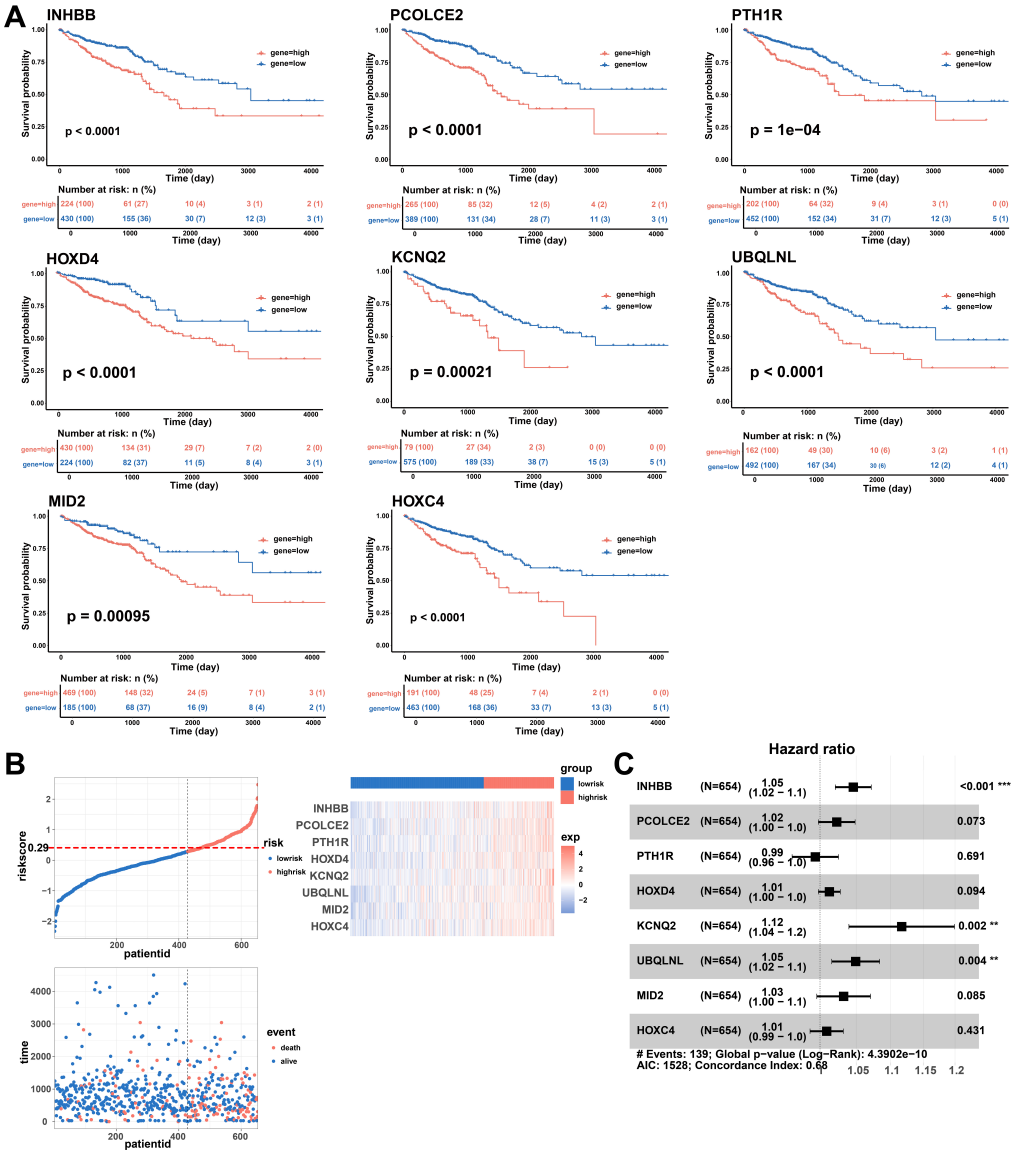


Fig. 3. Progression analysis of 8 differential expression genes. (A) Kaplan-Meier curves of OS for the 8 differentially expressed genes, all $p < 0.01$. (B) Expression heatmap of the 8 differentially expressed genes in two groups of CRC patients. Upper panel: Horizontal coordinates are patient Identification (ID); and the vertical coordinates are predicted risk scores from COX regression analysis. Middle panel: The horizontal coordinate is the patient ID, the vertical coordinate is the prognostic survival time, and a dotted line is used to differentiate between the two groups. Lower panel: Heatmap demonstrating the difference in expression of the 8 differentially expressed genes in patients in the low- and high-risk groups. (C) Forest plot of univariate COX regression analysis between the expression of the 8 differential genes and OS. ** $p < 0.01$, *** $p < 0.001$. OS, overall survival; *INHBB*, Inhibin Subunit Beta B; *PCOLCE2*, Procollagen C-Endopeptidase Enhancer 2; *PTH1R*, Parathyroid Hormone 1 Receptor; *HOXD4*, Homeobox D4; *KCNQ2*, Potassium Voltage-Gated Channel Subfamily Q Member 2; *UBQLNL*, Ubiquilin Like; *MID2*, Malignant T Cell Invasion and Dissemination 2; *HOXC4*, Homeobox C4; AIC, Akaike Information Criterion.

Identification of Independent Prognostic Factors in CRC Patients

To investigate the relationship between clinicopathologic characteristics and overall survival and identify independent prognostic factors in CRC patients, we compared risk scores across different variables, including age, gender, tumor grade, pathological T-stage, and pathological N-stage in the high-risk and low-risk groups using *t*-tests. Significant disparities in risk scores were observed for tumor grade ($p < 0.05$), pathological T-stage ($p < 0.0001$), and pathological N-stage ($p < 0.01$) (Fig. 4A).

To further assess the independent prognostic value of these factors for CRC, univariate and multivariate COX regression analyses were conducted. The findings showed that age (Hazard Ratio (HR) = 1.764, 95% Confidence Interval (CI): 1.139–2.734, $p = 0.011$), tumor stage (HR = 3.085, 95% CI: 2.131–4.466, $p < 0.001$), case T-stage (HR = 2.293, 95% CI: 1.232–4.267, $p = 0.009$), and pathological N-stage (HR = 2.624, 95% CI: 1.833–3.755, $p < 0.001$) were significantly associated with overall survival. Particularly, the pathologic N stage had an HR of 2.624, indicating that patients with a higher N stage had a significantly higher risk of death (Fig. 4B). Importantly, when these variables were subjected to the multivariate COX regression analysis, age (Log10HR = 0.360, 95% CI: 0.165–0.556, $p < 0.001$), tumor stage (Log10HR = 0.990, 95% CI: 0.609–1.372, $p < 0.001$), and pathological N stage (Log10HR = –0.502, 95% CI: –0.868–0.135, $p = 0.007$) were confirmed as independent prognostic factors (Fig. 4C). These factors remained significantly associated with overall survival after adjusting for other variables, indicating their significant impact on the prognosis of CRC patients.

Immune Infiltration Analysis

We utilized the Wilcoxon test to assess immune function. The results showed that there was no significant difference in stromal scores between the two groups ($p = 0.97$). However, immune scores were significantly elevated in the low-risk group compared to the high-risk group ($p = 0.041$), suggesting increased immune cell activity within the low-risk cohort (Fig. 5A). A deeper examination of the distribution of 28 distinct immune cell types uncovered significant differences in effector memory Cluster of Differentiation (CD)8 T cells ($p < 0.01$), activated CD8 T cells ($p < 0.05$), immature B cells ($p < 0.05$), type 1 T helper cells ($p < 0.05$), type 2 T helper cells ($p < 0.05$), and activated CD4 T cells ($p < 0.05$) (Fig. 5B). Furthermore, a correlation analysis between eight key prognostic genes and the 28 types of immune cells demonstrated varying levels of associations. Notably, the *INHBB* gene exhibited significant correlations with 21 immune cell types, while the *PCOLCE2* gene was significantly associated with 26 immune cell types (Fig. 5C). Collectively, these results imply that certain prognostic genes may influence tumor progression and patient outcomes by regulating the functionality and distribution of immune cells.

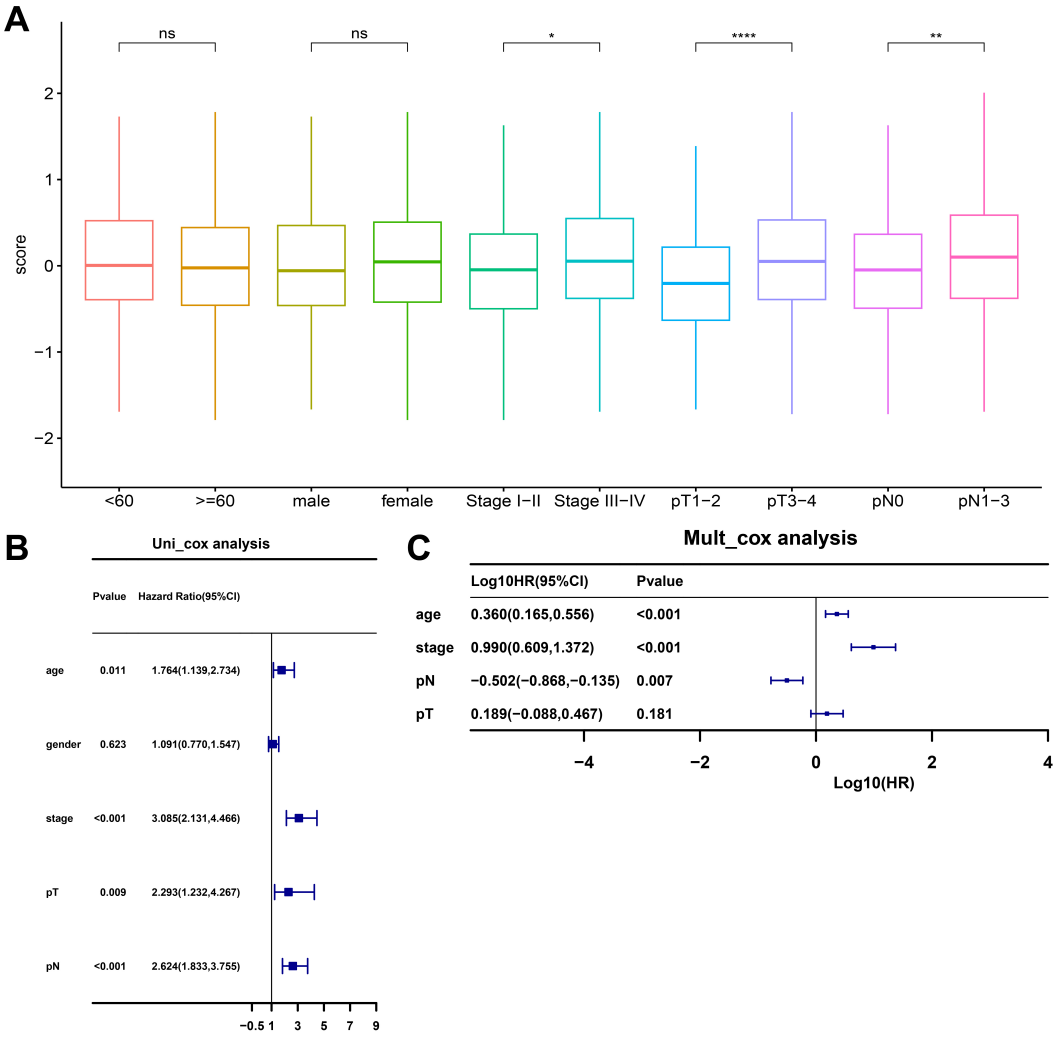


Fig. 4. Identification of independent prognostic factors in CRC patients. (A) *t*-test comparing the difference in risk scores between age, gender, tumor grade, pathological T-stage, and pathological N-stage in the high-risk and low-risk groups. ns, not significant; * $p < 0.05$, ** $p < 0.01$, and **** $p < 0.0001$. (B) Forest plot of one-way correlation between CRC patient characteristics and OS. (C) Forest plot of multivariate COX regression study of the correlation between clinicopathologic characteristics and OS in CRC patients. pN, pathological N-stage; pT, pathological T-stage.

Using Gene-Radiomics Combined Analysis to Establish CRC Prognostic Model

In our training dataset, we performed a COX regression analysis on curated radiomic features and pivotal genes, as delineated in Fig. 6A, which led to a CRC prognostic model containing the *INHBB* ($p = 0.027$), *KCNQ2* ($p = 0.009$) and *UBQLNL* ($p = 0.008$) genes. This model exhibited a significant risk score ($p = 0.015$) and an exemplary concordance index of 0.89. Subsequent analyses confirmed that the expression of *INHBB*, *KCNQ2*, and *UBQLNL* was substantially reduced in tumor tissues compared to normal tissues ($p < 0.0001$), as depicted in Fig. 6B. The nomogram, although primarily intended as a diagnostic tool and not a direct indication of the validity of the model, highlights the strong correlation between the risk scores and radiographic characteristics of these genes and

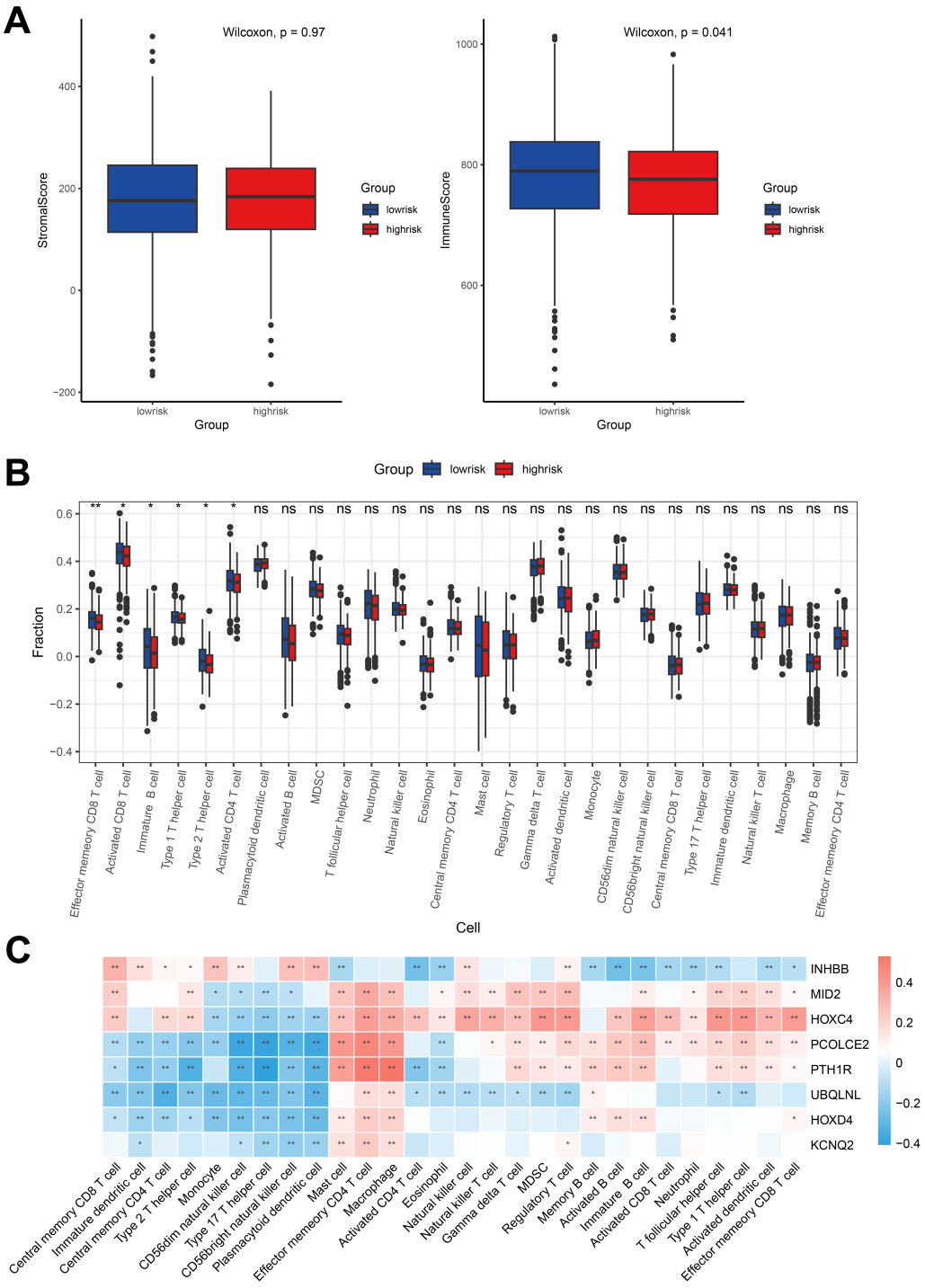


Fig. 5. Analysis of immune cell infiltration within tumor tissues. (A) Variation in stromal and immune scores between low- and high-risk groups were assessed via Wilcoxon tests. (B) Discrepancy of 28 immune cell types amid the low-danger and high-danger groups. (C) Correlation examination of the 28 immune cell types with 8 prognostic genes. ns indicates no significant difference, * $p < 0.05$, ** $p < 0.01$. CD, Cluster of Differentiation; MDSC, Myeloid-derived suppressor cells.

the prognosis of CRC (Fig. 6C). To rigorously evaluate the model's performance, we performed ROC curve analysis. Validation set results suggested that the model achieved AUC values of 0.80, 0.87, and 0.84 at 1, 3, and 5 years, respectively, un-

derscoring its ability to significantly improve prediction accuracy and prognostic reliability (Fig. 6E). Calibration plots further substantiated the model's predictive prowess, aligning the predicted survival rates closely with observed rates at the 1, 3, and 5-year marks, thus validating its robust prognostic capacity and clinical relevance (Fig. 6D).

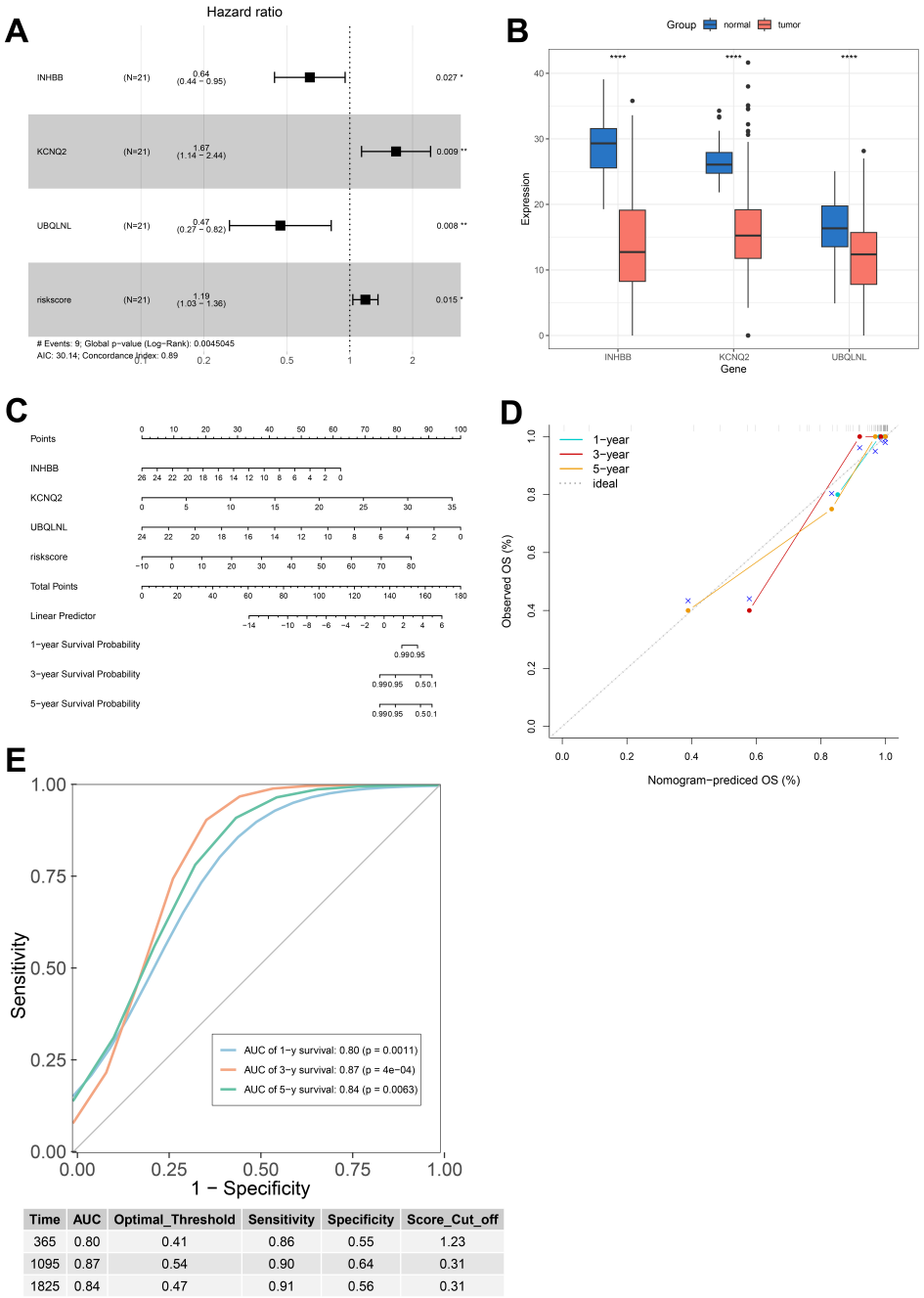
Discussion

CRC, a leading cause to global morbidity and mortality, is increasingly diagnosed in younger patients, raising significant concerns (Chan and Buczacki, 2021; Liu et al, 2024). In the context where current treatment modalities remain limited, it is imperative to conduct extensive research on CRC diagnosis, treatment, and prognostic assessment. This study focused on evaluating the value of radiomics and genetic data in CRC prognosis, yielding several meaningful results through rigorous analyses.

In imaging histological feature analysis, five key features identified through multi-step radiomics data processing from the TICA database effectively distinguished high-risk and low-risk CRC patients, with the high-risk group demonstrating a relatively lower survival rate. CT radiomics has proven valuable in various aspects of CRC, such as early diagnosis (Lv et al, 2022), treatment evaluation (Liu et al, 2017), risk stratification (Lovinfosse et al, 2018), and tumor heterogeneity assessment (González-Castro et al, 2020). The key radiomics features identified in this study highlight their potential in CRC prognostic assessment and may provide valuable insights into tumor microstructure, heterogeneity, and prognosis.

Transcriptome data analysis further identified eight key genes associated with CRC prognosis. Among them, the high expression of three genes, *INHBB*, *KCNQ2*, and *UBQLNL,* significantly increased the risk of CRC-related death. Prior research has established a strong association between elevated *INHBB* expression and advanced CRC characteristics, including increased invasion depth, distant metastasis, and higher Tumor Node Metastasis (TNM) stage. Moreover, this expression pattern correlates with diminished OS and disease-free survival (DFS), thereby suggesting its potential as a prognostic biomarker for CRC (Yuan et al, 2020). Furthermore, *INHBB* expression significantly promotes macrophages infiltration while inhibiting the infiltration of memory T cells, mast cells, and dendritic cells (Yin et al, 2024), aligning with our results. A study by Schell et al (2016), found that mutations in the *KCNQ2* gene may affect the prognosis of CRC, especially in microsatellite-stable (MSS) tumors, although its impact is relatively small. *UBQLNL*, an important member of the Ubiquilin (*UBQLN*) family, has been investigated less in the context of CRC prognosis and clinical diagnosis (Shah et al, 2022). These results suggest that *KCNQ2* and *UBQLNL* play significant roles in CRC prognostic research, warranting further investigation into their potential mechanisms and contribution to CRC progression, development, and prognostic assessment.

Our research on independent prognostic factors identified age, tumor stage, pathological T stage, and pathological N stage as crucial factors, which align with the current clinical understanding of CRC prognosis. These results underscore the



Time	AUC	Optimal_Threshold	Sensitivity	Specificity	Score_Cut_off
365	0.80	0.41	0.86	0.55	1.23
1095	0.87	0.54	0.90	0.64	0.31
1825	0.84	0.47	0.91	0.56	0.31

Fig. 6. Evaluation of selected genes and radiomics features in combination. (A) Forest plot showing the forward-backward stepwise regression analysis of risk scores for radiomics features and key genes. * $p < 0.05$, ** $p < 0.01$. (B) Differential expression of the three key genes amid the normal set and tumor set. **** $p < 0.0001$. (C) Nomogram illustrates the findings of the combined COX regression examination. (D) Calibration plot assessing the prognostic precision of the model. (E) Performance evaluation of the combined model using time ROC curves. ROC, receiver operating characteristic; AUC, area under the curve.

significance of integrating these factors into clinical decision-making and providing a basis for formulating subsequent precision medicine strategies, ultimately facilitating personalized treatment for patients at varying risk levels (Walther et al, 2009). Immune infiltration analysis revealed that the low-risk group showed a better im-

mune status, with significant differences in the distribution of specific immune cell types and correlations between key genes and immune cells. These results highlight the crucial role of the immune microenvironment in CRC prognosis. Prognostic genes may affect CRC progression by regulating the function and distribution of immune cells, supporting the findings of Wang et al (2021).

Based on the previous studies, we conducted COX regression analysis on selected radiomics features and key genes, constructing a CRC prognostic model incorporating *INHBB*, *KCNQ2*, and *UBQLNL*. This model showed a high concordance index, and its prediction accuracy was validated through AUC values and calibration plot results at various time points in the validation set. These findings highlight the advantages of integrating multimodal data, where radiomics features, and gene expression information complement each other, offering a more comprehensive and accurate understanding of CRC prognosis. This approach provides a more valuable prognostic prediction tool for clinical practice, potentially guiding the selection and optimization of clinical treatment regimens.

Numerous studies have shown that combining radiomics and genomics improves the accuracy of survival rate prediction for various cancers, including CRC (Chaddad et al, 2019; Wijethilake et al, 2020). For example, Badic et al (2019) integrated enhanced CT radiological features with gene expression data and found that changes in ATP-binding cassette subfamily C member 2 (*ABCC2*) gene expression correlated with radiological features, N stage, M stage, clinical stage, and progression-free survival of CRC patients. This further indicates that radiomics can enhance the prognostic capability of radiological examinations for CRC patients and strongly corroborate the significance of our integrated model in CRC prognostic assessment.

However, this study has certain limitations. Firstly, the data sources are mainly obtained from databases, which may introduce sample selection bias and data quality differences, potentially affecting the universality of the results. Additionally, while the model performs well in internal validation, it lacks external independent verification, and its applicability across different clinical scenarios needs further investigation. Future research should expand sample sources and include multi-center data to enhance the robustness of the model. Furthermore, investigating the potential interaction mechanisms between key genes and radiomics features and how they jointly affect CRC prognosis will contribute to developing more effective prognostic assessment models and treatment interventions, ultimately improving clinical outcomes for CRC patients and promoting the development of precision medicine.

Conclusion

This comprehensive analysis involving radiomics and transcriptomics develops a prognostic CRC model incorporating three specific genes. Furthermore, our model improves the predictive accuracy of CRC prognosis and efficacy of clinical treatment. However, this study has limitations that require further validation through external datasets and experiments.

Key Points

- This study evaluated the significance of integrating radiomics with genetic data in the prognostic assessment of colorectal cancer.
- Analysis of imaging and transcriptomic data of CRC patients yielded five important features and eight differentially expressed genes associated with the prognosis.
- This study constructed a combined gene-radiomics model incorporating *INHBB*, *KCNQ2*, and *UBQLNL* genes.
- This model exhibits potential clinical applications in CRC prognostic evaluation.

Availability of Data and Materials

The data used to support the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by YNC and XMW. The first draft of the manuscript was written by YNC and ZMF and all authors contributed to revising the manuscript critically for important intellectual content. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgement

We would like to express our sincerest gratitude to all those who provided invaluable support and assistance in the completion of this research project. We would like to extend our gratitude to our colleagues in the Department of Radiology and the Department of Cardiovascular Medicine at Yiwu City Central Hospital, Zhejiang Province, for their invaluable support and assistance. We are also indebted to our peers for their constructive feedback and insightful discussions, which significantly enhanced the quality of our work. Despite the absence of dedicated funding from external sources, we are immensely grateful for the comprehensive support provided by the institutions, which was pivotal to the success of our study.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

References

Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature Communications. 2014; 5: 4006. https://doi.org/10.1038/ncomms5006

Badic B, Hatt M, Durand S, Jossic-Corcos CL, Simon B, Visvikis D, et al. Radiogenomics-based cancer prognosis in colorectal cancer. Scientific Reports. 2019; 9: 9743. https://doi.org/10.1038/s41598-019-46286-6

Cao Q, Tian Y, Deng Z, Yang F, Chen E. Epigenetic Alteration in Colorectal Cancer: Potential Diagnostic and Prognostic Implications. International Journal of Molecular Sciences. 2024; 25: 3358. https://doi.org/10.3390/ijms25063358

Caruso D, Polici M, Zerunian M, Pucciarelli F, Guido G, Polidori T, et al. Radiomics in Oncology, Part 2: Thoracic, Genito-Urinary, Breast, Neurological, Hematologic and Musculoskeletal Applications. Cancers. 2021; 13: 2681. https://doi.org/10.3390/cancers13112681

Chaddad A, Daniel P, Sabri S, Desrosiers C, Abdulkarim B. Integration of Radiomic and Multi-omic Analyses Predicts Survival of Newly Diagnosed IDH1 Wild-Type Glioblastoma. Cancers. 2019; 11: 1148. https://doi.org/10.3390/cancers11081148

Chan DKH, Buczacki SJA. Tumour heterogeneity and evolutionary dynamics in colorectal cancer. Oncogenesis. 2021; 10: 53. https://doi.org/10.1038/s41389-021-00342-x

Chen B, Zhang RN, Fan X, Wang J, Xu C, An B, et al. Clinical diagnostic value of long non-coding RNAs in Colorectal Cancer: A systematic review and meta-analysis. Journal of Cancer. 2020; 11: 5518–5526. https://doi.org/10.7150/jca.46358

Fan G, Qin J, Liu H, Liao X. Commentary: Radiomics in oncology: A 10-year bibliometric analysis. Frontiers in Oncology. 2022; 12: 891056. https://doi.org/10.3389/fonc.2022.891056

González-Castro V, Cernadas E, Huelga E, Fernández-Delgado M, Porto J, Antunez JR, et al. CT Radiomics in Colorectal Cancer: Detection of KRAS Mutation Using Texture Analysis and Machine Learning. Applied Sciences. 2020; 10: 6214. https://doi.org/10.3390/app10186214

Inchingolo R, Maino C, Cannella R, Vernuccio F, Cortese F, Dezio M, et al. Radiomics in colorectal cancer patients. World Journal of Gastroenterology. 2023; 29: 2888–2904. https://doi.org/10.3748/wjg.v29.i19.2888

Klimeck L, Heisser T, Hoffmeister M, Brenner H. Colorectal cancer: A health and economic problem. Best Practice & Research. Clinical Gastroenterology. 2023; 66: 101839. https://doi.org/10.1016/j.bpg.2023.101839

Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nature Reviews. Clinical Oncology. 2017; 14: 749–762. https://doi.org/10.1038/nrclinonc.2017.141

Li X, Wu M, Wu M, Liu J, Song L, Wang J, et al. A radiomics and genomics-derived model for predicting metastasis and prognosis in colorectal cancer. Carcinogenesis. 2024; 45: 170–180. https://doi.org/10.1093/carcin/bgad098

Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuzé S, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. Annals of Oncology. 2017; 28: 1191–1206. https://doi.org/10.1093/annonc/mdx034

Liu J, Zhong F, Chen Y. *UCN*-Centric Prognostic Model for Predicting Overall Survival and Immune Response in Colorectal Cancer. Genes. 2024; 15: 1139. https://doi.org/10.3390/genes15091139

Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. Clinical Cancer Research. 2017; 23: 7253–7262. https://doi.org/10.1158/1078-0432.CCR-17-1038

Lovinfosse P, Polus M, Van Daele D, Martinive P, Daenen F, Hatt M, et al. FDG PET/CT radiomics for predicting the outcome of locally advanced rectal cancer. European Journal of Nuclear Medicine and Molecular Imaging. 2018; 45: 365–375. https://doi.org/10.1007/s00259-017-3855-5

Lv L, Xin B, Hao Y, Yang Z, Xu J, Wang L, et al. Radiomic analysis for predicting prognosis of colorectal cancer from preoperative ^{18}F-FDG PET/CT. Journal of Translational Medicine. 2022; 20: 66. https://doi.org/10.1186/s12967-022-03262-5

Mortezapour M, Tapak L, Bahreini F, Najafi R, Afshar S. Identification of key genes in colorectal cancer diagnosis by co-expression analysis weighted gene co-expression network analysis. Computers in Biology and Medicine. 2023; 157: 106779. https://doi.org/10.1016/j.compbiomed.2023.106779

Nunes L, Li F, Wu M, Luo T, Hammarström K, Torell E, et al. Prognostic genome and transcriptome signatures in colorectal cancers. Nature. 2024; 633: 137–146. https://doi.org/10.1038/s41586-024-07769-3

Schell MJ, Yang M, Teer JK, Lo FY, Madan A, Coppola D, et al. A multigene mutation classification of 468 colorectal cancers reveals a prognostic role for APC. Nature Communications. 2016; 7: 11743. https://doi.org/10.1038/ncomms11743

Shah PP, Saurabh K, Kurlawala Z, Vega AA, Siskind LJ, Beverly LJ. Towards a molecular understanding of the overlapping and distinct roles of UBQLN1 and UBQLN2 in lung cancer progression and metastasis. Neoplasia. 2022; 25: 1–8. https://doi.org/10.1016/j.neo.2021.11.010

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: a Cancer Journal for Clinicians. 2021; 71: 209–249. https://doi.org/10.3322/caac.21660

Walther A, Johnstone E, Swanton C, Midgley R, Tomlinson I, Kerr D. Genetic prognostic and predictive markers in colorectal cancer. Nature Reviews. Cancer. 2009; 9: 489–499. https://doi.org/10.1038/nrc2645

Wang Y, Li W, Jin X, Jiang X, Guo S, Xu F, et al. Identification of prognostic immune-related gene signature associated with tumor microenvironment of colorectal cancer. BMC Cancer. 2021; 21: 905. https://doi.org/10.1186/s12885-021-08629-3

Wei W, Li Y, Huang T. Using Machine Learning Methods to Study Colorectal Cancer Tumor Micro-Environment and Its Biomarkers. International Journal of Molecular Sciences. 2023; 24: 11133. https://doi.org/10.3390/ijms241311133

Wen L, Liu J, Hu P, Bi F, Liu S, Jian L, et al. MRI-Based Radiomic Models Outperform Radiologists in Predicting Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. Academic Radiology. 2023; 30: S176–S184. https://doi.org/10.1016/j.acra.2022.12.037

Wijethilake N, Islam M, Ren H. Radiogenomics model for overall survival prediction of glioblastoma. Medical & Biological Engineering & Computing. 2020; 58: 1767–1777. https://doi.org/10.1007/s11517-020-02179-9

Xu H, Zhao H, Ding C, Jiang D, Zhao Z, Li Y, et al. Celastrol suppresses colorectal cancer via covalent targeting peroxiredoxin 1. Signal Transduction and Targeted Therapy. 2023; 8: 51. https://doi.org/10.1038/s41392-022-01231-4

Xue T, Peng H, Chen Q, Li M, Duan S, Feng F. Preoperative prediction of KRAS mutation status in colorectal cancer using a CT-based radiomics nomogram. The British Journal of Radiology. 2022; 95: 20211014. https://doi.org/10.1259/bjr.20211014

Yin Y, Yang S, Huang Z, Yang Z, Zhang C, He Y. RNA methylation-related genes INHBB and SOWAHA are associated with MSI status in colorectal cancer patients and may serve as prognostic markers for predicting immunotherapy efficacy. Carcinogenesis. 2024; 45: 337–350. https://doi.org/10.1093/carcin/bgae004

Yuan J, Xie A, Cao Q, Li X, Chen J. INHBB Is a Novel Prognostic Biomarker Associated with Cancer-Promoting Pathways in Colorectal Cancer. BioMed Research International. 2020; 2020: 6909672. https://doi.org/10.1155/2020/6909672

Zhang Y, Luo J, Liu Z, Liu X, Ma Y, Zhang B, et al. Identification of hub genes in colorectal cancer based on weighted gene co-expression network analysis and clinical data from The Cancer Genome Atlas. Bioscience Reports. 2021; 41: BSR20211280. https://doi.org/10.1042/BSR20211280

Zhang Y, Wang Y, Zhang B, Li P, Zhao Y. Methods and biomarkers for early detection, prediction, and diagnosis of colorectal cancer. Biomedicine & Pharmacotherapy. 2023; 163: 114786. https://doi.org/10.1016/j.biopha.2023.114786