frontiers | Research Topics

Application of bioinformatics, machine learning, and artificial intelligence to improve diagnosis, prognosis and treatment of cancer

Edited by
Tinka Vidovic and Petar Ozretić

Published in
Frontiers in Aging

ISSN 1664-8714
ISBN 978-2-8325-7402-7
DOI 10.3389/978-2-8325-7402-7

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Application of bioinformatics, machine learning, and artificial intelligence to improve diagnosis, prognosis and treatment of cancer

Topic editors
Tinka Vidovic — Moltech Innovations, Croatia
Petar Ozretić — Rudjer Boskovic Institute, Croatia

Citation
Vidovic, T., Ozretić, P., eds. (2026). *Application of bioinformatics, machine learning, and artificial intelligence to improve diagnosis, prognosis and treatment of cancer*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-7402-7

Table of contents

frontiers | Frontiers in Aging

TYPE Editorial
PUBLISHED 02 January 2026
DOI 10.3389/fragi.2025.1763491

EDITED AND REVIEWED BY
Matthew Yousefzadeh,
Columbia University, United States

*CORRESPONDENCE
Tinka Vidović,
✉ tinka@moltech-innovations.com
Petar Ozretić,
✉ pozretic@irb.hr

†These authors have contributed equally to this work and share first authorship

‡These authors have contributed equally to this work and share senior authorship

RECEIVED 08 December 2025
ACCEPTED 15 December 2025
PUBLISHED 02 January 2026

CITATION
Vidović T and Ozretić P (2026) Editorial: Application of bioinformatics, machine learning, and artificial intelligence to improve diagnosis, prognosis and treatment of cancer.
Front. Aging 6:1763491.
doi: 10.3389/fragi.2025.1763491

Editorial: Application of bioinformatics, machine learning, and artificial intelligence to improve diagnosis, prognosis and treatment of cancer

Tinka Vidović[1]*†‡ and Petar Ozretić[2]*†‡

[1]Moltech Innovations, Zagreb, Croatia, [2]Laboratory for Hereditary Cancer, Division of Molecular Medicine, Ruđer Bošković Institute, Zagreb, Croatia

KEYWORDS

cancer biomarkers, cancer diagnosis and therapy, cancer drug targets, computational cancer biology, computational drug discovery artificial intelligence, bioinformatics, machine learning

Editorial on the Research Topic
Application of bioinformatics, machine learning, and artificial intelligence to improve diagnosis, prognosis and treatment of cancer

In recent years, omics approaches have yielded great advances in cancer research and have provided new in-depth insights into the processes involved in cancer development and progression. Practical use of the information contained within this huge amount of data requires computational approaches such as bioinformatics, machine learning (ML), and artificial intelligence (AI). These computational methods, together with omics data from large databases, such as The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO), can now be used to develop cancer biomarkers, novel anti-cancer drug targets, and both novel and repurposed treatment options for cancer. Considering the application of versatile computational methods in cancer research, we collected original research articles in this Research Topic to present the novel discovery of potential cancer drug targets, prognostic biomarkers, or therapeutic interventions.

Lower extremity deep vein thrombosis (DVT) is a frequent postoperative complication, occurring in up to 40% of patients with colorectal cancer. Liu et al. used an ML model optimized for predicting an individual's risk of DVT in colorectal cancer patients. Given the prevalence of DVT and that traditional risk assessments may not be accurate indicators of true risk, they showed that the XGBoost model (Chen and Guestrin, 2016) has strong potential for improving early detection and treatment in clinical settings.

Identifying novel biomarkers for predicting patient survival time is of crucial practical clinical significance, since it could lead to better patient stratification and treatment decisions. Li et al. used the reactive oxygen species (ROS)-related signature genes, which they identified using the TCGA data, to predict the prognosis and chemotherapy response of patients with bladder cancer. They did not only identify 17 ROS-related genes that exhibited good overall survival in bladder cancer patients, but also 11 potential small molecular drugs that target these ROS-related genes using the Connectivity Map (CMap) database (Lamb et al., 2006).

Using ML models to predict response to treatment could lead to the development of more personalized treatment, leading to significant improvement in patient outcomes. Guo et al. developed the Artificial Intelligence-Derived Anoikis Signature (AIDAS), a novel machine learning-based prognostic tool for breast cancer. AIDAS identifies key gene expression patterns related to anoikis, a form of programmed cell death triggered by detachment from the extracellular matrix. Using AIDAS, the authors found they could more accurately predict breast cancer outcomes compared to existing prognostic models. They discovered that patients with low AIDAS levels may be more responsive to immunotherapy, while those with high AIDAS levels are more susceptible to certain chemotherapies like methotrexate.

Zhu et al. showed the crucial role of autophagy in acute myeloid leukemia (AML) prognosis, identifying essential autophagy genes that correlate with patient survival. Using ML, they developed a predictive model that aids risk stratification and suggested potential therapeutic targets. Their findings also reveal a link between autophagy and the immune microenvironment, offering insights for future research and clinical applications.

Due to the large difference in survival of patients with moderately differentiated gastric adenocarcinoma (MDGA) with distant metastases and without metastases, it becomes important to predict the occurrence of distant metastases after surgical treatment, after morphological examination of all removed lymph nodes, and after final staging of the disease. Yang et al. collected data from MDGA patients from the Surveillance, Epidemiology, and End Results (SEER) database from 2010 to 2019, as well as data from MDGA patients in China. Based on these data they conducted univariate and multivariate analyses, and factors were identified that contribute to the occurrence of distant metastases and worsen the prognosis of the disease.

Prostate cancer is a highly metastatic tumor, and it is estimated that about 50% of patients with advanced disease will develop bone metastases. Once bone metastasis occurs, it is incurable and is significantly associated with mortality. The STING signaling pathway is an important transduction mechanism in innate immunity and viral defense, and it has been demonstrated that this pathway plays a key role in tumorigenesis and metastasis (Wang et al., 2025). In their study, Li et al. extrapolated three key STING pathway genes related to bone metastasis based on a machine learning algorithm. After comprehensive analysis, it was verified that these three genes have key roles in prostate cancer development, metastasis, and tumor immunity, while RELA or transcription factor p65 is a highly potential therapeutic target.

Ding et al. explored the role of FAT1, which is crucial for cellular adhesion and cell signaling, in lung cancer cell lines. The authors identified *FAT1* mutations in five out of thirty-seven individuals diagnosed with non-small cell lung cancer (NSCLC), using next-generation sequencing (NGS) technology. These mutations included four missense mutations and one splice variant. The frequency of *FAT1* mutations was the third highest, following those in *EGFR* and *TP53* genes. The study further demonstrated correlations between *FAT1* expression and methylation with the malignancy of certain cancer types. Knockdown of *FAT1* in A549 and H1299 lung cancer cell lines led to downregulated PD-L1 expression. Additionally, *FAT1* knockdown significantly inhibited cell proliferation, colony formation, and migration. It also affected the cell cycle and the FAK-YAP/TAZ signaling pathway, ultimately inhibiting the proliferation of lung cancer cells *in vivo*.

In their retrospective study, Huang et al. developed a radiomics-clinical predictive model for the response to neoadjuvant chemoimmunotherapy in patients with NSCLC. Their model integrates clinical and radiomic data from two institutions, drawing from a training and internal validation cohort of 105 patients and a second external validation cohort of 43 patients.

Sun et al. conducted an in-depth study on immune-related genes in hepatocellular carcinoma (HCC) using extensive datasets and robust bioinformatics methods, leading to the development of the Subtype-specific and Immune-Related Prognostic Signatures (SIR-PS) model. The SIR-PS model effectively predicted survival outcomes and immunotherapy responses in HCC patients, providing meaningful guidance for personalized immunotherapy.

High serum levels of hepatitis B surface antigen (HBsAg) increase the risk of developing HCC and have a worse prognosis for patients who have already developed HCC. Xiong et al. compared the effects of high and low levels of HBsAg in HCC patients undergoing transarterial chemoembolization (TACE) and sequential ablation and utilized propensity score matching to minimize selection bias. In addition, they created a nomogram to predict the prognosis of HCC patients with high levels of HBsAg after local treatment to more accurately guide the clinical decision.

Cancer incidence rises with aging, even though there are more senescent cells that have stopped dividing as we age. In their bioinformatics study, Ru et al. explored the molecular and immune landscape of cellular senescence in lung adenocarcinoma using publicly available TCGA and GEO datasets to gain deeper insights on the impact of cellular senescence on tumor progression. They showed that patients with low aging scores exhibited better survival, lower tumor mutation burden (TMB), lower somatic mutation frequency, lower tumor proliferation rate, and an immune-activated phenotype compared to patients with high aging scores.

Altogether, with this Research Topic, we primarily wanted to demonstrate that datasets from databases like TCGA and GEO, the former of which are being available and massively reanalyzed for more than a decade, are still relevant and useful for discovering new potential cancer drug targets, prognostic biomarkers, or therapeutic interventions, supported by new methods and ways of analyzing big data, especially now in the dawn of the development and application of AI models in basic cancer research. Even though such studies usually lack, at least *in vitro*, experimental validation, their results validated on external cohorts still present valuable and scientifically sound bases for further research and eventual translation into the clinical practice, while the amount of omics data continues to grow unstoppably. . .

Author contributions

TV: Conceptualization, Investigation, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review and editing. PO: Conceptualization, Investigation, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review and editing.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

Author TV was employed by Moltech Innovations.

The remaining author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

Chen, T., and Guestrin, C. (2016). "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (KDD '16)* (New York, NY, USA: Association for Computing Machinery), 785–794. doi:10.1145/2939672.2939785

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., et al. (2006). The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313 (5795), 1929–1935. doi:10.1126/science.1132939

Wang, Q., Yu, Y., Zhuang, J., Liu, R., and Sun, C. (2025). Demystifying the cGAS-STING pathway: precision regulation in the tumor immune microenvironment. *Mol. Cancer.* 24 (1), 178. doi:10.1186/s12943-025-02380-0

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 20 March 2024
DOI 10.3389/fonc.2024.1348678

EDITED BY
Petar Ozretić,
Rudjer Boskovic Institute, Croatia

REVIEWED BY
Luis Lara-Mejía,
National Institute of Cancerology (INCAN), Mexico
Zheng Yuan,
Shanghai Jiaotong University School of Medicine, China

*CORRESPONDENCE
Lei Shi
shilei@zjcc.org.cn
Hongjie Hu
hongjiehu@zju.edu.cn
Rui Xu
xurui@dlut.edu.cn

†These authors have contributed equally to this work

RECEIVED 03 December 2023
ACCEPTED 06 March 2024
PUBLISHED 20 March 2024

CITATION
Huang D, Lin C, Jiang Y, Xin E, Xu F, Gan Y, Xu R, Wang F, Zhang H, Lou K, Shi L and Hu H (2024) Radiomics model based on intratumoral and peritumoral features for predicting major pathological response in non-small cell lung cancer receiving neoadjuvant immunochemotherapy. *Front. Oncol.* 14:1348678. doi: 10.3389/fonc.2024.1348678

Radiomics model based on intratumoral and peritumoral features for predicting major pathological response in non-small cell lung cancer receiving neoadjuvant immunochemotherapy

Dingpin Huang[1,2,3†], Chen Lin[4†], Yangyang Jiang[1], Enhui Xin[5], Fangyi Xu[1], Yi Gan[6], Rui Xu[7,8*], Fang Wang[1], Haiping Zhang[1], Kaihua Lou[1], Lei Shi[4*] and Hongjie Hu[1,3*]

[1]Department of Radiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China, [2]Department of Radiology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China, [3]Medical Imaging International Scientific and Technological Cooperation Base of Zhejiang Province, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China, [4]Department of Radiology, Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, Zhejiang, China, [5]Department of Research and Development, Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China, [6]Department of Pathology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China, [7]DUT-RU International School of Information Science and Engineering, Dalian University of Technology, Dalian, Liaoning, China, [8]DUT-RU Co-Research Center of Advanced Information Computing Technology (ICT) for Active Life, Dalian University of Technology, Dalian, Liaoning, China

Objective: To establish a radiomics model based on intratumoral and peritumoral features extracted from pre-treatment CT to predict the major pathological response (MPR) in patients with non-small cell lung cancer (NSCLC) receiving neoadjuvant immunochemotherapy.

Methods: A total of 148 NSCLC patients who underwent neoadjuvant immunochemotherapy from two centers (SRRSH and ZCH) were retrospectively included. The SRRSH dataset (n=105) was used as the training and internal validation cohort. Radiomics features of intratumoral (T) and peritumoral regions (P1 = 0-5mm, P2 = 5-10mm, and P3 = 10-15mm) were extracted from pre-treatment CT. Intra- and inter- class correlation coefficients and least absolute shrinkage and selection operator were used to feature selection. Four single ROI models mentioned above and a combined radiomics (CR: T+P1+P2+P3) model were established by using machine learning algorithms. Clinical factors were selected to construct the combined radiomics-clinical (CRC) model, which was validated in the external center ZCH (n=43). The performance of the models was assessed by DeLong test, calibration curve and decision curve analysis.

Results: Histopathological type was the only independent clinical risk factor. The model CR with eight selected radiomics features demonstrated a good predictive performance in the internal validation (AUC=0.810) and significantly improved than the model T (AUC=0.810 vs 0.619, p<0.05). The model CRC yielded the best predictive capability (AUC=0.814) and obtained satisfactory performance in the independent external test set (AUC=0.768, 95% CI: 0.62-0.91).

Conclusion: We established a CRC model that incorporates intratumoral and peritumoral features and histopathological type, providing an effective approach for selecting NSCLC patients suitable for neoadjuvant immunochemotherapy.

KEYWORDS

lung neoplasms, machine learning, immunotherapy, neoadjuvant therapy, peritumor

1 Introduction

Lung cancer has emerged as the leading cause of cancer-related deaths worldwide (1). Among them, non-small cell lung cancer (NSCLC) accounts for approximately 85% (2). The past decade of lung cancer treatment history has demonstrated that preoperative administration of antitumor drugs can reduce tumor size, leading to downstaging and creating favorable conditions for surgery (3). Additionally, research has indicated that neoadjuvant therapy can help eliminate micrometastases and reduce the risk of post-operative recurrence (4). With the advancement of lung cancer treatment drugs, immune checkpoint inhibitors have emerged as a novel and promising class of antitumor agents (5, 6). Studies have shown that the addition of nivolumab to neoadjuvant chemotherapy in lung cancer significantly improves pathological response in patients compared to the use of chemotherapy alone (7, 8).

However, only part of NSCLC patients can benefit from neoadjuvant immunochemotherapy (7). In many cases, the tumor did not shrink significantly following neoadjuvant therapy, and these drugs can have notable side effects such as leukopenia and immune-related pneumonitis (3, 9). Therefore, it is crucial to identify patients who will truly benefit from neoadjuvant immunochemotherapy before initiating treatment (10). In fact, assessing the efficacy of neoadjuvant therapy in lung cancer poses certain challenges, as studying the survival outcomes of patients after neoadjuvant treatment typically requires a long time follow-up (11). The International Association for the Study of Lung Cancer (IASLC) in 2021 suggested that the major pathological response (MPR) in postoperative specimens can be used as an evaluation criterion for neoadjuvant therapy (12). MPR was defined as the viable tumor is less than or equal to 10% in the tumor bed, which provided a convenient approach to assessing treatment effectiveness after neoadjuvant therapy.

Some clinical trials have explored the use of biomarkers such as PD-L1 expression and tumor mutational burden (TMB) to predict MPR. However, their predictive effectiveness remained controversial and the detection of PD-L1 and TMB is invasive. To date, there is no reliable biomarker available to predict MPR following neoadjuvant immunochemotherapy in NSCLC. Thus, there is an urgent need for a credible and non-invasive pre-treatment assessment method.

Radiomics aims to capture the heterogeneity within tumors non-invasively by extracting high-throughput features from images for analysis (13). Numerous studies have demonstrated that radiomics plays a valuable role in tumor diagnosis, treatment, and prognosis assessment (14–16). Research has already utilized pre-treatment CT tumor features to build radiomics model and predict pathological response following neoadjuvant chemoradiation for lung cancer, yielding promising results (17). In fact, the microenvironment surrounding the tumor can also influence the response to immunotherapy, such as the distribution of tumor-infiltrating lymphocytes (TILs) (18). Studies have shown that the distribution of TILs is associated with survival outcomes and treatment response in various diseases (19, 20). Therefore, it is also necessary to further investigate the impact of the specificity of the tumor microenvironment on the effectiveness of neoadjuvant immunochemotherapy.

In this study, we constructed models to predict MPR following neoadjuvant immunochemotherapy for non-small cell lung cancer by extracting radiomic features from both the intratumor and the peritumor regions on CT images. Furthermore, the optimal prediction model was validated in an independent external cohort.

2 Methods and materials

2.1 Study population

This study was granted ethical approval by the institutional review board of Sir Run Run Shaw Hospital (SRRSH) and Zhejiang Cancer Hospital (ZCH), which was performed in accordance with the ethical standards of the 1964 Declaration of Helsinki. Informed consent was waived due to the retrospective nature of this study.

This research retrospectively included patients diagnosed with non-small cell lung cancer (NSCLC) who underwent neoadjuvant immunochemotherapy between June 2019 and December 2022 at two centers (SRRSH and ZCH). The inclusion criteria were as follows: 1) pathologically confirmed NSCLC through endoscopic bronchoscopy or CT-guided needle puncture, 2) preoperative neoadjuvant immunochemotherapy was received, and 3) pre-treatment chest CT was performed. Patients were excluded if any of the following conditions were met: 1) pre-treatment staging as stage I or stage IV; 2) less than two cycles of neoadjuvant treatment received; 3) unavailable enhanced chest CT; 4) time interval between chest CT and treatment initiation exceeds one month; 5) poor CT image quality. Patients from SRRSH were used as the model training and internal validation set, while patients from ZCH were used as the independent external test set. The detailed process of patient inclusion and exclusion is shown in Figure 1.

2.2 Treatment method

All patients underwent standard preoperative evaluations and tumor staging procedures before determining treatment, including tumor biopsy (via bronchoscopy or CT-guided fine-needle puncture), chest CT, abdominal ultrasound, head MRI, and whole-body nuclear imaging. The thoracic surgeons assessed the tumor staging of the patients according to the 8th edition of the Lung Cancer TNM staging system, published by the Union for International Cancer Control (UICC), and determined the neoadjuvant treatment strategy. The standard neoadjuvant immunochemotherapy regimen comprises 2 to 4 cycles of immunotherapy in conjunction with platinum-based chemotherapy. Following the completion of neoadjuvant treatment, comprehensive tumor resection is undertaken by the thoracic surgeons.

2.3 Pathological evaluation

According to the multidisciplinary recommendations from the IASLC regarding pathological assessment of lung cancer excision specimens after neoadjuvant therapy (12), pathologists are responsible for evaluating the pathological responses of surgical specimens. All specimens were re-evaluated by an experienced senior pathologist (Y. Gan) who has more than 10 years of experience in accordance with IASLC. If the initial pathology report is different from Dr. Gan's, Dr. Gan's opinion shall prevail. MPR is defined as the percentage of viable tumor cells in the tumor bed being less or equal to 10%. Non-MPR is defined as the percentage of residual tumor cells in the tumor bed more than 10%.

2.4 Image acquisition

The CT scanning parameters in the two centers are shown in Table 1. The contrast-enhanced scanning technique involved intravenous injection of nonionic contrast material (Ultravist 300 or Ultravist 370, Bayer; or ioversol 320, Hengrui) at a rate of 2.2 to 3 ml/s, based on a dosage of 1.2 ml/kg body weight. Bolus tracking technique was employed, with the arterial phase scan initiated 8 seconds after the descending aortic CT density reached 100 HU. All CT scans were retrieved from the picture archiving and communication system (PACS) for further feature extraction.

2.5 Radiomics procedures

The workflow of radiomics analysis consisted of five steps: region of interest (ROI) segmentation, radiomics features extraction and selection, model construction and evaluation. Radiomics analysis was performed with uAI Research Portal (United Imaging Intelligence, China) (21), which is a clinical research platform implemented by Python programming language (version 3.7.3), and widely used package PyRadiomics (https://pyradiomics.readthedocs.io/en/latest/index.html).

All images were imported into an open-source software ITK-SNAP (Version 3.8.0). The tumor ROI was manually segmented slice-by-slice by an experienced radiologist with over 10 years (DP. Huang), without knowledge of the pathological results. Then, the uAI Research Portal was applied for morphological expansion of intratumor ROI. Previous study showed that it would not reduce

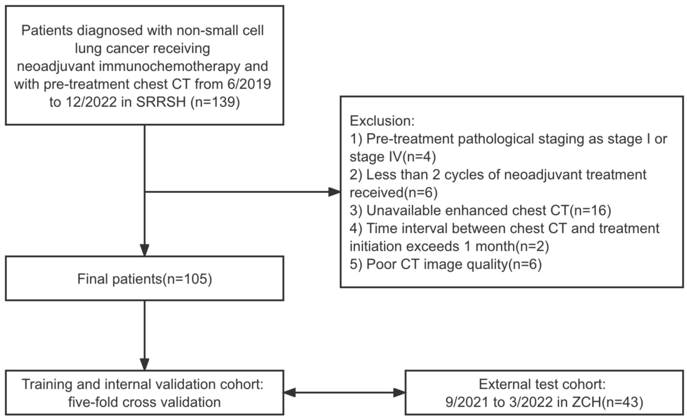


FIGURE 1
Patient selection and distribution flowchart.

TABLE 1 Scanning parameters and CT specifications in both hospitals.

	Sir Run Run Shaw Hospital					Zhejiang Cancer Hospital		
Brand	Siemens	Siemens	Siemens	GE	GE	Siemens	GE	Philips
Machine type	SOMATOM Definition Flash	SOMATOM Force	SOMATOM go. Top	Lightspeed VCT	Optima CT620	SOMATOM Definition Flash	Optima CT680	Ingenuity CT
Tube voltage (KV)	120	100/120	120	120	120	120	120	120
Tube current (mAs)	smart	smart	smart	smart	smart	smart	smart	smart
Rotation time (s)	0.5	0.5	0.5	0.4、0.5	0.5	0.5	0.5	0.5
Image matrix	512×512	512×512	512×512	512×512	512×512	512×512	512×512	512×512
Field of view (mm)	350	350	350	350	350	350	350	350
Reconstruction slice thickness and spacing	2mm/2mm	2mm/2mm	2mm/2mm	1.25mm/ 1.25mm	2mm/2mm	2mm/2mm	5mm/5mm	1.25mm/ 1.25mm
Reconstruction algorithm	B41f	B41f	B41f	Standard resolution	Standard resolution	B31f	Standard resolution	Standard resolution

the risk of recurrence when the tumor resection margin exceeded 15mm (22). Based on this, we performed peritumor expansion three times, 5mm each time, for a total of 15mm. During the delineation and dilation process of the ROIs, non-pulmonary regions were excluded. The peritumoral area beyond the lung outline was manually erased when the tumor located in paramediastinal, subpleural and other special areas. Therefore, in this study, a total of four ROIs were delineated showed in Figure 2, namely T (intratumor), P1 (peritumoral 0-5mm), P2 (peritumoral 5-10mm), and P3 (peritumoral 10-15mm). Subsequently, we established a combined radiomics (CR: T+P1+P2+P3) model by integrating intratumoral and three peritumoral ROI features.

In addition, to evaluate the reproducibility of image segmentation, we randomly selected 20 patients to be re-segmented by Dr. Huang and the other doctor (HP. Zhang, with 1 year of experience in imaging) one month later. Intra-observer and inter-observer reproducibility of radiomics features were assessed using intra- and inter- class correlation coefficient (ICC). A value of ICC ≥ 0.85 was considered indicative of good reproducibility. To eliminate index dimension difference, the extracted radiomics features were standardized into normal distributed z-scores. For feature selection, the least absolute shrinkage and selection operator (LASSO) regression was utilized.

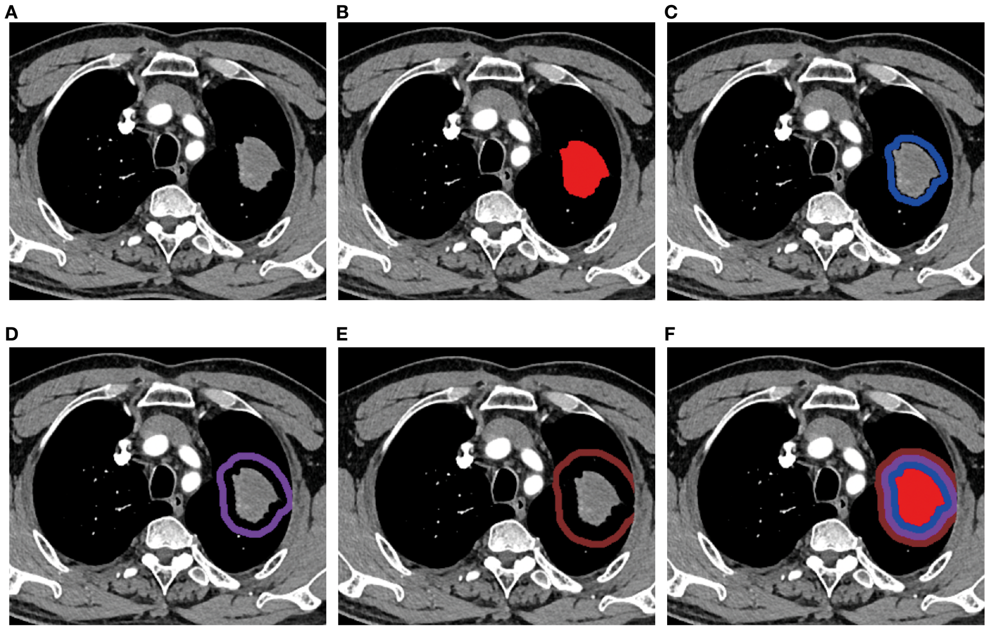


FIGURE 2
Region of interest (ROI) segmentation. **(A)** A mass showed in the upper lobe of the left lung. The ROIs of **(B)** intratumor(T), **(C)** peritumoral 0~5mm (P1), **(D)** peritumoral 5~10mm(P2), **(E)** peritumoral 10~15mm(P3), and **(F)** T+P1+P2+P3(CR).

With the selected optimal feature sets, we built prediction models for the MPR of neoadjuvant immunochemotherapy for lung cancer by using five machine learning algorithms, including decision tree, Gaussian process, logistic regression, random forest and support vector machine, and the model with the best predictive capability was reserved for external validation. The performance of the different prediction models in internal dataset was assessed by the cross-validation strategy to protect against overfitting due to the limited amount data. In this study, we used five-fold cross-validation (23): the feature set was split randomly, while, the same ratio of positive and negative patients was kept in each partition. Consequently, training on four-fifths of dataset and validating on the remaining partition in each fold, the process was repeated five times within different subgroups, and thus formed five unlike training/validation sets and obtained an average result.

2.6 Statistics

SPSS (version 25.0) and Python (version 3.5.6) were used for statistical analysis. Continuous data was presented as mean ± standard deviation or median (interquartile range), and the differences between groups were compared using independent sample t-tests or non-parametric tests. Categorical data was evaluated using chi-square tests or Fisher's exact tests to assess intergroup differences. Univariable and multivariable logistic regression were used to identify clinical risk factors with odds ratio (OR) and 95% confidence interval (CI). The performance of the model was evaluated using receiver operating curves (ROC), and the area under the curve (AUC), sensitivity, specificity and accuracy were quantified. The DeLong test was used for the performance comparison between different models. The LASSO was utilized for the radiomics features selection. Calibration curve was applied to determine whether the projected probability matches the actual probability. Decision curve analysis was used to assess the prediction models' clinical viability. A P-value less than 0.05 (P-value < 0.05) was considered statistically significant.

3 Results

3.1 Clinical characteristics

A total of 148 patients were enrolled retrospectively, and their baseline clinical characteristics were presented in Table 2. The training and internal validation sets consisted of 105 patients from SRRSH, of whom 76 achieved MPR (72.4%). The independent external test set (ZCH) included 43 patients, with 22 achieving MPR (51.2%). The average age of the entire cohort was 63.8 ± 6.3 years, predominantly male (94.6%), and most patients

TABLE 2 Clinical factors of the entire dataset.

Clinical factor	Entire	Training and internal validation (n=105)			External test (n=43)		
	N=148	MPR (n=76)	Non-MPR (n=29)	*P* value	MPR (n=22)	Non-MPR (n=21)	*P* value
Age	63.8 ± 6.3	64.0 ± 6.3	63.1 ± 6.3	0.5	63.1 ± 6.9	64.7 ± 6.3	0.43
Gender				0.25			0.58
Male	140(94.6)	74(97.4)	26(89.7)		20(90.9)	20(95.2)	
Female	8(5.4)	2(2.6)	3(10.3)		2(9.1)	1(4.8)	
Smoking history				0.60			0.96
Current or before	88(59.5)	35(46.1)	15(51.7)		20(90.9)	18(85.7)	
Never	60(40.5)	41(53.9)	14(48.3)		2(9.1)	3(14.3)	
Histopathological type				0.02*			0.02*
Adenocarcinoma	17(11.5)	4(5.3)	7(24.1)		1(4.5)	5(23.8)	
Squamous	115(77.7)	64(84.2)	20(69.0)		20(90.9)	11(52.4)	
Others	16(10.8)	8(10.5)	2(6.9)		1(4.5)	5(23.8)	
Pretreatment clinical stage				0.74			0.32
II	34(23.0)	18(23.7)	6(20.7)		7(31.8)	3(14.3)	
III	114(77.0)	58(76.3)	23(79.3)		15(68.2)	18(85.7)	
Clinical T stage				0.70			0.46
T1	18(12.2)	9(11.8)	2(6.9)		2(9.1)	5(23.8)	

(Continued)

TABLE 2 Continued

Clinical factor	Entire	Training and internal validation (n=105)			External test (n=43)		
	N=148	MPR (n=76)	Non-MPR (n=29)	*P* value	MPR (n=22)	Non-MPR (n=21)	*P* value
T2	61(41.2)	28(36.8)	13(44.8)		10(45.5)	10(47.6)	
T3	43(29.1)	24(31.6)	7(24.1)		8(36.4)	4(19.0)	
T4	26(17.6)	15(19.7)	7(24.1)		2(9.1)	2(9.5)	
Clinical N stage				0.41			0.55
N0	18(12.2)	11(14.5)	1(3.4)		4(18.2)	2(9.5)	
N1	34(23.0)	19(25.0)	8(27.6)		4(18.2)	3(14.3)	
N2	86(58.1)	39(51.3)	18(62.1)		13(59.1)	16(76.2)	
N3	10(6.8)	7(9.2)	2(6.9)		1(4.5)	0	
Treatment cycle				0.37			0.13
2	117(79.1)	61(80.3)	22(75.9)		15(68.2)	19(90.5)	
3	25(16.9)	12(15.8)	7(24.1)		4(18.2)	2(9.5)	
4	6(4.1)	3(3.9)	0		3(13.6)	0	
Platinum drugs				0.61			0.37
Carboplatin	82(55.4)	30(39.5)	14(48.3)		18(81.8)	20(95.2)	
Cisplatin	65(43.9)	45(59.2)	15(51.7)		4(18.2)	1(4.8)	
Nedaplatin	1(0.7)	1(1.3)	0		0	0	
ICIs				0.84			0.10
Pembrolizumab	43(29.1)	21(27.6)	12(41.4)		7(31.8)	3(14.3)	
Tislelizumab	41(27.7)	19(25.0)	6(20.7)		10(45.5)	6(28.6)	
Camrelizumab	35(23.6)	15(19.7)	5(17.2)		5(22.7)	10(47.6)	
Sintilimab	16(10.8)	10(13.2)	4(13.8)		0	2(9.5)	
Toripalimab	11(7.4)	9(11.8)	2(6.9)		0	0	
Durvalumab	1(0.7)	1(1.3)	0		0	0	
Penpulimab	1(0.7)	1(1.3)	0		0	0	

Data are presented as mean ± SD. Data in parentheses are percentages. $^{*}p<0.05$.
MPR, major pathological response; ICIs, immune checkpoint inhibitors.

had a history of smoking (59.5%). The majority of patients had squamous cell carcinoma as the histopathological type (77.7%). In both of the two cohorts, there was significant difference in histopathological type between the MPR and non-MPR groups ($p < 0.05$). In the entire cohort, the majority of patients undergoing neoadjuvant treatment were assessed as stage III lung cancer (77.0%). Moreover, T2 (41.2%) and N2 (58.1%) stage were accounted for the most. The main types of immunotherapy agents employed in the two hospitals include pembrolizumab, tislelizumab, and camrelizumab (31.4%, 23.8%, 19.0% in SRRSH and 23.2%, 37.2%, 34.9% in ZCH, respectively). There was no significant difference in the treatment modality for neoadjuvant therapy between the MPR and non-MPR groups in both cohorts.

After performing univariable and multivariable logistic regression analysis, the histopathological type was confirmed as an independent risk factor and then included in the clinical model ($p = 0.026$; OR = 3.328, 95% CI: 1.155-9.588) (Table 3).

3.2 Selection of the radiomics features

In total, 2264 radiomic features were extracted, including 104 original features grouped as: 18 first-order statistics, 72 texture and 14 shape, and other 2160 features based on images through 25 filters, such as boxmean, wavelet, laplacian, etc. A total of 1,067 features were retained after ICC analysis (Table 4). After the feature

TABLE 3 Univariable and multivariable logistic regression analyses of clinical factors.

Clinical factors	Univariable		Multivariable	
	OR (95% CI)	*P* value	OR (95%CI)	*P* value
Age	1.024(0.956-1.098)	0.499		
Gender	4.269(0.675-26.993)	0.123		
Smoking history	0.797(0.338-1.877)	0.603		
Histopathological type	3.328(1.155-9.588)	0.026*	3.328(1.155-9.588)	0.026*
Pretreatment clinical stage	0.841(0.296-2.384)	0.744		
Clinical T stage	0.930(0.587-1.472)	0.756		
Clinical N stage	0.758(0.435-1.321)	0.328		
Treatment cycle	0.981(0.410-2.348)	0.966		
Platinum drugs	0.794(0.349-1.810)	0.583		
ICIs	1.127(0.900-1.411)	0.297		

*$p<0.05$.
OR, odds ratio; CI, confidence interval; ICIs, immune checkpoint inhibitors.

selection processes as mentioned above, the top features of each radiomics model were selected and presented in Table 5.

The CR model incorporated a total of eight radiomics features as follows: 1) intratumor: glszm_wavelet-LLL-SZNUN; 2) peritumoral 0-5mm: firstorder_Maximum, glrlm_log-sigma-4-0-mm-3D-LRHGLE; 3) peritumoral 5-10mm: gldm_wavelet-LLL-SDLGLE, glcm_wavelet-LLH-Idmn; 4) peritumoral 10-15mm: Complexity, glrlm_log-sigma-4-0-mm-3D-LRHGLE, gldm_SDLGLE.

3.3 Development and validation of the prediction models

The predictive performance of each model was shown in Table 6 and Figure 3.

The clinical model showed relatively poor predictive performance in training and internal validation sets (AUC=0.612 and 0.563, respectively). The single ROI radiomics models based on intratumor(T), peritumoral 0-5mm(P1), peritumoral 5-10mm(P2), peritumoral 10-15mm(P3) showed higher AUCs (0.679, 0.882, 0.746, 0.777 and 0.619, 0.712, 0.662, 0.741, respectively) in training and internal validation sets than the clinical model.

The model CR based on Gaussian process demonstrated an AUC of 0.810 for MPR prediction in NSCLC neoadjuvant immunochemotherapy, which is superior than the four single ROI models and significantly improved than the model T (AUC=0.810 vs 0.619, p<0.05). The Delong test showed that the AUC of models CR and CRC was significantly improved compared to models T and P2. However, pairwise comparisons among the remaining models indicated no statistically significant differences in performance (Figure 3C). We fused CR model with the clinical model to create combined radiomics + clinical (CRC) model and obtained optimal predictive capability, which achieved an AUC of 0.814, sensitivity of 0.947, specificity of 0.567, precision of 0.851, and accuracy of 0.838 in the internal validation set (Table 6).

TABLE 4 Radiomics features distribution (Total and after ICC analysis).

	Features	Total (n=2264)	After ICC analysis (n=1067)
Original	First-order features	18	6
Original	Shape features	14	3
Original	GLCM based features	21	11
Original	GLRLM based features	16	6
Original	GLSZM based features	16	1
Original	GLDM based features	14	5
Original	NGTDM based features	5	2
Filtered	Box mean based features	90	35
Filtered	Additive Gaussian noise based features	90	32
Filtered	Binomial blur image based features	90	32
Filtered	Curvature flow based features	90	30
Filtered	Box sigma image based features	90	64
Filtered	Log based features	360	193
Filtered	Wavelet based features	720	395
Filtered	Normalize based features	90	14
Filtered	Laplacian sharpening based features	90	41
Filtered	Discrete Gaussian based features	90	34
Filtered	Mean based features	90	34

(Continued)

TABLE 4 Continued

	Features	Total (n=2264)	After ICC analysis (n=1067)
Filtered	Speckle noise based features	90	34
Filtered	Recursive Gaussian based features	90	34
Filtered	Shortnoise based features	90	61

GLCM, gray level co-occurrence matrix; GLRLM, gray level run length matrix; GLSZM, gray level size zone matrix; GLDM, gray level dependence matrix; NGTDM, neighbourhood gray-tone difference matrix.

Finally, the CRC model was validated in an independent external test set and achieved favorable predictive performance, with an AUC of 0.768 (95% CI, 0.62-0.91) (Figure 3D).

3.4 Calibration curve and decision curve analysis of the prediction models

The calibration curve of the model CR showed that the predicted probability had a good consistency in the internal validation set. And the fusion model CRC had the smallest Brier score loss, which means it has the best predictive performance (Figures 4A, B).

Decision curve analysis showed that the fusion model CRC provided a better net benefit than other radiomics models for the most of the threshold range (Figures 4C, D).

4 Discussion

Neoadjuvant immunochemotherapy has emerged as a promising therapeutic approach for non-small cell lung cancer (NSCLC) (7). However, the evaluation of neoadjuvant treatment efficacy relies on postoperative pathological assessment, leading to time delay. Additionally, the effects of immune checkpoint inhibitors on tumors are complex, and atypical responses such as hyperprogression or pseudoprogression may occur (24, 25), making it challenging to assess the efficacy of neoadjuvant immunochemotherapy through CT follow-up during treatment. Our research showed that the combined radiomics model based on intratumoral and peritumoral regions derived from pre-treatment CT images can predict MPR to neoadjuvant immunochemotherapy in NSCLC. After incorporating the independent risk factor of histopathological type, the model achieved the optimal predictive performance. Furthermore, its predictive efficacy was validated in an external center, indicating its robustness.

Squamous cell carcinoma was identified as an independent clinical risk factor for predicting MPR in neoadjuvant immunochemotherapy in our research, consistent with previous

TABLE 5 The selected radiomics features in different radiomics models.

Radiomics Models	The selected Radiomics Features
T(Intratumor) (n=5)	boxsigmaimage_glrlm_LongRunHighGrayLevelEmphasis wavelet_glcm_wavelet-LHH-Idn wavelet_gldm_wavelet-LHH-SmallDependenceLowGrayLevelEmphasis wavelet_glszm_wavelet-LLL-SizeZoneNonUniformityNormalized wavelet_glcm_wavelet-HHL-Idn
P1(Peritumoral 0-5mm) (n=8)	log_firstorder_log-sigma-2-0-mm-3D-Skewness log_glrlm_log-sigma-4-0-mm-3D-LongRunHighGrayLevelEmphasis wavelet_gldm_wavelet-HHL-SmallDependenceHighGrayLevelEmphasis boxsigmaimage_glszm_SmallAreaLowGrayLevelEmphasis shotnoise_glcm_Imc1 wavelet_glcm_wavelet-LLH-Idn wavelet_gldm_wavelet-LLH-SmallDependenceHighGrayLevelEmphasis mean_firstorder_Maximum
P2(Peritumoral 5-10mm) (n=4)	wavelet_gldm_wavelet-HLH-LargeDependenceEmphasis shotnoise_glcm_Idmn wavelet_glcm_wavelet-LLH-Idmn wavelet_gldm_wavelet-LLL-SmallDependenceLowGrayLevelEmphasis
P3(Peritumoral 10-15mm) (n=3)	laplaciansharpening_gldm_SmallDependenceLowGrayLevelEmphasis shotnoise_ngtdm_Complexity log_glrlm_log-sigma-4-0-mm-3D-LongRunHighGrayLevelEmphasis
CR (Combined radiomics) (n=8)	T_wavelet_glszm_wavelet-LLL-SizeZoneNonUniformityNormalized P1_mean_firstorder_Maximum P1_log_glrlm_log-sigma-4-0-mm-3D-LongRunHighGrayLevelEmphasis P2_wavelet_gldm_wavelet-LLL-SmallDependenceLowGrayLevelEmphasis P2_wavelet_glcm_wavelet-LLH-Idmn P3_shotnoise_ngtdm_Complexity P3_log_glrlm_log-sigma-4-0-mm-3D-LongRunHighGrayLevelEmphasis P3_laplaciansharpening_gldm_SmallDependenceLowGrayLevelEmphasis

TABLE 6 The performance of different models in training and internal validation sets.

Model	Training set					Internal validation set				
	AUC [95%CI]	Sen	Spe	Pre	Acc	AUC [95%CI]	Sen	Spe	Pre	Acc
Clinical	0.612[0.55, 0.67]	0.947	0.242	0.766	0.752	0.563[0.44, 0.69]	0.947	0.247	0.768	0.752
T	0.679[0.62, 0.74]	0.648	0.604	0.810	0.636	0.619[0.49, 0.74]	0.698	0.587	0.823	0.667
P1	0.882[0.84, 0.92]	0.967	0.347	0.799	0.795	0.712[0.59, 0.83]	0.947	0.200	0.760	0.743
P2	0.746[0.69, 0.80]	0.687	0.655	0.840	0.679	0.662[0.54, 0.78]	0.670	0.660	0.834	0.667
P3	0.777[0.72, 0.83]	0.937	0.293	0.777	0.760	0.741[0.63, 0.85]	0.934	0.273	0.775	0.752
CR	0.889[0.85, 0.93]	0.964	0.613	0.868	0.867	0.810[0.71, 0.91]	0.921	0.533	0.840	0.810
CRC	0.897[0.86, 0.94]	0.977	0.630	0.874	0.881	0.814[0.71, 0.92]	0.947	0.567	0.851	0.838

AUC, area under the curve; CI, confidence interval, Sen Sensitivity; Spe, Specificity, Pre, Precision; Acc, Accuracy; T, intratumor; P1, peritumoral 0-5mm; P2, peritumoral 5-10mm, P3, peritumoral 10-15mm; CR, combined radiomics; CRC, combined radiomics+clinical.

related research (26). A meta-analysis exploring the impact of histopathology on the efficacy of immune checkpoint inhibitors in treating NSCLC showed that immunotherapy can improve overall survival (OS) and progression-free survival (PFS) in both squamous cell carcinoma and non-squamous cell carcinoma, with squamous cell carcinoma patients benefiting more significantly (27). Studies have indicated that compared to non-squamous cell carcinoma, lung squamous cell carcinoma exhibits higher PD-L1 expression, higher tumor mutational burden (TMB), and a greater density of functional TILs in the tumor microenvironment, factors that collectively contribute to the enhanced therapeutic effects of immunotherapy in squamous cell carcinoma patients (28).

In this study, the radiomics model based on intratumoral region had an AUC of only 0.619 (sensitivity of 0.698, specificity of 0.587) in the internal validation group. The peritumoral models showed improvement in AUC compared to the intratumoral model

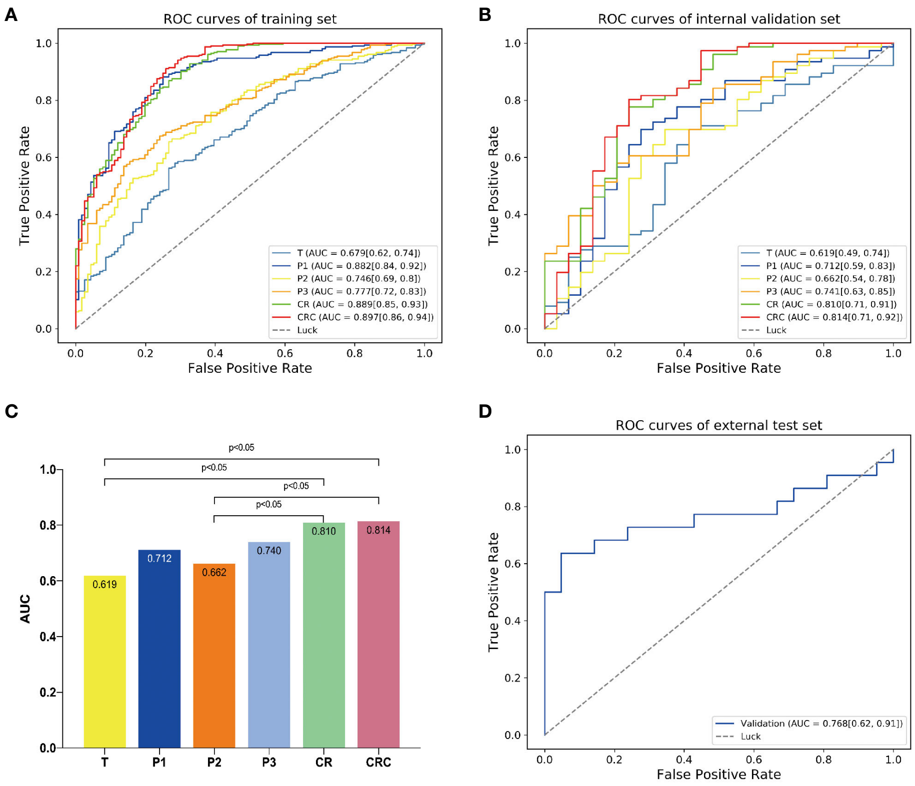


FIGURE 3
The predictive performance of different models. The AUCs of different models in **(A)** training and **(B)** internal validation sets. **(C)** Delong test showed that the model CR was significantly better than model T. By adding clinical independent risk factor to the model CR, the fusion model (CRC) obtained the best predictive performance [AUC=0.814 (0.71, 0.92)]. **(D)** Receiver operating characteristic (ROC) curve of the fusion model (CRC) in the external test set [AUC=0.768 (0.62, 0.91)].

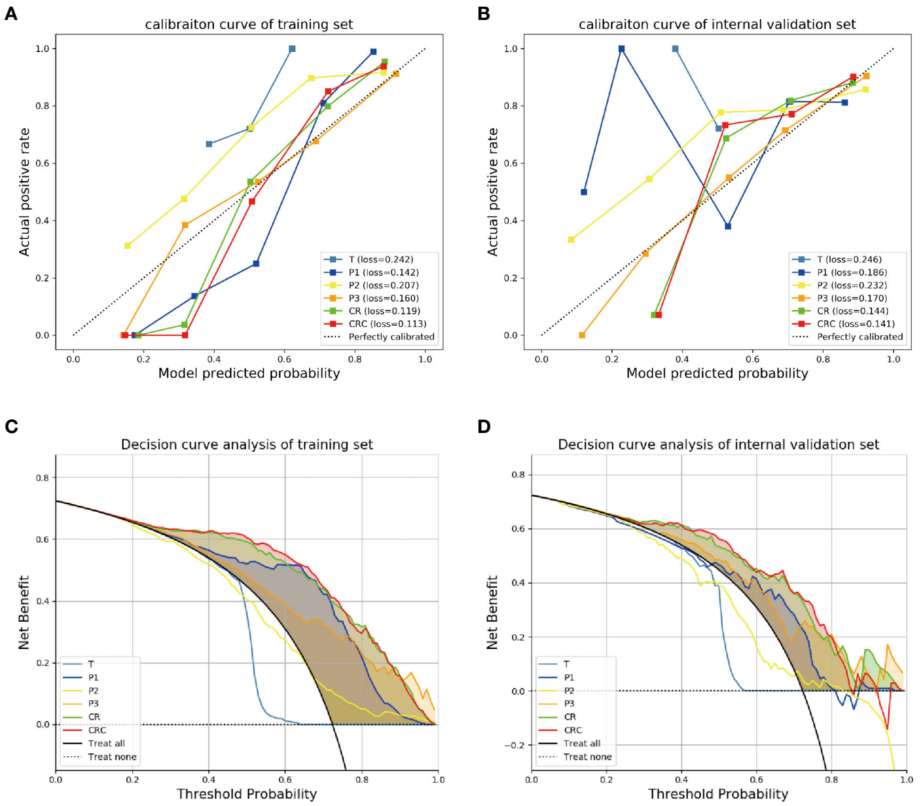


FIGURE 4
The calibration curves and decision curve analysis for different models. The calibration curves for different models in **(A)** training and **(B)** internal validation sets showed the fusion model CRC had the smallest Brier score loss, which means it has the best predictive performance. The decision curve analysis for the different models in **(C)** training and **(D)** internal validation sets showed that the fusion model CRC provided a better net benefit than other radiomics models for the most of the threshold range.

(ranging from 0.662 to 0.741) and higher sensitivity (ranging from 0.670 to 0.947) while their specificity was notably low (P1, P2, P3 = 0.2, 0.273, 0.66, respectively). This pointed out that any radiomics model based on single ROI either intratumoral region or peritumoral region cannot achieve the ideal prediction ability in predicting the effect of neoadjuvant immunochemotherapy in NSCLC.

Prior studies, including one in which our center participated, built prediction models focusing on intratumoral features to predict MPR in NSCLC following neoadjuvant therapy and achieved favorable results (26, 29). Considering that immune checkpoint inhibitors exert their anti-tumor effects by influencing the tumor and its surrounding immune microenvironment (30), it is crucial to investigate the peritumoral microenvironment's features and their impact on immunotherapy. We extracted radiomics features from both intratumoral and peritumoral regions in arterial-phase, establishing a combined radiomics model with eight top-level radiomics features (one feature from the intratumoral region and the remaining seven from the peritumoral regions). Furthermore, among the seven peritumoral top-level radiomic features, six were texture features, including the common feature (log-glrlm-LRHGLE) from P1 and P3. These results highly suggested that the heterogeneity of the peritumoral microenvironment plays a crucial role in neoadjuvant immunochemotherapy for NSCLC. Moreover, when combining the intratumoral and peritumoral models, there was a significant increase in predicting MPR (AUC=0.810), achieving a relative balance between sensitivity and specificity (0.921 and 0.533, respectively), resulting in a satisfactory accuracy of 0.810.

The tumor microenvironment is composed of fibroblasts, immune and inflammatory cells, as well as interstitial components and microvessels (31). Several studies indicated a correlation between peritumoral texture features and tumor-infiltrating lymphocyte (TIL) density, and higher TIL levels are associated with immune system activation for tumor suppression, indicating a greater likelihood of responding to immunotherapy (32, 33). The distribution of blood vessels in the peritumoral environment also influences the efficacy of chemotherapy and immunotherapy (34). Research by Vaidya P et al. demonstrated that peritumoral texture features can reflect biological pathways such as tumor vascular invasion and neovascularization (35). Disorganized and irregular peritumoral blood vessels promote tumor growth, inhibit the anti-tumor effects of drugs, and are often associated with more heterogeneous radiomic features (17, 35). Additionally, it was demonstrated that different ranges of peritumoral regions are associated with differences in treatment response (36) and exhibit distinct texture feature expressions (37). Our study confirmed a strong correlation between the peritumoral microenvironment and neoadjuvant immunochemotherapy in NSCLC. By incorporating different ranges of peritumoral microenvironment texture features, the prediction model got obvious improvement in predicting MPR.

Although the addition of independent clinical risk factor to model CR resulted in a slight improvement in prediction performance, there was no statistically significant difference in AUC between CR and the best model CRC. This may suggest that information contained within combined intratumoral and peritumoral radiomics adequately capture the efficacy of neoadjuvant immunochemotherapy for NSCLC, thus constraining the representation of clinical factor in the model. However, this requires further verification.

In addition, there is a clear difference in the proportion of MPR patients between the two hospitals included in our study. Indeed, according to a review on neoadjuvant therapy for non-small cell lung cancer, the attainment of MPR varies significantly across different studies, ranging approximately from 36.9% to 84.6% after neoadjuvant immunochemotherapy (10). This variability may be attributed to differences in the patient demographics, disease stages at presentation, and the specific neoadjuvant immunochemotherapy regimens. Based on the aforementioned understanding, we consider the MPR proportions in both hospitals in our study to still fall within a reasonable range. On the other hand, despite the differences in patients and treatment regimens at the two hospitals in our study, our research results still demonstrate that the combined intratumoral and peritumoral radiomics model achieves favorable predictive performance at external center, possibly indicating the effectiveness and robustness of this model.

Our research has several limitations. Firstly, the study was retrospective and might be subject to selection bias. Secondly, while the study included patients receiving neoadjuvant immunochemotherapy, there were variations in the selection of chemotherapy drugs and immune checkpoint inhibitors, as well as differences in the treatment cycles. Therefore, it is essential to unified treatment protocols or conduct a stratified study focusing on different regimens in future research. Thirdly, the imaging data from the two centers were obtained from different manufacturers and multiple models of CT machines, which may introduce inconsistencies in equipment parameters. Lastly, the sample size of this study is limited, and it is necessary to further expand the sample for future research.

In conclusion, our study constructed a CRC model comprising intratumoral and peritumoral features and independent clinical risk factors for predicting MPR in NSCLC patients receiving neoadjuvant immunochemotherapy. The combined model achieved an optimal predictive performance (AUC=0.814), and successfully validated in an external center (AUC=0.768). This provides a non-invasive and effective predictive approach for clinical physicians to identify suitable NSCLC patients for neoadjuvant immunochemotherapy.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving humans were approved by Institutional review board of Sir Run Run Shaw Hospital and Zhejiang Cancer Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from a by- product of routine care or industry. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

DH: Writing – review & editing, Writing – original draft, Resources, Methodology, Formal analysis, Data curation, Conceptualization. CL: Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. YJ: Writing – original draft, Methodology, Investigation, Data curation. EX: Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. FX: Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. YG: Writing – original draft, Methodology, Data curation, Conceptualization. RX: Writing – review & editing, Visualization, Methodology, Investigation, Funding acquisition. FW: Writing – original draft, Methodology, Data curation. HZ: Writing – original draft, Methodology, Data curation. KL: Writing – original draft, Investigation, Data curation. LS: Writing – review & editing, Validation, Supervision, Resources, Methodology, Funding acquisition. HH: Writing – review & editing, Visualization, Supervision, Software, Resources, Funding acquisition.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by Grant 82071988 and 61772106 from National Natural Science Foundation of China, Grant 2019C03064 from Key Research and Development Program of Zhejiang Province, Grant WKJ-ZJ-1926 from Program Co-sponsored by Province and Ministry, Grant 2021FZZX003-02-17 and 226-2022-00160 from Fundamental Research Funds for the Central Universities, Grant 2022KY097 from Medical Science and Technology Project of Zhejiang Province, Grant Y2020172 from Wenzhou Municipal Science and Technology Bureau Project, Grant LQ22H160051 from Zhejiang Provincial Natural Science Foundation of China.

Conflict of interest

EX was employed by the company Shanghai United Imaging Intelligence Co., Ltd.

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

Publisher's note

References

1. 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 Cancer J Clin*. (2021) 71:209–49. doi: 10.3322/caac.21660

2. Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. *Lancet*. (2021) 398:535–54. doi: 10.1016/S0140-6736(21)00312-3

3. Shu CA, Gainor JF, Awad MM, Chiuzan C, Grigg CM, Pabani A, et al. Neoadjuvant atezolizumab and chemotherapy in patients with resectable non-small-cell lung cancer: an open-label, multicentre, single-arm, phase 2 trial. *Lancet Oncol*. (2020) 21:786–95. doi: 10.1016/S1470-2045(20)30140-6

4. Chen Y, Yan B, Xu F, Hui Z, Zhao G, Liu J, et al. Neoadjuvant chemoimmunotherapy in resectable stage IIIA/IIIB non-small cell lung cancer. *Transl Lung Cancer Res*. (2021) 10:2193–204. doi: 10.21037/tlcr-21-329

5. Borghaei H, Gettinger S, Vokes EE, Chow LQM, Burgio MA, de Castro Carpeno J, et al. Five-year outcomes from the randomized, phase III trials CheckMate 017 and 057: nivolumab versus docetaxel in previously treated non-small-cell lung cancer. *J Clin Oncol*. (2021) 39:723–33. doi: 10.1200/JCO.20.01605

6. Herbst RS, Garon EB, Kim D-W, Cho BC, Gervais R, Perez-Gracia JL, et al. Five year survival update from KEYNOTE-010: pembrolizumab versus docetaxel for previously treated, programmed death-ligand 1-positive advanced NSCLC. *J Thorac Oncol*. (2021) 16:1718–32. doi: 10.1016/j.jtho.2021.05.001

7. Forde PM, Spicer J, Lu S, Provencio M, Mitsudomi T, Awad MM, et al. Neoadjuvant nivolumab plus chemotherapy in resectable lung cancer. *N Engl J Med*. (2022) 386:1973–85. doi: 10.1056/NEJMoa2202170

8. Rothschild SI, Zippelius A, Eboulet EI, Savic Prince S, Betticher D, Bettini A, et al. SAKK 16/14: durvalumab in addition to neoadjuvant chemotherapy in patients with stage IIIA(N2) non-small-cell lung cancer-A multicenter single-arm phase II trial. *J Clin Oncol*. (2021) 39:2872–80. doi: 10.1200/JCO.21.00276

9. Provencio M, Nadal E, Insa A, García-Campelo MR, Casal-Rubio J, Dómine M, et al. Neoadjuvant chemotherapy and nivolumab in resectable non-small-cell lung cancer (NADIM): an open-label, multicentre, single-arm, phase 2 trial. *Lancet Oncol*. (2020) 21:1413–22. doi: 10.1016/S1470-2045(20)30453-8

10. Saw SPL, Ong B-H, Chua KLM, Takano A, Tan DSW. Revisiting neoadjuvant therapy in non-small-cell lung cancer. *Lancet Oncol*. (2021) 22:e501–16. doi: 10.1016/S1470-2045(21)00383-1

11. Hellmann MD, Chaft JE, William WN, Rusch V, Pisters KMW, Kalhor N, et al. Pathological response after neoadjuvant chemotherapy in resectable non-small-cell lung cancers: proposal for the use of major pathological response as a surrogate endpoint. *Lancet Oncol*. (2014) 15:e42–50. doi: 10.1016/S1470-2045(13)70334-6

12. Travis WD, Dacic S, Wistuba I, Sholl L, Adusumilli P, Bubendorf L, et al. IASLC MULTIDISCIPLINARY RECOMMENDATIONS FOR PATHOLOGIC ASSESSMENT OF LUNG CANCER RESECTION SPECIMENS FOLLOWING NEOADJUVANT THERAPY. *J Thorac Oncol*. (2020) 15:709–40. doi: 10.1016/j.jtho.2020.01.005

13. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036

14. Zhong Y, She Y, Deng J, Chen S, Wang T, Yang M, et al. Deep learning for prediction of N2 metastasis and survival for clinical stage I non-small cell lung cancer. *Radiology*. (2022) 302:200–11. doi: 10.1148/radiol.2021210902

15. Zhu Z, Chen M, Hu G, Pan Z, Han W, Tan W, et al. A pre-treatment CT-based weighted radiomic approach combined with clinical characteristics to predict durable clinical benefits of immunotherapy in advanced lung cancer. *Eur Radiol*. (2023) 33:3918–30. doi: 10.1007/s00330-022-09337-7

16. Beig N, Khorrami M, Alilou M, Prasanna P, Braman N, Orooji M, et al. Perinodular and intranodular radiomic features on lung CT images distinguish adenocarcinomas from granulomas. *Radiology*. (2019) 290:783–92. doi: 10.1148/radiol.2018180910

17. Khorrami M, Jain P, Bera K, Patil P, Ahmad U, Murthy S, et al. Predicting pathologic response to neoadjuvant chemoradiation in resectable stage III non-small cell lung cancer patients using computed tomography radiomic features. *Lung Cancer*. (2019) 135:1–9. doi: 10.1016/j.lungcan.2019.06.020

18. Mazzaschi G, Milanese G, Pagano P, Madeddu D, Gnetti L, Trentini F, et al. Integrated CT imaging and tissue immune features disclose a radio-immune signature with high prognostic impact on surgically resected NSCLC. *Lung Cancer*. (2020) 144:30–9. doi: 10.1016/j.lungcan.2020.04.006

19. Kim HR, Park HJ, Son J, Lee JG, Chung KY, Cho NH, et al. Tumor microenvironment dictates regulatory T cell phenotype: Upregulated immune checkpoints reinforce suppressive function. *J Immunother Cancer*. (2019) 7:339. doi: 10.1186/s40425-019-0785-8

20. Khorrami M, Prasanna P, Gupta A, Patil P, Velu PD, Thawani R, et al. Changes in CT radiomic features associated with lymphocyte distribution predict overall survival and response to immunotherapy in non-small cell lung cancer. *Cancer Immunol Res*. (2020) 8:108–19. doi: 10.1158/2326-6066.CIR-19-0476

21. Wu J, Xia Y, Wang X, Wei Y, Liu A, Innanje A, et al. uRP: An integrated research platform for one-stop analysis of medical images. *Front Radiol*. (2023) 3:1153784. doi: 10.3389/fradi.2023.1153784

22. Mohiuddin K, Haneuse S, Sofer T, Gill R, Jaklitsch MT, Colson YL, et al. Relationship between margin distance and local recurrence among patients undergoing wedge resection for small (≤2 cm) non-small cell lung cancer. *J Thorac Cardiovasc Surg*. (2014) 147:1169–75. doi: 10.1016/j.jtcvs.2013.11.056

23. Chong H-H, Yang L, Sheng R-F, Yu Y-L, Wu D-J, Rao S-X, et al. Multi-scale and multi-parametric radiomics of gadoxetate disodium-enhanced MRI predicts microvascular invasion and outcome in patients with solitary hepatocellular carcinoma ≤ 5 cm. *Eur Radiol*. (2021) 31:4824–38. doi: 10.1007/s00330-020-07601-2

24. Kim CG, Kim KH, Pyo K-H, Xin C-F, Hong MH, Ahn B-C, et al. Hyperprogressive disease during PD-1/PD-L1 blockade in patients with non-small-cell lung cancer. *Ann Oncol*. (2019) 30:1104–13. doi: 10.1093/annonc/mdz123

25. Ferrara R, Mezquita L, Texier M, Lahmar J, Audigier-Valette C, Tessonnier L, et al. Hyperprogressive disease in patients with advanced non-small cell lung cancer treated with PD-1/PD-L1 inhibitors or with single-agent chemotherapy. *JAMA Oncol*. (2018) 4:1543–52. doi: 10.1001/jamaoncol.2018.3676

26. Liu C, Zhao W, Xie J, Lin H, Hu X, Li C, et al. Development and validation of a radiomics-based nomogram for predicting a major pathological response to neoadjuvant immunochemotherapy for patients with potentially resectable non-small cell lung cancer. *Front Immunol*. (2023) 14:1115291. doi: 10.3389/fimmu.2023.1115291

27. Li F, Zhai S, Lv Z, Yuan L, Wang S, Jin D, et al. Effect of histology on the efficacy of immune checkpoint inhibitors in advanced non-small cell lung cancer: A systematic review and meta-analysis. *Front Oncol*. (2022) 12:968517. doi: 10.3389/fonc.2022.968517

28. Tian Y, Zhai X, Yan W, Zhu H, Yu J. Clinical outcomes of immune checkpoint blockades and the underlying immune escape mechanisms in squamous and adenocarcinoma NSCLC. *Cancer Med*. (2021) 10:3–14. doi: 10.1002/cam4.3590

29. She Y, He B, Wang F, Zhong Y, Wang T, Liu Z, et al. Deep learning for predicting major pathological response to neoadjuvant chemoimmunotherapy in non-small cell lung cancer: A multicentre study. *EBioMedicine*. (2022) 86:104364. doi: 10.1016/j.ebiom.2022.104364

30. Zhou F, Qiao M, Zhou C. The cutting-edge progress of immune-checkpoint blockade in lung cancer. *Cell Mol Immunol*. (2021) 18:279–93. doi: 10.1038/s41423-020-00577-5

31. Altorki NK, Markowitz GJ, Gao D, Port JL, Saxena A, Stiles B, et al. The lung microenvironment: an important regulator of tumour growth and metastasis. *Nat Rev Cancer*. (2019) 19:9–31. doi: 10.1038/s41568-018-0081-9

32. Parra ER, Behrens C, Rodriguez-Canales J, Lin H, Mino B, Blando J, et al. Image analysis-based assessment of PD-L1 and tumor-associated immune cells density supports distinct intratumoral microenvironment groups in non-small cell lung carcinoma patients. *Clin Cancer Res*. (2016) 22:6278–89. doi: 10.1158/1078-0432.CCR-15-2443

33. Reynders K, De Ruysscher D. Tumor infiltrating lymphocytes in lung cancer: a new prognostic parameter. *J Thorac Dis*. (2016) 8:E833–835. doi: 10.21037/jtd.2016.07.75

34. Hendry SA, Farnsworth RH, Solomon B, Achen MG, Stacker SA, Fox SB. The role of the tumor vasculature in the host immune response: implications for therapeutic strategies targeting the tumor microenvironment. *Front Immunol*. (2016) 7:621. doi: 10.3389/fimmu.2016.00621

35. Vaidya P, Bera K, Gupta A, et al. CT derived radiomic score for predicting the added benefit of adjuvant chemotherapy following surgery in Stage I, II resectable Non-Small Cell Lung Cancer: a retrospective multi-cohort study for outcome prediction. *Lancet Digit Health*. (2020) 2:e116–28. doi: 10.1016/s2589-7500(20)30002-9

36. Chang R, Qi S, Zuo Y, et al. Predicting chemotherapy response in non-small-cell lung cancer via computed tomography radiomic features: Peritumoral, intratumoral, or combined? *Front Oncol*. (2022) 12:915835. doi: 10.3389/fonc.2022.915835

37. Algohary A, Shiradkar R, Pahwa S, et al. Combination of peri-tumoral and intra-tumoral radiomic features on bi-parametric MRI accurately stratifies prostate cancer risk: A multi-site study. *Cancers (Basel)*. (2020) 12:2200. doi: 10.3390/cancers12082200

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 27 March 2024
DOI 10.3389/fimmu.2024.1357496

OPEN ACCESS

EDITED BY
Tinka Vidovic,
University of Zagreb, Croatia

REVIEWED BY
Cemil Çolak,
İnönü University, Türkiye
Dillip Kumar Bishi,
Rama Devi Women's University, India

*CORRESPONDENCE
Yonghong Zhang
zhangyh@ccmu.edu.cn
Ronghua Jin
ronghuajin@ccmu.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 18 December 2023
ACCEPTED 19 March 2024
PUBLISHED 27 March 2024

CITATION
Xiong Y, Qiao W, Wang Q, Li K, Jin R and Zhang Y (2024) Construction and validation of a machine learning-based nomogram to predict the prognosis of HBV associated hepatocellular carcinoma patients with high levels of hepatitis B surface antigen in primary local treatment: a multicenter study. *Front. Immunol.* 15:1357496.
doi: 10.3389/fimmu.2024.1357496

Construction and validation of a machine learning-based nomogram to predict the prognosis of HBV associated hepatocellular carcinoma patients with high levels of hepatitis B surface antigen in primary local treatment: a multicenter study

Yiqi Xiong[1†], Wenying Qiao[2,3†], Qi Wang[4†], Kang Li[5], Ronghua Jin[2,3*] and Yonghong Zhang[1*]

[1]Interventional Therapy Center for Oncology, Beijing You'an Hospital, Capital Medical University, Beijing, China, [2]Research Center for Biomedical Resources, Beijing You'an Hospital Capital Medical University, Beijing, China, [3]National Center for Infectious Diseases, Beijing Ditan Hospital, Capital Medical University, Beijing, China, [4]Interventional Radiology Department, Beijing Friendship Hospital, Capital Medical University, Beijing, China, [5]Research Center for Biomedical Resources, Beijing You'an Hospital, Capital Medical University, Beijing, China

Background: Hepatitis B surface antigen (HBsAg) clearance is associated with improved long-term outcomes and reduced risk of complications. The aim of our study was to identify the effects of levels of HBsAg in HCC patients undergoing TACE and sequential ablation. In addition, we created a nomogram to predict the prognosis of HCC patients with high levels of HBsAg (≥1000U/L) after local treatment.

Method: This study retrospectively evaluated 1008 HBV-HCC patients who underwent TACE combined with ablation at Beijing Youan Hospital and Beijing Ditan Hospital from January 2014 to December 2021, including 334 patients with low HBsAg levels and 674 patients with high HBsAg levels. The high HBsAg group was divided into the training cohort (N=385), internal validation cohort (N=168), and external validation cohort (N=121). The clinical and pathological features of patients were collected, and independent risk factors were identified using Lasso-Cox regression analysis for developing a nomogram. The performance of the nomogram was evaluated by C-index, receiver operating characteristic (ROC) curves, calibration curves, and decision curve analysis (DCA) curves in the training and validation cohorts. Patients were classified into high-risk and low-risk groups based on the risk scores of the nomogram.

Result: After PSM, mRFS was 28.4 months (22.1-34.7 months) and 21.9 months (18.5-25.4 months) in the low HBsAg level and high HBsAg level groups (P<0.001). The content of the nomogram includes age, BCLC stage, tumor size, globulin, GGT, and bile acids. The C-index (0.682, 0.666, and 0.740) and 1-, 3-, and 5-year AUCs of the training, internal validation, and external validation cohorts proved good discrimination of the nomogram. Calibration curves and DCA curves suggested accuracy and net clinical benefit rates. The nomogram enabled to classification of patients with high HBsAg levels into low-risk and high-risk groups according to the risk of recurrence. There was a statistically significant difference in RFS between the two groups in the training, internal validation, and external validation cohorts (P<0.001).

Conclusion: High levels of HBsAg were associated with tumor progression. The nomogram developed and validated in the study had good predictive ability for patients with high HBsAg levels.

KEYWORDS

hepatocellular carcinoma, hepatitis B surface antigen (HBsAg), TACE, ablation, nomogram, recurrence

1 Introduction

Primary liver cancer is the sixth most common cancer and the second leading cause of cancer death worldwide, which poses a huge economic and disease burden worldwide due to its high morbidity and mortality rates (1, 2). China is the country with the highest hepatocellular carcinoma (HCC) occurrence and the overall incidence of HCC is expected to continue to climb (3). HCC occurs most often in the setting of chronic liver inflammation and is mainly induced by hepatitis B virus (HBV) infection (4), which is a key risk factor for liver cirrhosis and HCC, capable of increasing the risk of HCC approximately 20-fold (5–7). For early HCC, surgical resection, liver transplantation, and ablation are recommended treatments. Studies have shown that ablation has similar five-survival rates compared to surgical treatment, and fewer complications than surgery (8, 9). However, the recurrence rate after ablation remains high, with a five-year recurrence rate of 50-70% (10). Transcatheter arterial chemoembolization (TACE) is the only guideline-recommended global standard of care for intermediate-stage HCC, and the median progression-free survival time (mPFS) is only 5 months (11). Therefore, diagnosis and treatment of HCC is an increasingly important public health problem.

The first serologic marker of HBV infection is Hepatitis B surface antigen (HBsAg), which can be detected from 2 to 12 weeks after infection with HBV (12). HBsAg clearance, which is currently regarded as the functional cure of chronic hepatitis (CHB), is associated with improved long-term outcomes and reduced risk of complications (13, 14). The decline in HBsAg during antiviral therapy is relatively slow, and the seroclearance rate is faster at low serum HBsAg expression (<1000U/L) (15, 16). Previous studies revealed that high serum levels of HBsAg increase the risk of developing HCC and have a worse prognosis for patients who have already developed HCC (17). Nevertheless, the prognostic impact of serum HBsAg levels in patients after TACE sequential ablation therapy needs to be further confirmed.

HBV-HCC prognosis is linked to several factors, including tumor burden, AFP, disease stage, ALBI, and NLR (18, 19), and there are also nomograms about HBV-HCC (20–22). However, no nomogram for HCC patients with high HBsAg expression after local treatment has been available to our knowledge. We compared the effects of high levels of HBsAg (≥1000U/L) and low levels of HBsAg (<1000U/L) in HCC patients undergoing TACE and sequential ablation and utilized propensity score matching to minimize selection bias. In addition, we created a nomogram to predict the prognosis of HCC patients with high levels of HBsAg after local treatment to more accurately guide the clinical decision.

2 Materials and methods

2.1 Patient selection

This study retrospectively evaluated 1008 HBV-HCC patients who underwent TACE combined with ablation at Beijing Youan Hospital and Beijing Ditan Hospital from January 2014 to December 2021. The diagnosis of HCC was based on the guideline of the America Association for the Study of Liver Diseases (ASSLD) (1, 23). The patients at Youan Hospital consisted of 553 patients with a high level of HBsAg and 334 patients with a low level of HBsAg. In order to build a reliable model, the patients from Youan Hospital were divided into the training cohort (N=385) and the validation cohort (N=168). Furthermore, 121 patients from Ditan Hospital were used as an independent external verification cohort to verify the external applicability of the nomogram. The inclusion criteria of patients were as follows (1): Aged 18-80 years (2). received TACE combined ablation (3). Child-Pugh classification was class A or B (4). all patients had not received any other therapeutics before ablation. Exclusion criteria were listed as follows (1): with second primary malignant tumors (2). clinical follow-up data incomplete (3). advanced HCC. (Figure 1).

The study was approved by the Medical Ethics Committee of Youan Hospital and Ditan Hospital and was performed in compliance with the standards of the Helsinki Declaration. The requirement for informed consent was waived because the study was deemed to pose no additional risk to patients and the data were deidentified.

2.2 Clinicopathologic characteristics

The demographic, clinical, and histopathologic data of patients were collected. Demographics included age, sex, drinking history, smoking history, hypertension and diabetes. Clinical and pathological data was composed of tumor size, tumor number, alpha-fetoprotein (AFP), aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma glutamyl transferase (GGT), albumin (ALB), neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and gamma glutamyl transferase to lymphocyte ratio (GLR).

2.3 Treatment received

2.3.1 TACE procedure

TACE was conducted by experienced interventional radiologists. Under local anesthesia, percutaneous right femoral artery puncture with a modified Seldinger technique was performed. Angiography was conducted by the 5-F (Terumo, Tokyo, Japan) catheter to identify arterial supply to tumors and to assess the patency of the portal vein. When applicable, a microcatheter was inserted into the blood-supply artery of the carcinoma to inject a mixture of doxorubicin (Pfizer Inc., New York, NY, USA) and lipiodol (Guerbet, Villepinte, France), followed by embolization using embolic materials, such as gelfoam or polyvinyl alcohol particles. The blood flow was monitored until complete vessel occlusion was observed. TACE was repeated thereafter if the lesion is not completely necrotic and the active portion exceeds 50% of the baseline value.

2.3.2 Ablation procedure

Performed under the guidance of computed tomography (CT) and magnetic resonance imaging (MRI) by a qualified interventionalist. The size of the tumor decided the number of electrodes. Routine disinfection and intravenous anesthesia were applied around the puncture points. During RFA, after measuring the baseline impedance, the power was gradually increased from 80w to 200w to reach the maximum impedance. The electrode tip temperature was kept below 20°C by the pump injected cold brine into the electrode chamber. Moreover, to achieve complete ablation, the safe margin for complete ablation of the tumor was 0.5cm. After ablation, the needle track was ablated to prevent postoperative bleeding and tumor implantation along the needle track. Arteriography-enhanced CT was performed immediately after treatment to evaluate the success of the procedure and its complications.

2.3 Follow-up

All patients underwent regular follow-ups at the outpatient clinics. Tumor responses were evaluated at approximately 4-6 weeks after ablation by using CT or MRI. For the follow-up protocol, patients were examined every 3 months during the first year and every 6 months thereafter. The contents of the follow-up included blood tests, liver function, and imaging examination to detect tumor recurrence. The study endpoint was recurrence-free

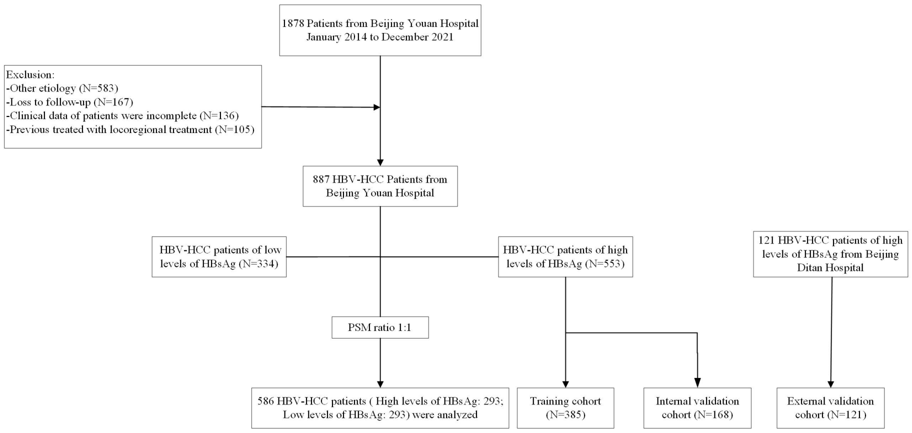


FIGURE 1
Screening flow chart of enrolled patients.

survival (RFS), defined as the time from ablation to the first recurrence.

2.4 Statistical analysis

Differences between the groups were compared through the t-test, chi-square test, Mann-Whitney U test, and Kruskal-Wallis test, with the purpose of providing median or counts and percentages to summarize baseline variables. Survival and recurrence were calculated using the Kaplan-Meier method, and the log-rank test was used for comparison. Lasso regression was performed for risk factor selection and identified independent risk factors for tumor recurrence were used in Multivariate Cox regression analysis. A nomogram based on independent risk factors to predict recurrence. Subsequently, the performance of the nomogram was validated in the internal validation and the external validation cohort. According to the nomogram scores, the patients were classified as low-risk and high-risk groups, and their recurrence rates were predicted. The receiver operating characteristic (ROC) curves were plotted and the area under the curves (AUCs) was calculated to evaluate prognostic value. Calibration curves and the Hosmer-Lemeshow test were conducted to assess the predictive ability of the nomogram. To estimate the clinical utility of the nomogram, decision curve analysis (DCA) was conducted by calculating the net benefits for a range of threshold probabilities.

To reduce the potential selection bias, 1:1 propensity score matching (PSM) was conducted, with a matching tolerance was 0.1. Matches were made in baseline variables that were previously considered clinically relevant in the literature, comprising age, sex, Child-pugh classification, BCLC stage, tumor size, tumor number, ALT, AST, and AFP.

All data were analyzed with SPSS (version 26.0, IBM, Armonk, NY, USA) and R software (version 4.1.3) in this study, and a P-value less than 0.05 was considered statistically significant (two-tailed tests).

3 Result

A total of 1008 HBV-HCC patients from Beijing Youan Hospital and Beijing Ditan Hospital were screened between January 1, 2014, to December 31, 2021, including 334 patients with low HBsAg levels and 674 patients with high HBsAg levels. After PSM, 293 patients were included in each group (Figure 1). The high levels of HBsAg groups were divided into the training cohort (N=385), internal validation cohort (N=168), and external validation cohort (N=121). The last follow-up until July 1, 2023, and the median follow-up time was 4.05 years (25~75th percentiles, 2.68~7.05 years).

Before PSM, baseline data showed that compared to the low HBsAg level group, the high HBsAg level group had a younger age (55.9 ± 9.03 VS. 58.3 ± 8.34, P<0.001), lower levels of TBIL (18.64 ± 9.42 VS. 20.84 ± 11.37, P=0.002), and shorter TT (15.77 ± 2.12VS. 15.95 ± 2.39, P=0.008). After PSM, all demographic and clinicopathologic data were well balanced between the two groups (Table 1).

The internal validation cohort and the external validation cohort had similar baseline characteristics to the training cohort. In the three cohorts, the majority of the patients were male (81.0% VS. 77.4%VS. 76.9%, p=0.466), and the average age was over 50 years(56.1 ± 9.10 VS. 56.6 ± 8.46 VS. 57.9 ± 8.57, P=0.466). Most patients were Child-Pugh A (76.9%VS. 80.4% VS., P=0.427), suggesting that the patients had good liver function. BCLC A had

TABLE 1 Demographics and clinical characteristics before and after PSM.

	Before PSM			After PSM		
	Low HBsAg level (N=334)	High HBsAg level (N=553)	P value	Low HBsAg level (N=293)	High HBsAg level (N=293)	P value
Age	58.3 ± 8.34	55.9 ± 9.03	<0.001	58.4 ± 8.40	58.6 ± 7.71	0.728
Sex			0.614			0.918
male	271 (81.1%)	439 (79.5%)		233 (79.5%)	235 (80.2%)	
female	63 (18.9%)	114 (20.5%)		60 (20.5%)	58 (19.8%)	
Diabetes			0.429			0.136
Yes	252 (75.4%)	479 (75.9%)		220 (75.1%)	236 (80.5%)	
No	82 (24.6%)	74 (24.1%)		73 (24.9%)	57 (19.5%)	
Child-Pugh class			0.244			1
A	246 (73.7%)	427 (77.3%)		216 (73.7%)	216 (73.7%)	
B	88 (26.3%)	126 (22.7%)		77 (26.3%)	77 (26.3%)	
Cirrhosis			0.989			0.718
Yes	43 (12.9%)	72 (13.1%)		38 (13.0%)	42 (14.3%)	
No	291 (87.1%)	481 (86.9%)		255 (87.0%)	251 (85.7%)	
BCLC stage			0.604			0.315
0	97 (29.0%)	174 (31.4%)		86 (29.4%)	85 (29.0%)	
A	187 (56.0%)	291 (52.6%)		168 (57.3%)	156 (53.2%)	
B	50 (15.0%)	88 (16.0%)		39 (13.3%)	52 (17.7%)	

(Continued)

TABLE 1 Continued

	Before PSM			After PSM		
	Low HBsAg level (N=334)	High HBsAg level (N=553)	P value	Low HBsAg level (N=293)	High HBsAg level (N=293)	P value
T.N			0.436			0.783
Single	237 (71.0%)	378 (68.3%)		213 (72.7%)	209 (71.3%)	
Multiple	97 (29.0%)	175 (31.7%)		80 (27.3%)	84 (28.7%)	
T.S			1			0.730
<30mm	215 (64.4%)	357 (64.5%)		192 (65.5%)	187 (63.8%)	
≥30mm	119 (35.6%)	196 (35.5%)		101 (34.5%)	106 (36.2%)	
WBC (10^9/L)	5.06±2.17	5.22±2.16	0.976	5.00 ± 2.10	5.22 ± 2.22	0.208
NLR	3.17±2.62	3.28±2.84	0.850	3.21 ± 2.67	3.43 ± 2.90	0.342
MLR	0.38±0.23	0.37±0.21	0.342	0.38 ± 0.234	0.39 ± 0.216	0.711
Hb (g/L)	128±20	131±19	0.073	129 ± 20.4	131 ± 19.4	0.268
PLR	111.61±61.43	108.78±54.90	0.283	113 ± 62.6	108 ± 51.7	0.359
TBIL (umol/L)	20.84±11.37	18.64±9.42	0.002	21.2 ± 11.6	19.4 ± 9.26	0.046
GGT (U/L)	67.48±64.43	68.15±58.89	0.617	65.6 ± 58.7	63.5 ± 53.7	0.651
GLR	68.06±90.12	69.12±84.74	0.928	67.5 ± 90.8	64.0 ± 60.7	0.582
Fib (g/L)	2.81±1.02	2.84±0.88	0.064	2.78 ± 0.99	2.88 ± 0.92	0.236
TT (s)	15.95±2.39	15.77±2.12	0.008	16.0 ± 2.38	15.8 ± 2.12	0.451
Alb (g/L)	37.19±4.78	37.21±4.88	0.914	37.3 ± 4.85	36.8 ± 4.94	0.201
Palb (g/L)	139.09±60.38	139.32±58.59	0.653	140 ± 60.7	131 ± 56.8	0.081
ALT (U/L)	29.28±16.95	32.55±19.96	0.111	29.4 ± 17.0	31.2 ± 19.6	0.228
AST (U/L)	31.51±14.09	32.27±15.66	0.247	31.6 ± 13.9	32.4 ± 15.4	0.515
AFP (umol/L)	382.96±1930.96	341.67±1871.85	0.495	337 ± 1800	357 ± 1730	0.895

ALD, alcoholic liver cancer; BCLC, Barcelona Clinic Liver Cancer; T.N, tumor number; T.S, tumor size; WBC, leukocyte; Hb, hemoglobin; NLR, neutrophil-to-lymphocyte ratio; MLR, monocyte-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TBIL, total bilirubin; ALB, albumin; GGT, gamma glutamyl transferase; GLR, gamma glutamyl transferase to lymphocyte ratio; Palb, prealbumin; Fib, fibrous protein; TT, thrombin time; AFP, alpha-fetoprotein.

the highest percentage of patients (51.7% VS. 55.4% VS. 71.7%, P=0.182). Regarding tumor characteristics, most tumors were solitary (70.6% vs.71.4% VS. 67.8%, P=0.780) and tumor size was less than 3cm (70.6% vs. 69.0% VS. 67.8%, P=0.150) (Table 2).

3.1 Efficacy

After PSM, mRFS was 28.4 months (22.1-34.7 months) and 21.9 months (18.5-25.4 months) in the high HBsAg level and low HBsAg level groups, respectively (Figure 2). Because mRFS were significantly shorter in the high HBsAg level (P<0.001), a nomogram for predicting recurrence needs to be developed for the high HBsAg group in order to prompt clinical interventions.

3.2 The prediction model was built based on the Lasso-Cox regression

3.2.1 Independent prognostic factors of RFS

The cohort in Beijing Youan Hospital was randomly split in a 7:3 ratio into the training (N=385) and internal validation (N=168) sets. The external validation cohort consisted of patients from Beijing Ditan Hospital. There were no statistical differences between the three groups (P<0.05), which showed that the data grouping was random and reasonable. Lasso regression was used to screen parameters, and the variation characteristics of the coefficient of these variables were shown in Figure 3A. The model exhibited outstanding performance and the least number of independent variables (Figure 3B). The screened variables included age, BCLC stage, tumor size, ALB, Palb, GLB, GGT, and bile acids. Variables screened based on Lasso regression were further subjected to multifactorial COX regression analysis to screen independent risk factors associated with recurrence (Table 3). The final results obtained were age (HR: 1.02, 95% CI: 1.01-1.04), BCLC stage (HR: 1.53, 95% CI: 1.22-1.91), tumor size (HR: 1.44, 95% CI: 1.06-1.94), globulin (HR: 1.02, 95% CI: 1-1.04), GGT (HR: 1.01, 95% CI: 1-1.01), and bile acids (HR: 1, 95% CI: 1-1.01).

3.2.2 Develop the nomogram

The independent predictors found by the Lasso-Cox regression analysis were used to construct a nomogram (Figure 4). In the training cohort, the C-index was 0.682(95%CI: 0.639-0.725), and the time-dependent ROC curve demonstrated that AUCs of 1-, 3-, and 5-year were 0.741, 0.723, and 0.687 (Figure 5). It indicated the good predicting ability of our nomogram. The calibration curves of 1-, 3-, and 5-year demonstrated satisfactory accordance between the nomogram prediction and actual observation. In addition, the clinical value of the nomogram was evaluated using DCA, which provided the net benefits in reasonable threshold probability (Figure 6).

Patients were classified into two groups according to the score of the nomogram: low-risk group and high-risk group. In the training cohort, there were apparent variances in RFS (Figure 7) between the low-risk group (N=193) and high-risk group (N=192) (P<0.001).

3.2.3 Validate the nomogram

To further test the efficacy of the reliability and robustness of our prognostic nomogram, internal and external validations were conducted on the nomogram. In the internal and external

TABLE 2 Demographics and clinical characteristics for training and validation sets.

	Training Cohort (N=385)	Internal Validation Cohort (N=168)	External Validation Cohort (N=121)	P-value
Age				
Mean ± SD	56.1 ± 9.10	56.6 ± 8.46	57.9 ± 8.57	0.150
Sex				0.466
Male	312 (81.0%)	130 (77.4%)	93 (76.9%)	
Female	73 (19.0%)	38 (22.6%)	28 (23.1%)	
Hypertension				0.617
No	292 (75.8%)	129 (76.8%)	97 (80.2%)	
Yes	93 (24.2%)	39 (23.2%)	24 (19.8%)	
Diabetes				0.559
No	310 (80.5%)	141 (83.9%)	101 (83.5%)	
Yes	75 (19.5%)	27 (16.1%)	20 (16.5%)	
Antiviral				0.137
No	158 (41.0%)	71 (42.3%)	62 (51.2%)	
Yes	227 (59.0%)	97 (57.7%)	59 (48.8%)	
Smoking				0.467
No	234 (60.8%)	94 (56.0%)	68 (56.2%)	
Yes	151 (39.2%)	74 (44.0%)	53 (43.8%)	
Cirrhosis				0.434
No	48 (12.5%)	27 (16.1%)	14 (11.6%)	
Yes	337 (87.5%)	141 (83.9%)	107 (88.4%)	
ChildPugh				0.182
A	296 (76.9%)	135 (80.4%)	86 (71.1%)	
B	89 (23.1%)	33 (19.6%)	35 (28.9%)	
BCLC				0.288
0	119 (30.9%)	56 (33.3%)	31 (25.6%)	
A	199 (51.7%)	93 (55.4%)	70 (57.9%)	
B	67 (17.4%)	19 (11.3%)	20 (16.5%)	
T.N				0.780
Single	272 (70.6%)	120 (71.4%)	82 (67.8%)	
Multiple	113 (29.4%)	48 (28.6%)	39 (32.2%)	
T.S				0.150
<3cm	234 (60.8%)	116 (69.0%)	73 (60.3%)	
≥3cm	151 (39.2%)	52 (31.0%)	39 (32.2%)	
WBC (10^9/L)				0.648
Mean ± SD	5.29 ± 2.22	5.15 ± 2.00	5.10 ± 2.17	
NLR				0.892
Age				
Mean ± SD	3.31 ± 2.93	3.19 ± 2.80	3.34 ± 3.10	
MLR				0.124
Mean ± SD	0.377 ± 0.216	0.355 ± 0.196	0.41 ± 0.24	
PLR				0.173
Mean ± SD	112 ± 58.8	102 ± 45.9	107 ± 56.2	
ALT (U/L)				0.101
Mean ± SD	33.1 ± 20.6	30.7 ± 17.7	29.3 ± 15.6	
AST (U/L)				0.180
Mean ± SD	33.0 ± 16.6	30.6 ± 13.0	33.5 ± 14.6	
TBIL (umol/L)				0.042
Mean ± SD	18.7 ± 9.59	18.6 ± 9.04	21.2 ± 12.2	
DBIL (umol/L)				0.777
Mean ± SD	6.41 ± 4.56	6.58 ± 4.40	6.19 ± 4.82	
Total.alb (g/L)				0.473
Mean ± SD	65.0 ± 8.52	64.6 ± 5.95	65.7 ± 6.74	
Alb (g/L)				0.887
Mean ± SD	37.1 ± 5.11	37.3 ± 4.56	37.0 ± 4.72	
Globulin (g/L)				0.041
Mean ± SD	28.4 ± 5.65	27.4 ± 4.99	28.9 ± 6.11	
GGT (umol/L)				0.484
Mean ± SD	69.8 ± 60.3	70.0 ± 67.6	77.7 ± 74.7	
GLR				0.362
Mean ± SD	69.8 ± 90.6	70.5 ± 77.9	83.6 ± 126	
Bile.acid				0.852
Mean ± SD	21.8 ± 30.2	20.5 ± 26.0	22.2 ± 23.9	
Fib (g/L)				0.348
Mean ± SD	2.85 ± 0.889	2.83 ± 0.896	2.71 ± 0.91	
AFP (umol/L)				0.707
Mean ± SD	412 ± 2240	266 ± 770	432 ± 2531	

BCLC, Barcelona Clinic Liver Cancer; ALD, alcoholic liver cancer; BCLC, Barcelona Clinic Liver Cancer; T.N, tumor number; T.S, tumor size; WBC, leukocyte; NLR, neutrophil-to-lymphocyte ratio; MLR, monocyte-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TBIL, total bilirubin; DBIL, direct bilirubin; ALB, albumin; GGT, gamma glutamyl transferase; GLR, gamma glutamyl transferase to lymphocyte ratio; Fib, fibrous protein; AFP, alpha-fetoprotein.

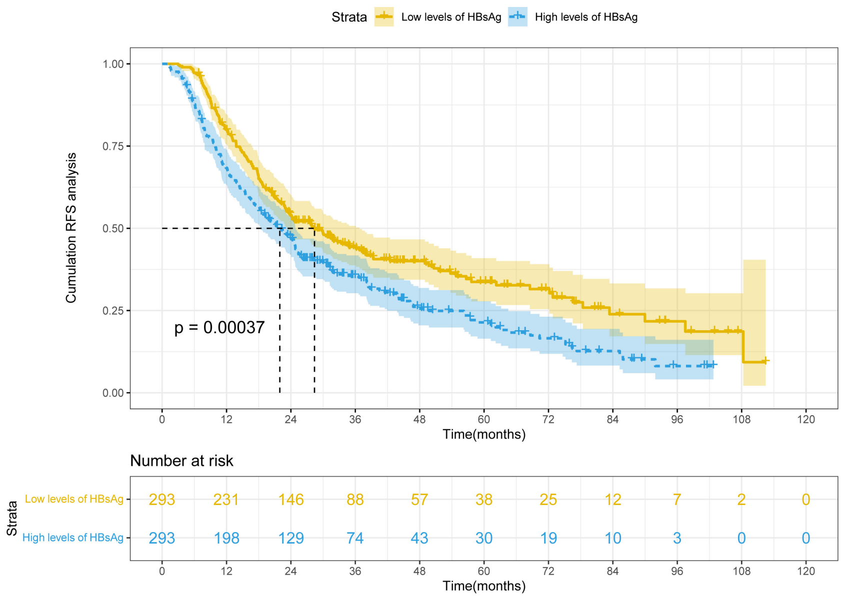


FIGURE 2
Kaplan-Meier plot of RFS for HBV-HCC after PSM.

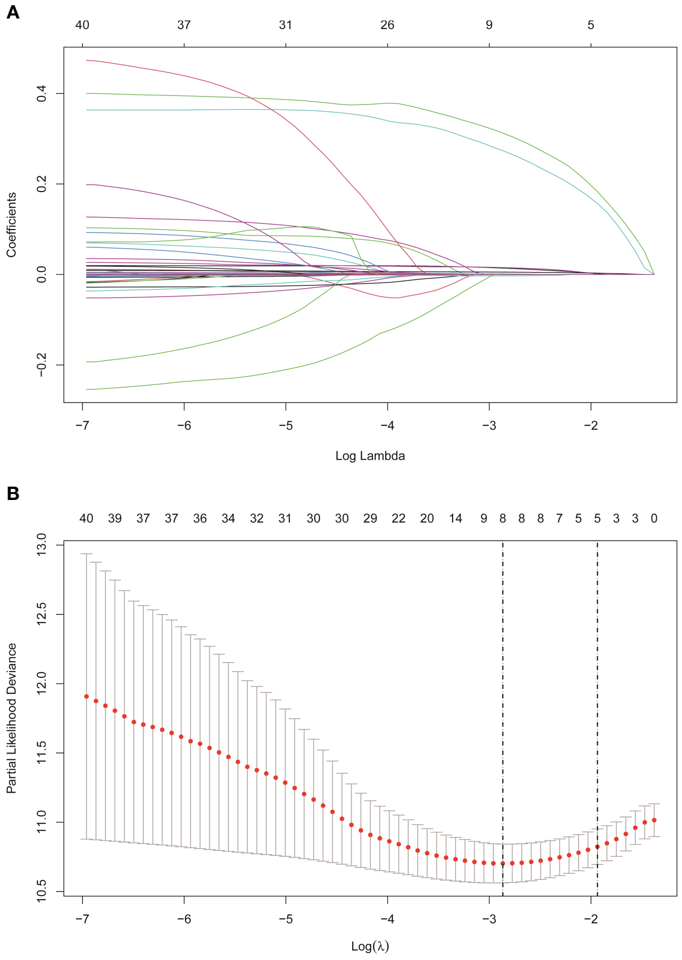


FIGURE 3
Screening of variables based on Lasso regression. **(A)** The variation characteristics of the coefficient of variables. **(B)** the selection process of the optimum value of the parameter λ in the Lasso regression model by cross-validation method.

TABLE 3 Cox proportional hazards regression to predict recurrence based on Lasso regression.

Variables	β	Z	HR (95%CI)	P value
Age	0.025	3.51	1.02 (1.01-1.04)	<0.001
BCLC	0.423	3.69	1.53 (1.22-1.91)	<0.001
T.S Alb Palb	0.364 -0.004 -0.001	2.37 -0.25 -0.895	1.44 (1.06-1.94) 0.99 (0.96-1.03) 0.99 (0.99-1)	0.018 0.804 0.371
Globulin GGT Bile.acid	0.022 0.006 0.005	2.02 6.5 2.68	1.02 (1-1.04) 1.01 (1-1.01) 1.00 (1-1.01)	0.043 <0.001 0.007

BCLC, Barcelona Clinic Liver Cancer; T.S, tumor size; Palb, prealbumin; GGT, gamma glutamyl transferase.

validation cohorts, the C-indexes of the nomogram for predicting the RFS were 0.666 (95%CI: 0.613-0.719) and 0.74 (95%CI: 0.696-0.783). The time-dependent ROC revealed that the AUCs of 1-, 3-, and 5-year were 0.702, 0.704, 0.684, 0.792, 0.734, and 0.770 in the internal and external validation cohorts (Supplementary Figure S1). The calibration curves also matched well (Supplementary Figure S2), and the DCA curves of 1-, 3-, and 5-year had good clinical practicability (Supplementary Figure S3).

The patients in two validation cohorts were also divided into high-risk and low-risk groups. The recurrence rates in the high-risk groups were significantly higher in the low-risk groups (P<0.001) (Supplementary Figure S4).

4 Discussion

HCC is one of the most common malignant tumors in the world. In China, the major etiology of the HCC is the HBV infection, which can promote the development and metastasis of the HCC (10, 24, 25). With the use of 1:1 PSM, our study found that the high level of HBsAg had a higher risk of recurrence than the low level of HBsAg. Consequently, our study is the first to focus on the high level of HBsAg patients who underwent TACE combined ablation to develop and validate a nomogram, which will hopefully predict the recurrence in H-HBsAg patients (High level of HBsAg). At present, there is a lack of a recurrence prediction model for H-HBsAg. We simultaneously created a nomogram by Lasso-Cox regression to accurately predict the prognosis of H-HBsAg patients.

The nomogram contains seven factors to produce the probability of an individual-specific clinical event, including age, tumor size, BCLC stage, globulin, GGT, and bile acid. The scores of the nomogram were obtained by drawing a vertical line at the location of the corresponding total score so that it intersected the three lines predicting the risk of recurrence, and the values shown at the intersection were predicted RFS at 1, 3, and 5 years. The C-index and AUCs of the training cohort and validation cohorts were similar, demonstrating adequate discrimination ability. The calibration curves presented the good prediction performance of the nomogram. Moreover, the nomogram indicated reliable clinical applicability by DCA curves. Patients were divided into two different risk groups according to the nomogram, and RFS was clearly different(P<0.001), which illustrated that our nomogram had a better ability to distinguish H-HBsAg patients to determine the risk of relapse after ablation therapy.

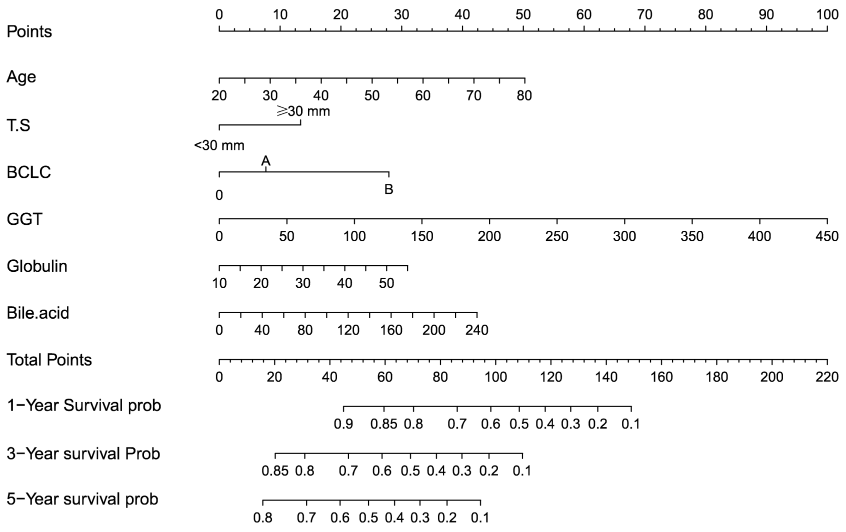


FIGURE 4
Nomogram, including Age, tumor number, BCLC stage, Globulin, GGT, and Bile acid for 1-, 3-, and 5- years recurrence free survival (RFS) in HCC patients with high HBsAg levels in AFP. The nomogram is valued to obtain the probability of 1-, 3-, and 5- years recurrence by adding up the points identified on the points scale for each variable. T.S, tumor size; BCLC, Barcelona Clinic Liver Cancer; GGT, gamma glutamyl transferase.

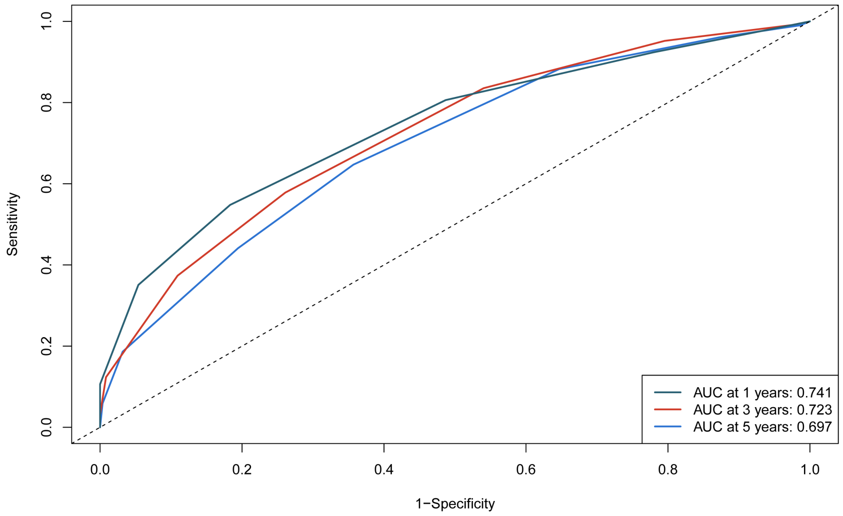


FIGURE 5
1-, 3-, and 5-year ROC curves of the nomogram in the training cohort.

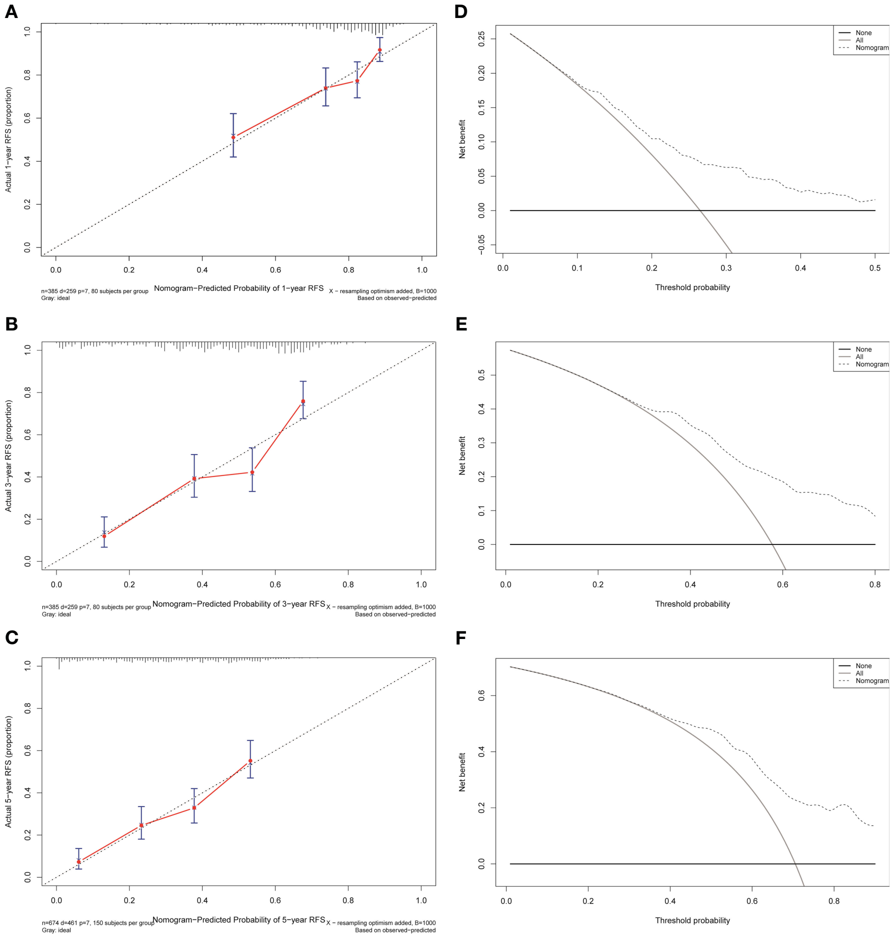


FIGURE 6
Calibration curves and decision curves analysis for recurrence of the nomogram in the training cohort. **(A)** One-year calibration curve in the training cohort. **(B)** Three-year calibration curve in the training cohort. **(C)** Five-year calibration curve in the training cohort. **(D)** One-year decision curve analysis in the training cohort. **(E)** Three-year decision curve analysis in the training cohort. **(F)** Five-year decision curve analysis in the training cohort.

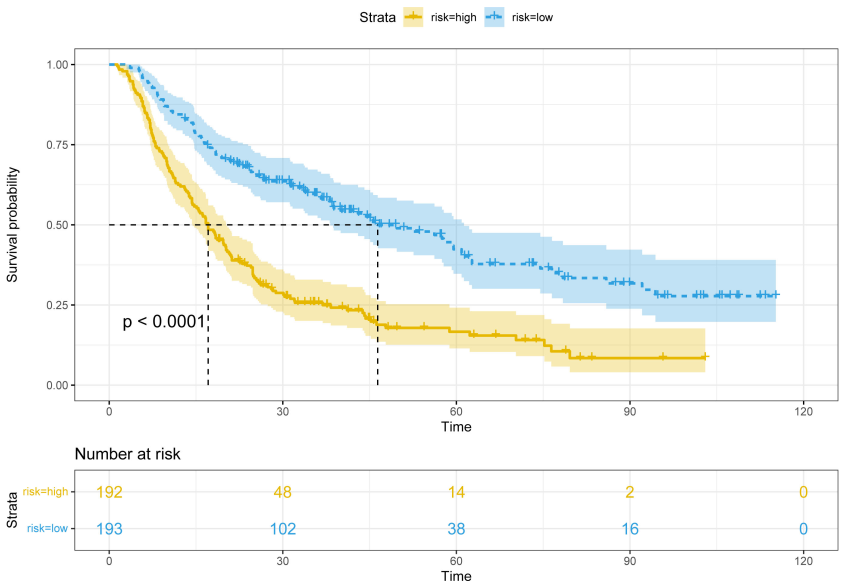


FIGURE 7
Kaplan-Meier plots of RFS for the low-risk group and high-risk group in the training cohort.

The number and size of tumors suggested strong tumor aggressiveness and poor prognosis of HCC, which was currently uncontroversial and needed not to be described here. Liver weight and portal blood flow velocity are reduced in the elderly, resulting in less reparability of the young body. Elderly people have lower immunity and faster tumor progression after treatment, leading to higher recurrence rates and worse prognosis (26, 27). At present, the BCLC system is regarded as an optimal staging system for tumor stage, treatment regimens, and expected survival. The expected survival rate is 50-70% for patients who are BCLC A at 5 years (28, 29). When we combined BCLC with other independent prognostic factors, the predictive value for prognosis could improve. GGT may be involved in the balance of oxidant and anti-oxidation, leading to sustained oxidative stress in tumor cells, which can contribute to the process of cancer (30, 31). Various proinflammatory proteins, including immunoglobulins, C-reactive protein, α2 macroglobulin, and fibrinogen are globulins (32, 33). Since human immunoglobulins are mainly metabolized by the liver, patients with severe hepatic dysfunction have a reduced ability to clear immunoglobulins, causing hyperglobulinemia (34, 35). Bile acid synthesis occurs in liver cells and is the end product of cholesterol metabolism (36). The Systemic homeostasis of bile acid mainly depends on its enterohepatic circulation process, which is of great significance for nutrient absorption and distribution, metabolic regulation, and homeostasis (37). Bile acid metabolism is implicated in tumor progression and hydrophobic bile acids are promoters of HCC (38, 39). Besides, reduced Farnesoid X (FXR) receptor signaling during hepatic inflammation induces to decrease in bile acid transporter proteins, resulting in elevated bile acids and persistent hepatic inflammation, which promote the development of HCC (40, 41).

The presence of HBsAg is a serologic marker of HBV infection and is used in clinical diagnosis (42, 43). HBsAg appears 1-2 weeks after exposure to HBV and precedes the onset of clinical symptoms and other serologic biochemical indicators of infection. There are still 257 million carriers of HBsAg despite the availability of antiviral therapeutics (44, 45). Many studies showed that the spontaneous HBsAg seroconversion rate was 1% and the presence of persistent HBsAg was associated with a high risk of HCC and a worse prognosis (46, 47). Previous studies by our team have also reported that the prognosis of HCC patients with negative HBsAg expression was better than that with positive HBsAg expression (48). In our study, we investigated the role of HBsAg levels in the recurrence of HCC after local treatment and used PSM to reduce bias. The results revealed that HBV-HCC patients with high HBsAg levels have worse prognosis than those with low HBsAg levels.

In the BCLC Guideline, TACE is recommended for BCLC intermediate stage B HCC. For early-stage HCC, TACE can mark the tumor and achieve tumor downstaging, thereby declining the time and increasing the success rate of ablation (49). Foreign and domestic studies have suggested that combination therapy by TACE and ablation improved overall and progression-free survival compared with TACE alone (50, 51). Unlike the conventional univariate analysis, the LASSO regression that we used aimed to select variables for Cox regression to avoid overfitting. Also, the nomogram can be validated by both internal and external validation because our study was a multicenter retrospective study. Simultaneous examination of comprehensive patient features covering demographics, liver function, tumor load, tumor markers, and inflammatory markers was a major strength of our study. The consists of our nomogram are simple and easy to obtain so that the clinicians are able to evaluate the patient's condition in a timely and effective manner.

Several limitations of our study should be addressed. The first one of them is the retrospective nature and it is necessary to strengthen the conclusions by further validations in large prospective studies. Because as a retrospective study, there is inevitable selection bias. Although internal and external validations were conducted by a larger multicenter sample, external validations from other centers are still required in the future. Besides, the patients included in our study all received TACE combined with ablation. Whether the nomogram would be suitable for other treatments such as surgery and liver transplantation requires further investigation. Lastly, the study was conducted only in China, where hepatitis B virus is the principal cause of HCC. Thus, generalizing to other populations in which HBV is not a major causative factor for HCC must be carried out with caution. Nevertheless, we used up to eight years of follow-up to create an accurate and reliable nomogram to better guide clinical practice for this group of HCC patients with high levels of HBsAg. In general, high-risk patients needed more frequent clinical surveillance and appropriate interventions to prevent recurrence and progression.

5 Conclusion

In summary, high levels of HBsAg were associated with tumor progression and poor prognosis. For high levels of HBsAg patients, we created an accurate and reliable nomogram to predict recurrence based on the Lasso-Cox regression analysis. The nomogram, including age, BCLC stage, tumor size, globulin, GGT, and bile acids, demonstrated adequate discrimination ability, which could better guide the clinical decisions.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

YX: Data curation, Writing – original draft, Writing – review & editing. WQ: Validation, Visualization, Writing – review & editing. QW: Data curation, Writing – review & editing. KL: Supervision, Writing – review & editing. RJ: Resources, Supervision, Writing – review & editing. YZ: Resources, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was funded by a grant Beijing Municipal Natural Science Foundation (7191004), Capital health development project (2020-1-2182 and 2020-2-1153), Beijing Key Laboratory (BZ0373).

Acknowledgments

The authors are highly grateful to all the patients who participated in this study and to our team from Beijing Youan Hospital and Beijing Ditan Hospital.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1357496/full#supplementary-material

SUPPLEMENTARY FIGURE 1
A. 1-, 3-, and 5-year ROC curves of the nomogram in the internal validation cohort. B. 1-, 3-, and 5-year ROC curves of the nomogram in the external validation cohort.

SUPPLEMENTARY FIGURE 2
Calibration curves for recurrence of the nomogram in the validation cohorts. **(A)** One-year calibration curve in the internal validation cohort. **(B)** Three-year calibration curve in the internal validation cohort. **(C)** Five-year calibration curve in the internal validation cohort. **(D)** One-year calibration curve in the external validation cohort. **(E)** Three-year calibration curve in the external validation cohort. **(F)** Five-year calibration curve in the external validation cohort.

SUPPLEMENTARY FIGURE 3
Decision curves analysis for recurrence in the internal validation cohort. **(A)** Decision curve analysis for one-year RFS in the internal validation cohort. **(B)** Decision curve analysis for three-year RFS in the internal validation cohort. **(C)** Decision curve analysis for five- year RFS in the internal validation cohort. **(D)** Decision curve analysis for one-year RFS in the external validation cohort. **(E)** Decision curve analysis for three-year RFS in the external validation cohort. **(F)** Decision curve analysis for five- year RFS in the external validation cohort.

SUPPLEMENTARY FIGURE 4
(A) Kaplan-Meier plots of RFS for the low-risk group and high-risk group in the internal validation cohort. **(B)** Kaplan-Meier plots of RFS for the low-risk group and high-risk group in the external validation cohort.

References

1. Vogel A, Meyer T, Sapisochin G, Salem R, Saborowski A. Hepatocellular carcinoma. *Lancet*. (2022) 400:1345–62. doi: 10.1016/s0140-6736(22)01200-4

2. 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 Cancer J Clin*. (2021) 71:209–49. doi: 10.3322/caac.21660

3. Zeng H, Chen W, Zheng R, Zhang S, Ji JS, Zou X, et al. Changing cancer survival in China during 2003-15: A pooled analysis of 17 population-based cancer registries. *Lancet Glob Health*. (2018) 6:e555–e67. doi: 10.1016/s2214-109x(18)30127-x

4. Asafo-Agyei KO, Samant H. Hepatocellular carcinoma. In: *Statpearls*. StatPearls Publishing LLC, Treasure Island (FL (2023). StatPearls Publishing Copyright © 2023.

5. Iannacone M, Guidotti LG. Immunobiology and pathogenesis of hepatitis B virus infection. *Nat Rev Immunol*. (2022) 22:19–32. doi: 10.1038/s41577-021-00549-4

6. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology*. (2007) 132:2557–76. doi: 10.1053/j.gastro.2007.04.061

7. Xue M, Lin X, Lin QX, Pu X, Liu J, Li XF, et al. Association between hepatitis B and E virus infection and hepatocellular carcinoma risk. *Int J Cancer*. (2021) 148:2974–81. doi: 10.1002/ijc.33505

8. Bai XM, Cui M, Yang W, Wang H, Wang S, Zhang ZY, et al. The 10-year survival analysis of radiofrequency ablation for solitary hepatocellular carcinoma 5 cm or smaller: primary versus recurrent hcc. *Radiology*. (2021) 300:458–69. doi: 10.1148/radiol.2021200153

9. Kang TW, Kim JM, Rhim H, Lee MW, Kim YS, Lim HK, et al. Small hepatocellular carcinoma: radiofrequency ablation versus nonanatomic resection–propensity score analyses of long-term outcomes. *Radiology*. (2015) 275:908–19. doi: 10.1148/radiol.15141483

10. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, et al. Hepatocellular carcinoma. *Nat Rev Dis Primers*. (2021) 7:6. doi: 10.1038/s41572-020-00240-3

11. Reig M, Forner A, Rimola J, Ferrer-Fàbrega J, Burrel M, Garcia-Criado Á, et al. Bclc strategy for prognosis prediction and treatment recommendation: the 2022 update. *J Hepatol*. (2022) 76:681–93. doi: 10.1016/j.jhep.2021.11.018

12. Broquetas T, Carrión JA. Past, present, and future of long-term treatment for hepatitis B virus. *World J Gastroenterol*. (2023) 29:3964–83. doi: 10.3748/wjg.v29.i25.3964

13. Terrault NA, Bzowej NH, Chang KM, Hwang JP, Jonas MM, Murad MH. Aasld guidelines for treatment of chronic hepatitis B. *Hepatology*. (2016) 63:261–83. doi: 10.1002/hep.28156

14. Easl 2017 clinical practice guidelines on the management of hepatitis B virus infection. *J Hepatol*. (2017) 67:370–98. doi: 10.1016/j.jhep.2017.03.021

15. García-López M, Lens S, Pallett LJ, Testoni B, Rodríguez-Tajes S, Mariño Z, et al. Viral and immune factors associated with successful treatment withdrawal in hbeag-negative chronic hepatitis B patients. *J Hepatol*. (2021) 74:1064–74. doi: 10.1016/j.jhep.2020.11.043

16. Sonneveld MJ, Chiu SM, Park JY, Brakenhoff SM, Kaewdech A, Seto WK, et al. Probability of hbsag loss after nucleo(S)Tide analogue withdrawal depends on hbv genotype and viral antigen levels. *J Hepatol*. (2022) 76:1042–50. doi: 10.1016/j.jhep.2022.01.007

17. Anderson RT, Choi HSJ, Lenz O, Peters MG, Janssen HLA, Mishra P, et al. Association between seroclearance of hepatitis B surface antigen and long-term clinical outcomes of patients with chronic hepatitis B virus infection: systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. (2021) 19:463–72. doi: 10.1016/j.cgh.2020.05.041

18. Liang X, Liangliang X, Peng W, Tao Y, Jinfu Z, Ming Z, et al. Combined prognostic nutritional index and albumin-bilirubin grade to predict the postoperative prognosis of hbv-associated hepatocellular carcinoma patients. *Sci Rep*. (2021) 11:14624. doi: 10.1038/s41598-021-94035-5

19. Zheng X, Ye B, Gou Y, Li Z, Chen C, Liao F, et al. Neutrophil to lymphocyte and platelet to lymphocyte ratios as biomarkers to predict relapse and survival in posthepatectomy hbv-related hepatocellular carcinoma: A meta-analysis and preliminary immune perspective. *Transl Cancer Res*. (2021) 10:1261–72. doi: 10.21037/tcr-20-3125

20. Su K, Shen Q, Tong J, Gu T, Xu K, Li H, et al. Construction and validation of a nomogram for hbv-related hepatocellular carcinoma: A large, multicenter study. *Ann Hepatol*. (2023) 28:101109. doi: 10.1016/j.aohep.2023.101109

21. Wang Q, Qiao W, Zhang H, Liu B, Li J, Zang C, et al. Nomogram established on account of lasso-cox regression for predicting recurrence in patients with early-stage hepatocellular carcinoma. *Front Immunol*. (2022) 13:1019638. doi: 10.3389/fimmu.2022.1019638

22. Wang Q, Guo D, Gao W, Yuan C, Li J, Zhang Y, et al. Individual surveillance by competing risk model for patients with hepatocellular carcinoma occurrence in all-cause cirrhosis. *J Cancer Res Clin Oncol*. (2023) 149:13403–16. doi: 10.1007/s00432-023-04911-y

23. Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LR, et al. Aasld guidelines for the treatment of hepatocellular carcinoma. *Hepatology*. (2018) 67:358–80. doi: 10.1002/hep.29086

24. Qiu H, Cao S, Xu R. Cancer incidence, mortality, and burden in China: A time-trend analysis and comparison with the United States and United Kingdom based on the global epidemiological data released in 2020. *Cancer Commun (Lond)*. (2021) 41:1037–48. doi: 10.1002/cac2.12197

25. Lu L, Mullins CS, Schafmayer C, Zeißig S, Linnebacher M. A global assessment of recent trends in gastrointestinal cancer and lifestyle-associated risk factors. *Cancer Commun (Lond)*. (2021) 41:1137–51. doi: 10.1002/cac2.12220

26. Schmucker DL. Aging and the liver: an update. *J Gerontol A Biol Sci Med Sci*. (1998) 53:B315–20. doi: 10.1093/gerona/53a.5.b315

27. Zoli M, Iervese T, Abbati S, Bianchi GP, Marchesini G, Pisi E. Portal blood velocity and flow in aging man. *Gerontology*. (1989) 35:61–5. doi: 10.1159/000213000

28. Tsilimigras DI, Bagante F, Sahara K, Moris D, Hyer JM, Wu L, et al. Prognosis after Resection of Barcelona Clinic Liver Cancer (Bclc) Stage 0, a, and B Hepatocellular Carcinoma: A Comprehensive Assessment of the Current Bclc Classification. *Ann Surg Oncol*. (2019) 26:3693–700. doi: 10.1245/s10434-019-07580-9

29. Wang J, Wang K, Chen C, Xiong Y, Guo C, Wang C, et al. Survival analysis and development of a prognostic nomogram for patients with hepatitis B virus-associated hepatocellular carcinoma. *Heliyon*. (2023) 9:e20850. doi: 10.1016/j.heliyon.2023.e20850

30. Yang JG, He XF, Huang B, Zhang HA, He YK. Rule of changes in serum ggt levels and ggt/alt and ast/alt ratios in primary hepatic carcinoma patients with different afp levels. *Cancer biomark*. (2018) 21:743–6. doi: 10.3233/cbm-170088

31. Zhang LX, Lv Y, Xu AM, Wang HZ. The prognostic significance of serum gamma-glutamyltransferase levels and ast/alt in primary hepatic carcinoma. *BMC Cancer*. (2019) 19:841. doi: 10.1186/s12885-019-6011-8

32. Deng Y, Pang Q, Miao RC, Chen W, Zhou YY, Bi JB, et al. Prognostic significance of pretreatment albumin/globulin ratio in patients with hepatocellular carcinoma. *Onco Targets Ther*. (2016) 9:5317–28. doi: 10.2147/ott.S109736

33. Xing Y, Guo ZN, Yan S, Jin H, Wang S, Yang Y. Increased globulin and its association with hemorrhagic transformation in patients receiving intra-arterial thrombolysis therapy. *Neurosci Bull*. (2014) 30:469–76. doi: 10.1007/s12264-013-1440-x

34. Doi H, Hayashi E, Arai J, Tojo M, Morikawa K, Eguchi J, et al. Enhanced B-cell differentiation driven by advanced cirrhosis resulting in hyperglobulinemia. *J Gastroenterol Hepatol*. (2018) 33(9):1667–76. doi: 10.1111/jgh.14123

35. Li J, Li Z, Hao S, Wang J, Chen W, Dai S, et al. Inversed albumin-to-globulin ratio and underlying liver disease severity as a prognostic factor for survival in hepatocellular carcinoma patients undergoing transarterial chemoembolization. *Diagn Interv Radiol*. (2023) 29:520–8. doi: 10.5152/dir.2022.211166

36. Boyer JL. Bile formation and secretion. *Compr Physiol*. (2013) 3:1035–78. doi: 10.1002/cphy.c120027

37. Wahlström A, Sayin SI, Marschall HU, Bäckhed F. Intestinal crosstalk between bile acids and microbiota and its impact on host metabolism. *Cell Metab*. (2016) 24:41–50. doi: 10.1016/j.cmet.2016.05.005

38. Režen T, Rozman D, Kovács T, Kovács P, Sipos A, Bai P, et al. The role of bile acids in carcinogenesis. *Cell Mol Life Sci*. (2022) 79:243. doi: 10.1007/s00018-022-04278-2

39. Nguyen PT, Kanno K, Pham QT, Kikuchi Y, Kakimoto M, Kobayashi T, et al. Senescent hepatic stellate cells caused by deoxycholic acid modulates Malignant behavior of hepatocellular carcinoma. *J Cancer Res Clin Oncol*. (2020) 146:3255–68. doi: 10.1007/s00432-020-03374-9

40. Fuchs CD, Trauner M. Role of bile acids and their receptors in gastrointestinal and hepatic pathophysiology. *Nat Rev Gastroenterol Hepatol*. (2022) 19:432–50. doi: 10.1038/s41575-021-00566-7

41. Shi Q, Yuan X, Xue C, Gu X, Li L. Establishment and validation of a novel risk score for hepatocellular carcinoma based on bile acid and bile salt metabolism-related genes. *Int J Mol Sci*. (2023) 24(10):8579. doi: 10.3390/ijms24108597

42. Terrault NA, Lok ASF, McMahon BJ, Chang KM, Hwang JP, Jonas MM, et al. Update on prevention, diagnosis, and treatment of chronic hepatitis B: aasld 2018 hepatitis B guidance. *Hepatology*. (2018) 67:1560–99. doi: 10.1002/hep.29800

43. Liaw YF. Clinical utility of hepatitis B surface antigen quantitation in patients with chronic hepatitis B: A review. *Hepatology*. (2011) 53:2121–9. doi: 10.1002/hep.24364

44. Schweitzer A, Horn J, Mikolajczyk RT, Krause G, Ott JJ. Estimations of worldwide prevalence of chronic hepatitis B virus infection: A systematic review of data published between 1965 and 2013. *Lancet*. (2015) 386:1546–55. doi: 10.1016/s0140-6736(15)61412-x

45. Yuen MF, Chen DS, Dusheiko GM, Janssen HLA, Lau DTY, Locarnini SA, et al. Hepatitis B virus infection. *Nat Rev Dis Primers*. (2018) 4:18035. doi: 10.1038/nrdp.2018.35

46. Trépo C, Chan HL, Lok A. Hepatitis B virus infection. *Lancet*. (2014) 384:2053–63. doi: 10.1016/s0140-6736(14)60220-8

47. Moini M, Fung S. Hbsag loss as a treatment endpoint for chronic hbv infection: hbv cure. *Viruses*. (2022) 14(4):657. doi: 10.3390/v14040657

48. Liu B, Wang Q, Mei T, Zheng J, Gao W, Yuan C, et al. Effect of hbsag expression in liver tissue on prognosis of hepatocellular carcinoma after minimally invasive interventional therapy. *Front Oncol.* (2023) 13:1106333. doi: 10.3389/fonc.2023.1106333

49. Wang Q, Qiao W, Liu B, Li J, Yuan C, Long J, et al. The Monocyte to Lymphocyte Ratio Not Only at Baseline but Also at Relapse Predicts Poor Outcomes in Patients with Hepatocellular Carcinoma Receiving Locoregional Therapy. *BMC Gastroenterol.* (2022) 22:98. doi: 10.1186/s12876-022-02180-6

50. Chen S, Zeng X, Su T, Xiao H, Lin M, Peng Z, et al. Combinatory local ablation and immunotherapies for hepatocellular carcinoma: rationale, efficacy, and perspective. *Front Immunol.* (2022) 13:1033000. doi: 10.3389/fimmu.2022.1033000

51. Sun Y, Xiong Y, Wang Q, Qiao W, Zhang H, Zhang Y. Development and validation of a nomogram to predict the recurrence of hepatocellular carcinoma patients with dynamic changes in afp undergoing locoregional treatments. *Front Oncol.* (2023) 13:1206345. doi: 10.3389/fonc.2023.1206345

TYPE Original Research
PUBLISHED 21 May 2024
DOI 10.3389/fmed.2024.1372495

EDITED BY
Tinka Vidovic,
University of Zagreb, Croatia

REVIEWED BY
JuanJuan Yin,
National Cancer Institute (NIH), United States
Chen Hao Lo,
Moffitt Cancer Center and Research Institute,
United States

*CORRESPONDENCE
Dagang Tang
✉ tdg518718@163.com
Zefang Li
✉ zefangli@foxmail.com
Weiyang He
✉ weiyang361@163.com

†These authors have contributed equally to this work

RECEIVED 18 January 2024
ACCEPTED 29 April 2024
PUBLISHED 21 May 2024

CITATION
Li G, Zhao R, Xie Z, Qu X, Duan Y, Zhu Y, Liang H, Tang D, Li Z and He W (2024) Mining bone metastasis related key genes of prostate cancer from the STING pathway based on machine learning.
Front. Med. 11:1372495.
doi: 10.3389/fmed.2024.1372495

Mining bone metastasis related key genes of prostate cancer from the STING pathway based on machine learning

Guiqiang Li[1,2†], Runhan Zhao[3†], Zhou Xie[3†], Xiao Qu[3], Yingtao Duan[3], Yafei Zhu[3], Hao Liang[3], Dagang Tang[3,4*], Zefang Li[3,5*] and Weiyang He[1*]

[1]Department of Urology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China, [2]Department of Urology, Chongqing Traditional Chinese Medicine Hospital, Chongqing, China, [3]Department of Orthopedics, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China, [4]Department of Orthopedics, Chongqing Traditional Chinese Medicine Hospital, Chongqing, China, [5]Department of Orthopedics, Qianjiang Hospital Affiliated with Chongqing University, Chongqing, China

Background: Prostate cancer (PCa) is the second most prevalent malignant tumor in male, and bone metastasis occurs in about 70% of patients with advanced disease. The STING pathway, an innate immune signaling mechanism, has been shown to play a key role in tumorigenesis, metastasis, and cancerous bone pain. Hence, exploring regulatory mechanism of STING in PCa bone metastasis will bring novel opportunities for treating PCa bone metastasis.

Methods: First, key genes were screened from STING-related genes (SRGs) based on random forest algorithm and their predictive performance was evaluated. Subsequently, a comprehensive analysis of key genes was performed to explore their roles in prostate carcinogenesis, metastasis and tumor immunity. Next, cellular experiments were performed to verify the role of RELA in proliferation and migration in PCa cells, meanwhile, based on immunohistochemistry, we verified the difference of RELA expression between PCa primary foci and bone metastasis. Finally, based on the key genes to construct an accurate and reliable nomogram, and mined targeting drugs of key genes.

Results: In this study, three key genes for bone metastasis were mined from SRGs based on the random forest algorithm. Evaluation analysis showed that the key genes had excellent prediction performance, and it also showed that the key genes played a key role in carcinogenesis, metastasis and tumor immunity in PCa by comprehensive analysis. In addition, cellular experiments and immunohistochemistry confirmed that overexpression of RELA significantly inhibited the proliferation and migration of PCa cells, and RELA was significantly low-expression in bone metastasis. Finally, the constructed nomogram showed excellent predictive performance in Receiver Operating Characteristic (ROC, AUC = 0.99) curve, calibration curve, and Decision Curve Analysis (DCA) curve; and the targeted drugs showed good molecular docking effects.

Conclusion: In sum, this study not only provides a new theoretical basis for the mechanism of PCa bone metastasis, but also provides novel therapeutic targets and novel diagnostic tools for advanced PCa treatment.

KEYWORDS

prostate cancer, bone metastasis, random forest, STING, nomogram

1 Introduction

Prostate cancer (PCa) is the second most prevalent malignant tumor in the male population, with about 20% of male individuals developing the disease at some point. As a highly aggressive tumor, most advanced PCa patients are diagnosed with multiple metastasis throughout the body, mainly in lymph nodes near the prostate, distal lymph nodes, bone, as well as internal organs such as the liver, lungs, and brain (1, 2). Among the distal metastasis of PCa, bone is most common site of colonization, and approximately 70% of advanced PCa patients are diagnosed with bone metastasis (3). Once bone metastasis occurs, the disease is incurable and is significantly associated with mortality (4–6). Tumor growth in the bone can cause pain, hypercalcemia, anemia, fracture, and other adverse events, all of which severely impact the patient's survival status and quality of life (7, 8). The tumor bone pain usually presents as a persistent dull ache that increases in intensity over time and reaches a level not relieved by opioids (9, 10). Neurological dysfunction, pain, anxiety and depression due to bone metastasis are devastating for patients, severely affecting their quality of life and significantly increasing mortality (7, 9, 11–13). With the continuous iterative updating of treatment regimens, the median survival of PCa patients has been significantly prolonged, leading to an elevated incidence of bone metastasis, making this phenomenon even more clinically relevant (14). However, the current detection of metastasis is minimal, and it is estimated that only about 0.02% of cancer cells entering the blood circulation will produce clinically detectable metastasis (15). Once metastasis occurs, it is responsible for approximately 90% of all deaths (16). Therefore, there is an urgent need to increase our understanding of the cellular and molecular mechanisms involved in PCa bone metastasis to improve the prognosis of patients with bone metastasis.

As studies have progressed in recent years, researchers have found a strong link between tumor immunity and bone metastasis. Due to the strong similarities with inflammation, cancer has long been described as a wound that cannot be healed (17, 18). During routine wound healing, the body terminates the immune response in a timely manner through various immune regulatory mechanisms (19), whereas in tumors, the uncontrolled inflammatory response becomes a powerful driver of tumorigenesis (20–22). The STING pathway is a vital transduction mechanism in innate immunity and viral defense, and it also plays a crucial role in carcinogenesis and development (23). Substantial evidence that STING activators (DMXAA and ADU-S100) can inhibit tumor progression and increase survival in an adaptive immune cell-dependent manner (24–27). Studies have shown an essential link between chromosomal instability and tumor metastasis and that cell membrane dsDNA produced by persistent chromosome segregation errors is sensed by the STING pathway (28, 29). In brain metastasis from breast cancer, tumor cells communicate with adjacent astrocytes by producing cAMP signals that activate the STING signaling pathway to release inflammatory factors, leading to tumor and metastasis progression (30). In addition, in prostate, breast, and lung cancers, STING signaling can also promote or inhibit the onset and progression of bone metastasis by modulating immune cells (31–35). Based on synergistic effects on injury receptors, immune cells, and osteoclasts, the STING pathway is also significant in regulating cancerous bone pain (36). Since the STING pathway plays a critical role in bone metastasis, exploring its regulatory mechanism in PCa bone metastasis will bring novel opportunities for treating PCa bone metastasis.

In this study, we mined three key genes related to PCa bone metastasis from STING-related genes (SRGs) based on the random forest machine learning algorithm, constructed an accurate nomogram, and discovered several targeted drugs for key genes. These findings provide novel ideas to improve treatment strategies for patients with advanced PCa.

2 Methods

2.1 Cell culture and transfection

The human PCa cell line (DU145 and PC3) was purchased from the America Type Culture Collection (ATCC, United States). Meanwhile, the cells were inoculated in 1640 medium (Saimikebio; China) containing 10% FBS (ExCell Biology, Inc., Shanghai, China) and 1% penicillin–streptomycin (100 IU/mL; Hyclone; Cytiva), and then placed in certain environments (37°C and 5% CO2) to culture. The RELA-overexpression plasmid was purchased from YouBio Biotechnology Co., Ltd. (Hunan, China). Transfection was performed using lipofectamine 3000 (Invitrogen) according to the manufacturer's instructions.

2.2 Cell proliferation assay

Cell proliferation was measured using the CCK-8 assay. Inoculate DU145 and PC3 cells transfected with RELA-overexpression plasmid and transfected with GFP control plasmid in 96-well plates (5,000 cells/well). After a period of incubation, 10ul of CCK-8 solution was added to each well. Optical density (OD) value was evaluated using a microplate reader at 450 nm 2 h later.

2.3 Cell migration assay

Cell migration was assessed by transwell assay. 2.5×10^4 DU145 and PC3 cells transfected with RELA-overexpression plasmid and transfected with GFP control plasmid were inoculated into the upper chamber of the transwell. After 24 h of incubation in the incubator, use a moistened cotton swab to carefully wipe off the cells that did not pass through the holes, and then add crystal violet to stain the chamber and take photographs.

Cell migration was also measured using wound healing assay. DU145 and PC3 cells transfected with RELA-overexpression plasmid and transfected with GFP control plasmid were inoculated in 6-well culture plates. After the cell fusion rate reached 90%, the cell layer was scratched using the tip of a sterile lance tip. Next, after washing with PBS, the culture was continued with serum-free 1,640 medium, and the fixed sites were photographed using a light microscope at 0, 12, and 24 h.

2.4 Immunohistochemistry

Paraffin sections of PCa primary foci and bone metastasis were obtained from the Department of Pathology with the approval of the Ethics Committee of our Medical Center. First, the paraffin sections were dewaxed in an environmentally friendly dewaxing solution and hydrated in a gradient ethanol series, and heat-mediated antigenic repair was performed on them in a microwave oven using the citric acid antigen repair solution; next, the sections were incubated with an endogenous peroxidase blocking agent (3% H_2O_2) for 25 min at room temperature and protected from light; subsequently, circles were drawn with a immunohistochemical pen, and section sealing was operated with 3% BSA; Finally, immunohistochemical staining was performed according to standard procedures.

As for the immunohistochemical staining operation. First, paraffin sections were incubated with RELA primary antibody (Servicebio, GB11997, 1:1,000) overnight at 4°C; next, sections were incubated with secondary antibody (goat anti-rabbit IgG) for 50 min at room temperature; subsequently, sections were stained sequentially using DAB and hematoxylin; finally, the stained sections were dehydrated in ethanol and xylene and sealed with neutral resin. Immunohistochemical images were obtained based on a light microscope (E100, Nikon, Japan).

2.5 Data collection and preprocessing

GEO[1] is a free and publicly available gene expression database containing many diseases. We downloaded the gene expression dataset (GSE32269) containing PCa primary foci and bone metastasis from this database. Based on the annotation files of the corresponding platforms, we matched probes to their gene symbols on the dataset. Those with larger mean values were selected for retention when duplicate probes existed. Subsequently, based on a clustering algorithm, outlier samples were removed before subsequent analysis. Finally, 103 SRGs were obtained from the GeneCards database.[2]

2.6 Identification and evaluation of key genes

The SRGs were initially downscaled by a univariate logistic regression algorithm, and the genes that satisfied the p-value < 0.05 were considered candidate genes. Next, the candidate genes were assigned MDA and MDG values based on the random forest algorithm. MDA and MDG are two key indicators for assessing the importance of a variable by the random forest algorithm, and the larger of these two indicates the higher importance of the corresponding variable. Subsequently, genes ranked in the top 5 of MDA and MDG were cross-analyzed to filter out key genes. Finally, the predictive performance of the key genes was assessed by the Receiver Operating Characteristic (ROC) curve and confusion matrix.

2.7 Biological function exploration of key genes

First, Gene Ontology (GO) analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, and Gene set enrichment analysis (GSEA) analysis were performed to initially explore the key genes' biological functions. Subsequently, correlation and differential expression analyses were performed to explore the internal connections of the key genes and their function in PCa with bone metastasis. Then, based on the intOGen database[3] and the Depmap database,[4] we explored the role of key genes in PCa carcinogenesis. Subsequently, through the Gene Set Cancer Analysis database (GSCA),[5] we explored the regulation of key genes by copy number variation (CNV) and methylation. Finally, regulatory miRNAs for key genes included in miRecords, miRTarBase, and TarBase databases were obtained based on the "multiMiR" R package, and miRNA-mRNA regulatory networks were mapped using the Cytoscape software (V 3.9.1).

2.8 Immune analysis of key genes

Here, the single sample gene set enrichment analysis (ssGSEA) algorithm was used to calculate the degree of infiltration for 28 types of immune cells for each sample in the dataset. And by observing the immune microenvironment differences between primary and bone metastatic foci, the correlation between immune cells and metastasis, and the correlation between key genes and immune cells, we aimed to unearth the influence of key genes on PCa bone metastasis at the immune level. Meanwhile, the differential expression of several common immune checkpoints between primary and metastatic foci and the correlation between key genes and immune checkpoints were explored, aiming to explore the value of key genes in immunotherapy.

2.9 Molecular functional validation of RELA

Bioinformatic analysis showed that all three key genes play critical roles in prostate cancer bone metastasis, among which RELA is of particular attention. Therefore, in this study, the molecular function of RELA in prostate cancer metastasis was verified based on cell proliferation and migration assays after overexpression of RELA in prostate cancer cells using the RELA plasmid. Finally, the protein expression of RELA was explored by immunohistochemistry between PCa primary foci and bone metastasis foci, aiming to validate the differential expression of this gene between different pathological tissues at the protein level.

2.10 Construction of nomogram and prediction of targeted drugs

At the end of this study, we constructed a nomogram based on key genes, aiming to provide a new diagnostic tool for clinicians. Meanwhile, the accuracy and reliability of the nomogram were evaluated and validated by ROC curve, Calibration curve, and Decision Curve Analysis (DCA) curve. In addition, to further validate the predictive performance of the three key genes and the nomogram,

---

[1] www.ncbi.nlm.nih.gov/geo/

[2] www.genecards.org

[3] www.intogen.org/search

[4] www.depmap.org/portal/

[5] https://guolab.wchscu.cn/GSCA/#/

TABLE 1 Univariate logistic regression results of 24 candidate genes.

Gene	OR	CI (5–95%)	*P*-value
XRCC6	0.03	0–0.47	0.01
TRIM56	0.1	0.02–0.44	0
TREX1	0	0–0.15	0
TRADD	0.08	0.01–0.56	0.01
STAT6	0.36	0.15–0.83	0.02
RELA	0	0–0.04	0
PRKDC	8.66	2.02–37.12	0
POLR3K	6.82	1.26–37.01	0.03
POLR3C	13.6	1.7–108.7	0.01
POLR2L	12.78	2.01–81.3	0.01
POLR1D	3.03	1.04–8.83	0.04
NFKB2	3.81	1.06–13.76	0.04
NFKB1	0.01	0–0.23	0
MRE11	11.26	2.35–54.03	0
MAVS	8.79	1.58–48.96	0.01
IL6	0.07	0.01–0.44	0
IKBKE	5.05	1.32–19.24	0.02
IFNA7	3.71	1–13.81	0.05
IFNA21	3.5	1.18–10.4	0.02
IFNA10	5.06	1.59–16.15	0.01
IFI16	3.49	1.32–9.19	0.01
DTX4	9.65	1.97–47.17	0.01
DHX9	2.53	1.25–5.11	0.01
CASP8	4.75	1.26–17.91	0.02

we performed 200 times 5-Fold cross-validation. Next, the targeted drugs for the key genes were predicted based on the Enrichr database[6] (37). Subsequently, molecular docking was performed between targeted drugs and key genes with top 5 combined score using the Ledock software. The 3D structure data of proteins and targeted drugs were obtained from the RSCB-PDB database[7] and the PubChem database,[8] respectively. Finally, molecular docking was visualized using the Pymol software.

3 Results

3.1 Identification and evaluation of key genes

Here, we obtained gene expression data of 84 SRGs from GSE32269 dataset. Meanwhile, 40 samples were included by cluster analysis for this analysis (Supplementary Figure 1). Based on the univariate logistic regression algorithm, we performed an initial downscaling of the 84 SRGs and 24 genes with p-value < 0.05 were identified as candidate genes (Table 1). Subsequently, the MDA and MDG values of the 24 genes were calculated using the random forest algorithm, and genes with negative MDA or MDG values were excluded (Figure 1A). Then, we performed cross-tabulation analysis on the genes ranked in the top five of MDA and MDG and finally obtained three key genes (TREX1, RELA, and CASP8), which was shown in Figure 1B.

Finally, we evaluated the prediction accuracy of the three key genes based on the ROC and confusion matrix. The ROC curves showed that the AUC values of the three key genes were 0.9123, 0.9023, and 0.7744, respectively, which indicated that the three key genes had a high prediction accuracy (Figures 1C–E). In addition, as shown in Figures 1F–H, the confusion matrix also indicated that the key genes had good prediction performance.

3.2 Biological function exploration of key genes

To explore the biological functions of the key genes, we performed GO analysis, KEGG analysis, and GSEA analysis. As shown in Figure 2A, GO analysis showed that the key genes were mainly involved in response to tumor cell, regulation of T cell receptor signaling pathway, regulation of innate immune response, mismatch repair (BP, Biological Process); replication fork, oligosaccharyltransferase complex, nuclear replication fork (CC, Cell Component); tumor necrosis factor receptor superfamily binding, mismatch repair complex binding, DNA binding, bending, chromatin DNA binding (MF, Molecular Function). The results suggest that key genes are involved in the regulation of tumors as well as immune-related mechanisms. Furthermore, KEGG also confirmed this result from another aspect. The KEGG analysis showed that the key genes were mainly enriched in tumor and immune-related signaling pathways, such as the TNF signaling pathway, Viral carcinogenesis, p53 signaling pathway, Pancreatic cancer, RIG−I−like receptor signaling pathway, IL-17 signaling pathway, etc. (Figure 2B). It is well known that Ecm-Receptor Interaction, Cell Cycle, and Homologous Recombination are key signaling pathways in tumorigenesis. And the Ecm-Receptor Interaction also plays a crucial role in PCa bone metastasis (38–40). Subsequent GSEA analysis showed that TREX1 and RELA inhibited the activation of Ecm-Receptor Interaction, Cell Cycle, and Homologous Recombination, while CASP8 was involved in the activation of these three signaling pathways (Figures 2C–E).

Correlation analysis showed a strong internal correlation between the three key genes and was closely associated with bone metastasis (Figure 3A). Differential expression analysis showed that three key genes had significant expression differences in primary foci and bone metastatic foci tissues (Figures 3B–D), demonstrating that the key genes play critical roles in developing PCa bone metastasis. Tumor driver genes play a crucial role in tumorigenesis. Chronos Score is a metric in the Depmap database to assess the degree of impact on cell proliferation after gene knockdown, and a more negative value indicates a greater impact of the gene on cell proliferation (41). As shown in Figures 3E,F, the three key genes were significantly correlated with multiple PCa driver genes; and the Chronos Score of TREX1 and RELA, except CASP8, were negative in prostate tumor cells. The above

---

[6] https://maayanlab.cloud/Enrichr/#

[7] https://www.rcsb.org/

[8] https://pubchem.ncbi.nlm.nih.gov/

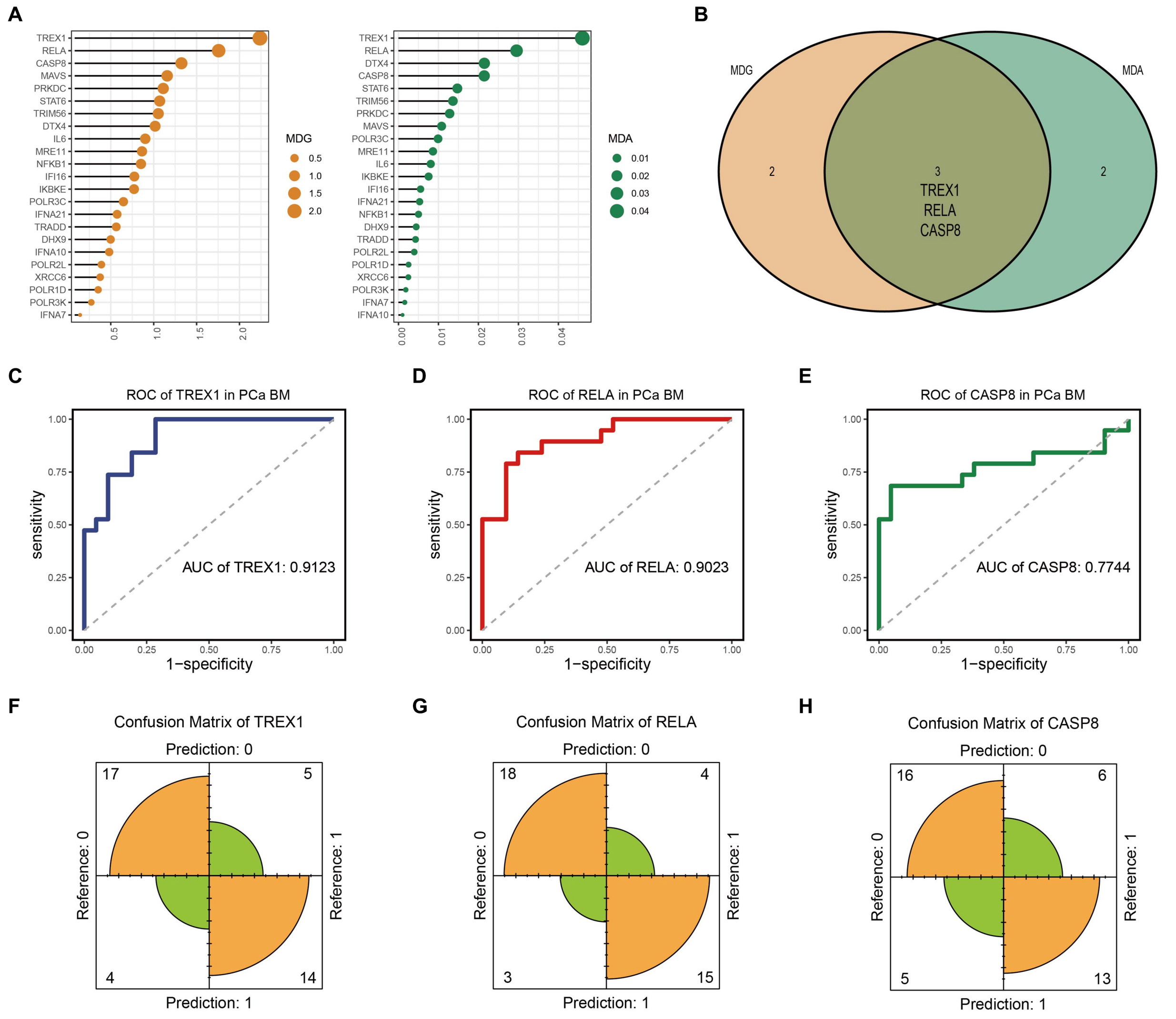


FIGURE 1
Identification and evaluation of key genes. **(A)** MDA and MDG ordering of candidate genes; **(B)** Identification of key genes; **(C–E)** ROC curve of TREX1, RELA, and CASP8; **(F–H)** Confusion matrix of TREX1, RELA, and CASP8.

results demonstrated that the three key genes also play critical roles in PCa development.

The CNV, methylation, and miRNA are critical parts of the gene expression regulatory network, hence we also analyzed the effects of CNV, methylation, and miRNA on key genes. The results showed that in CNV, the three key genes were significantly positively correlated with CNV; while in methylation, TREX1 and RELA were significantly negatively correlated with methylation, except for CASP8, which was positively correlated with methylation (Figure 3G); meanwhile, the miRNA-mRNA regulatory network showed that a total of 128 miRNAs were involved in the regulation of three key genes, and 9 miRNAs were involved in the regulation of several key genes, among which has-miR-155-5P and has-miR-16-5P were noteworthy (Figure 3H).

In conclusion, three key genes play an essential role in the development and metastasis of PCa. And in the process of bone metastasis, TREX1 and RELA are protective factors, while CASP8 is a risk factor.

3.3 Immune analysis of key genes

The above studies revealed that key genes were involved in the immune regulation of tumors, so we explored the key genes in the immune-related mechanisms of PCa bone metastasis. Analysis of immune infiltration differences showed that 14 immune cells had significant infiltration differences between primary and metastatic foci (Figure 4A). Studies have shown that four types of immune cells, Macrophage, Myeloid derived suppressor cell, Regulatory T cell, and Plasmacytoid dendritic cell promote PCa bone metastasis (32, 42–44). Meanwhile, correlation analysis showed that TREX1 was closely associated with a variety of immune cells and was significantly negatively correlated with three types of bone-metastasis-promoting immune cells (Regulatory T cells, Myeloid derived suppressor cells, and Plasmacytoid dendritic cells); RELA was significantly negatively correlated with two types of bone-metastasis-promoting immune cells (Regulatory T cells and Plasmacytoid dendritic cells); whereas CASP8 showed an opposite correlation trends (Figure 4B). Of course, tumor

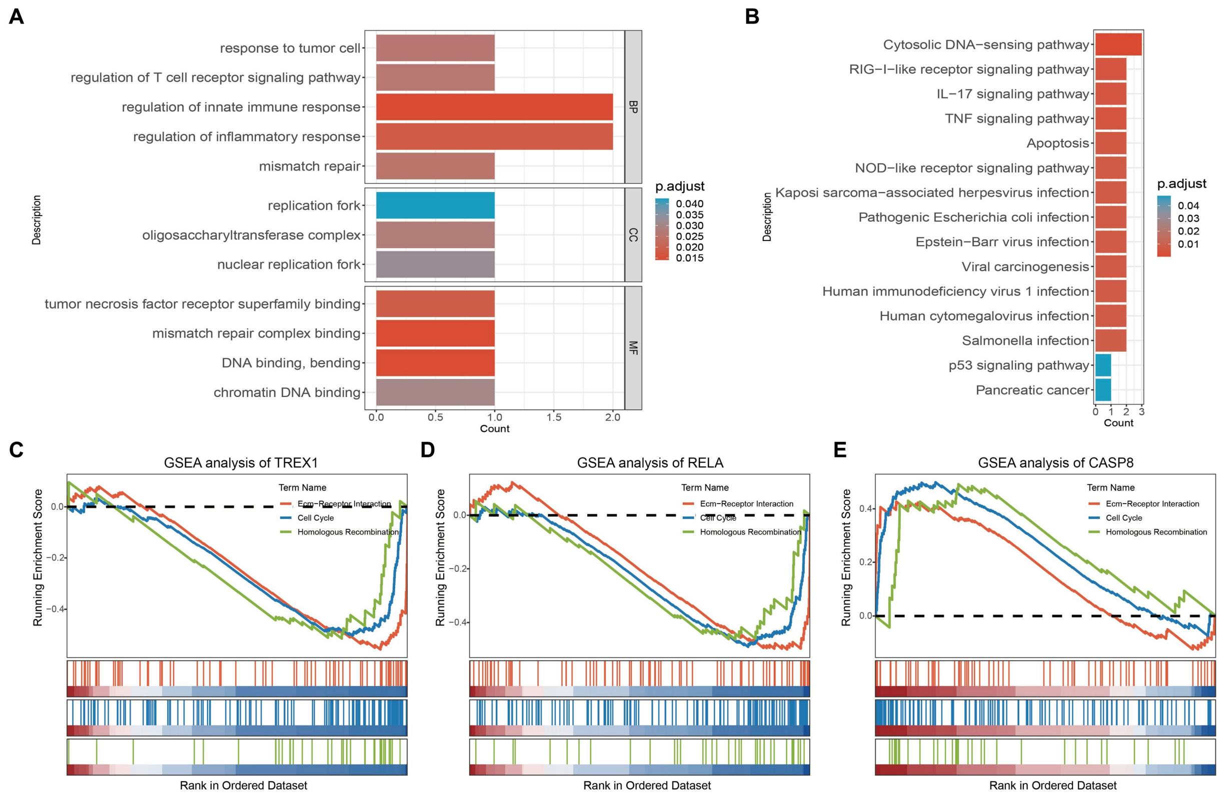


FIGURE 2
GO, KEGG and GSEA analysis of key genes. **(A)** GO analysis of key genes; **(B)** KEGG analysis of key genes; **(C–E)** GSEA analysis of key genes.

immunity is a very complex biological regulatory network, and a variety of immune cells (CD8 T cells, Natural killer cells, and Monocytes) have been also shown to play important roles in tumor bone metastasis (45–47). The molecular mechanisms between the three key genes and the four bone-metastasis-promoting immune cells as well as other types of immune cells are complex and need to be explored by subsequent in-depth studies.

Immune checkpoints have been a hot research topic in the field of oncology, and in this study, we found that CD-200 and PD-1 were significantly highly expressed in bone metastatic foci (Figure 4C); TREX1 and RELA were negatively correlated with CD-200 and PD-1, while CASP8 was positively correlated with CD-200 and CTLA-4 (Figure 4D). Thus, key genes also play a role in the immunoregulation of bone metastasis in PCa.

3.4 Molecular functional validation of RELA

The results of several analyses indicate that RELA plays a crucial role in the ontogenesis, progression, and metastasis of PCa. Hence, we validated its molecular function in proliferation and metastasis. Cell proliferation assay showed that the proliferation ability of DU145 and PC3 cells was significantly inhibited after overexpression of RELA (Figure 5A). Meanwhile, cell migration assay showed the same trend. As shown in Figures 5B,C, the migration of DU145 and PC3 cells were also significantly inhibited after overexpression of RELA. In addition, we verified the difference in RELA expression between primary foci and bone metastasis by immunohistochemistry. The results showed that the expression of RELA was significantly lower in bone metastasis compared with primary foci, which was consistent with the results of our data analysis (Figure 5D). Hence, this greatly confirms the critical role of RELA in PCa development and metastasis, and is most likely a novel therapeutic target.

3.5 Construction of nomogram and prediction of targeted drugs

Based on the three key genes, we constructed a nomogram designed to help clinicians make clinical predictions (Figure 6A). Subsequently, we evaluated the accuracy and reliability of the nomogram. The results showed that the AUC value of the ROC curve was 0.99, and both the Calibration curve and the DCA curve performed well (Figures 6B–D). In addition, a 200 times 5-fold cross-validation analysis suggested high AUC and C-Index values for all three key genes and the nomogram (Supplementary Figure 2). Therefore, the nomogram constructed in this study was accurate and reliable. Next, based on the Enrichr database, we excavated 22 potential TREX1 and RELA targeted overexpression drugs (Figure 6E; Supplementary File 1). Finally, molecular docking was performed using the Ledock software for the targeted drugs with a top 5 ranked TREX1 and RELA combined score. The results showed that the targeting pockets of the 2 key genes were successfully occupied by the drugs (Figure 6F). In conclusion, the nomogram constructed and the new targeted drugs discovered in this study provided a new strategy for the treatment of advanced PCa.

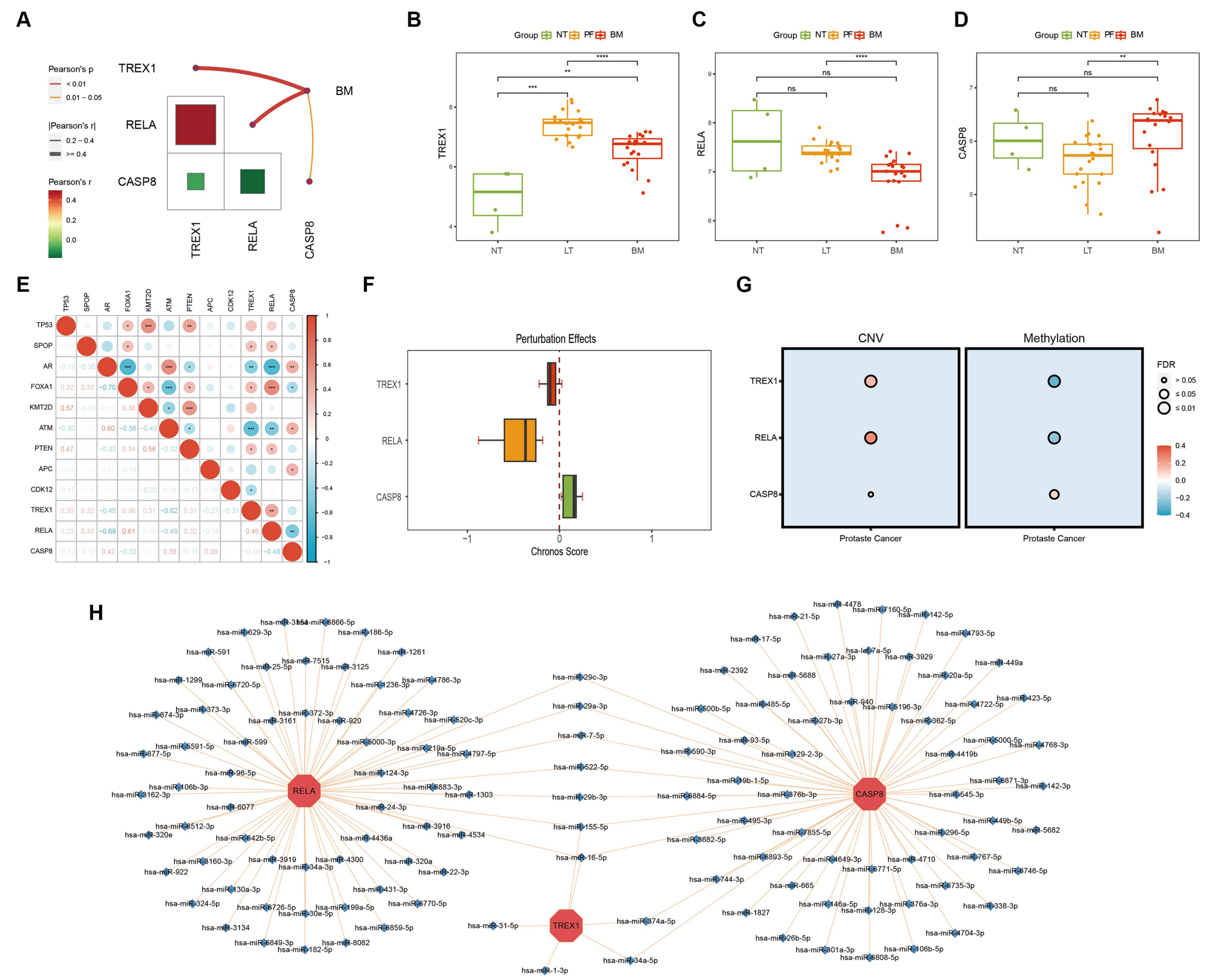


FIGURE 3
Biological function exploration of key genes. **(A)** Internal correlation analysis of key genes; **(B–D)** Differential expression analysis of key genes (NT, normal tissue; PF, primary foci; BM, bone metastasis); **(E)** Correlation analysis of key with PCa driver genes; **(F)** Chronos Score of key genes; **(G)** Correlation analysis of key genes with CNV/methylation; **(H)** miRNA-mRNA regulatory network of key genes.

4 Discussion

Bone metastasis occur in more than 1.5 million cancer patients worldwide, and the tumors most at risk for this complication are PCa and breast cancer (48). Although there are well-established treatment strategies for early-stage PCa, including surgical resection, chemotherapy, and androgen deprivation (49), most PCa continues to progress and develop bone metastasis. Studies have shown that the incidence of bone metastasis in PCa is about 70%, and the most common site is the vertebrae (50, 51). The development of bone metastasis can be catastrophic for PCa patients, who often suffer from fractures, spinal cord compression, and disability (9, 48), significantly reducing survival quality. Unfortunately, there is no effective treatment for PCa bone metastasis at this stage (52, 53). It is urgent to deeply explore the molecular mechanisms of PCa bone metastasis and tap new therapeutic targets.

It is well known that the immune system plays an indispensable role in maintaining normal bone homeostasis and in various bone-related diseases, mainly through inflammation. Recently, it has been found that pro-inflammatory cytokines in tumors cause homeostatic abnormalities in osteoclasts and osteoblasts, leading to the development of bone metastasis (54, 55). Furthermore, there is also evidence that immune cells can influence the colonization and progression of tumor cells in bone metastatic foci (56). The STING pathway, as an innate immune regulatory mechanism in the human body, has been shown to play an essential role in the development and metastasis of various tumors by mediating inflammatory responses (23, 30). Therefore, exploring the relevant molecular mechanisms of the STING pathway in PCa bone metastasis will bring new opportunities for the treatment of PCa bone metastasis.

The random forest algorithm is an excellent machine learning algorithm that has been commonly used in the biosciences field. Three key genes were finally screened in this study based on the MDA and MDG values assigned to each gene by this algorithm. The MDA indicates the extent to which model accuracy decreases when a variable is excluded, and the MDG indicates the extent to which a variable contributes to the reduction of the Gini in a random forest, with a higher Gini indicating a higher prediction error rate for that node (57). Therefore, the higher the two metrics are, the more critical their corresponding genes are. The subsequent comprehensive analysis

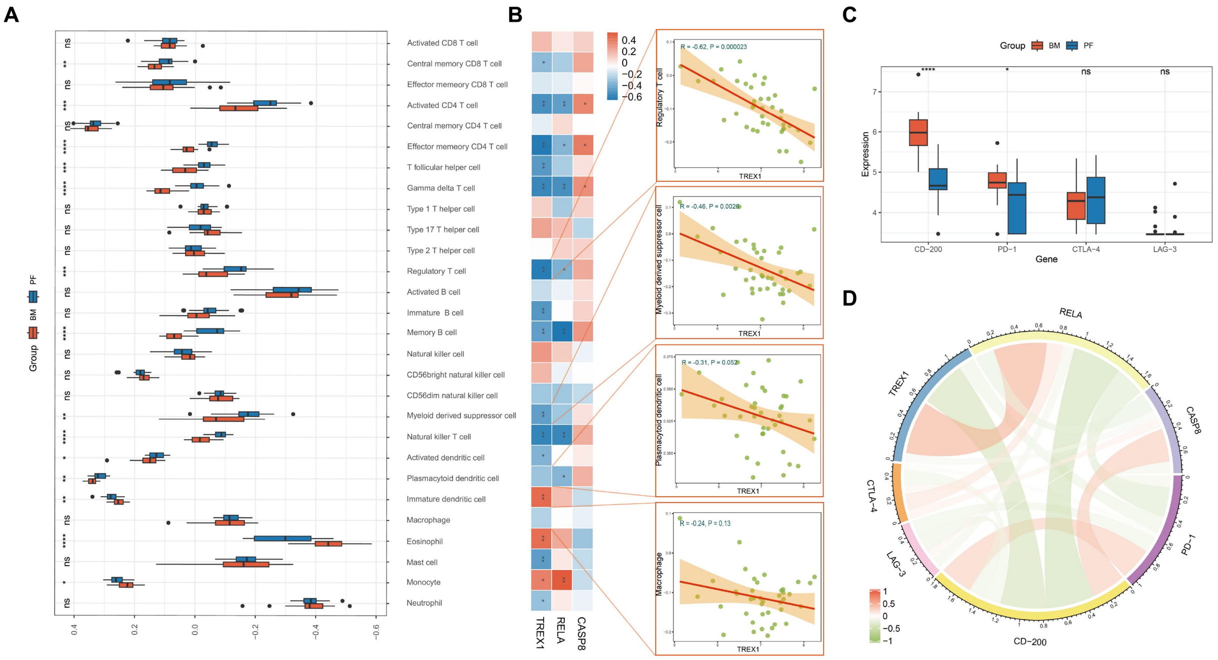


FIGURE 4

Immune analysis of key genes. **(A)** Differential analysis of the degree of immune cell infiltration in primary foci and bone metastasis (PF, primary foci; BM, bone metastasis); **(B)** Correlation analysis of key genes and immune cells; **(C)** Differential expression analysis of immune checkpoints in primary foci and bone metastasis; **(D)** Correlation analysis of key genes and immune checkpoints.

demonstrated that the three key genes play critical roles in PCa development and metastasis.

TREX1 (Three Prime Repair Exonuclease 1) is a major cytoplasmic nuclease, prevalently expressed in mammalian cells, that acts primarily on double-stranded DNA, specializing in the excision of oligonucleotides that are mismatched at their 3′-ends. The enzyme is mainly involved in immune regulation and DNA damage repair in the body (58). Numerous studies have shown that loss-of-function mutations in TREX1 lead to abnormal accumulation of cytoplasmic DNA, which in turn over-activates the natural immune response and ultimately leads to the development of autoimmune diseases, including systemic lupus erythematosus and vascular diseases, retinopathy, and cerebral leukoencephalopathy (59, 60). Recent studies have shown that TREX1 also plays a key role in tumors (61). In breast and colon cancers, high expression of TREX1 leads to blockage of type I IFN pathway activation, inhibiting the anti-tumor immune response and the anti-tumor response of immune checkpoint inhibitors (61, 62). However, this phenomenon seems to be reversed in PCa, where it was shown that the expression level of TREX1 was not associated with the degree of anti-tumor immune response to Radiotherapy-induced activation of the type I IFN pathway in three PCa cell lines (63). In this study, TREX1 was found to be negatively correlated with the degree of infiltration of multiple PCa bone metastasis-promoting immune cells (including Regulatory T cell, Myeloid derived suppressor cell, and Plasmacytoid dendritic cell), and was negatively correlated with PD-1 and CD-200. In addition, high expression of TREX1 can prevent the activation of the Ecm-Receptor Interaction pathway, which plays a crucial role in prostate bone metastasis (38–40). Therefore, TREX1 is not only valuable in the prediction of PCa bone metastasis but also has great potential in its immunotherapy. The related molecular mechanisms deserve to be explored in depth.

RELA (RELA Proto-Oncogene, NF-KB Subunit) is a ubiquitous transcription factor involved in a variety of biological processes. Previous studies have shown that RELA plays a key role in the initial stages of Pca (64), and this study found that RELA has an important role in the carcinogenesis, development and metastasis of PCa. Correlation analysis showed that RELA was strongly associated with multiple prostate cancer driver genes (SPOP, AR, FOXA1, ATM, and PTEN); the Depmap database results showed that RELA had a high effect on the proliferation of PCa cell lines; bio-functional analyses showed that RELA may hinder the occurrence of bone metastasis by inhibiting the Ecm-Receptor Interaction pathway; subsequent cell and tissue experiments have also demonstrated that overexpression of RELA can significantly inhibited the proliferation and migration of PCA cells (DU145 and PC3), as well as expression of RELA is significantly decreased in bone metastasis; Immune analysis results showed that RELA was significantly negatively correlated with 2 types of bone metastasis-promoting immune cells (regulatory T cells and plasmacytoid dendritic cells), and significantly negatively correlated with PD-1 and CD-200. These results strongly suggest that RELA plays a crucial role in prostate cancer occurrence, progression and metastasis based on multiple molecular mechanisms. Therefore, RELA is likely to be a new therapeutic target for advanced PCa, and its complex biomolecular function is worthy of subsequent in-depth study.

As for CASP8 (Caspase 8), this gene is a member of the cysteine-aspartate protease (cysteinyl aspartate) family and plays a central role in apoptosis and necroptosis (65). Studies have shown that this gene is closely associated with PCa and its recurrence and can be used as a

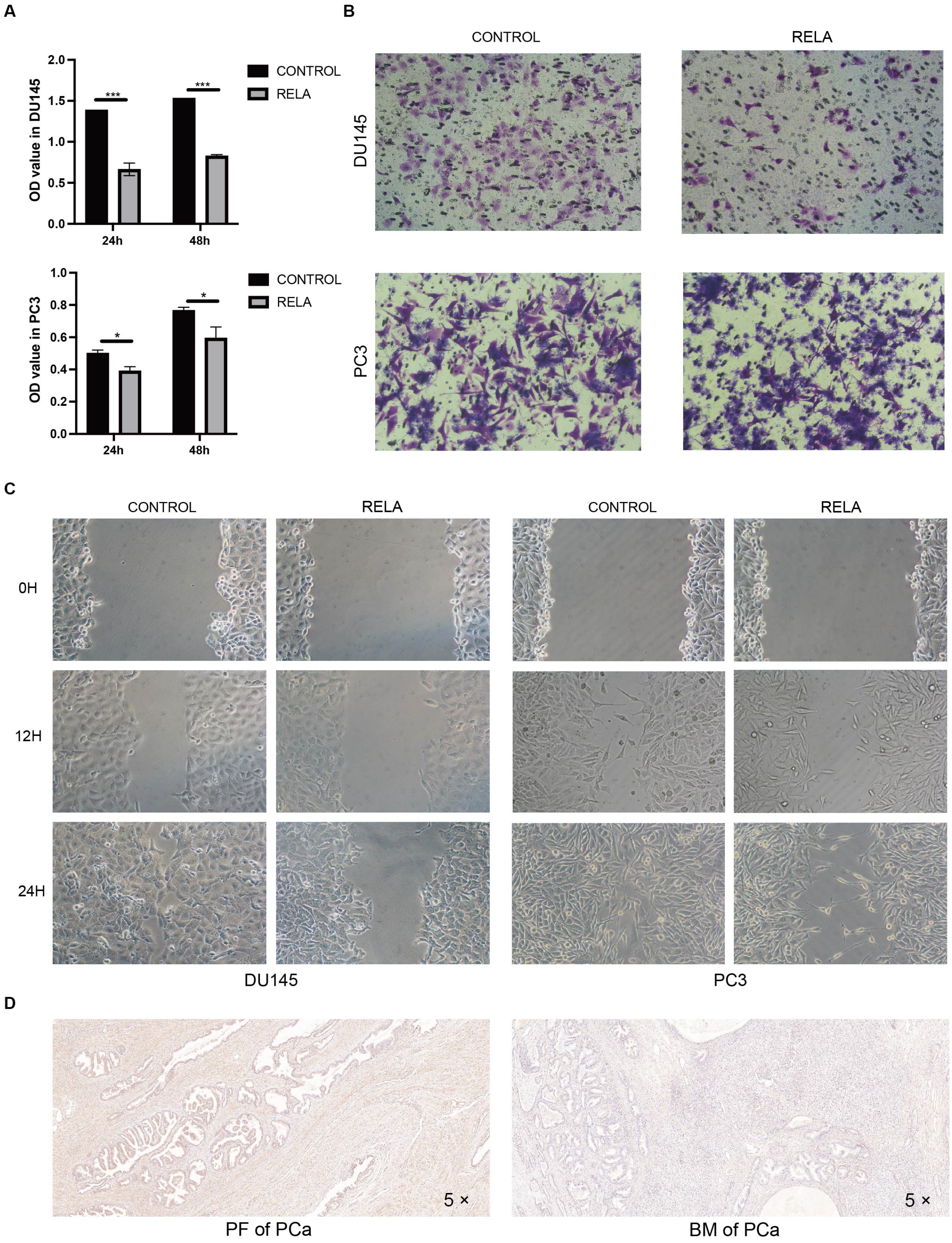


FIGURE 5
Molecular functional validation of RELA. **(A)** Cell proliferation assessment of DU145 and PC3 cells based on CCK-8 assay (RELA, PCa cells treated with RELA-overexpression plasmid; **p*-value < 0.05; ****p*-value < 0.001); **(B)** Cell migration assessment of DU145 and PC3 cells based on transwell assay; **(C)** Cell migration assessment of DU145 and PC3 cells based on wound healing assay; **(D)** Immunohistochemistry of RELA in the PCa primary foci (PF) and bone metastasis (BM).

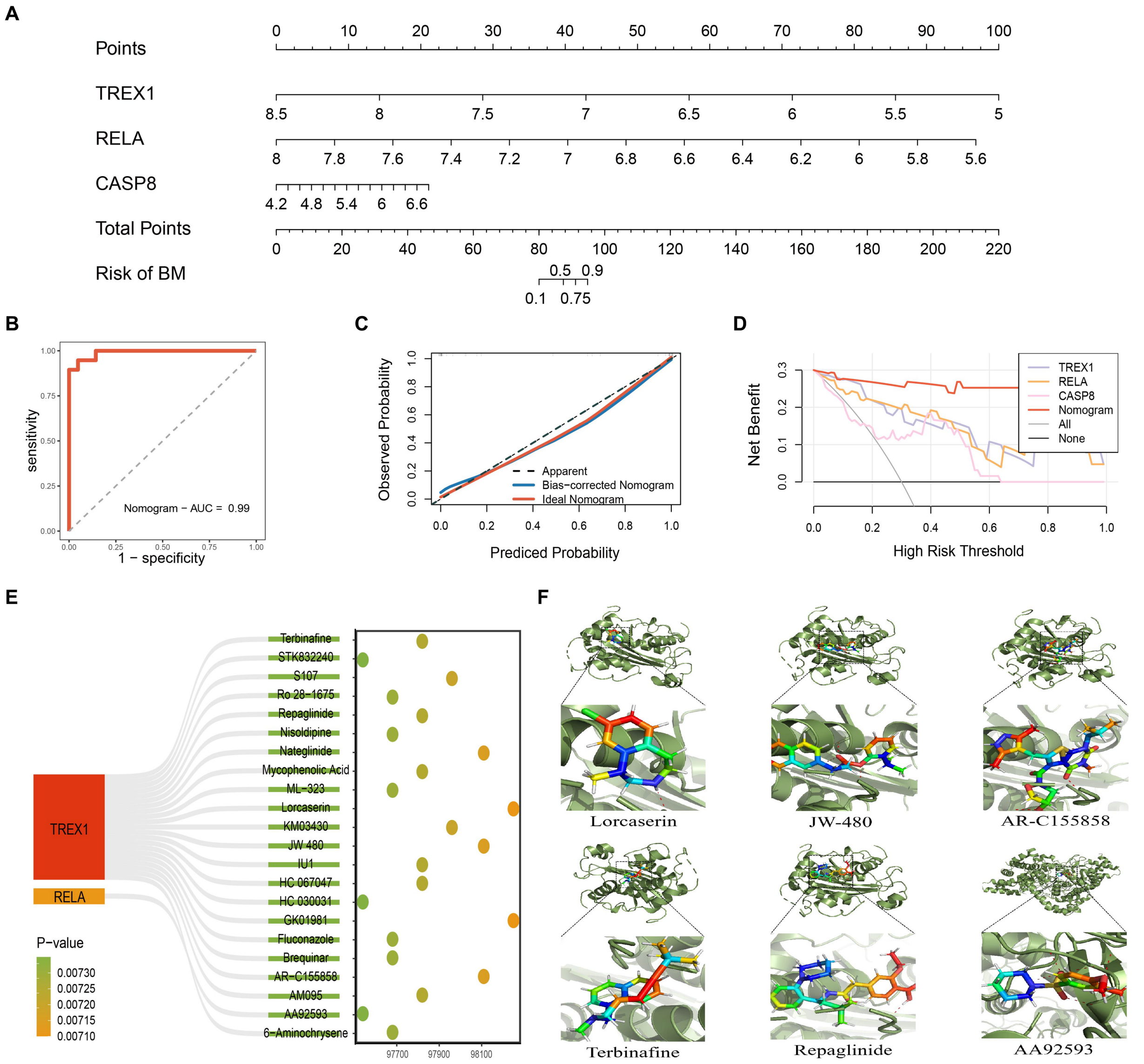


FIGURE 6
Construction and evaluation of nomogram. **(A)** Nomogram constructed based on key genes; **(B)** ROC curve of the nomogram; **(C)** Calibration curve of the nomogram; **(D)** DCA curve of the nomogram; **(E)** 22 kinds of potential targeted drugs for TREX1 and RELA; **(F)** Molecular docking results for the top 5 TREX1 and RELA-targeted drugs in the combined score.

biomarker for bone metastasis in high-risk PCa (66, 67). This study found that CASP8 can activate three pathways, including Ecm-Receptor Interaction, Cell Cycle, and Homologous Recombination. In addition, CASP8 was positively associated with three types of pro-bone metastatic cells (Myeloid derived suppressor cell, Plasmacytoid dendritic cell, and Regulatory T cell) and two immune checkpoints (CD-200 and CTLA-4). Therefore, this gene is important in PCa risk prediction and bone metastasis prediction and also deserves attention.

At the end of this study, we constructed a nomogram and verified its reliability and accuracy in several approaches. This tool provides a reliable diagnostic tool for clinicians' diagnosis and treatment. In addition, we also identified 22 potential targeted drugs of key genes with satisfactory molecular docking effects. In conclusion, this study mined three bone metastasis related key genes for PCa based on machine learning, explored their related molecular mechanisms, constructed a reliable nomogram and discovered 22 kinds of potential targeted drugs. This study can improve the treatment of advanced PCa patients and provides a theoretical basis for subsequent research.

Of course, this study has some limitations. First, this study is essentially a retrospective analysis, which needs to be corroborated by subsequent prospective studies; second, only one dataset was included in this study to construct the nomogram, and the generalization ability of the model performance needs to be validated by the subsequent including of more datasets; third, this study did not perform more in-depth basic experimental validation, which needs to be supplemented in subsequent studies; and finally, it is difficult to define a fixed threshold for the expression of various genes in the clinic, which poses an obstacle to the popularization of diagnostic tools.

5 Conclusion

In this study, we explored the relevant mechanisms of the STING pathway in PCa bone metastasis based on bioinformatics analysis techniques. Based on the random forest algorithm mined three key genes, and the critical roles of the key genes in PCa development, metastasis, and tumor immunity were explored by multiple analyses. Finally, based on the key genes, a reliable nomogram was constructed and potential targeted drugs were discovered. In conclusion, this study provides new therapeutic targets and a reliable diagnostic tool for clinical treatment and provides a theoretical basis for follow-up studies.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

This study was approved by the Clinical Research Ethics Committee of the First Affiliated Hospital of Chongqing Medical University (ID: K2024-185-01). The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because the experiment was performed using pathological tissues left over from postoperative pathological examinations.

Author contributions

GL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. RZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. ZX: Conceptualization, Data curation, Formal analysis, Investigation, Resources, Validation, Visualization, Writing – original draft. XQ: Data curation, Investigation, Resources, Visualization, Writing – original draft. YD: Data curation, Investigation, Resources, Visualization, Writing – original draft. YZ: Data curation, Investigation, Resources, Visualization, Writing – original draft. HL: Data curation, Investigation, Resources, Visualization, Writing – original draft. DT: Methodology, Project administration, Supervision, Writing – review & editing. ZL: Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing. WH: Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was supported by Fundamental Research Funds for the Central Universities (Grant NO. 2022CDJYGRH-019).

Acknowledgments

The authors would like to thank the GeneCards, GEO IntOGen, DepMap, GSCA, Enrichr, RSCB-PDB, and PubChem databases for providing the data.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1372495/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Removal of outliers based on clustering algorithm. **(A)** Sample clustering to detect outliers; **(B)** Sample dendrogram and trait heatmap.

SUPPLEMENTARY FIGURE 2
200 times 5-fold cross validation of 3 key genes and the nomogram.

SUPPLEMENTARY DATA SHEET 1
22 potential TREX1 and RELA targeted overexpression drugs.

References

1. Gandaglia G, Abdollah F, Schiffmann J, Trudeau V, Shariat SF, Kim SP, et al. Distribution of metastatic sites in patients with prostate cancer: a population-based analysis. *Prostate*. (2014) 74:210–6. doi: 10.1002/pros.22742

2. Riihimäki M, Thomsen H, Sundquist K, Sundquist J, Hemminki K. Clinical landscape of cancer metastases. *Cancer Med*. (2018) 7:5534–42. doi: 10.1002/cam4.1697

3. Macedo F, Ladeira K, Pinho F, Saraiva N, Bonito N, Pinto L, et al. Bone Metastases: An Overview. *Oncol Rev*. (2017) 11:321. doi: 10.4081/oncol.2017.321

4. Yin JJ, Pollock CB, Kelly K. Mechanisms of cancer metastasis to the bone. *Cell Res*. (2005) 15:57–62. doi: 10.1038/sj.cr.7290266

5. Clark PE, Torti FM. Prostate cancer and bone metastases: medical treatment. *Clin Orthop Relat Res*. (2003) 415:S148–57. doi: 10.1097/01.blo.0000093840.72468.e1

6. Roodman GD. Mechanisms of bone metastasis. *N Engl J Med*. (2004) 350:1655–64. doi: 10.1056/NEJMra030831

7. Coleman RE. Clinical features of metastatic bone disease and risk of skeletal morbidity. *Clin Cancer Res*. (2006) 12:6243s–9s. doi: 10.1158/1078-0432.CCR-06-0931

8. Mercadante S. Malignant bone pain: pathophysiology and treatment. *Pain*. (1997) 69:1. doi: 10.1016/S0304-3959(96)03267-8

9. Hofbauer LC, Rachner TD, Coleman RE, Jakob F. Endocrine aspects of bone metastases. *Lancet Diabetes Endocrinol*. (2014) 2:500–12. doi: 10.1016/S2213-8587(13)70203-1

10. Mundy GR. Metastasis to bone: causes, consequences and therapeutic opportunities. *Nat Rev Cancer*. (2002) 2:584–93. doi: 10.1038/nrc867

11. Coleman RE, Rubens RD. The clinical course of bone metastases from breast cancer. *Br J Cancer*. (1987) 55:61–6. doi: 10.1038/bjc.1987.13

12. Costa L, Badia X, Chow E, Lipton A, Wardley A. Impact of skeletal complications on patients' quality of life, mobility, and functional independence. *Support Care Cancer*. (2008) 16:879–89. doi: 10.1007/s00520-008-0418-0

13. Paice JA, Mulvey M, Bennett M, Dougherty PM, Farrar JT, Mantyh PW, et al. AAPT diagnostic criteria for chronic cancer pain conditions. *J Pain*. (2017) 18:233–46. doi: 10.1016/j.jpain.2016.10.020

14. Turpin A, Duterque-Coquillaud M, Vieillard M-H. Bone metastasis: current state of play. *Transl Oncol*. (2020) 13:308–20. doi: 10.1016/j.tranon.2019.10.012

15. Luzzi KJ, MacDonald IC, Schmidt EE, Kerkvliet N, Morris VL, Chambers AF, et al. Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am J Pathol*. (1998) 153:865–73. doi: 10.1016/S0002-9440(10)65628-3

16. Chaffer CL, San Juan BP, Lim E, Weinberg RA. EMT, cell plasticity and metastasis. *Cancer Metastasis Rev*. (2016) 35:645–54. doi: 10.1007/s10555-016-9648-7

17. Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? *Lancet*. (2001) 357:539–45. doi: 10.1016/S0140-6736(00)04046-0

18. Dvorak HF. Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. *N Engl J Med*. (1986) 315:1650–9.

19. Henson PM. Dampening inflammation. *Nat Immunol*. (2005) 6:1179–81. doi: 10.1038/ni1205-1179

20. Colotta F, Allavena P, Sica A, Garlanda C, Mantovani A. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis*. (2009) 30:1073–81. doi: 10.1093/carcin/bgp127

21. Coussens LM, Werb Z. Inflammation and cancer. *Nature*. (2002) 420:860–7. doi: 10.1038/nature01322

22. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. *Cell*. (2010) 140:883–99. doi: 10.1016/j.cell.2010.01.025

23. Kwon J, Bakhoum SF. The cytosolic DNA-sensing cGAS-STING pathway in cancer. *Cancer Discov*. (2020) 10:26–39. doi: 10.1158/2159-8290.CD-19-0761

24. Corrales L, Glickman LH, McWhirter SM, Kanne DB, Sivick KE, Katibah GE, et al. Direct activation of STING in the tumor microenvironment leads to potent and systemic tumor regression and immunity. *Cell Rep*. (2015) 11:1018–30. doi: 10.1016/j.celrep.2015.04.031

25. Guerin MV, Regnier F, Feuillet V, Vimeux L, Weiss JM, Bismuth G, et al. TGFβ blocks IFNα/β release and tumor rejection in spontaneous mammary tumors. *Nat Commun*. (2019) 10:4131. doi: 10.1038/s41467-019-11998-w

26. Weiss JM, Guérin MV, Regnier F, Renault G, Galy-Fauroux I, Vimeux L, et al. The STING agonist DMXAA triggers a cooperation between T lymphocytes and myeloid cells that leads to tumor regression. *Onco Targets Ther*. (2017) 6:e1346765. doi: 10.1080/2162402X.2017.1346765

27. Sivick KE, Desbien AL, Glickman LH, Reiner GL, Corrales L, Surh NH, et al. Magnitude of therapeutic STING activation determines CD8+ T cell-mediated anti-tumor immunity. *Cell Rep*. (2018) 25:3074–3085.e5. doi: 10.1016/j.celrep.2018.11.047

28. Bakhoum SF, Ngo B, Laughney AM, Cavallo J-A, Murphy CJ, Ly P, et al. Chromosomal instability drives metastasis through a cytosolic DNA response. *Nature*. (2018) 553:467–72. doi: 10.1038/nature25432

29. Mackenzie KJ, Carroll P, Martin C-A, Murina O, Fluteau A, Simpson DJ, et al. cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature*. (2017) 548:461–5. doi: 10.1038/nature23449

30. Chen Q, Boire A, Jin X, Valiente M, Er EE, Lopez-Soto A, et al. Carcinoma-astrocyte gap junctions promote brain metastasis by cGAMP transfer. *Nature*. (2016) 533:493–8. doi: 10.1038/nature18268

31. Bouchet M, Lainé A, Boyault C, Proponnet-Gueurault M, Meugnier E, Bouazza L, et al. ERRα expression in bone metastases leads to an exacerbated antitumor immune response. *Cancer Res*. (2020) 80:2914–26. doi: 10.1158/0008-5472.CAN-19-3584

32. Wu AC, He Y, Broomfield A, Paatan NJ, Harrington BS, Tseng H-W, et al. CD169(+) macrophages mediate pathological formation of woven bone in skeletal lesions of prostate cancer. *J Pathol*. (2016) 239:218–30. doi: 10.1002/path.4718

33. Zhu L, Narloch JL, Onkar S, Joy M, Broadwater G, Luedke C, et al. Metastatic breast cancers have reduced immune cell recruitment but harbor increased macrophages relative to their matched primary tumors. *J Immunother Cancer*. (2019) 7:265. doi: 10.1186/s40425-019-0755-1

34. Fend L, Accart N, Kintz J, Cochin S, Reymann C, Le Pogam F, et al. Therapeutic effects of anti-CD115 monoclonal antibody in mouse cancer models through dual inhibition of tumor-associated macrophages and osteoclasts. *PLoS One*. (2013) 8:e73310. doi: 10.1371/journal.pone.0073310

35. Hiraoka K, Zenmyo M, Watari K, Iguchi H, Fotovati A, Kimura YN, et al. Inhibition of bone and muscle metastases of lung cancer cells by a decrease in the number of monocytes/macrophages. *Cancer Sci*. (2008) 99:1595–602. doi: 10.1111/j.1349-7006.2008.00880.x

36. Wang K, Donnelly CR, Jiang C, Liao Y, Luo X, Tao X, et al. STING suppresses bone cancer pain via immune and neuronal modulation. *Nat Commun*. (2021) 12:4558. doi: 10.1038/s41467-021-24867-2

37. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with Enrichr. *Curr Protoc*. (2021) 1:e90. doi: 10.1002/cpz1.90

38. Tang L, Xu M, Zhang L, Qu L, Liu X. Role of αVβ3 in prostate cancer: metastasis initiator and important therapeutic target. *Onco Targets Ther*. (2020) 13:7411–22. doi: 10.2147/OTT.S258252

39. San Martin R, Pathak R, Jain A, Jung SY, Hilsenbeck SG, Piña-Barba MC, et al. Tenascin-C and integrin α9 mediate interactions of prostate cancer with the bone microenvironment. *Cancer Res*. (2017) 77:5977–88. doi: 10.1158/0008-5472.CAN-17-0064

40. Karkampouna S, De Filippo MR, Ng CKY, Klima I, Zoni E, Spahn M, et al. Stroma Transcriptomic and proteomic profile of prostate cancer metastasis Xenograft models reveals prognostic value of Stroma signatures. *Cancers*. (2020) 12:3786. doi: 10.3390/cancers12123786

41. Dempster JM, Boyle I, Vazquez F, Root DE, Boehm JS, Hahn WC, et al. Chronos: a cell population dynamics model of CRISPR experiments that improves inference of gene fitness effects. *Genome Biol*. (2021) 22:343. doi: 10.1186/s13059-021-02540-7

42. Wen J, Huang G, Liu S, Wan J, Wang X, Zhu Y, et al. Polymorphonuclear MDSCs are enriched in the stroma and expanded in metastases of prostate cancer. *J Pathol Clin Res*. (2020) 6:171–7. doi: 10.1002/cjp2.160

43. Zhao E, Wang L, Dai J, Kryczek I, Wei S, Vatan L, et al. Regulatory T cells in the bone marrow microenvironment in patients with prostate cancer. *Onco Targets Ther*. (2012) 1:152–61. doi: 10.4161/onci.1.2.18480

44. Sawant A, Hensel JA, Chanda D, Harris BA, Siegal GP, Maheshwari A, et al. Depletion of plasmacytoid dendritic cells inhibits tumor growth and prevents bone metastasis of breast cancer cells. *J Immunol*. (2012) 189:4258–65. doi: 10.4049/jimmunol.1101855

45. Zhang K, Kim S, Cremasco V, Hirbe AC, Collins L, Piwnica-Worms D, et al. CD8+ T cells regulate bone tumor burden independent of osteoclast resorption. *Cancer Res*. (2011) 71:4799–808. doi: 10.1158/0008-5472.CAN-10-3922

46. Bidwell BN, Slaney CY, Withana NP, Forster S, Cao Y, Loi S, et al. Silencing of Irf7 pathways in breast cancer cells promotes bone metastasis through immune escape. *Nat Med*. (2012) 18:1224–31. doi: 10.1038/nm.2830

47. Taverna S, Pucci M, Giallombardo M, Di Bella MA, Santarpia M, Reclusa P, et al. Amphiregulin contained in NSCLC-exosomes induces osteoclast differentiation through the activation of EGFR pathway. *Sci Rep*. (2017) 7:3170. doi: 10.1038/s41598-017-03460-y

48. Weilbaecher KN, Guise TA, McCauley LK. Cancer to bone: a fatal attraction. *Nat Rev Cancer*. (2011) 11:411–25. doi: 10.1038/nrc3055

49. Nuhn P, De Bono JS, Fizazi K, Freedland SJ, Grilli M, Kantoff PW, et al. Update on systemic prostate cancer therapies: Management of Metastatic Castration-resistant Prostate Cancer in the era of precision oncology. *Eur Urol*. (2019) 75:88–99. doi: 10.1016/j.eururo.2018.03.028

50. Galasko CS. Monitoring of bone metastases. *Schweiz Med Wochenschr*. (1981) 111:1873–5.

51. Peng S, Yi Z, Liu M. Ailanthone: a new potential drug for castration-resistant prostate cancer. *Chin J Cancer*. (2017) 36:25. doi: 10.1186/s40880-017-0194-7

52. Body J-J, Casimiro S, Costa L. Targeting bone metastases in prostate cancer: improving clinical outcome. *Nat Rev Urol*. (2015) 12:340–56. doi: 10.1038/nrurol.2015.90

53. Norum J, Nieder C. Treatments for metastatic prostate cancer (mPC): a review of costing evidence. *PharmacoEconomics*. (2017) 35:1223–36. doi: 10.1007/s40273-017-0555-8

54. Xiang L, Gilkes DM. The contribution of the immune system in bone metastasis pathogenesis. *Int J Mol Sci*. (2019) 20:999. doi: 10.3390/ijms20040999

55. Owen KL, Parker BS. Beyond the vicious cycle: the role of innate osteoimmunity, automimicry and tumor-inherent changes in dictating bone metastasis. *Mol Immunol*. (2019) 110:57–68. doi: 10.1016/j.molimm.2017.11.023

56. Clézardin P, Coleman R, Puppo M, Ottewell P, Bonnelye E, Paycha F, et al. Bone metastasis: mechanisms, therapies, and biomarkers. *Physiol Rev*. (2021) 101:797–855. doi: 10.1152/physrev.00012.2019

57. Wang H, Yang F, Luo Z. An experimental study of the intrinsic stability of random forest variable importance measures. *BMC Bioinformatics*. (2016) 17:60. doi: 10.1186/s12859-016-0900-5

58. Grieves JL, Fye JM, Harvey S, Grayson JM, Hollis T, Perrino FW. Exonuclease TREX1 degrades double-stranded DNA to prevent spontaneous lupus-like inflammatory disease. *Proc Natl Acad Sci USA*. (2015) 112:5117–22. doi: 10.1073/pnas.1423804112

59. Yan N. Immune diseases associated with TREX1 and STING dysfunction. *J Interf Cytokine Res*. (2017) 37:198–206. doi: 10.1089/jir.2016.0086

60. Richards A, van den Maagdenberg AMJM, Jen JC, Kavanagh D, Bertram P, Spitzer D, et al. C-terminal truncations in human 3′-5' DNA exonuclease TREX1 cause autosomal dominant retinal vasculopathy with cerebral leukodystrophy. *Nat Genet*. (2007) 39:1068–70. doi: 10.1038/ng2082

61. Vanpouille-Box C, Alard A, Aryankalayil MJ, Sarfraz Y, Diamond JM, Schneider RJ, et al. DNA exonuclease Trex1 regulates radiotherapy-induced tumour immunogenicity. *Nat Commun*. (2017) 8:15618. doi: 10.1038/ncomms15618

62. Fuertes MB, Kacha AK, Kline J, Woo S-R, Kranz DM, Murphy KM, et al. Host type I IFN signals are required for antitumor CD8+ T cell responses through CD8{alpha}+ dendritic cells. *J Exp Med*. (2011) 208:2005–16. doi: 10.1084/jem.20101159

63. Xanthopoulou ET, Koukourakis IM, Kakouratos C, Nanos C, Kalaitzis C, Giatromanolaki A, et al. Irradiation-induced IFN-type-I pathway activation in prostate cancer cell lines. *Cytokine*. (2023) 169:156252. doi: 10.1016/j.cyto.2023.156252

64. Rodriguez M, Luo W, Weng J, Zeng L, Yi Z, Siwko S, et al. PSGR promotes prostatic intraepithelial neoplasia and prostate cancer xenograft growth through NF-κB. *Oncogenesis*. (2014) 3:e114. doi: 10.1038/oncsis.2014.29

65. Fritsch M, Günther SD, Schwarzer R, Albert M-C, Schorn F, Werthenbach JP, et al. Caspase-8 is the molecular switch for apoptosis, necroptosis and pyroptosis. *Nature*. (2019) 575:683–7. doi: 10.1038/s41586-019-1770-6

66. Lubahn J, Berndt SI, Jin CH, Klim A, Luly J, Wu WS, et al. Association of CASP8 D302H polymorphism with reduced risk of aggressive prostate carcinoma. *Prostate*. (2010) 70:646–53. doi: 10.1002/pros.21098

67. Liu S, Garcia-Marques F, Zhang CA, Lee JJ, Nolley R, Shen M, et al. Discovery of CASP8 as a potential biomarker for high-risk prostate cancer through a high-multiplex immunoassay. *Sci Rep*. (2021) 11:7612. doi: 10.1038/s41598-021-87155-5

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 24 May 2024
DOI 10.3389/fimmu.2024.1369073

OPEN ACCESS

EDITED BY
Tinka Vidovic,
University of Zagreb, Croatia

REVIEWED BY
Snehal Dinkar Nirgude,
Children's Hospital of Philadelphia,
United States
Shinya Hasegawa,
Hoshi University, Japan

*CORRESPONDENCE
Yongwen Li
✉ leeywen@163.com
Hongyu Liu
✉ liuhongyu123@hotmail.com
Jun Chen
✉ huntercj2004@qq.com

†These authors have contributed
equally to this work and share
first authorship

RECEIVED 11 January 2024
ACCEPTED 13 May 2024
PUBLISHED 24 May 2024

CITATION
Ding C, Huang H, Wu D, Chen C, Hua Y, Liu J, Li Y, Liu H and Chen J (2024) Pan-cancer analysis predict that FAT1 is a therapeutic target and immunotherapy biomarker for multiple cancer types including non-small cell lung cancer. *Front. Immunol.* 15:1369073. doi: 10.3389/fimmu.2024.1369073

Pan-cancer analysis predict that FAT1 is a therapeutic target and immunotherapy biomarker for multiple cancer types including non-small cell lung cancer

Chen Ding[1†], Hua Huang[1†], Di Wu[1†], Chen Chen[2], Yu Hua[1], Jinghao Liu[1], Yongwen Li[2*], Hongyu Liu[2*] and Jun Chen[1,2*]

[1]Department of Lung Cancer Surgery, Tianjin Medical University General Hospital, Tianjin, China,
[2]Tianjin Key Laboratory of Lung Cancer Metastasis and Tumor Microenvironment, Tianjin Lung Cancer Institute, Tianjin Medical University General Hospital, Tianjin, China

FAT1, a substantial transmembrane protein, plays a pivotal role in cellular adhesion and cell signaling. Numerous studies have documented frequent alterations in FAT1 across various cancer types, with its aberrant expression being linked to unfavorable survival rates and tumor progression. In the present investigation, we employed bioinformatic analyses, as well as *in vitro* and *in vivo* experiments to elucidate the functional significance of FAT1 in pan-cancer, with a primary focus on lung cancer. Our findings unveiled FAT1 overexpression in diverse cancer types, including lung cancer, concomitant with its association with an unfavorable prognosis. Furthermore, FAT1 is intricately involved in immune-related pathways and demonstrates a strong correlation with the expression of immune checkpoint genes. The suppression of FAT1 in lung cancer cells results in reduced cell proliferation, migration, and invasion. These collective findings suggest that FAT1 has potential utility both as a biomarker and as a therapeutic target for lung cancer.

KEYWORDS

FAT1, pan-cancer, prognosis, bioinformatics, immune

Introduction

Cancer stands as a prominent global cause of mortality. Among various cancer types, lung cancer emerges as the primary contributor to cancer-related fatalities (1). Non-small cell lung cancer (NSCLC) constitutes approximately 85% of all lung cancer cases, making it the most prevalent histological subtype. Despite significant advancements in our understanding of the biological foundations of NSCLC, the integration of predictive biomarkers, and the refinement of treatment strategies, substantial progress has been achieved over the past two decades, resulting in improved outcomes for numerous patients

(2). However, individuals afflicted with advanced-stage disease still confront a grim prognosis. Hence, the identification of novel biomarkers and therapeutic targets capable of enhancing patient outcomes represents an urgent imperative.

FAT atypical cadherin 1 (FAT1), a gene encoding protocadherin, ranks among the most frequently mutated genes in human cancer. As a transmembrane protein, FAT1 assumes a pivotal role in cellular adhesion and signaling pathways (3). Located on chromosome 4q35, it functions as a tumor-promoting gene, exerting regulatory control over cell proliferation, migration, and invasion (4–7). Notably, FAT1 undergoes frequent alterations in various cancer types, including lung cancer, where its mutations correlate with unfavorable survival rates and tumor progression (8, 9). In a prior study, our research unveiled the potential of FAT1 mutations as predictive biomarkers in NSCLC, aiding in the identification of patients less likely to derive sustained clinical benefits from immune checkpoint blockade (ICB). We proposed an FAT1 mutation-based model for the screening of NSCLC patients more suitable for ICB, thereby contributing to individualized immunotherapy (10). Recent investigations have further illuminated the role of FAT1 alterations in multiple signaling pathways and critical cellular processes, such as the Hippo and Wnt signaling pathways, along with epithelial–mesenchymal transition (EMT), all of which play pivotal roles in tumorigenesis (11). Studies exploring FAT1's significance in cancer have examined its prognostic value, association with immune infiltration, and impact on the tumor microenvironment (12, 13). However, the functional implications of FAT1 alterations across various cancer types, as well as their potential as therapeutic biomarkers, require further elucidation.

This study employs bioinformatic methodologies, as well as the *in vitro* and *in vivo* experiments to explore the functional role of FAT1 across NSCLC, aiming to elucidate its expression patterns, genetic alterations, and functional networks. Our analysis provides a comprehensive examination of FAT1, aiming to elucidate the potential mechanisms through which FAT1 mediates tumorigenesis and assess its clinical relevance in a pan-cancer context. This encompasses an in-depth exploration of FAT1's biological role in non-small cell lung cancer, with a focus on investigating its impact on cell proliferation and migration, revealing its potential significance in the progression of NSCLC.

Materials and methods

NSCLC samples and next-generation sequencing

We collected 37 fresh tumor tissues from patients who were diagnosed with NSCLC at Tianjin Medical University General Hospital. Tissues were collected by surgery, and routinely processed with formalin fixation, embedded with paraffin. All cases were confirmed postoperative histological pathology with NSCLC. All patients provided written consent, and the research was approved by the Institutional Ethics Committee of the General Hospital of Tianjin Medical University. Tumor DNA was extracted from 5 to 10 10μm FFPE curls, and DNA quantification was performed using Qubit™ dsDNA HS and BR Assay Kits (Thermo Fisher Scientific, MA, USA). Target gene capture NGS technology was employed to detect 1267 genes related to lung cancer treatment plans, high-throughput sequencing data were obtained, and somatic mutations were identified by comparing with matched adjacent lung tissues. The NGS sequencing process was completed by Yuce Biological Technology Co., Ltd.

Data sources and processing for FAT1 mRNA expression

FAT1 expression was analyzed in 34 different tumor types and their corresponding normal tissues using a combination of The Cancer Genome Atlas (TCGA) and GTEx cohorts. For this analysis, SangerBox, a web-based tool (http://sangerbox.com/), was utilized to obtain FAT1 expression levels in different pathological stages (stages I–IV) of selected TCGA tumors through the "Pathological Stage Plot" module of Gene Expression Profiling Interactive Analysis. Violin plots were generated to depict the relationship between FAT1 expression levels and pathological stages. Survival analysis was performed using the Kaplan–Meier method and Cox proportional hazards regression analysis. Furthermore, protein expression levels were investigated using the UALCAN portal (http://ualcan.path.uab.edu/index.html) (14), which offers an interactive web resource for CPTAC analysis. The UALCAN portal utilizes the CPTAC database and normalizes logged expression values to standard deviations from the median in each proteomic profile.

Genetic analysis

To explore FAT1 genetic alterations in various cancer types, we analyzed somatic mutation data retrieved from cBioPortal (available at https://www.cbioportal.org/) (15). Through this analysis, information regarding the frequency and types of mutations in FAT1 across cancer types was acquired.

FAT1-related gene enrichment analysis

To conduct gene enrichment analysis related to FAT1, the STRING website was utilized (available at https://string-db.org/) (16). Furthermore, the "Similar Gene Detection" module of GEPIA2 was employed to generate a list of the top 100 FAT1-associated genes in TCGA tumors. Subsequently, the correlation analysis module of GEPIA2 was employed to explore the correlation between FAT1 and these top five FAT1-associated targeting genes. To perform pathway and process enrichment analysis for the identified top 100 FAT1-associated targeting genes, the Metascape web-based tool (17) was utilized, and specific parameters were selected, including $P < 0.01$, a minimum count of three for the terms, and an enrichment factor > 1.5 for canonical pathways.

Analysis of tumor immune and immunosuppressive cell infiltration

By using the TIMER2 server, we examined the correlation between FAT1 expression and the infiltration of various immune cell types. To evaluate the effect of genetic and epigenetic alterations of FAT1 on dysfunctional T-cell phenotypes, the QUERY module of the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm was utilized (18).

Epigenetic methylation analysis

To examine the differences in FAT1 methylation levels between tumor and paired normal tissues across various TCGA cancer types, the TCGA methylation module within the UALCAN interactive web resource was employed (19, 20). Furthermore, the TIDE server was utilized to investigate the effect of FAT1 methylation on dysfunctional T-cell phenotypes and prognoses.

Analysis of gene expression correlations with therapeutic responses

To assess the therapeutic potential of FAT1 as a target in various cancers, the drug sensitivity data obtained from the ROC Plotter was examined (http://www.rocplot.org/). The ROC Plotter is a transcriptome-based tool that enables the prediction of biomarkers by establishing connections between gene expressions and responses to therapy among patients with cancer (21).

Cell culture and transfection

The human cancer cell lines OVCAR3, Hep3B, PANC1, H1299, A549, as well as the embryonic kidney 293T were obtained from the American Type Culture Collection. These cell lines were cultured in DMEM supplemented with 10% fetal bovine serum and 1% penicillin–streptomycin in a humidified incubator at 37°C with 5% CO_2. To perform the transfection, siRNA specifically targeting FAT1 or a negative control siRNA (Ribobio, China) was introduced into the cells using Lipofectamine 2000 (Invitrogen, USA) following the manufacturer's instructions. The siRNAs targeting FAT1 were sense: 5′-GCACCACAAUUUCGAGCAATT-3′, antisense: 5′-UUGCUCGAAAUUGUGGUGCTT-3′.

Real-time polymerase chain reaction

According to the manufacturer's instructions, the cells were digested down with trypsin, and the total RNA was extracted using the SPARKeasy Cell RNA Rapid Extraction Kit (Sparkjade, Shandong, China). RNA concentration was measured, and a total of 2 μg of RNA was reverse-transcribed into complementary DNA using a reverse-transcription kit (Takara, Beijing, China) according to the manufacturer's instructions. Real-time PCR used a fully automated PCR analyzer SLAN-96P, with GAPDH as the internal reference. The mRNA primer sequences used were as follows: 5′-GGAGCGAGATCCCTCCAAAAT-3′ and 5′- GGCTGTTGTCATACTTCTCATGG -3′ for GAPDH; 5′-CATCCTGTCAAGATGGGTGTTT-3′ and 5′- TCCGAGAATGTACTCTTCAGCTT-3′ for FAT1.

Cell proliferation assay

Cell proliferation was assessed using the Cell Counting Kit-8 (CCK8) assay. Briefly, cells were seeded into 96-well plates and incubated for 24, 48, and 72 h. At each time point, 10 μL of the CCK8 solution was added to each well and incubated for 2 h. The absorbance was measured at 450 nm using a microplate reader.

5-ethynyl-2′-deoxyuridine assay

Cell proliferation was also assessed using the EdU assay. Briefly, the cells were subsequently stained with cell-Light™ EdU Apollo567 *In Vitro* Kit (Ribobio, Guangzhou, China) following the manufacturer's instructions. The images were captured using a fluorescence microscope, and the percentage of EdU-positive cells was calculated using ImageJ.

Colony formation assay

The cells were seeded into 6-well plates at a density of 500 cells per well, cultured for 14 days, and fixed with 4% paraformaldehyde and stained with crystal violet. The area of colonies was counted using ImageJ.

Transwell assay

Cell migration and invasion were evaluated using Transwell chambers. For the migration assay, cells were seeded into the upper chamber with a serum-free medium, whereas the lower chamber was filled with a medium containing 10% fetal bovine serum. After incubation for 24 h, the cells that migrated to the lower chamber were fixed with 4% paraformaldehyde and stained with crystal violet. For the invasion assay, Transwell chambers were coated with 20% Matrigel before cell seeding.

Scratch wound-healing assay

The cells were seeded into 6-well plates and cultured to confluence. A scratch wound was created using a sterile 200 μL pipette tip, and the cells were washed with phosphate-buffered saline to remove the debris. Then, the cells were incubated in serum-free medium, and wound closure was monitored at different time points using an inverted microscope.

Protein extraction and western blot analysis

Total protein was extracted, and protein concentration was measured by the BCA method. The proteins to be isolated were separated with 8% to 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis, transferred to polyvinylidene fluoride membrane, and incubated overnight at 4°C under the following primary antibodies: anti-β-TUBULIN (66240–1-Ig, Proteintech), anti-FAT1 (E95869, Sigma), anti-YAP (13584–1-AP, Proteintech), anti-P-TAZ (AF4315, Affinity), Anti-FAK (12636–1-AP, Proteintech), and anti-Src (11097–1-AP, Proteintech). Then, the membrane was incubated with anti-rabbit/mouse IgG (Abclonal, anti-rabbit: AS014, anti-mouse: AS003) secondary antibody at room temperature for 1 h, and bands developed on the membrane using Syngene G-Box and GeneSnap software (Syngene, Cambridge, UK).

Multicolor immunofluorescence

Formalin-fixed paraffin-embedded sections were fractionated alcohol dewaxed and rehydrated, and EDTA antigen retrieval buffer was subjected to antigen retrieval at 98°C for 8 min. Then the slides were soaked in 3% hydrogen peroxide for 15 minutes. CD63, CD168, α-SMA (Servicebio, Wuhan, Hubei, China) and FAT1 (Abcam, ab-190242) were applied at 4°C overnight, followed by incubation with secondary antibodies for 90 min at 37°C. Digital slide scanner (Pannoramic 250, 3DHistech, Hungary) were performed to scan samples.

Animal xenograft tumor experiment

The animals used were 4-week-old female nude mice with BALB/c (Hfkbio, Beijing, China). In a specific environment, nude mice were randomly divided into two groups, control group and sh-FAT1 group (n=5). A total of 2×10^6 cells were implanted subcutaneously in the right groin of nude mice, and when the tumor was clearly palpated, tumor volume measurement was started, and thereafter, tumor volume was measured every 2 days, and at the end of observation, the mouse subcutaneous tumor was excised, peeled and resected and images were collected. The tumor volume calculation formula used in this study is $(L \times W^2)/2$.

Statistical analysis

R software was used in the statistical analysis. Student's t-test was used to compare the expression levels of FAT1 between different groups, and the Wilcoxon rank-sum test was used to analyze non-normally distributed data. Pearson correlation analysis was employed to evaluate the correlation between FAT1 expression and immune infiltration.

Results

FAT1 is frequently mutated in NSCLC tissues

Our research team previously established that FAT1 mutations may serve as predictive biomarkers for identifying NSCLC patients who may not derive sustained clinical benefits from ICB, thereby laying the groundwork for the potential application of individualized immunotherapy screening (10). In this present study, we first evaluated FAT1 mutation status in NSCLC patients. By collecting FFPE tumor tissues from 37 individuals who diagnosed with NSCLC and utilizing NGS technology for targeted gene capture, we analyzed a panel of 1267 genes for those patients. Our analysis identified the presence of FAT1 mutations in five out of thirty seven samples, constituting a mutation frequency of 14%. Among these mutations, four were missense mutations, while one was a splice variant (Figure 1A). Notably, within the 1267 genes analyzed, FAT1 ranked third in terms of mutation frequency, following EGFR and P53. This finding suggests that FAT1 may play a significant role in the pathogenesis and progression of NSCLC, warranting further in-depth investigation.

FAT1 expression and its clinical significance in pan-cancer

To further explore whether aberrant expression of FAT1 associated with clinical significance in pan-cancer, we conducted an extensive analysis of FAT1 expression across various cancer types, utilizing publicly available databases, including the TCGA and GTEx databases. Our findings align with prior reports (22), demonstrating a frequent upregulation of FAT1 expression in most cancer types when compared to their normal counterparts, including lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) (Figure 1B). Notably, protein expression analysis revealed significant variations in FAT1 expression across various tumors (Figure 1C), highlighting its potential role as a target in tumorigenesis and development. Furthermore, our investigation unveiled a substantial association between elevated FAT1 expression and adverse overall survival (OS) outcomes in patients with several tumor types, including acute lymphocytic leukemia (ALL), adenoid cystic carcinoma (ACC), mesothelioma (MESO), LUAD, head and neck squamous cell carcinoma (HNSC), thyroid carcinoma (THCA), pancreatic adenocarcinoma (PAAD), and cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC) (Figure 2A; log-rank test, $P < 0.05$). Additionally, we observed a positive correlation between FAT1 expression and advanced pathological stages in ACC and Skin Cutaneous Melanoma (SKCM) (Figure 2B). Notably, metastatic testicular germ cell tumors (TGCT) and ACC exhibited higher FAT1 expression levels compared to their corresponding primary tumors (Figure 2C).

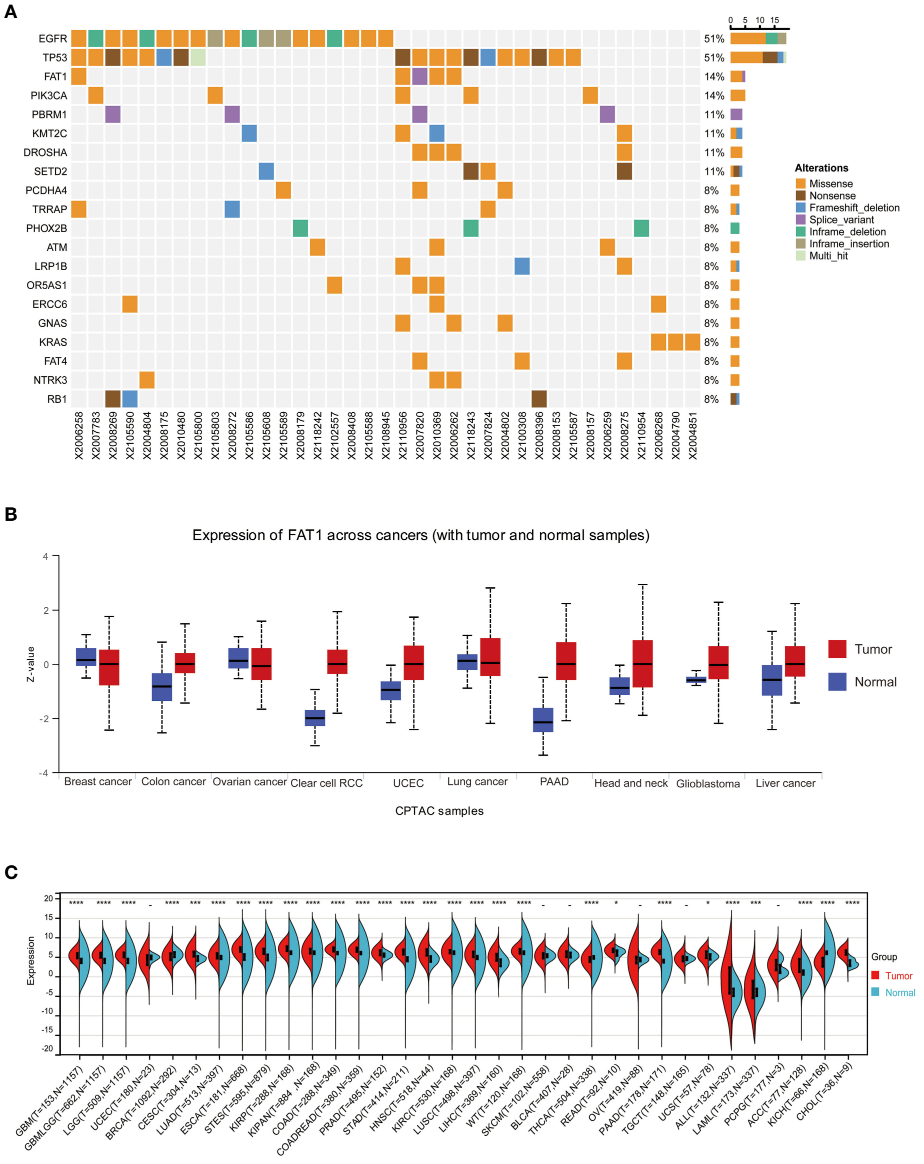


FIGURE 1
FAT1 expression in cancer. **(A)** Results of target gene capture second-generation sequencing in 37 lung cancer tissues. **(B)** FAT1 mRNA expression levels in various cancer types based on TCGA data. **(C)** Protein expression level of FAT1 in pan-cancer. *P < 0.05,***P < 0.001,****P < 0.0001.

DNA methylation analysis of FAT1 in pan-cancer

The promoter region of FAT1 displayed frequent hypermethylation in various cancer types, including kidney renal clear cell carcinoma (KIRC), LUAD, CESC, breast invasive carcinoma (BRCA), and LUSC. In contrast, the levels of FAT1 promoter methylation in Prostate adenocarcinoma (PRAD), liver hepatocellular carcinoma (LIHC), and colon adenocarcinoma (COAD) were lower when compared to their adjacent normal tissues (Figure 3A). This suggests that epigenetic silencing of FAT1 may play a role in its expression regulation in cancer. We conducted an in-depth analysis to examine the impact of FAT1 methylation on different cancer types and made intriguing observations. We found a correlation between FAT1 hypomethylation and dysfunctional T-cell phenotypes, as well as shorter survival in brain, lymphoma, and uveal cancers (Figure 3B). However, it's noteworthy that in the context of kidney cancer, FAT1 hypomethylation was associated with a favorable prognosis (Figure 3C).

Correlation analysis and pathway enrichment of FAT1 in pan-cancer

We employed the STRING tool to systematically screen for FAT1-binding proteins, aiming to uncover the potential role of FAT1 in tumor pathogenesis. The interaction network reveals associations with RXRA, PPARA, MYC, RELA, NFKR, NOS2, SP1, NR2F1, TNF, and JUN (Figure 4A). To gain deeper insights into FAT1's potential functions in cancer, we conducted correlation

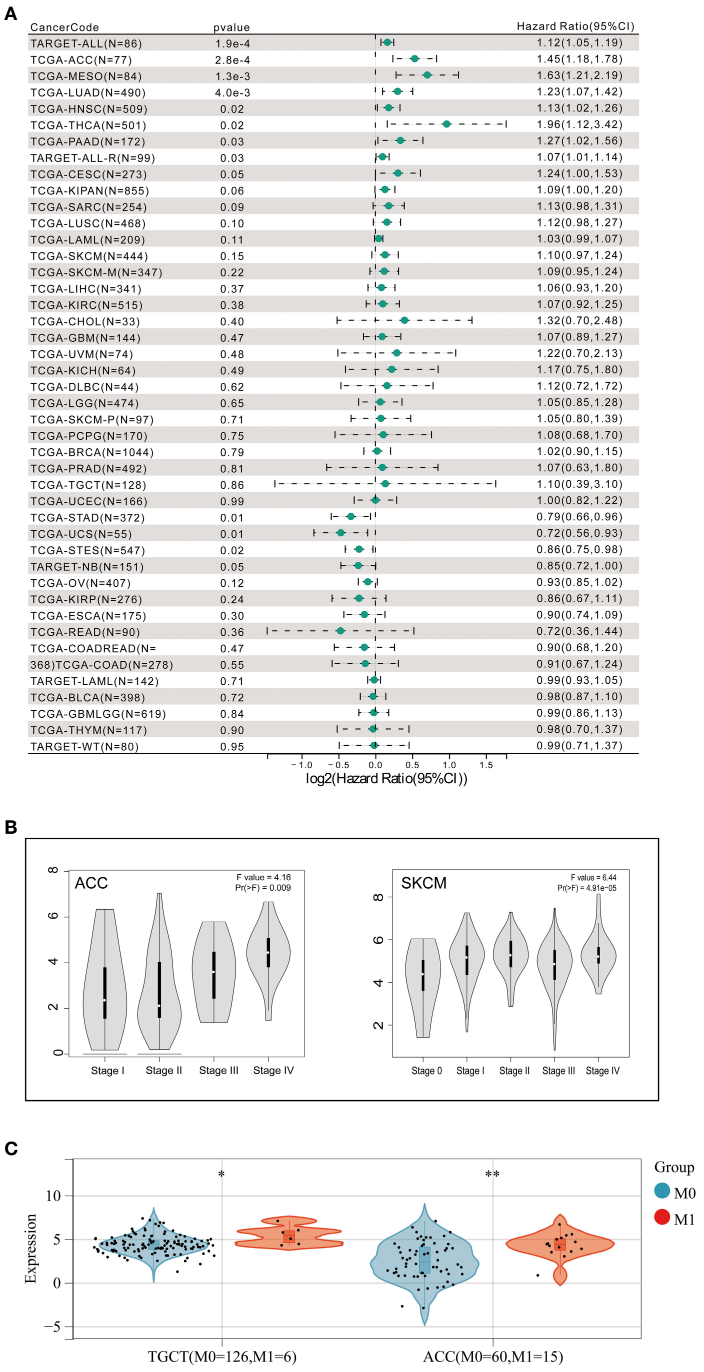
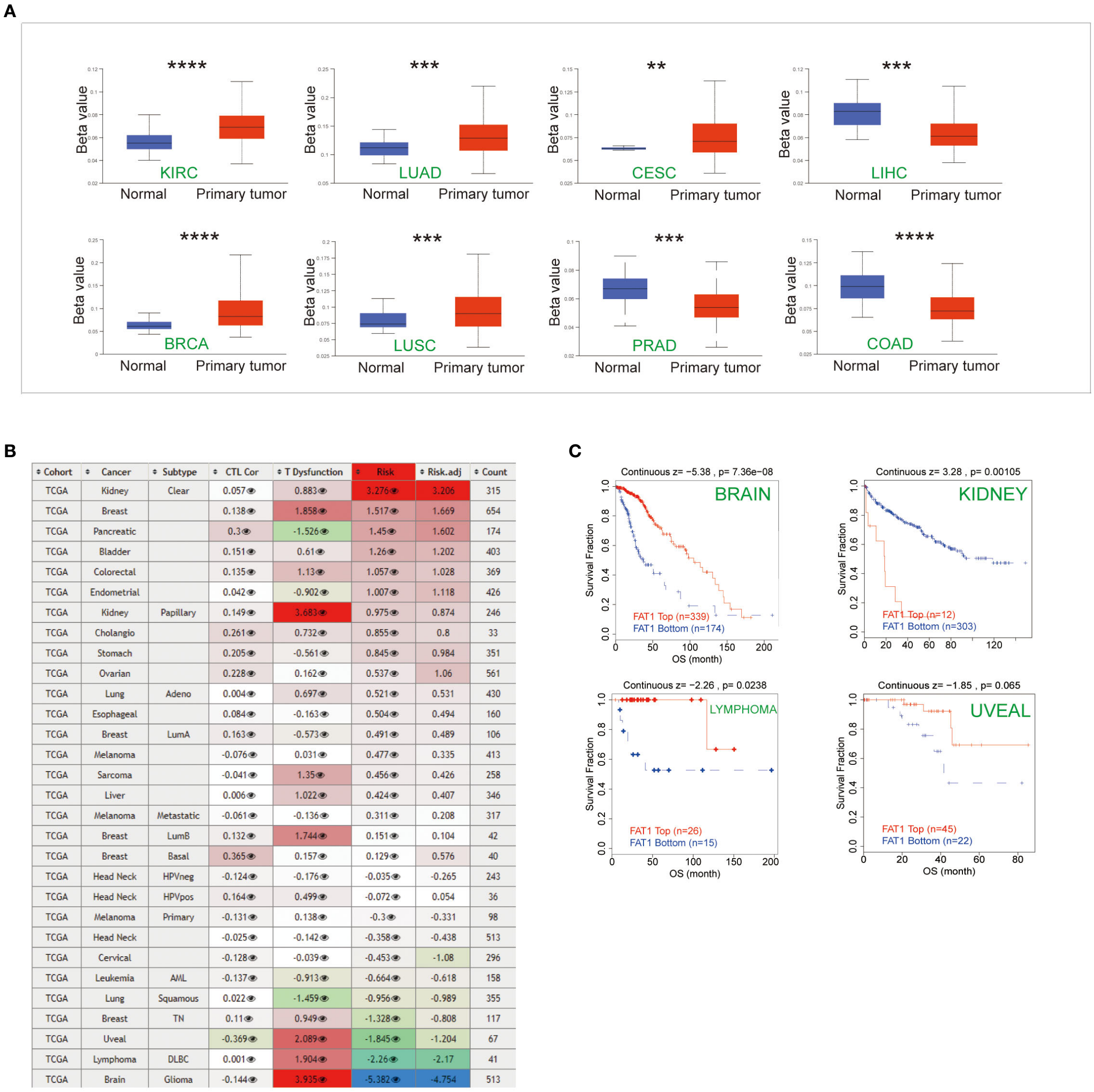


FIGURE 2
(A) Correlation between FAT1 expression and prognosis in various types of cancer using SangerBox. **(B)** Association of FAT1 gene expression levels with pathological stages and **(C)** metastasis. *P < 0.05,**P < 0.01

and pathway enrichment analyses using gene expression data from TCGA. Our analysis identified the top five genes significantly correlated with FAT1 expression: CARD10, CTTNBP2NL, F2RL1, MYO1E, and SPATS2L (Figure 4B). Additionally, an exploration of the top 100 FAT1-associated genes revealed significant associations with multiple cancer-related signaling pathways, encompassing cell–cell junction organization, integrin-mediated signaling pathways, and enzyme-linked receptor protein signaling pathways (Figure 4C).

Distinct immune microenvironment based on FAT expression

Tumor-infiltrating immune cells wield a pivotal influence within the tumor microenvironment, impacting cancer initiation, progression, and metastasis (23, 24). In our prior study, we unveiled the potential of FAT1 aberrant as predictive biomarkers in NSCLC, indicating the potential role of FAT1 regulating immune microenvironment. To explore the connection between FAT1



Cohort	Cancer	Subtype	CTL Cor	T Dysfunction	Risk	Risk.adj	Count
TCGA	Kidney	Clear	0.057	0.883	3.276	3.206	315
TCGA	Breast		0.138	1.858	1.517	1.669	654
TCGA	Pancreatic		0.3	-1.526	1.45	1.602	174
TCGA	Bladder		0.151	0.61	1.26	1.202	403
TCGA	Colorectal		0.135	1.13	1.057	1.028	369
TCGA	Endometrial		0.042	-0.902	1.007	1.118	426
TCGA	Kidney	Papillary	0.149	3.683	0.975	0.874	246
TCGA	Cholangio		0.261	0.732	0.855	0.8	33
TCGA	Stomach		0.205	-0.561	0.845	0.984	351
TCGA	Ovarian		0.228	0.162	0.537	1.06	561
TCGA	Lung	Adeno	0.004	0.697	0.521	0.531	430
TCGA	Esophageal		0.084	-0.163	0.504	0.494	160
TCGA	Breast	LumA	0.163	-0.573	0.491	0.489	106
TCGA	Melanoma		-0.076	0.031	0.477	0.335	413
TCGA	Sarcoma		-0.041	1.35	0.456	0.426	258
TCGA	Liver		0.006	1.022	0.424	0.407	346
TCGA	Melanoma	Metastatic	-0.061	-0.136	0.311	0.208	317
TCGA	Breast	LumB	0.132	1.744	0.151	0.104	42
TCGA	Breast	Basal	0.365	0.157	0.129	0.576	40
TCGA	Head Neck	HPVneg	-0.124	-0.176	-0.035	-0.265	243
TCGA	Head Neck	HPVpos	0.164	0.499	-0.072	0.054	36
TCGA	Melanoma	Primary	-0.131	0.138	-0.3	-0.331	98
TCGA	Head Neck		-0.025	-0.142	-0.358	-0.438	513
TCGA	Cervical		-0.128	-0.039	-0.453	-1.08	296
TCGA	Leukemia	AML	-0.137	-0.913	-0.664	-0.618	158
TCGA	Lung	Squamous	0.022	-1.459	-0.956	-0.989	355
TCGA	Breast	TN	0.11	0.949	-1.328	-0.808	117
TCGA	Uveal		-0.369	2.089	-1.845	-1.204	67
TCGA	Lymphoma	DLBC	0.001	1.904	-2.26	-2.17	41
TCGA	Brain	Glioma	-0.144	3.935	-5.382	-4.754	513

FIGURE 3
Epigenetic methylation analysis. **(A)** Boxplots illustrate the differential FAT1 methylation levels (beta values) across TCGA. **(B)** Heatmap demonstrating the effect of FAT1 methylation on cytotoxic T-cell levels (CTLs), dysfunctional T-cell phenotypes, and risk factors within TCGA cohorts. **(C)** FAT1 methylation levels at Kaplan-Meier survival curves. In kidney cancer, FAT1 hypomethylation is associated with a better prognosis, and in brain cancer, lymphoma, and uveal cancer, FAT1 hypomethylation has a shorter survival. **P < 0.01,***P < 0.001,****P < 0.0001

expression and immune cell infiltration, we conducted a comprehensive analysis across 39 cancer types. Our investigation revealed a significant positive correlation between FAT1 expression and the infiltration of six immune cell types—namely, B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and dendritic cells—in KIRC and Pheochromocytoma and Paraganglioma (PCPG) (Supplementary Figure S1A). Furthermore, we delved into the relationship between FAT1 expression levels and the infiltration of three immunosuppressive cell types known to promote T-cell exclusion: myeloid-derived suppressor cells (MDSCs), cancer-associated fibroblasts, and T-regulatory cells. Our findings indicated a positive correlation between FAT1 expression and the infiltration of MDSCs in several cancer types, including ACC, BRCA, BRCA-LumA, BRCA-LumB, Glioblastoma (GBM), Head and Neck Squamous Cell Carcinoma-Human Papillomavirus, Kidney Chromophobe, Kidney Renal Papillary Cell Carcinoma, and Low-Grade Gliomas (Supplementary Figure S1B). In our quest to assess FAT1's relevance as a biomarker, we compared it to established biomarkers with regard to their ability to predict response outcomes and OS in ICB subcohorts. Notably, FAT1 exhibited an area under the receiver operating characteristic curve (AUC) exceeding 0.5 in 6 out of 18 ICB subcohorts (Supplementary Figure S1C). Additionally, FAT1 expression displayed positive correlations with immune-related gene signatures, encompassing immune cell infiltration, immune checkpoint genes, and major histocompatibility complex class I expression, suggestive of FAT1's potential role in modulating the immune microenvironment of tumors. We conducted extensive correlation analyses between FAT1 expression and various genes,

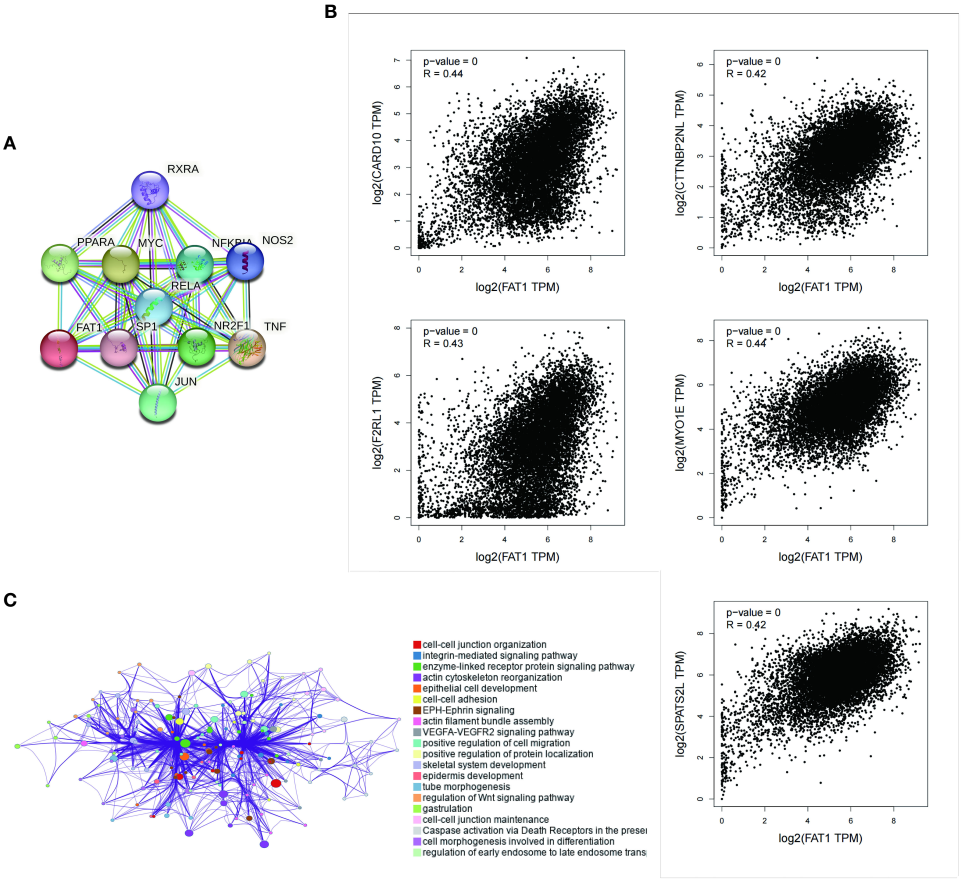


FIGURE 4
Enrichment analysis of FAT1-related genes across pan-cancers. **(A)** Displayed here are the FAT1-binding proteins identified using the STRING tool. **(B)** The top five FAT1-correlated genes are displayed across pan-cancers, and the relationships between FAT1 expression and these selected genes are analyzed using the GEPIA2 website. **(C)** Enriched terms with a similarity >0.3 are connected by edges.

including chemokines and their receptors (e.g., CXC and CC family) as well as major histocompatibility complex classes I and II (Supplementary Figure S2). Among the 26 cancer types with high FAT1 expression, we noted elevated expression of CD274 (PD-L1), particularly in Diffuse Large B Cell Lymphoma (DLBCL), Uveal Melanoma (UVM), GBM, and PAAD. Furthermore, FAT1 expression positively correlated with PD-L1 and CTLA-4 expression in LUAD, suggesting the potential involvement of FAT1 in regulating the expression of immune checkpoint genes in lung cancer.

Prediction of therapeutic response based on FAT1 expression

To explore the potential clinical utility of FAT1 as a biomarker, we conducted an assessment of its predictive value in gauging therapeutic responses among cancer patients. Our findings revealed a notable correlation between high FAT1 expression and an unfavorable survival to immune checkpoint inhibitors (Figure 5A). This observation suggests that FAT1 holds promise as a potential predictive biomarker for assessing responses to these treatments. Further analysis showed that elevated FAT1 expression was associated with shorter OS in patients treated with ICB for bladder and melanoma cancers. Moreover, heightened levels of FAT1 expression exhibited a negative association with cytotoxic T lymphocyte (CTL) (Figure 5B).

Altered FAT1 expression changes tumor immune microenvironment

To explore the functional implications of FAT1 in the regulation of the tumor immune microenvironment, we initiated our investigation by assessing FAT1 expression levels across various cell lines, including OVCAR3, Hep3B, PANC1, H1299, A549, and 293T, utilizing Western blot analysis. Our findings revealed the highest expression of FAT1 in the A549 and H1299 cell lines, as illustrated in Figure 6A. Subsequently, we focused our *in vitro* experiments on these two lung cancer cell lines. To ensure the specificity and efficacy of the siRNAs targeting FAT1, we employed quantitative real-time polymerase chain reaction (qRT-PCR) and Western blot analyses. As expected, the siRNAs demonstrated significant reduction in FAT1 expression in both A549 and H1299 cells compared to the control group (Figures 6B, C). Cancer immunotherapy, with a particular emphasis on targeting

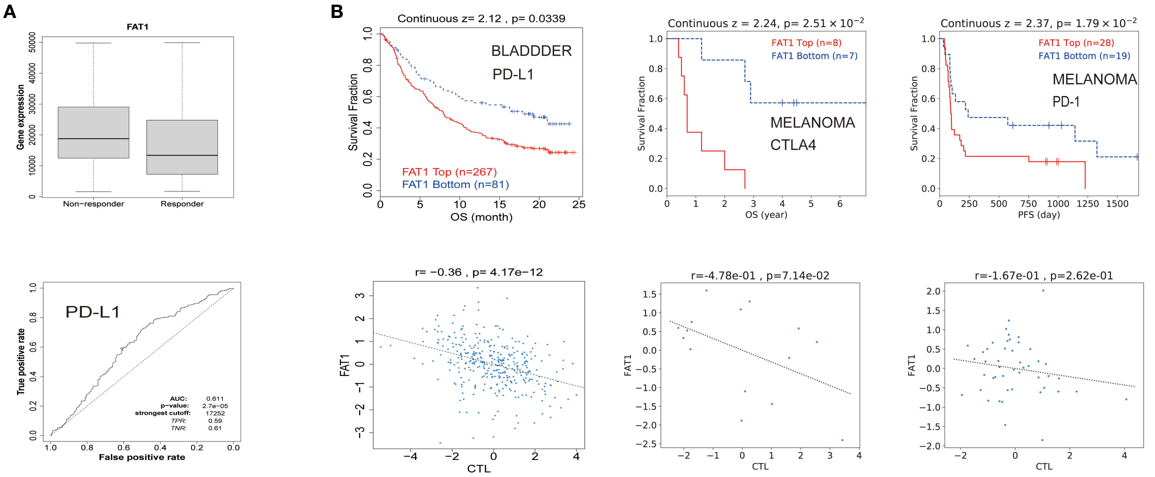


FIGURE 5

FAT1 expression is associated with therapeutic responses in cancers. **(A)** Receiver operating characteristic curve plot of the association between FAT1 expression and responses to PD-L1 in cancers. **(B)** The Kaplan-Meier survival curve showed shorter OS with high FAT1 expression in bladder cancer and melanoma receiving immunotherapy. The lower panel displays the correlation between FAT1 expression and cytotoxic T lymphocytes.

the programmed death 1/programmed death ligand 1 (PD-1/PDL1) pathway, has exhibited remarkable therapeutic efficacy in lung cancer patients (25). The level of PD-L1 expression in tumor cells has emerged as a pivotal indicator of the effectiveness of PD-1/PD-L1 blockade (26), we elucidate the relationship between FAT1 expression and PD-L1. Our results indicated a substantial decrease in PD-L1 expression following the knockdown of FAT1 expression (Figure 6D). Flow cytometry analysis of PD-L1 expression also yielded similar results (Figures 6E, F). Recognizing the pivotal role of the immune microenvironment in tumor development (27), we further explored the correlation of FAT1 expression with myeloid-derived suppressor cells (MDSC), macrophage M1, macrophage M2, and cancer-associated fibroblasts (CAF) in lung cancer patient tissues through immunofluorescence assay. Intriguingly, our findings demonstrated a positive correlation between FAT1 expression and the markers CD68, CD163, and α-SMA, indicative of the essential role of FAT1 in immune regulation within NSCLC (Figure 6G). These results collectively underscore the critical involvement of FAT1 in orchestrating immune responses in the context of NSCLC.

FAT1 accelerated the proliferation and migration of lung cancer cells *in vitro*

To explore the biological role of FAT1 in NSCLC, we focused on investigating its effects on cell proliferation and migration. Employing the CCK-8 assay, we observed a pronounced inhibition of cell proliferation in both A549 and H1299 cells upon FAT1 knockdown, as evidenced by significantly reduced cell viability compared to the control group (Figure 7A). This anti-proliferative effect was further corroborated by the EdU assay, revealing a notable decrease in the proportion of EdU-positive cells following FAT1 knockdown, indicative of suppressed cell proliferation (Figure 7B). Furthermore, our investigation extended to colony formation assays, where FAT1 knockdown exhibited a significant impediment to both the number of colonies in both cell lines. These findings underscore the pivotal role of FAT1 in promoting not only lung cancer cell growth but also colony formation (Figure 7C). Delving into the realm of cell migration and invasion, the Transwell assay demonstrated a marked reduction in the number of migrating and invading cells in both A549 and H1299 cell lines following FAT1 knockdown, implicating FAT1 in the facilitation of lung cancer cell migration and invasion (Figure 7D). This observation was further supported by the cell scratch assay, where FAT1 knockdown attenuated the wound-healing ability of cells, indicating a weakened capacity for migration compared to the control group (Figure 7E). Moreover, our exploration revealed that FAT1 knockdown induced G0/G1 phase cell cycle arrest (Figure 8A). To shed light on the underlying molecular mechanisms, Western blot experiments targeting integrin-related pathways were conducted based on pathway enrichment results. The outcomes suggested that knocking down FAT1 may exert its effects on cell growth and proliferation through the FAK–YAP/TAZ pathway (Figure 8B). To further validate the impact of FAT1 on the FAK-YAP/TAZ pathway, immunohistochemical staining was performed on mouse tumor tissues, revealing consistent results (Figure 8E). This intricate interplay emphasizes the multifaceted involvement of FAT1 in orchestrating cellular processes crucial for the progression of NSCLC.

Knockdown of FAT1 inhibited the proliferation of lung cancer cells *in vivo*

To further confirm the effect of inhibiting FAT1 expression on cell proliferation observed *in vitro*, we conducted xenograft tumor experiments to assess the effect of FAT1 knockdown on lung cancer cell growth *in vivo*. The results demonstrated that knocking down FAT1 inhibited the growth of lung cancer cells *in vivo* (Figures 8C, D). These comprehensive findings collectively underscore the role

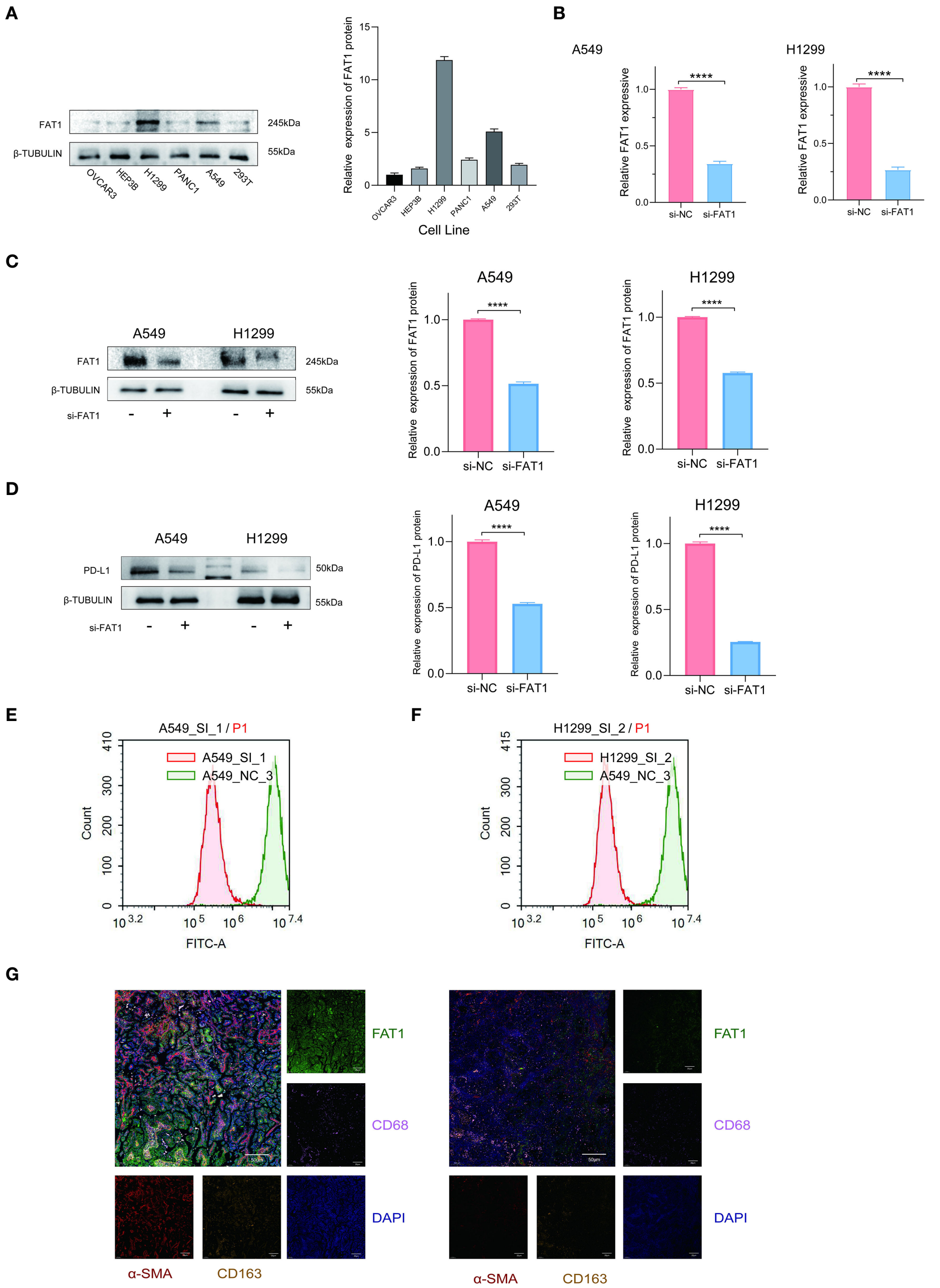


FIGURE 6
FAT1 knockdown effects in lung cancer cells. **(A)** Western blotting detects FAT1 expression in various cancer cell lines. **(B, C)** Western blot and qRT-PCR confirming FAT1 knockdown efficiency in lung cancer cells. Two-sided t-tests. **(D)** Western blot results verified the relationship between FAT1 expression and PD-L1. **(E, F)** Flow cytometry analysis to investigate the relationship between FAT1 and PD-L1 expression. **(G)** Multicolor immunofluorescence compared CD68, CD163 and α-SMA between FAT1 High (left) and Low (right) NSCLC tissue samples; ****P < 0.0001.

of FAT1 as a tumor promoter in lung cancer cells, with its downregulation leading to the inhibition of cell proliferation.

Discussion

FAT1, a sizable transmembrane protein, plays a significant role in diverse biological processes, encompassing cell adhesion, migration, and proliferation (28, 29). Extensive research has unveiled frequent upregulation of FAT1 in numerous cancer types, including lung cancer, suggesting its potential as a pivotal contributor to tumorigenesis (30, 31). Consequently, FAT1 emerges as a promising candidate for both therapeutic intervention and prognostic assessment in cancer treatment. Under normal circumstances, FAT1 functions as a molecular 'brake' on mitochondrial respiration and serves as a receptor involved in the

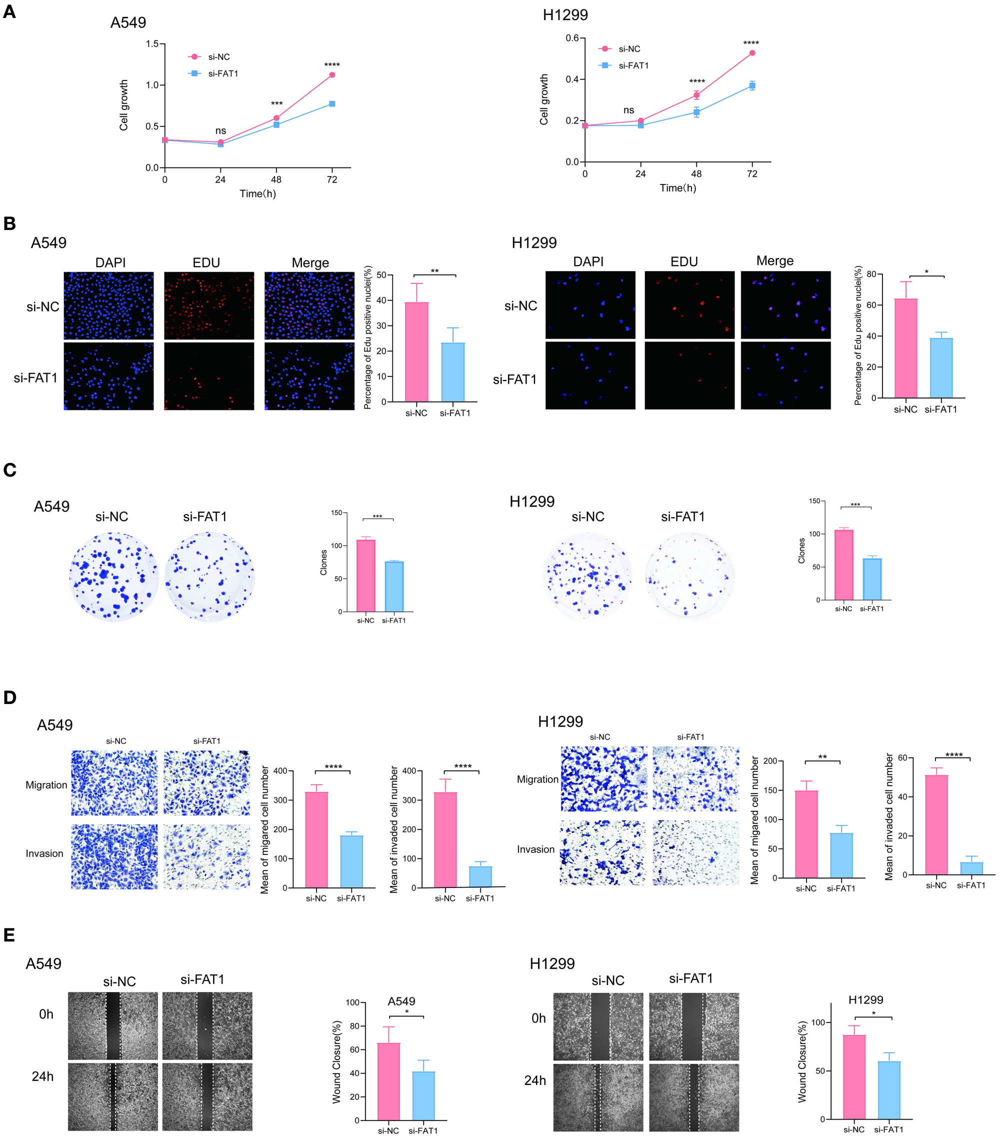


FIGURE 7
FAT1 knockdown effects in lung cancer cells. **(A)** CCK-8 assay measuring cell proliferation in FAT1-knockdown and control cells, two-sided t-tests. **(B)** EdU assay measuring cell proliferation in FAT1-knockdown and control cells, two-sided t-tests. **(C)** Colony formation assay showed that FAT1 knockdown inhibits the colony formation ability of cells. Two-sided t-tests; **(D)** Transwell migration and invasion assays showing decreased migration and invasion of FAT1-knockdown cells compared with control cells, two-sided t-tests. **(E)** Wound-healing experiments confirmed that knocking down FAT1 inhibits cell wound-healing ability, two-sided t-tests. $^{ns}P > 0.05$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

regulation of cell–cell contact interactions and planar cell polarity (22, 29, 32). In several cancer types, the loss of FAT1 function contributes to epithelial-mesenchymal transition (EMT) and the emergence of cancer-initiating or stem-like cells (8, 33–35). Conversely, in specific cancer types, FAT1 overexpression induces EMT (30). However, the precise roles of FAT1 in cancer progression remain intricate and contingent on the specific cancer type. Consequently, further investigations are warranted to attain a comprehensive understanding of FAT1's function within distinct cancer contexts. In this study, we conducted a comprehensive investigation into the expression patterns and clinical implications of FAT1 across a spectrum of cancer types employing bioinformatics analysis. Additionally, we substantiated its pro-tumorigenic role through experimental validation in lung cancer cells. Our findings, in alignment with prior research, consistently demonstrated a prevalent upregulation of FAT1 expression in a majority of cancer types, including LUAD and LUSC. Notably, we established a significant association between elevated FAT1 expression and reduced OS in LUAD patients, thus hinting at the potential prognostic utility of FAT1 in lung cancer. Furthermore, our study revealed a positive correlation between FAT1 expression and advanced pathological stage as well as metastasis in various cancer types, providing additional substantiation for the putative role of FAT1 in tumor progression and metastasis. Additionally, we illuminated potential mechanisms underpinning FAT1 upregulation in cancer, such as promoter

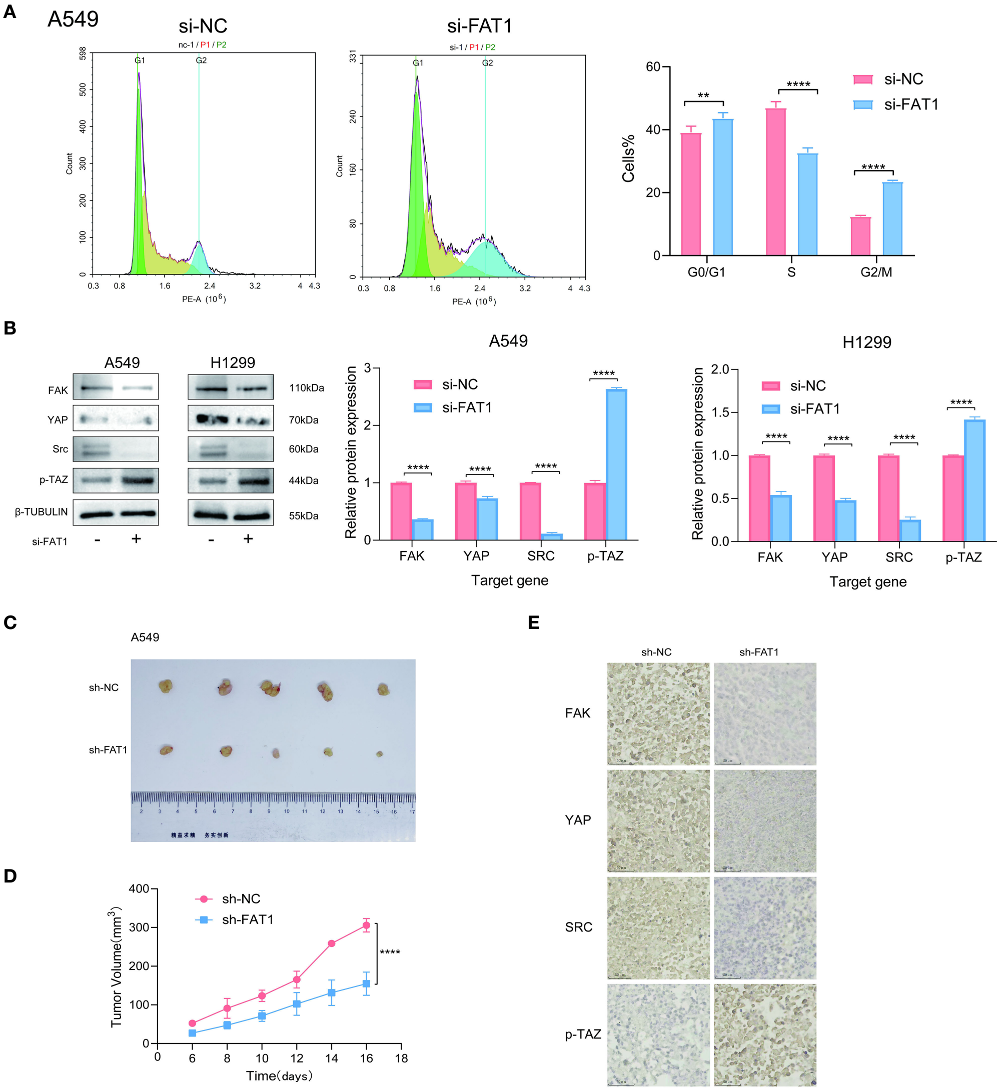


FIGURE 8

FAT1 knockdown effects in lung cancer cells. **(A)** Cell cycle assay showed that FAT1 knockdown induced G0/G1 phase arrest. **(B)** Western blot results verifying the effect of knockdown FAT1 on the FAK–YAP/TAZ signaling pathway. **(C)** Immunohistochemical validation of the effect of FAT1 knockdown on the FAK-YAP/TAZ signaling pathway. **(D, E)** Pictures of dissected mouse tumors and knocking down FAT1 inhibits tumor growth in mice. **P < 0.01, ****P < 0.0001.

hypermethylation—a phenomenon observed across multiple cancer types. Interestingly, our investigation also unveiled a positive correlation between FAT1 expression and immune-related gene signatures, suggesting a potential involvement of FAT1 in modulating the immune microenvironment within tumors. Furthermore, our analysis revealed a distinctive set of genes and signaling pathways that exhibited significant correlations with FAT1 expression, thereby providing additional insights into the potential functional roles of FAT1. Remarkably, our investigation unveiled FAT1's association with pathways linked to tumor occurrence and development, including cell–cell junction organization, the integrin-mediated signaling pathway, and the enzyme-linked receptor protein signaling pathway. Integrin-mediated cell migration predominantly relies on the activation of the FAK/Src signaling pathway, which, in turn, contributes to the regulation of several key signaling cascades governing cell motility (36). Noteworthy, previous studies have identified YAP/TAZ as downstream molecules of the Hippo signaling pathway, exerting control over cell proliferation and apoptosis, thus playing a pivotal role in tumor growth regulation. Integrin-FAK/Src activation has been shown to enhance YAP activation, leading to the accumulation and activation of YAP/TAZ, further promoting the proliferation and metastasis of malignant tumors (37, 38). Additionally, our assessment of the correlation between FAT1 expression and immune checkpoint genes hints at the potential utility of FAT1 as a predictive marker for immunotherapy efficacy. To gain deeper insights into FAT1's implications in cancer, we conducted *in vitro* functional assays utilizing lung cancer cell lines. Targeting FAT1 expression through knockdown in these cells resulted in a notable reduction in cell proliferation, migration, and invasion, accompanied by cell cycle inhibition. These findings strongly substantiate an oncogenic role for FAT1 in lung cancer and suggest that targeting FAT1 may hold promise in enhancing the effectiveness of chemotherapy in lung cancer treatment.

In conclusion, our study conducted a comprehensive bioinformatics analysis to investigate the functional role of FAT1 across diverse cancer types, with a particular focus on lung cancer. The insights derived from our findings underscore the potential utility of FAT1 as both a biomarker and a therapeutic target not only in lung cancer but also in various other cancer types.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving humans were approved by Institutional Ethics Committee of the General Hospital of Tianjin Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin. The animal study was approved by Ethics Committee of Tianjin Medical University General Hospital. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

CD: Validation, Writing – review & editing. HH: Writing – review & editing. DW: Data curation, Writing – original draft. CC: Writing – review & editing. YH: Writing – original draft. JL: Funding acquisition, Writing – review & editing. YL: Methodology, Writing – review & editing. HL: Methodology, Writing – review & editing. JC: Methodology, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Project Foundation of Tianjin Education Committee (2020KJ152).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1369073/full#supplementary-material

References

1. Bade BC, Dela Cruz CS. Lung cancer 2020: epidemiology, etiology, and prevention. *Clin Chest Med*. (2020) 41:1–24. doi: 10.1016/j.ccm.2019.10.001

2. da Cunha Santos G, Shepherd FA, Tsao MS. EGFR mutations and lung cancer. *Annu Rev Pathol.* (2011) 6:49–69. doi: 10.1146/annurev-pathol-011110-130206

3. Dunne J, Hanby AM, Poulsom R, Jones TA, Sheer D, Chin WG, et al. Molecular cloning and tissue expression of FAT, the human homologue of the Drosophila fat gene that is located on chromosome 4q34-q35 and encodes a putative adhesion molecule. *Genomics.* (1995) 30:207–23. doi: 10.1006/geno.1995.9884

4. Sadeqzadeh E, de Bock CE, Thorne RF. Sleeping giants: emerging roles for the fat cadherins in health and disease. *Med Res Rev.* (2014) 34:190–221. doi: 10.1002/med.21286

5. Nusse R, Clevers H. Wnt/beta-catenin signaling, disease, and emerging therapeutic modalities. *Cell.* (2017) 169:985–99. doi: 10.1016/j.cell.2017.05.016

6. Perugorria MJ, Olaizola P, Labiano I, Esparza-Baquer A, Marzioni M, Marin JJG, et al. Wnt-beta-catenin signalling in liver development, health and disease. *Nat Rev Gastroenterol Hepatol.* (2019) 16:121–36. doi: 10.1038/s41575-018-0075-9

7. Morin PJ, Sparks AB, Korinek V, Barker N, Clevers H, Vogelstein B, et al. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science.* (1997) 275:1787–90. doi: 10.1126/science.275.5307.1787

8. Pastushenko I, Mauri F, Song Y, de Cock F, Meeusen B, Swedlund B. Fat1 deletion promotes hybrid EMT state, tumour stemness and metastasis. *Nature.* (2021) 589:448–55. doi: 10.1038/s41586-020-03046-1

9. Feng Z, Yin Y, Liu B, Zheng Y, Shi D, Zhang H, et al. Prognostic and immunological role of FAT family genes in non-small cell lung cancer. *Cancer Control.* (2022) 29:10732748221076682. doi: 10.1177/10732748221076682

10. Zhu G, Ren D, Lei X, Shi R, Zhu S, Zhou N, et al. Mutations associated with no durable clinical benefit to immune checkpoint blockade in non-S-cell lung cancer. *Cancers (Basel).* (2021) 13:1397. doi: 10.3390/cancers13061397

11. Peng Z, Gong Y, Liang X. Role of FAT1 in health and disease. *Oncol Lett.* (2021) 21:398. doi: 10.3892/ol

12. Irshad K, Srivastava C, Malik N, Arora M, Gupta Y, Goswami S, et al. Upregulation of atypical cadherin FAT1 promotes an immunosuppressive tumor microenvironment via TGF-beta. *Front Immunol.* (2022) 13:813888. doi: 10.3389/fimmu.2022.813888

13. Zhang W, Tang Y, Guo Y, Kong Y, Shi F, Sheng C, et al. Favorable immune checkpoint inhibitor outcome of patients with melanoma and NSCLC harboring FAT1 mutations. *NPJ Precis Oncol.* (2022) 6:46. doi: 10.1038/s41698-022-00292-6

14. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVSK, et al. UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia*. (2017) 19:649–58. doi: 10.1016/j.neo.2017.05.002

15. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*. (2012) 2:401–4. doi: 10.1158/2159-8290.CD-12-0095

16. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. (2021) 49: D605–12. doi: 10.1093/nar/gkaa1074

17. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. (2019) 10:1523. doi: 10.1038/s41467-019-09234-6

18. Fu J, Li K, Zhang W, Wan C, Zhang J, Jiang P, et al. Large-scale public data reuse to model immunotherapy response and resistance. *Genome Med*. (2020) 12:21. doi: 10.1186/s13073-020-0721-z

19. Shinawi T, Hill VK, Krex D, Schackert G, Gentle D, Morris MR, et al. DNA methylation profiles of long- and short-term glioblastoma survivors. *Epigenetics*. (2013) 8:149–56. doi: 10.4161/epi.23398

20. Men C, Chai H, Song X, Li Y, Du H, Ren Q. Identification of DNA methylation associated gene signatures in endometrial cancer via integrated analysis of DNA methylation and gene expression systematically. *J Gynecol Oncol*. (2017) 28:e83. doi: 10.3802/jgo.2017.28.e83

21. Fekete JT, Gyorffy B. ROCplot.org: Validating predictive biomarkers of chemotherapy/hormonal therapy/anti-HER2 therapy using transcriptomic data of 3,104 breast cancer patients. *Int J Cancer*. (2019) 145:3140–51. doi: 10.1002/ijc.32369

22. Hou R, Liu L, Anees S, Hiroyasu S, Sibinga NE. The Fat1 cadherin integrates vascular smooth muscle cell growth and migration signals. *J Cell Biol*. (2006) 173:417–29. doi: 10.1083/jcb.200508121

23. Mao X, Xu J, Wang W, Liang C, Hua J, Liu J, et al. Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. *Mol Cancer*. (2021) 20:131. doi: 10.1186/s12943-021-01428-1

24. Petitprez F, Meylan M, de Reyniès A, Sautès-Fridman C, Fridman WH. The tumor microenvironment in the response to immune checkpoint blockade therapies. *Front Immunol*. (2020) 11:784. doi: 10.3389/fimmu.2020.00784

25. Cortez MA, Ivan C, Valdecanas D, Wang X, Peltier HJ, Ye Y, et al. PDL1 Regulation by p53 via miR-34. *J Natl Cancer Inst*. (2016) 108:djv303. doi: 10.1093/jnci/djv303

26. Yi M, Niu M, Xu L, Luo S, Wu K. Regulation of PD-L1 expression in the tumor microenvironment. *J Hematol Oncol*. (2021) 14:10. doi: 10.1186/s13045-020-01027-5

27. Zhang J, Tavakoli H, Ma L, Li X, Han L, Li X. Immunotherapy discovery on tumor organoid-on-a-chip platforms that recapitulate the tumor microenvironment. *Adv Drug Delivery Rev*. (2022) 187:114365. doi: 10.1016/j.addr.2022.114365

28. Chen ZG, Saba NF, Teng Y. The diverse functions of FAT1 in cancer progression: good, bad, or ugly? *J Exp Clin Cancer Res*. (2022) 41:248. doi: 10.1186/s13046-022-02461-8

29. Riascos-Bernal DF, Maira A, Sibinga NES. The atypical cadherin FAT1 limits mitochondrial respiration and proliferation of vascular smooth muscle cells. *Front Cardiovasc Med*. (2022) 9:905717. doi: 10.3389/fcvm.2022.905717

30. Srivastava C, Irshad K, Dikshit B, Chattopadhyay P, Sarkar C, Gupta DK, et al. FAT1 modulates EMT and stemness genes expression in hypoxic glioblastoma. *Int J Cancer*. (2018) 142:805–12. doi: 10.1002/ijc.31092

31. Zhang W, Ji K, Min C, Zhang C, Yang L, Zhang Q, et al. Oncogenic LINC00857 recruits TFAP2C to elevate FAT1 expression in gastric cancer. *Cancer Sci*. (2023) 114:63–74. doi: 10.1111/cas.15394

32. Cao LL, Riascos-Bernal DF, Chinnasamy P, Dunaway CM, Hou R, Pujato MA, et al. Control of mitochondrial function and cell growth by the atypical cadherin Fat1. *Nature*. (2016) 539:575–8. doi: 10.1038/nature20170

33. Jiang S, Zhu Y, Chen Z, Huang Z, Liu B, Xu Y, et al. S100A14 inhibits cell growth and epithelial-mesenchymal transition (EMT) in prostate cancer through FAT1-mediated Hippo signaling pathway. *Hum Cell*. (2021) 34:1215–26. doi: 10.1007/s13577-021-00538-8

34. Meng P, Zhang YF, Zhang W, Chen X, Xu T, Hu S, et al. Identification of the atypical cadherin FAT1 as a novel glypican-3 interacting protein in liver cancer cells. *Sci Rep*. (2021) 11:40. doi: 10.1038/s41598-020-79524-3

35. Hu X, Zhai Y, Kong P, Cui H, Yan T, Yang J, et al. FAT1 prevents epithelial mesenchymal transition (EMT) via MAPK/ERK signaling pathway in esophageal squamous cell cancer. *Cancer Lett*. (2017) 397:83–93. doi: 10.1016/j.canlet.2017.03.033

36. Fournier AK, Campbell LE, Castagnino P, Liu WF, Chung BM, Weaver VM, et al. Rac-dependent cyclin D1 gene expression regulated by cadherin- and integrin-mediated adhesion. *J Cell Sci*. (2008) 121:226–33. doi: 10.1242/jcs.017012

37. Johnson R, Halder G. The two faces of Hippo: targeting the Hippo pathway for regenerative medicine and cancer treatment. *Nat Rev Drug Discovery*. (2014) 13:63–79. doi: 10.1038/nrd4161

38. Ma H, Wang J, Zhao X, Wu T, Huang Z, Chen D, et al. Periostin promotes colorectal tumorigenesis through integrin-FAK-src pathway-mediated YAP/TAZ activation. *Cell Rep*. (2020) 30:793–806e6. doi: 10.1016/j.celrep.2019.12.075

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 29 August 2024
DOI 10.3389/fimmu.2024.1347770

OPEN ACCESS

EDITED BY
Kevin M. Mcbride,
University of Texas MD Anderson Cancer Center, United States

REVIEWED BY
Petar Ozretić,
Rudjer Boskovic Institute, Croatia
Ivana Samarzija,
Rudjer Boskovic Institute, Croatia

*CORRESPONDENCE
Miao Q. Zhao
zhaomqsd@163.com

†These authors have contributed equally to this work

RECEIVED 01 December 2023
ACCEPTED 08 August 2024
PUBLISHED 29 August 2024

CITATION
Ru K, Cui L, Wu C, Tan XX, An WT, Wu Q, Ma YT, Hao Y, Xiao X, Bai J, Liu X, Xia XF and Zhao MQ (2024) Exploring the molecular and immune landscape of cellular senescence in lung adenocarcinoma. *Front. Immunol.* 15:1347770. doi: 10.3389/fimmu.2024.1347770

Exploring the molecular and immune landscape of cellular senescence in lung adenocarcinoma

Kun Ru[1†], Liang Cui[2†], Cong Wu[1], Xin X. Tan[3,4], Wen T. An[5], Qiang Wu[6], Yu T. Ma[1], Yu Hao[1], Xiao Xiao[7], Jing Bai[2], Xiang Liu[8], Xue F. Xia[2] and Miao Q. Zhao[1*]

[1]Department of Pathology, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, China, [2]Geneplus-Beijing Institute, Beijing, China, [3]Geneplus-Shenzhen Clinical Laboratory, Shenzhen, China, [4]Xiangya School of Pharmaceutical Sciences, Central South University, Changsha, China, [5]Key Laboratory Experimental Teratology of the Ministry of Education, Department of Physiology, School of Basic Medical Sciences, Cheeloo College of Medicine, Shandong University, Jinan, China, [6]School of Tropical Medicine and Laboratory Medicine, Hainan Medical University, Haikou, China, [7]Geneplus-Shenzhen Institute, Shenzhen, China, [8]Department of Thoracic Surgery, The Second Affiliated Hospital, Hengyang Medical School, University of South China, Hengyang, China

Introduction: The connection between aging and cancer is complex. Previous research has highlighted the association between the aging process of lung adenocarcinoma (LUAD) cells and the immune response, yet there remains a gap in confirming this through single-cell data validation. Here, we aim to develop a novel aging-related prognostic model for LUAD, and verify the alterations in the genome and immune microenvironment linked to cellular senescence.

Methods: We integrated a comprehensive collection of senescence genes from the GenAge and CellAge databases and employed the least absolute shrinkage and selection operator (LASSO) Cox analysis to construct and validate a novel prognostic model for LUAD. This model was then utilized to examine the relationship between aging, tumor somatic mutations, and immune cell infiltration. Additionally, we explored the heterogeneity of senescence and intercellular communication within the LUAD tumor microenvironment (TME) through single-cell transcriptomic data analysis.

Results: By exploring the expression profiles of 586 cellular senescence-related genes in 428 LUAD patients, we constructed an aging-related genes (ARGs) risk model included 10 ARGs and validated it as an independent prognostic predictor

---

Abbreviations: ARGs, aging-related marker genes; ARKGs, aging-related key genes; ARRSs, aging-related risk scores; CI, confidence interval; CNVs, copy number variations; DDR, DNA damage repair; DEGs, differentially expressed genes; HR, hazard ratio; IQR, interquartile range; LUAD, lung adenocarcinoma; SASP, senescence-associated secretory phenotype; SRS, senescence-related signature; TMB, tumor mutation burden; TME, tumor microenvironment.

for LUAD patients. Notably, patients with low aging scores (LAS group) exhibited better survival, lower tumor mutation burden (TMB), lower somatic mutation frequency, lower tumor proliferation rate, and an immune activated phenotype compared to patients with high aging scores (HAS group). While the HAS group was enriched in tumor cells and showed a lower infiltration of CD8-CCR7, CD8-CXCL13, CD8-GNLY, FCGR3A NK cells, XCL1 NK cells, plasma cell (PC) and other immune subsets. Furthermore, the SPP1 and TENASCIN pathways, associated with tumor immune escape and tumor progression, were also enriched in the HAS group. Additionally, our study also indicated that senescence levels were heterogeneous in the LUAD tumor microenvironment (TME), especially with tumor cells in the LAS group showing higher age scores compared to those in the HAS group.

Conclusions: Collectively, our findings underscore that ARRS through ARGs serves as a robust biomarker for the prognosis in LUAD.

KEYWORDS

cellular senescence, lung adenocarcinoma, tumor microenvironment, heterogeneity, machine learning

1 Introduction

Cancer with complex molecular characteristics (1), remains a significant global health challenge, accounting for a substantial number of deaths and impacting life expectancy worldwide. Amid the array of cancer types, lung cancer emerges as the second most prevalent contributor to cancer-related mortality, marked by a discouraging 5-year relative survival rate of just 23% (2). Lung adenocarcinoma (LUAD), the predominant histological subtype within non-small cell lung cancer, constitutes over 40% of all lung cancer cases (3). Notably, LUAD continues to rise in incidence among current smokers, former smokers, and even non-smokers, and its five-year survival rate remains dishearteningly low at approximately 15%, as a significant majority of patients are typically diagnosed at advanced stages of the disease (4). Hence, there is still a compelling need to formulate a novel prognostic model for predicting the outcomes of LUAD to advance more potent strategies for diagnosis and treatment.

Aging is a ubiquitous biological process that results in a progressive and irreversible decline in physical function across all organ systems, which presents with genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, disabled macroautophagy, deregulated nutrient-sensing, mitochondrial dysfunction, stem cell exhaustion, chronic inflammation, altered intercellular communication, cellular senescence, and dysbiosis (5–8). Cellular senescence refers to the essentially irreversible arrest of cell proliferation (growth) that occurs when cells experience potentially oncogenic stress (damage to DNA, strong mitogenic signals, damage or disruptions to the epigenome, and ectopic expression of certain tumor suppressors) (9, 10). Several evidences have shown that cellular senescence plays a double-edged role in initiation, growth, and progression of tumor (11, 12). Senescent tumor cells wield influence over the tumor microenvironment (TME) via the senescence-associated secretory phenotype (SASP). On one hand, by emitting pro-inflammatory cytokines, chemokines, growth factors, and proteases like IL-6, IL-8, and TGF-β, senescent cells can trigger paracrine senescence, transforming neighboring non-senescent cells into senescent counterparts. This process recruits and activates immune cells within the TME, leading to outcomes that can either hinder or foster tumor growth. M1 macrophages and natural killer cells, for instance, can eliminate tumor cells and foster their senescence through the secretion of IFN-γ and TNF-α, thereby restraining tumor expansion. On the other hand, senescent tumor cells may activate myeloid-derived suppressor cells and M2 macrophages via SASP, affecting the clearance of senescent tumor cells, in turn, driving tumor progression and vascularization (9, 13, 14). Given the role of cellular senescence in constraining tumor development, it emerges as a potential target for tumor therapy. Hence, unraveling the impact of senescence in tumorigenesis is paramount importance.

In recent years, several studies have focused on the role of senescence in LUAD (15–20). For example, Lin et al. constructed a cellular senescence-related signature (SRS) by leveraging senescence-related genes. They found that SRS involved in the regulation of the tumor immune microenvironment through SASP was a robust biomarker for the immunotherapeutic response and prognosis in LUAD (15). In a similar vein, another research by Lin et al. explored cellular senescence patterns within LUAD by analyzing mRNA expression profiles of 278 cellular senescence-related genes, demonstrating the association between cellular senescence patterns and tumor immune infiltration in LUAD (16). Besides, Liu et al. developed a 12-gene signature for LUAD using 91 cancer-related senescence genes to assess survival outcome (19). Nonetheless, prior investigations were marked by limitations. Firstly, all focused on only a subset of senescence genes. Secondly, the assessment of the TME was largely confined to the bulk transcriptomic level. As a result, the role of senescence in LUAD has yet to undergo systematic evaluation, and the intricate interplay between senescence and LUAD prognosis has remained obscure.

This current study seeks to overcome these limitations by integrating a comprehensive collection of 586 senescence genes sourced from the GenAge and CellAge databases. Employing the least absolute shrinkage and selection operator (LASSO) Cox analysis, a novel prognostic model for LUAD was constructed and validated. This model was further investigated the relationship between aging and tumor somatic mutation or immune cell infiltration. Furthermore, this study delved into the senescent heterogeneity and intercellular communication of various cells within the LUAD TME through the analysis of single-cell transcriptomic data. In summary, this study enriches our understanding of the profound impact of cell senescence on the survival outcomes of patients with LUAD, which unravels the complex associations between senescence, the immune landscape, and the intricate genetic makeup of the tumor, ultimately illuminating novel avenues for therapeutic interventions and prognostic assessments.

2 Materials and methods

2.1 Data source and processing

In the training cohort, bulk RNA sequencing (RNA-seq) data, somatic mutation data and clinical information for LUAD were downloaded from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/) (21). After excluding non-primary cases and patients with incomplete follow-up information, we analyzed 428 patients from the TCGA dataset as the training set. For the validation cohort (GSE31210, GSE50081, and GSE30219) (22–25), transcriptome data were obtained from data series in the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) (26). Single-cell RNA-seq (scRNA-seq) data (GSE189357) comprising nine patients with LUAD was also download from the GEO database (27). Fragments per kilobase million (FPKMs) values or raw gene expression counts were normalized to transcripts per kilobase million (TPMs) in both the training and validation cohorts. Genes that were not expressed in more than half of the samples were excluded from the expression profiles. The clinical features of 428 patients are listed in Table 1.

2.2 Aging gene set and screening

The set of aging-related marker genes (ARGs) was obtained from two databases, GenAge and CellAge. Initially, 279 ARGs were selected from CellAge (https://genomics.senescence.info/cells/) (28), and an additional 307 ARGs were obtained from GenAge (https://genomics.senescence.info/genes/index.html) (29) (Supplementary Table S1). Univariate Cox analysis was conducted by survival (version 3.3-1) packages to preliminarily identify ARGs associated with the overall survival (OS) of LUAD patients in the TCGA cohort (30), resulting in a final gene set comprising 102 ARGs (Supplementary Table S2).

2.3 Construction and validation of an ARGs risk model

We utilized the "glmnet" (version 4.1-8) package in R software (version 4.1.2) to perform the LASSO Cox regression analyses (family="cox") to screen out the prominent genes (31, 32). The "lambda.1se" value, determined through tenfold cross-validation, was employed as the lambda for model fitting. Ten genes were ultimately selected to construct the risk model. The prognostic capability of the ten genes was assessed using Kaplan-Meier survival curves generated with the survminer (version 0.4.9) and survival (version 3.3-1) R packages (30). Subsequently, we calculated a risk score for each sample, as a linear combination of gene expression levels within the signature set, weighted by their respective LASSO Cox regression coefficients, using the following formula:

$$\mathbf{Aging-related\ risk\ scores\ (ARRSs)}$$
$$=\sum\nolimits_i^n Expre(gene_i)*Coef(gene_i)$$

Here, "Coef (gene$_i$)", signifies the LASSO Cox regression coefficient, "Expre (gene$_i$)", represents the expression level of each gene, and "n" denotes the number of genes included in the model. In addition, the R package "survival" (version 3.3-1) was used to construct multiple multivariate Cox analysis to determine the independent prognostic factor in LUAD patients (30).

In the TCGA training cohort, LUAD patients were classified into high aging score group (HAS group) and low aging score group (LAS group) based on the median value of ARRSs. The prognostic capability of the risk model in terms of OS and progression-free survival (PFS) was assessed using Kaplan-Meier survival curves generated with the survminer (version 0.4.9) and survival (version 3.3-1) R packages (30). Additionally, we also compared the clinicopathological characteristics of TCGA-LUAD patients between the HAS group and the LAS group using Fisher's Exact Test.

To validate the ARGs Risk Model, we calculated the risk score for patients in the validation cohort (GSE31210, GSE50081, and GSE30219) using the same formula as applied to the TCGA-LUAD cohort. Patients in the validation cohort were also categorized into high and low-risk groups based on the median value of ARRS. Kaplan-Meier curves were generated to assess the relationship between ARRS and OS in the validation cohort.

2.4 Functional enrichment analysis of differentially expressed genes based on HAS and LAS groups

We used the "DESeq2" (version 1.36.0) R package to calculate fold-changes and identify differentially expressed genes (DEGs)

TABLE 1 Patient characteristics for TCGA_LUAD cohort.

		Total (n = 428)	HAS group (n = 214)	LAS group (n = 214)	Fisher's Exact Test (P value)
Age	<=65	206	111	95	0.142
	>65	212	98	114	
	NA	10	5	5	
Gender	female	238	111	127	0.144
	male	190	103	87	
race	american indian or alaska native	1	1	0	0.154
	asian	8	6	2	
	black or african american	47	19	28	
	white	330	168	162	
	NA	42	20	22	
OS	Alive	321	146	175	**0.002**
	Dead	107	68	39	
AJCC	I	245	105	140	**0.004**
	II	103	59	44	
	III	59	38	21	
	IV	14	9	5	
	NA	7	3	4	
T stage	T1	149	56	93	**0.002**
	T2	231	130	101	
	T3	36	22	14	
	T4	11	5	6	
	TX	1	1	0	
N stage	N0	292	130	162	**0.001**
	N1	77	47	30	
	N2	50	35	15	
	N3	2	1	1	
	NX	6	1	5	
	NA	1	0	1	
M stage	M0	286	147	139	0.313
	M1	14	9	5	
	MX	124	56	68	
	NA	4	2	2	

The "NA" represents sample with missing clinical information. Samples with missing clinical information were not considered in Fisher's Exact Test statistics.

based on the two risk groups (false discovery rate (FDR) < 0.05 and |Log2FC| > 1) (33). Subsequently, we conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses on these DEGs using the "clusterProfiler" (version 4.7.1.002) R package (34). Pathways with adjusted p-values less than 0.05 were considered significant.

2.5 Immune infiltration between the HAS-group and LAS-group from TCGA-LUAD cohort

The "estimate" R package, a powerful tool for quantifying the immune stromal, and ESTIMATE scores, which was based on the expression of related molecular biomarkers in immune and stromal cells, to predict the TME (35). The "xCell" is a robust algorithm that analyzes the infiltration levels of 64 immune and stroma cell types, including extracellular matrix cells, epithelial cells, hematopoietic progenitors, innate and adaptive immune cells (36). Herein, we utilized the R package estimate (version 1.0.13) and xCell (version 1.1.0) to evaluate the immune infiltration score and immune cell infiltration in each patient between HAS and LAS subgroups. Additionally, the T cell-inflamed gene expression profile (GEP) was calculated as a weighted sum of standardized expression values of 18 genes (*CCL5, CD27*, *CD274*, *CD276*, *CD8A*, *CMKLR1*, *CXCL9*, *CXCR6*, *HLA-DQA1*, *HLA-DRB1*, *HLA-E*, *IDO1*, *LAG3*, *NKG7*, *PDCD1LG2*, *PSMB10*, *STAT1*, *TIGIT*) as described in previous literature (37–39). The single sample gene set enrichment analysis (ssGSEA) algorithm in "gsva" (version 1.42.0) R package was performed to compare differences in 13 gene sets associated immune function and 4 gene sets related to angiogenesis, matrix, matrix remodeling, and tumor proliferation rate from previous studies (40–42). Box plots were developed using ggplot2 software (version 3.4.3) in R to display the differences between the two groups (43).

2.6 The genetic landscapes of HAS-group and LAS-group

Genetic landscapes were analyzed and visualized using the "maftools" (version 2.12.0) R package (44). Tumor Mutation Burden (TMB) was defined as the number of somatic, non-silent, protein-coding mutations in the coding regions per megabase (mut/Mb) and counted using 'maftools' (version 2.12.0). The mutated samples of tumor-related and DNA damage repair (DDR) pathways in HAS and LAS groups were compared using Fisher's exact test (with $p < 0.05$ indicates a significant difference) and visualized using "ggradar" (version 2.12.0) and ggplot2 (version 3.4.3) R packages (43).

2.7 Single-cell RNA-seq analysis

Raw matrix data were obtained from the GEO database for subsequent analysis (27). Initially, cells with low quality were filtered out based on the following criteria: 1) fewer than 200 expressed genes, 2) total molecule count per cell less than 800, and 3) greater than 10% of reads mapped to the mitochondrial genome. Additionally, the "DoubletFinder" R package (45) was utilized to identify and remove doublet cells using default parameters.

The "Seurat" package (version 4.3.0) (46) was employed to normalize the single-cell gene expression data using the "NormalizeData" and "ScaleData" functions, respectively. Subsequently, the top 2,000 highly variable genes for each sample were selected using the "FindVariableFeatures" function. Principal component analysis (PCA) was performed using the "RunPCA" function, and the first 20 principal components were used for Uniform Manifold Approximation and Projection (UMAP) analysis with the "RunUMAP" function. Following UMAP analysis, cells were clustered using an unsupervised method with a resolution parameter of 1.5 employing the "FindNeighbors" function. Differential expression analysis was conducted on the original log-normalized data by comparing cells within each cluster to all other clusters using the "FindAllMarkers" function. Clusters were annotated based on the expression of well-known markers and differentially highly expressed genes.

Subgroup analysis of each cell group, including T/NK cells, B cells, and myeloid cells, was performed using the standard Seurat pipeline. Specific markers were used for grouping and are listed in Supplementary Table S3. Bar plots were generated to illustrate the percentage of cells between the two groups. Additionally, cell occupancy differences were assessed using Fisher's exact test. The cytotoxic and exhausted scores for T cell subgroups, as well as the hallmark pathways compared between HAS and LAS groups, were calculated using the ssGSEA algorithm in the "gsva" package (version 1.42.0) based on different sets of genes (42).

2.8 Identification of cancer cells

To identify cancer cells, we utilized the inferCNV (version 1.13.0) tool (https://github.com/broadinstitute/inferCNV), as previously described in studies by Liu, He, et al. and Chen et al. (47, 48). The inferCNV package compares gene expression profiles of each cell to reference gene expression profiles from other cells. Initially, raw count data and cell type annotations for all cells were extracted from the Seurat object. Immune cells and stromal cells were chosen as reference cells. A gene ordering file was generated from the human GRCh38 assembly, containing chromosomal start and end positions for each gene. These files were used to create an inferCNV object using the "CreateInfercnvObject" function, followed by running inferCNV with default parameters. The calculated copy number variation (CNV) signal was defined as the mean square of CNV estimates across all genomic locations. CNV R-scores were calculated as the Pearson correlation coefficient between each cell's CNV pattern and the average CNV pattern of the top 5% of cells from the same tumor based on CNV signal. Cells with CNV R-scores ≥0.3 were classified as tumor cells.

2.9 Aging-related risk scores based on pseudo-bulks

The Seurat object was transformed into a "SingleCellExperiment" object, followed by the computation of pseudo-bulks. Pseudo-bulks, which represent the sum of counts, were calculated using aggregation-based methods in the muscat (version 1.10.1) R package (https://github.com/HelenaLC/muscat). The ARRSs were then derived using the previously described formula based on the pseudo-bulks. Patients were stratified into two groups, HAS and LAS, based on the median value of ARRSs. Additionally, age scores for each cell were calculated based on ten ARKGs at the single-cell level using the ssGSEA algorithm.

2.10 Cell-cell interactions

CellChat (version 1.5.0) is an open-source R package (https://github.com/sqjin/CellChat) utilized for the analysis, comparison, and visualization of single-cell RNA sequencing data intercellular communication (49). In this study, CellChat was employed to infer cell-cell interactions across 24 immune subgroups, fibroblasts, normal epithelial cells, tumor cells, and endothelial cells for both the HAS and LAS groups. Subsequently, major signaling changes between the HAS and LAS groups were computed.

2.11 Statistical analysis

The Wilcoxon test was conducted to examine differences in variables between two groups, while the Kruskal-Wallis test was used to assess differences among groups greater than two. Gene mutation differences between the HAS and LAS groups were determined using Fisher's exact test.

3 Results

3.1 Construction and validation of aging-related risk score

The workflow of the whole study was graphically presented in Figure 1A. We compiled a comprehensive list of 586 aging-associated genes sourced from the CellAge and GenAge databases. Among these genes, 102 were significantly associated with clinical survival ($p < 0.05$) based on univariate Cox analysis (detailed results shown in Supplementary Table S2), conducted on the expression matrix and clinical survival information of 428 LUAD samples obtained from the TCGA dataset. Subsequently, to construct the ARGs risk model, we performed LASSO Cox regression analysis on the aforementioned 102 genes and the gene expression profiles of the training cohort (Figures 1B, C). Through this analysis, we successfully identified 10 aging-related key genes (ARKGs), including. *BRCA2*, *CSNK1E*, *EEF1E1*, *GAPDH*, *IGFBP3*, *IL1A*, *PSEN1*, *XRCC5*, *XRCC6*, and *YWHAZ*. And low RNA expression for the 10 ARKGS was correlated with longer survival time in LUAD (Supplementary Figure S1). Utilizing these ten ARKGs and their corresponding risk coefficients, we established an aging risk signature. The risk score of every patient was calculated using this formula. Patients in the training cohort were stratified into two groups: the high aging score group (HAS group) and the low aging score group (LAS group) based on median values of ARRSs. Upon investigating the expression levels of the ten ARKGs, we found that they were significantly higher in HAS group than LAS group (Supplementary Figure S2, Supplementary Table S4).

We compared the clinicopathological characteristics, including age, gender, race, OS, TNM tumor grade, and AJCC tumor grade, of TCGA-LUAD patients between the HAS group and the LAS group (Table 1). The results showed significant differences in OS status (P = 0.002), T grade (P = 0.002), N grade (P = 0.001), and AJCC tumor grade (P = 0.004) between the groups. Survival analysis demonstrated that the LAS-group exhibited significantly improved overall survival (OS) (hazard ratio (HR) = 0.45, 95% confidence interval (CI) = 0.31 – 0.66, P = 0.000044) and progression-free survival (PFS) (HR =55, 95% CI = 0.41 – 0.74, P = 0.000052) compared to the HAS-group (Figures 1D, E). Upon integrating age, gender, TNM tumor grade, and AJCC tumor grade into the multivariate Cox regression analysis, ARRSs emerged as the sole significant survival-related risk factor (HR = 9.32, 95% CI = 4.50 – 19.29, P = 0.0000000018) (Figure 1F), suggesting that ARRSs was an independent prognostic factor for LUAD.

To validate the prognostic roles of the above risk model, we applied the same stratification method to three independent datasets form the GEO database. Consistent with the findings from the training cohort, patients with in the high ARRSs group displayed significantly worse survival outcomes compared with the low ARRSs group in all three cohorts, namely GSE50081 (HR = 0.32, 95% CI = 0.18 – 0.56, P = 0.000081), GSE30219 (HR = 0.52, 95% CI = 0.29 – 0.95, P = 0.038), and GSE31210 (HR = 0.36, 95% CI = 0.19 – 0.71, P = 0.005) (Figures 1G-I).

3.2 The genetic characteristics of HAS-group and LAS-group

To explore the genetic features in LUAD with different ARRSs, we further investigated the genomic differences between the HAS group and the LAS group based on somatic mutation data in the TCGA-LUAD cohort (Figure 2A; Supplementary Figures S3A-C). We observed that HAS group had a higher mutation frequency than the LAS group, particularly in the top 20 genes such as, *TP53*, *TTN*, *CSMD3*, *ZFHX4*, *RYR3*, *CSMD2*, *SI*, *LRRC7*, and *PAPPA2* (detailed P values shown in Supplementary Table S5) between HAS and LAS groups (Figure 2B). Additionally, the HAS group displayed a higher tumor mutation burden (TMB) but a lower occurrence of co-occurring mutations between genes, indicating distinct genomic alteration patterns (Figure 2C; Supplementary Figure S3D). Further analysis of ten tumor-related pathways revealed significantly higher mutation frequencies in the Hippo (P = 0.011), NOTCH (P = 0.013), and TP53 (P = 0.011) pathways in the HAS-group compared to the LAS-group (Figures 2D, F; Supplementary Figure S4B). Similarly, higher mutation rates were observed in the HAS group among the eight DDR pathways, with five of them being statistically significant (Figures 2E, G; Supplementary Figure S4A).

3.3 ARRSs is associated with cell proliferation and immune function

Differential expression analysis of gene expression data based on the HAS group and LAS group identified a total of 1664 differentially expressed genes (DEGs) under a threshold of adjusted $p < 0.05$, comprising 707 up-regulated and 957 down-regulated genes (Figure 3A). GO enrichment analysis for DEGs revealed that in the HAS-group, biological processes were predominantly enriched in cell cycle, cell division, and cell development, indicating a potential involvement in regulating normal cell function and organismal development (Figure 3B). Furthermore, based on gene sets from

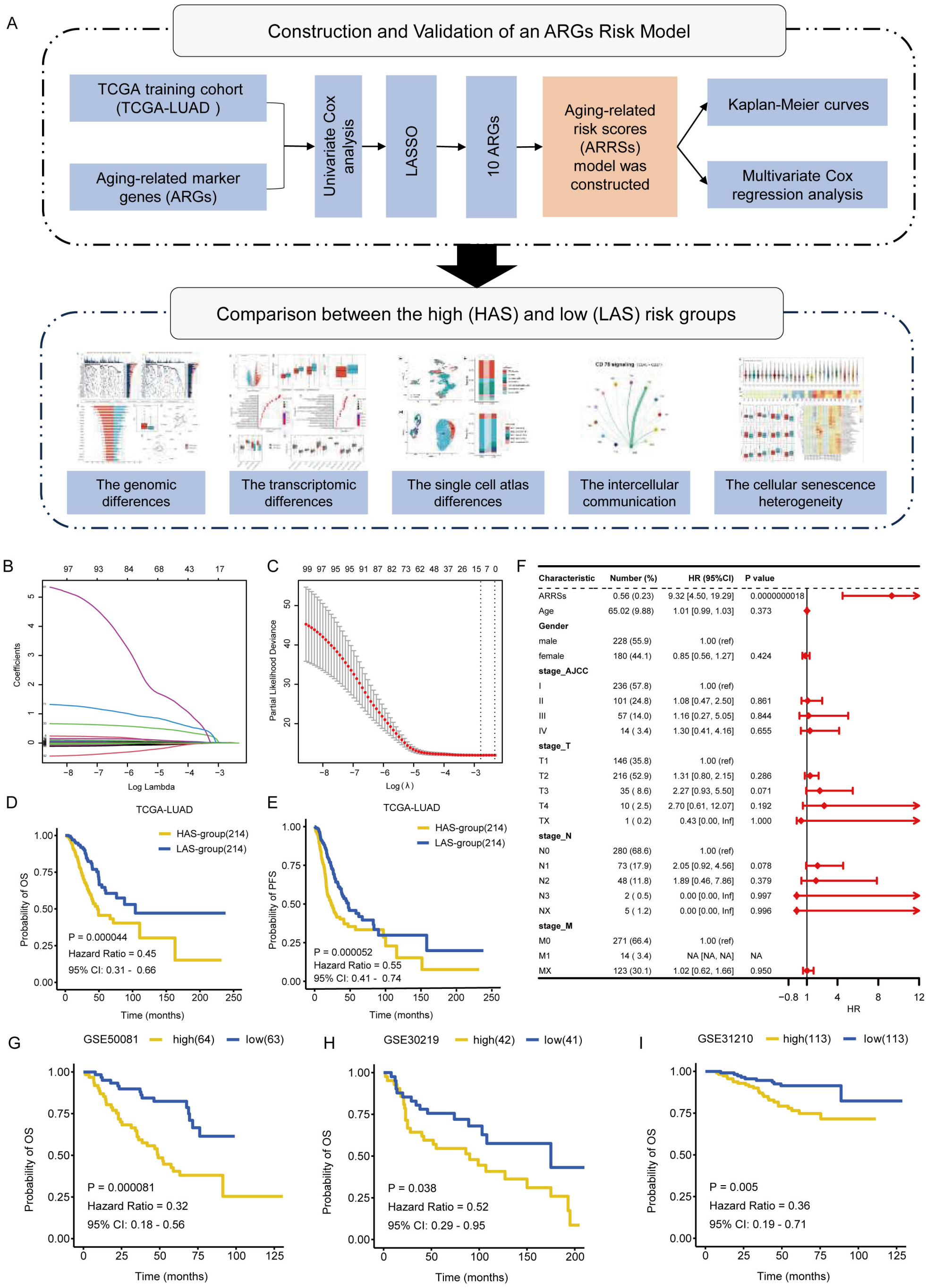


FIGURE 1
Identification of ARKGs related to prognosis in the TCGA training cohort. **(A)** The workflow of the present study. **(B)** Selection of optimal candidate genes in the LASSO model. **(C)** LASSO coefficients of prognosis-associated ARKGs. **(D, E)** Kaplan-Meier curves for overall survival **(D)** and progression free survival **(E)** of the TCGA-LUAD cohort in the HAS and LAS groups. **(F)** Forest plots showing results of multivariate Cox regression analysis between Risk score, clinical information and overall survival. **(G-I)** Kaplan-Meier curves for overall survival of validation cohorts in the high and low groups: GSE50081 **(G)**, GSE30219 **(H)**, GSE31210 **(I)**.

Bagaev, et al. (40), we found that the tumor proliferation rate, and matrix remodeling of the HAS group were significantly higher than those of the LAS group (Figure 3F, detailed P values were shown in the Supplementary Table S4).

In contrast, the LAS-group exhibited enrichment in immune response mechanisms, encompassing cell activation, signal transduction, and production of immune mediators (Figure 3C). Based on another gene set related to immune functions (41), we observed that nine of the 13 immune function gene sets had significantly higher ssGSEA scores in all LAS groups than the HAS group (Figure 3G, detailed P values were shown in the Supplementary Table S4), especially type II IFN response, T cell co-stimulation, and HLA. Immune estimations for LUAD patients within the training set (TCGA-LUAD) showed notably increased

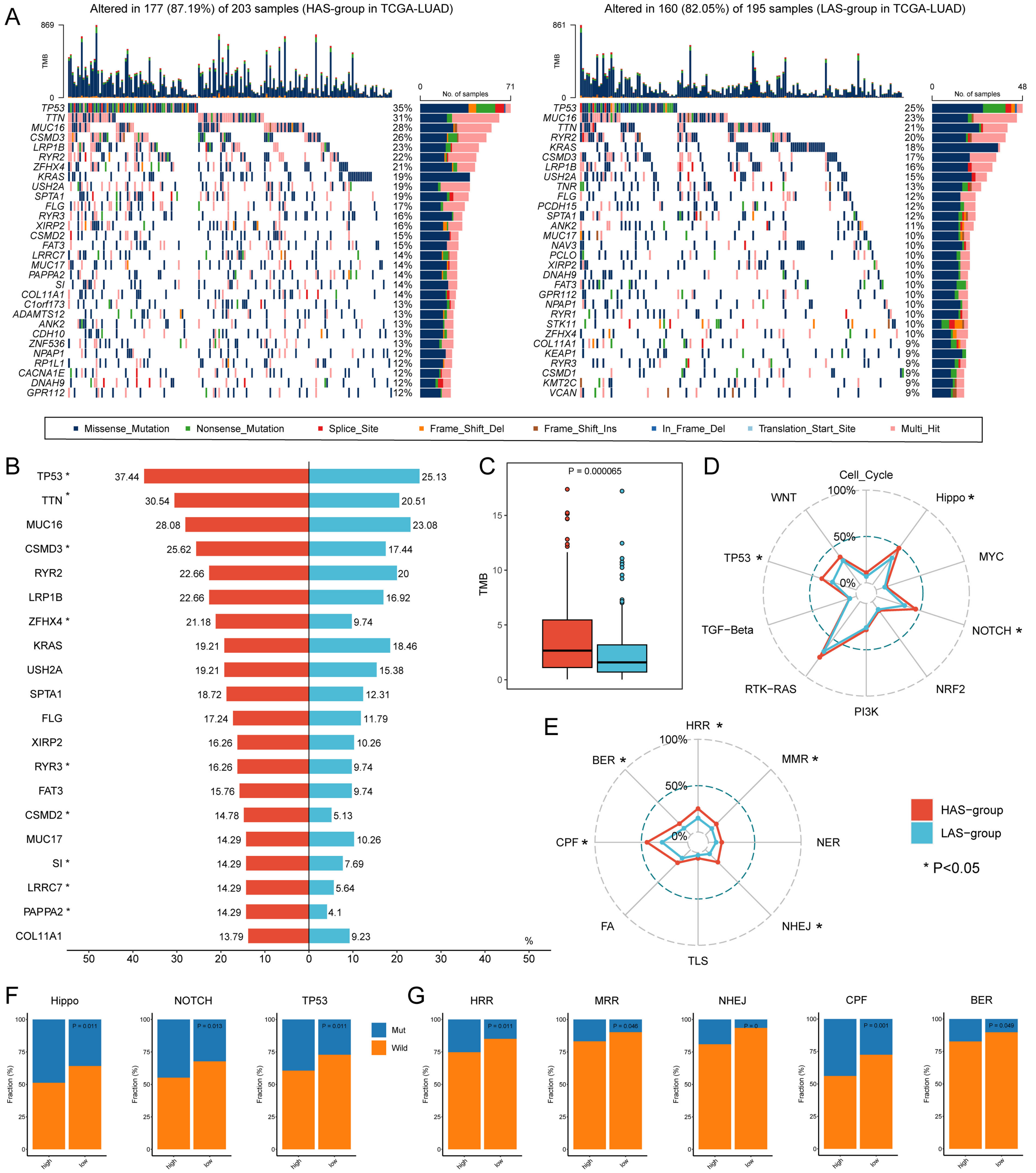


FIGURE 2

Genomic alterations differences between the HAS and LAS group from the TCGA-LUAD cohort. **(A)** Genomic alterations landscape between the HAS (left) and LAS (right) group. **(B)** Mutation frequency differences of the top 20 mutation genes in the HAS group compared to the LAS group. The asterisk to the right of the gene indicates that the mutations in the gene were significantly different in the two groups, as determined by Fisher's exact test. **(C)** The TMB between HAS and LAS groups. The HAS group had a higher TMB (2.66 (IQR: 0.04, 11.985) compared to the LAS group (1.58 (IQR: 0.02, 6.9)) with P value = 0.000065 compared by the Wilcoxon test. The frequency of mutated genes in each tumor-related pathway **(D, F)** and DDR pathway **(E, G)** difference between two groups. The asterisks in **(D, E)** denote significant differences of mutated genes in different pathways identified by Fisher's exact test which showed in **(F, G)**.

StromalScore, ImmuneScore, ESTIMATEScore, and GEP score in the LAS group when compared to the HAS group (Figures 3D, E). Xcell analysis revealed the immune infiltration of TME (36). The results indicated that LAS group had an activated TME, with significantly increased numbers of T cells, such as CD8+ T cells, CD8+ Tcm, CD4+ Tem, and CD4+ Tcm, and significantly decreased numbers of Th1 and Th2 (Figure 3H, detailed p values were shown in the Supplementary Table S4). Additionally, B cells such as plasma cells (Figure 3H), and myeloid cells such as Mast cells, and various DCs (Figure 3I, detailed p values were shown in

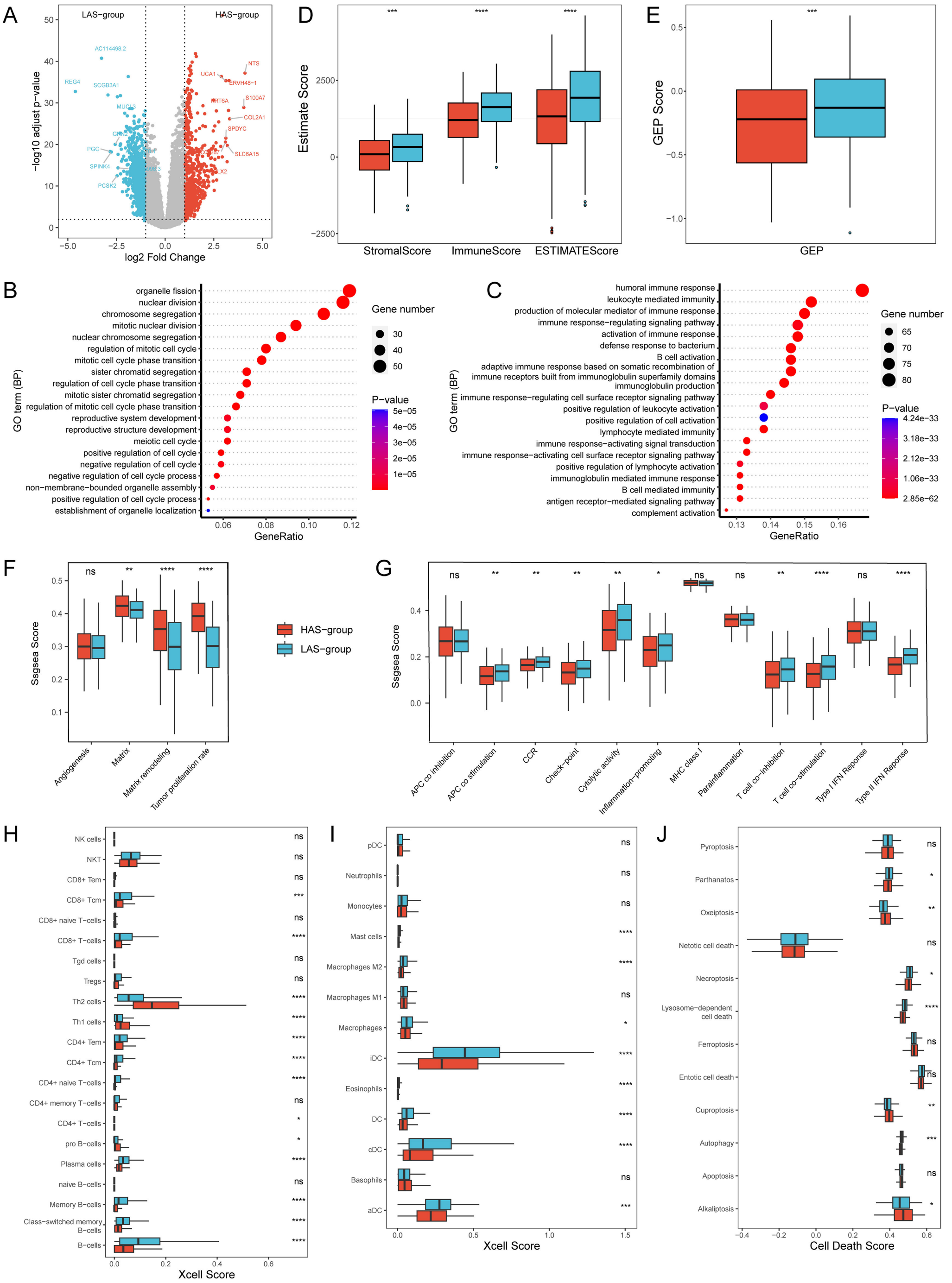


FIGURE 3
Transcriptomic differences between the HAS and LAS group from the TCGA-LUAD cohort. **(A)** Volcano Plot of DEGs between the HAS and LAS group. **(B, C)** Top 20 biological processes of GO enrichment results between the HAS **(B)** and LAS **(C)** group. **(D)** Stromal score, immune score and ESTIMATE score between the two groups. **(E)** GEP score between the two groups. **(F, G)** Boxplots of gene sets related to tumor proliferation **(F)** and immune-related functions **(G)**. **(H, I)** Box plot of T cells **(H)**, B cells **(H)**, and myeloid cells **(I)** infiltration in "Xcell" between the two groups. **(J)** Box plot of cell death between the two groups. "ns" indicates $p > 0.05$, * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$, *** indicates $p \leq 0.001$, and **** indicates $p \leq 0.0001$. The actual P determined by the Wilcoxon test, and the medians (IQR) in Figures 2D-F were all displayed in Supplementary Table S4. All abbreviations presented in Figure 3 showed as following: GEP, T cell-inflamed gene expression profile; CCR, cytokine and cytokine receptor; HLA, human leukocyte antigen; MHC, major histocompatibility complex.

the Supplementary Table S4), were also significantly increased in the LAS group. Furthermore, we explored the relationship between ARRSs and various cell death pathways. The findings revealed that significantly elevated scores for Alkaliptosis, Cuproptosis, and Oxeiptosis in the HAS-group, whereas Autophagy, Lysosome-dependent cell death, Necroptosis, and Parthanatos scores were markedly higher in the LAS-group (Figure 3J; Supplementary Table S4).

3.4 The single cell alta of HAS-group and LAS-group

To further investigate whether the ARRSs is heterogeneous in the TME, we utilized a single-cell dataset (GSE189357) containing over 10,000 cells from 9 patients. Initially, the single-cell dataset was converted to pseudo-bulks and then ARRSs were calculated. Subsequently, the 9 patients were divided into HAS (n = 5) and LAS (n = 4) groups based on the median value of ARRSs. Notably, two of the three invasive adenocarcinoma (IAC) samples were categorized into the HAS group, exhibiting significantly higher aging scores compared to the LAS group (Figures 4A, B). Employing the standard pipeline in Seurat (46), we identified six major cell types, including T/NK cells, B cells, myeloid cells, fibroblasts, endothelial cells, and epithelial cells (Figures 4C, D). Subsequently, the epithelial cells were further subdivided into tumor cells and normal epithelial cells (Figure 4E). Interestingly, we observed an enrichment of tumor cells and endothelial cells in the HAS (P = 6.44E-66, odds ratio (95% CI) = 1.57 (1.49, 1.66), Supplementary Table S6) and LAS (p = 0, odds ratio (95% CI) = 4.56 (4.2, 4.95)) groups, respectively (Figure 4F).

Furthermore, we conducted subtype annotation specifically for immune cells including T/NK cells, B cells, and myeloid cells (Figures 4G-I; Supplementary Table S2). T/NK cells were subdivided into eight T cell subpopulations and two NK cell subpopulations (Figure 4G). Functional scoring of T-cell subsets revealed that FCGR3A NK cells (T09) and CD8-GNLY (T08) had the highest cytotoxic scores, while CD8-CXCL13 (T06) had the highest exhausted score (Supplementary Figure S5). We compared the cellular infiltration in the HAS and LAS groups and found that the T and NK cell subpopulations were significantly differed between the HAS and LAS groups (Supplementary Table S6). Specifically, CD4-CCR7 (T01, P = 1.53E-134, odds ratio (95% CI) = 1.81 (1.72, 1.9)), and CD4-FOXP3 (T03, P = 4.55E-21, odds ratio (95% CI) = 1.41 (1.31, 1.52)) were enriched in the HAS group, whereas CD8-CCR7 (T05, P = 6.42E-14, odds ratio (95% CI) = 1.34 (1.24, 1.45)), CD8-CXCL13 (T06, P = 2.35E-64, odds ratio (95% CI) = 5.66 (4.51, 7.16)), CD8-GNLY (T08, P = 2.94E-23, odds ratio (95% CI) = 1.39 (1.3, 1.48)), FCGR3A NK cells (T09, P = 4.29E-98, odds ratio (95% CI) = 1.95 (1.83, 2.08)), and XCL1 NK cells (T10, P = 2.65E-34, odds ratio (95% CI) = 2.02 (1.8, 2.27)) were enriched in the LAS group. For B cell subsets, naive and memory B cells were more prevalent in the HAS group, whereas plasma cell (PC) subsets (B03 P = 1.33E-24, odds ratio (95% CI) = 2.3 (1.96, 2.69); B04, P = 3.51E-48, odds ratio (95% CI) = 2.96 (2.55, 3.43)) and stressed PC (B05, P = 7.24E-13, odds ratio (95% CI) = 3.02 (2.21, 4.12)) were more prevalent in the LAS group. The Mast cells (M01, P = 7.44E-292, odds ratio (95% CI) = 2.77 (2.62, 2.92)) showed a tendency to increase in the LAS group compared to the HAS group, while neutrophils (M02, P = 4.84E-164, odds ratio (95% CI) = 5.25 (4.55, 6.08)), S100B DC (M06, P = 1.08E-66, odds ratio (95% CI) = 2.05 (1.88, 2.23)), TXN DC (M07, P = 9.61E-15, odds ratio (95% CI) = 2.47 (1.93, 3.19)), and proliferation myeloid cells (M09, P = 1.51E-16, odds ratio (95% CI) = 1.87 (1.6, 2.2)) were significantly more prevalent in the HAS group. These results provide further evidence of heterogeneity in immune cell infiltration between groups with differing ARRs at the single-cell level, especially the LAS enriched more cytotoxic T/NK cells and antibody-secreting B cells.

3.5 Inference of cell-cell interactions

Given that senescence alters intercellular communication, we conducted a comparative analysis of intercellular communication between the HAS and LAS groups for each cell subset based on single-cell data. Significant differences were observed in several signaling networks between the HAS and LAS group (Supplementary Figure S6). Notably, SPP1 was exclusively present in in the HAS group (Supplementary Figure S6; Figure 4J). Especially, the interaction of SPP1-CD44 has been reported to inhibit T-cell activation and promote tumor immune evasion (50, 51). Additionally, TENASCIN was frequently observed in the HAS group, with tumor cells in this group interacting with other cells, including tumor cells themselves, via TNC - SDC1/SDC4 or TNC - ITGA8_ITGB1/ITGAV_ITGB6 (Supplementary Figure S6; Figure 4K). TNC is an extracellular matrix glycoprotein known to contribute to tumor progression, and increased TNC expression in LUAD tissues correlates with an unfavorable clinical outcome for patients (52). Conversely, certain pathways were exclusively or more frequently observed in the LAS group (Supplementary Figure S6). For example, the secreted signaling BAG, and CD70 pathways were uniquely found in the LAS group (Supplementary Figure S6). The BAG6-NCR3 interaction targeting T09 might trigger NK cell cytotoxicity (Figure 4L). Furthermore CD70-CD27 interaction was observed between B02 and PC or between B02 and T cells. CD27 receptor activation provides a costimulatory signal promoting T cell and B cell activity to enhance anti-tumor and anti-infection immunity (Figure 4M) (53).

3.6 Cellular senescence heterogeneity in the tumor microenvironment

Using single-cell data, we evaluated the senescence levels of individual cells and compared the senescence levels among different cell subpopulations (Figure 5A). We observed lower age scores in B01, B03, B05, M02, and endothelial cell subpopulations, while M03 and M05 exhibited higher age scores (Figure 5A). Subsequently, we compared the senescence levels of cell subpopulations between the HAS and LAS groups (Figure 5B). Most T cell subsets (e.g., T05, T07) displayed higher age scores in the HAS group than in the LAS group (Figures 5B, C). Moreover, endothelial and fibroblast cells exhibited higher age scores in the HAS group, whereas normal epithelial cells and tumor cells showed higher age scores in the LAS group (Figures 5B, C). Age scores for different subpopulations of myeloid and B cells varied between the HAS and LAS groups (Figures 5B, C). For instance, the age scores of B01, B02, M05 and M06 were significantly lower in the HAS group than in the LAS group, while B04 and M02 showed higher scores in the HAS group (Figures 5B, C). As cellular damage caused by reactive oxygen species (ROS) is a major trigger for senescence (54), we assessed and

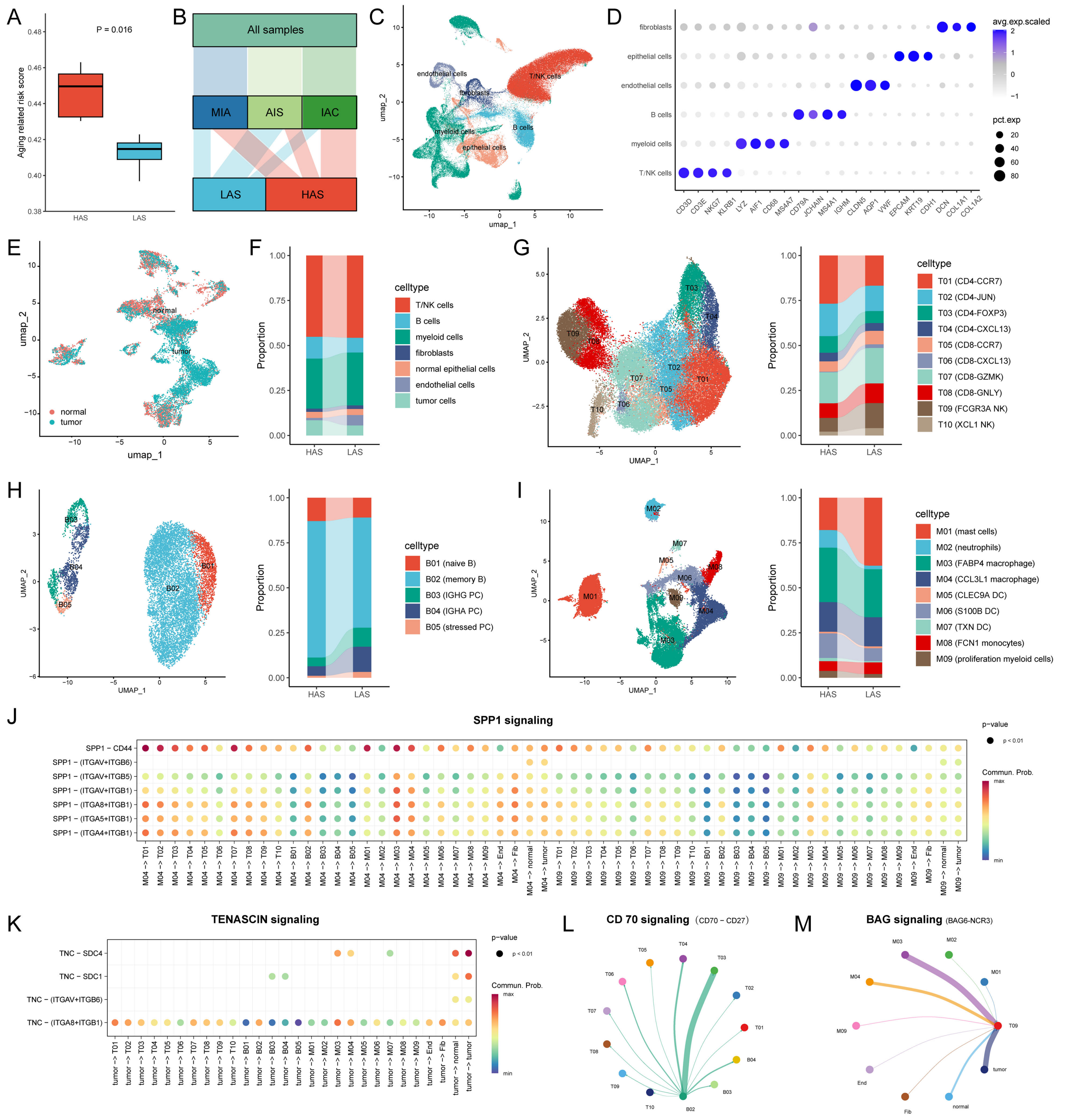


FIGURE 4
Single cell atlas and cell-cell interactions between the HAS (n = 5) and LAS (n = 4) group. **(A)** ARRSs based on single cell pseudo-bulks differed between the HAS and LAS group. The HAS group had a higher ARRS 0.4496 (IQR: 0.4304, 0.463) compared to the LAS group 0.4147 (IQR: 0.3969, 0.4228) with P value = 0.0159 compared by the Wilcoxon test. **(B)** Alluvial diagram showed the grouping of the nine samples. **(C)** UMAP plot for cells displaying the six major cell types from patients. **(D)** Dot plot depicting mean expression levels and percentage of cells expressing signature genes across the six major cell types. **(E)** Distribution of normal and tumor cells in epithelial cells from LUAD. **(F)** The composition of the cell compartment, displaying the average frequencies of cell subsets for HAS and LSA groups. **(G-I)** The UMAP plot and the average frequencies of different T cell, B cell and myeloid cell subgroups. **(J, K)** Comparison of the significant ligand-receptor pairs of SPP1 signaling **(J)** and TENASCIN signaling **(K)** for the HAS group. Dot color reflects communication probabilities and dot size represents computed p-values. Empty space means the communication probability is zero. p-values are computed from one-sided permutation test. **(L, M)** Circle plot showed cell–cell communication mediated by CD70-CD27 **(L)** and BAG6-NCR3 **(M)** in the LAS group. All abbreviations presented in Figure 4 showed as following: ARRS, aging related risk score; IQR, interquartile range; AIS, lung adenocarcinoma *in situ*; MIA, minimally invasive adenocarcinoma; IAC, invasive adenocarcinoma; tumor, tumor cells; normal, normal epithelial cells; Fib, fibroblasts; End, endothelial cells.

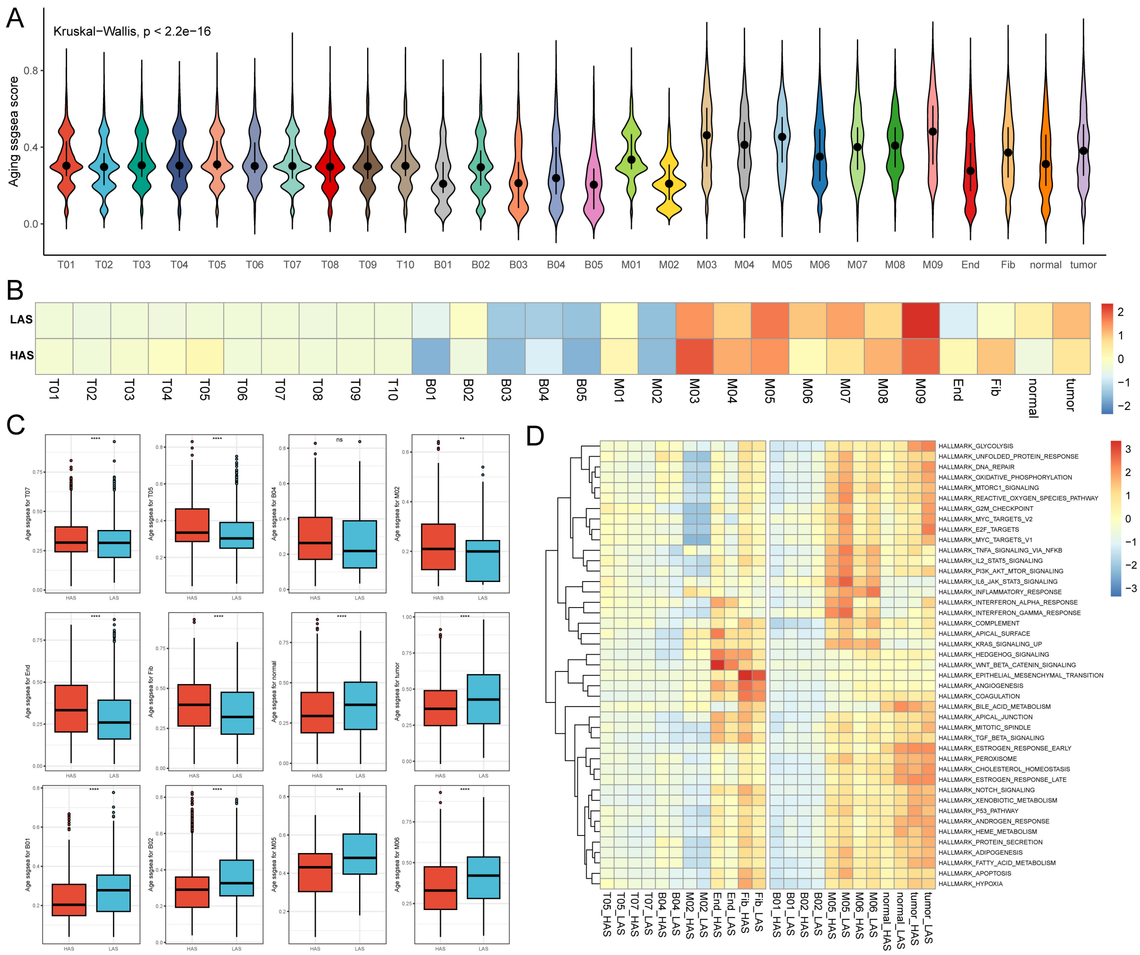


FIGURE 5

Single cell age score. **(A)** The age score for each cell. **(B)** The medians of age scores for each cell type in the HAS and LAS groups. **(C)** The box plots demonstrating between-group differences in the HAS and LAS groups for age scores for specific cell types. We used the following notation for statistical significance: "ns" indicates p > 0.05, ** indicates p ≤ 0.01, *** indicates p ≤ 0.001, and **** indicates p ≤ 0.0001. The actual P determined by the Wilcoxon test, and the medians (IQR) in **(C)** were all displayed in Supplementary Table S4. **(D)** Heatmap showed the activity of hallmarks between the HAS and the LAS groups for different cell types. All abbreviations presented in Figure 5 showed as following: tumor, tumor cells; normal, normal epithelial cells; Fib, fibroblasts; End, endothelial cells.

compared the "reactive oxygen species pathway". Our results revealed higher scores for this pathway in the HAS group for B01, B02, M05, M06, normal epithelial cells, and tumor cells, whereas the HAS group for B04, M02, T05, T07, endothelial, and fibroblast cells exhibited lower scores (Figure 5D), consistent with the trend observed in age scores (Figure 5C).

4 Discussion

Cellular senescence involves the cessation of cell-cycle and the release of inflammatory cytokines with autocrine, paracrine and endocrine activities (55). The SASP represents a significant feature of senescent cells, encompassing the release of various cytokines, chemokines, growth factors and proteases (56). The impact of cellular senescence on cancer is intricate, displaying both advantageous and detrimental effects. Nevertheless, the extent to which the senescent heterogeneity of immune infiltration cells within tumors, as well as the interplay between tumor senescence and immune infiltration in LUAD, remains incompletely documented. In the current investigation, we constructed an independent prognostic model based on cellular senescence-related genes, and comprehensively analyzed the role of aging in genomic alterations and immune landscape in LUAD, which might hold the potential to facilitate the development of personalized immunotherapy.

This study successfully identified a novel and independent prognostic risk model incorporating ten significantly upregulated genes in LUAD. Ten genes were selected from a comprehensive list of 586 aging-associated genes obtained from the CellAge and GenAge databases. These genes also have been previously reported as positive regulators of tumor development. For example, *CSNK1E*, a member of the serine/threonine protein kinase family, controls circadian rhythms, which is closely related to the animals longevity (57). Inhibition of *CSNK1E* has been show to selectively inhibit tumor cell development (58), and elevated *CSNK1E* expression is associated with poor prognosis in patients with ovarian cancer and malignant melanoma (59, 60). *EEF1E1*, a tumor suppressor, plays a role in ATM/ATR-mediated p53 activation (61), and serves as a poor prognosis predictor in lung cancer (62). Overexpression of *EEF1E1* in transgenic mice resulted in a significantly shorter mean lifespan (63). *GAPDH* directly participates in tumor progression, invasiveness, and metastasis (64), and conditions such as oxidative stress impair *GAPDH* catalytic activity, leading to cellular aging and apoptosis (65). Increased expression of *PSEN1* in colorectal cancer is associated with enhanced tumor development through heightened EGFR signaling via NOTCH1 processing and activation of the COX-2-PGE2 pathway (66). PSEN1-null mice die shortly after birth (67), although *PSEN1*'s role in human aging remains largely unknown. *YWHAZ* is an adapter protein implicated in several signal transduction pathways (68) and interacts with numerous proteins associated with aging, such as the INS/IGF1 pathway (69, 70). *YWHAZ* has also been shown to mediate lung cancer malignancy and β-catenin protein through its complex with β-catenin (71). *IL1A*, a pivotal inflammatory cytokine, is thought to be one of the critical upstream regulators of other SASP-related genes (72, 73) and drives tumor growth and metastasis (74). *IGFBP3*, a member of the insulin-like growth factor-binding protein (IGFBP) family, regulates IGF1 and IGF2 by altering the interaction of IGFs with their cell surface receptors. Interestingly, the cell growth regulator *IGFBP3* exhibits a unique pattern, as elevated levels are associated with a good prognosis in patients with advanced NSCLC (75). *BRCA2*, *XRCC5*, and *XRCC6* are all DDR related genes, involved in DNA damage and repair. Mice deficient for *BRCA2* and *XRCC5* have a reduced lifespan (76, 77). *XRCC5/6* are associated with poor prognosis and can be used as diagnostic and prognostic biomarkers for LUAD (78). *BRCA2*'s role in cancer well-established, as elevated *BRCA2* expression is associated with a significantly reduced number of stromal cells and high infiltration of both beneficial and detrimental immune cells in breast cancer (79). *BRCA2* has also been demonstrated to exhibit increased mRNA levels and poor prognosis in lung cancer (80). These findings collectively provide compelling evidence that this newly proposed prognostic risk model has the potential to reflect LUAD prognosis by considering genomic alterations and the immune landscape.

Genetic instability is a common characteristic of both aging and cancer (81), encompassing changes in DNA damage, DNA damage response and repair, mutations, replication stress, transposition, chromosome aberrations, telomere shortening, micronuclei, and DNA fragments (82). In our study, we found that the HAS group exhibited more frequent gene mutations and higher TMB, indication the presence of an unstable genome and immunogenic potential in patients with HAS. Furthermore, the mutation frequency of the Hippo, NOTCH, TP53, and DDR pathways in the HAS group were also significantly increased. Hippo is an important pathway regulating differentiation, stem cell renewal, and oncogenic transformation (83). In cancer research, the activated Hippo pathway is considered as a tumor suppressor pathway due to its ability to impede cell proliferation and facilitate apoptosis (84). Similarly, NOTCH (85) and TP53 (86) pathway mutations have also been reported to associate with unfavorable prognosis in lung cancer. DNA damage response plays a significant role in maintaining genomic integrity and closely associated with lung cancer progression and treatment (87, 88). These researches provide additional insights into our observed outcomes that patients with HAS experience poorer survival when compared to those with LAS patients.

Cellular senescence functions as a stress response characterized by a halt in proliferation and heightened secretion of pro-inflammatory cytokines (89). Senescent cells recruit immune cells, facilitating their own immune clearance, thereby restoring tissue homeostasis. In the context of cancer, various stressors such as oncogenic signaling, replication stress, hypoxia, reactive oxygen species, nutrient deprivation, and exposure to cytokines within the tumor microenvironment can trigger senescence. This underscores the significant link between tumor cell senescence and immune cell infiltration. Through a bulk-transcriptome analysis, we observed that senescence-associated genes exert a strong influence on the immune microenvironment in LUAD. Specifically, the LAS group showed an activated TME, this manifested as a noteworthy increase in the quantities of CD8+ T cells, CD8+ Tcm, CD4+ Tem, CD4+ Tcm, plasma cells, mast cells and DC, alongside heightened ImmuneScore, GEP score and type II IFN response, T cell co-stimulation, and HLA scores, in addition to enriched immune response pathways. These findings were further corroborated though single-cell analysis, which revealed that CD8-CCR7 (T05), CD8-CXCL13 (T06), CD8-GNLY (T08), FCGR3A NK cells (T09), XCL1 NK cells (T10), plasma cell sets (B03, B04, B05), and mast cells (M01) were more enriched in the LAS group (Figure 4). In contrast, the HAS group displayed an immunosuppressive microenvironment with lower immune function scores and a higher tumor proliferation rate (Figure 4). Additionally, based on the cellular communication results, we identified some signaling pathways specific to the HAS group, such as SPP1 and TENASCIN (Figure 4), which contribute to tumor immune escape and tumor progression (50–52). These results suggest that the HAS group might promote tumor cell invasion by evading immune surveillance, enhancing proliferation and immune escape, leading to poor prognosis in LUAD.

In addition to bulk-level senescence assessment, we also compared senescence at the single-cell level and found significant heterogeneity in cellular senescence. Interestingly, we found that the age scores for tumor cells in the HAS group were significantly lower than that in the LAS group (Figure 5C), suggesting that senescence at the bulk-level is not the same as senescence at the cellular level. Senescent tumor cells might augment the immune response against tumors (90), which is entirely consistent with the highly senescent tumor cells and activated immune microenvironment in the LAS group. However, it's worth noting that these senescent cells could also reinforce the tumor's resistance to immunotherapy through potent immunosuppressive mechanisms (91, 92). Therefore, more in-depth studies at the cellular level remain essential.

Herein, we also explore the relationship between senescence and other modes of cell death. Patients in the HAS-group demonstrated a propensity for Alkaliptosis and ROS cell death mechanisms such as Oxeiptosis (93) and Cuproptosis. These endogenous damages, coupled with certain exogenous factors, induced a wide array of genetic injuries, including point mutations and deletions (94), ultimately leading to significantly higher TMB in the HAS-group compared to the LAS-group. To counteract DNA damage, the HAS-group employed a series of intricate DNA repair and maintenance mechanisms associated with cell proliferation and differentiation, ensuring the preservation of proper chromosomal structure and stability (8, 94). Conversely, the interactions among lysosome-dependent cell death, autophagy, and apoptosis played a more significant role in the LAS-group. Meanwhile, the LAS-group exhibited immunological functions in response to cellular senescence, engaging in tissue repair through immune cell recruitment and immune clearance of senescent cells.

More novel analyses were added to our study, although studies related to senescence in LUAD have been reported (15–18, 20). Firstly, although previous studies have also compared differences between aging subgroups in terms of mutations, or TMB (15–18). Patients with higher risk scores had noticeably increased TMB and mutated more frequently for TP53 (15, 16, 18), which is consistent with the results we found. Furthermore, our study was the first to compare at the pathway level which showed significant differences in patients with different ARRs. Second, existing researches related to senescence in LUAD have found that the lower risk scores group embodies an immune-activated microenvironment. Lin, et al., 2023 showed that the ASRS was positively correlated with most immunomodulator-related mRNAs, including chemokines, and immune inhibitors, and receptors (18). This study collected a previously reported set of 13 immune-related gene sets (41) and comprehensively compared the immunity of different subgroups. We found that nine of the 13 immune function gene sets were positively correlated with ARRS score, including APC to stimulation, cytokine and cytokine receptor (CCR), Check-point, cytolytic activity, inflammation-promoting, HLA, T cell co-stimulation, T cell co-stimulation, and type II IFN response (Figure 3). Thirdly, previous studies based on different datasets and different methods have been performed to show the association between immune infiltration and senescence. However, sometimes inconsistent results were obtained by different software. Our study evaluates the association between immune infiltration and senescence for the first time at the single cell level, and using scRNA-seq, this study compared cellular communication between different senescence groups, revealing possible alterations in cellular communication caused by senescence (Figure 4). Finally, we assessed senescence at the cellular level for the first time and found significant inter-cellular heterogeneity for senescence. In particular, we found an opposite trend between the overall senescence score and the tumor cell senescence score. This study still had some limitations, the limited availability of single-cell samples and immune cohort samples may introduce some bias in our model validation. Although we validated the aging score model using several external independent public datasets, prospective clinical trials verification of our model is still necessary. Nevertheless, we hope that this model can contribute to the comprehension of the molecular mechanisms of cellular senescence and TME in LUAD.

In conclusion, our study identified and validated a senescence-related signature based on 10 senescence-related genes as an independent prognostic significance for patients with LUAD, indicating that the senescence levels are heterogeneous in LUAD immune microenvironment, and the HAS group might promote tumor cell invasion by evading immune surveillance, enhancing proliferation and immune escape, leading to poor prognosis in LUAD.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article.

Author contributions

KR: Writing – original draft, Conceptualization. LC: Writing – original draft, Conceptualization, Data curation. CW: Investigation, Methodology, Software, Writing – original draft. XT: Data curation, Investigation, Methodology, Software, Writing – original draft. WA: Writing – review & editing, Methodology. QW: Writing – review & editing. YM: Investigation, Methodology, Software, Writing – original draft. YH: Project administration, Supervision, Writing – original draft. XX: Writing – original draft. JB: Project administration, Supervision, Writing – original draft. XL: Writing – original draft. XFX: Supervision, Writing – original draft. MZ: Supervision, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. National Natural Science Foundation of China (No. 82071035,82371165). Natural Science Foundation of Shandong Province (No. ZR2022LZL001). Hunan Provincial Department of Science and Technology Clinical Medical Technology Innovation Guidance Project Foundation (No. 2021SK51711).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1347770/full#supplementary-material

References

1. Hanahan D. Hallmarks of cancer: new dimensions. *Cancer Discovery*. (2022) 12:31–46. doi: 10.1158/2159-8290.CD-21-1059

2. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA: Cancer J Clin*. (2023) 73:17–48. doi: 10.3322/caac.21763

3. Duma N, Santana-Davila R, Molina JR. Non-small cell lung cancer: epidemiology, screening, diagnosis, and treatment. *Mayo Clinic Proc*. (2019) 94:1623–40. doi: 10.1016/j.mayocp.2019.01.013

4. Spella M, Stathopoulos GT. Immune resistance in lung adenocarcinoma. *Cancers*. (2021) 13:384. doi: 10.3390/cancers13030384

5. Guo J, Huang X, Dou L, Yan M, Shen T, Tang W, et al. Aging and aging-related diseases: from molecular mechanisms to interventions and treatments. *Signal Transduct Target Ther*. (2022) 7:391. doi: 10.1038/s41392-022-01251-0

6. Mittelbrunn M, Kroemer G. Hallmarks of T cell aging. *Nat Immunol*. (2021) 22:687–98. doi: 10.1038/s41590-021-00927-z

7. Cho SJ, Stout-Delgado HW. Aging and lung disease. *Annu Rev Physiol*. (2020) 82:433–59. doi: 10.1146/annurev-physiol-021119-034610

8. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: An expanding universe. *Cell*. (2023) 186:243–78. doi: 10.1016/j.cell.2022.11.001

9. Campisi J. Aging, cellular senescence, and cancer. *Annu Rev Physiol*. (2013) 75:685–705. doi: 10.1146/annurev-physiol-030212-183653

10. Campisi J, d'Adda di Fagagna F. Cellular senescence: when bad things happen to good cells. *Nat Rev Mol Cell Biol*. (2007) 8:729–40. doi: 10.1038/nrm2233

11. Schmitt CA, Wang B, Demaria M. Senescence and cancer - role and therapeutic opportunities. *Nat Rev Clin Oncol*. (2022) 19:619–36. doi: 10.1038/s41571-022-00668-4

12. Anczuków O, Airhart S, Chuang JH, Coussens LM, Kuchel GA, Korstanje R, et al. Challenges and opportunities for modeling aging and cancer. *Cancer Cell*. (2023) 41:641–5. doi: 10.1016/j.ccell.2023.03.006

13. Hernandez-Segura A, Nehme J, Demaria M. Hallmarks of cellular senescence. *Trends Cell Biol*. (2018) 28:436–53. doi: 10.1016/j.tcb.2018.02.001

14. Calcinotto A, Kohli J, Zagato E, Pellegrini L, Demaria M, Alimonti A. Cellular senescence: aging, cancer, and injury. *Physiol Rev*. (2019) 99:1047–78. doi: 10.1152/physrev.00020.2018

15. Lin W, Wang X, Wang Z, Shao F, Yang Y, Cao Z, et al. Comprehensive analysis uncovers prognostic and immunogenic characteristics of cellular senescence for lung adenocarcinoma. *Front Cell Dev Biol*. (2021) 9:780461. doi: 10.3389/fcell.2021.780461

16. Lin W, Wang X, Xu Z, Wang Z, Liu T, Cao Z, et al. Identification and validation of cellular senescence patterns to predict clinical outcomes and immunotherapeutic responses in lung adenocarcinoma. *Cancer Cell Int*. (2021) 21:652. doi: 10.1186/s12935-021-02358-0

17. Zhang W, Li Y, Lyu J, Shi F, Kong Y, Sheng C, et al. An aging-related signature predicts favorable outcome and immunogenicity in lung adenocarcinoma. *Cancer Sci*. (2022) 113:891–903. doi: 10.1111/cas.15254

18. Lin T, Wang H, Liu Y, Zhao F, He X. Gene set variation analysis-based aging and senescence score as a prognostic indicator and therapeutic guide in lung adenocarcinoma. *Front Genet*. (2023) 14:1176292. doi: 10.3389/fgene.2023.1176292

19. Liu X, Lin L, Cai Q, Sheng H, Zeng R, Zhao Y, et al. Construction and validation of a prognostic model based on novel senescence-related genes in non-small cell lung cancer patients with drug sensitivity and tumor microenvironment. *Adv Biol (Weinh)*. (2023) 7:e2300190. doi: 10.1002/adbi.202300190

20. Xu Q, Chen Y. An aging-related gene signature-based model for risk stratification and prognosis prediction in lung adenocarcinoma. *Front Cell Dev Biol*. (2021) 9:685379. doi: 10.3389/fcell.2021.685379

21. Collisson EA, Campbell JD, Brooks AN, Berger AH, Lee W, Chmielecki J, et al. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. (2014) 511:543–50. doi: 10.1038/nature13385

22. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res*. (2012) 72:100–11. doi: 10.1158/0008-5472.CAN-11-1403

23. Yamauchi M, Yamaguchi R, Nakata A, Kohno T, Nagasaki M, Shimamura T, et al. Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma. *PloS One*. (2012) 7:e43923. doi: 10.1371/journal.pone.0043923

24. Der SD, Sykes J, Pintilie M, Zhu CQ, Strumpf D, Liu N, et al. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. *J Thorac Oncol: Off Publ Int Assoc Study Lung Cancer*. (2014) 9:59–64. doi: 10.1097/JTO.0000000000000042

25. Rousseaux S, Debernardi A, Jacquiau B, Vitte AL, Vesin A, Nagy-Mignotte H, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci Trans Med*. (2013) 5:186ra66. doi: 10.1126/scitranslmed.3005723

26. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res*. (2013) 41:D991–5. doi: 10.1093/nar/gks1193

27. Zhu J, Fan Y, Xiong Y, Wang W, Chen J, Xia Y, et al. Delineating the dynamic evolution from preneoplasia to invasive lung adenocarcinoma by integrating single-cell RNA sequencing and spatial transcriptomics. *Exp Mol Med*. (2022) 54:2060–76. doi: 10.1038/s12276-022-00896-9

28. Avelar RA, Ortega JG, Tacutu R, Tyler EJ, Bennett D, Binetti P, et al. A multidimensional systems biology analysis of cellular senescence in aging and disease. *Genome Biol*. (2020) 21:91. doi: 10.1186/s13059-020-01990-9

29. Tacutu R, Thornton D, Johnson E, Budovsky A, Barardo D, Craig T, et al. Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Res*. (2018) 46:D1083–d1090. doi: 10.1093/nar/gkx1042

30. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. New York: Springer. (2000).

31. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. (2010) 33:1–22. doi: 10.18637/jss.v033.i01

32. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for cox's proportional hazards model via coordinate descent. *J Stat Softw*. (2011) 39:1–13. doi: 10.18637/jss.v039.i05

33. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. (2014) 15:550. doi: 10.1186/s13059-014-0550-8

34. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. (2012) 16:284–7. doi: 10.1089/omi.2011.0118

35. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. (2013) 4:2612. doi: 10.1038/ncomms3612

36. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. (2017) 18:220. doi: 10.1186/s13059-017-1349-1

37. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, et al. IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest*. (2017) 127:2930–40. doi: 10.1172/JCI91190

38. Cristescu R, Mogg R, Ayers M, Albright A, Murphy E, Yearley J, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Sci*. (2018) 362:eaar3593. doi: 10.1126/science.aar3593

39. Wang Y, Weng W, Liang R, Zhou Q, Hu H, Li M, et al. Predicting T cell-inflamed gene expression profile in hepatocellular carcinoma based on dynamic contrast-enhanced ultrasound radiomics. *J Hepatocell Carcinoma*. (2023) 10:2291–303. doi: 10.2147/JHC.S437415

40. Bagaev A, Kotlov N, Nomie K, Svekolkin V, Gafurov A, Isaeva O, et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. *Cancer Cell*. (2021) 39:845–865.e7. doi: 10.1016/j.ccell.2021.04.014

41. He Y, Jiang Z, Chen C, Wang X. Classification of triple-negative breast cancers based on Immunogenomic profiling. *J Exp Clin Cancer Res: CR*. (2018) 37:327. doi: 10.1186/s13046-018-1002-1

42. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf*. (2013) 14:7. doi: 10.1186/1471-2105-14-7

43. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. (2016).

44. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. (2018) 28:1747–56. doi: 10.1101/gr.239244.118

45. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst*. (2019) 8:329–337.e4. doi: 10.1016/j.cels.2019.03.003

46. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al. Comprehensive integration of single-cell data. *Cell*. (2019) 177:1888–1902 e21. doi: 10.1016/j.cell.2019.05.031

47. Liu Y, He S, Wang XL, Peng W, Chen QY, Chi DM, et al. Tumour heterogeneity and intercellular networks of nasopharyngeal carcinoma at single cell resolution. *Nat Commun*. (2021) 12:741. doi: 10.1038/s41467-021-21043-4

48. Chen K, Wang Y, Hou Y, Wang Q, Long D, Liu X, et al. Single cell RNA-seq reveals the CCL5/SDC1 receptor-ligand interaction between T cells and tumor cells in pancreatic cancer. *Cancer Lett*. (2022) 545:215834. doi: 10.1016/j.canlet.2022.215834

49. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*. (2021) 12:1088. doi: 10.1038/s41467-021-21246-9

50. Klement JD, Paschall AV, Redd PS, Ibrahim ML, Lu C, Yang D, et al. An osteopontin/CD44 immune checkpoint controls CD8+ T cell activation and tumor immune evasion. *J Clin Invest*. (2018) 128:5549–60. doi: 10.1172/JCI123360

51. Nallasamy P, Nimmakayala RK, Karmakar S, Leon F, Seshacharyulu P, Lakshmanan I, et al. Pancreatic tumor microenvironment factor promotes cancer stemness via SPP1-CD44 axis. *Gastroenterology*. (2021) 161:1998–2013.e7. doi: 10.1053/j.gastro.2021.08.023

52. Schlensog M, Ruehlmann AC, Haeberle L, Opitz F, Becher AK, Goering W, et al. Tenascin-C affects invasiveness of EGFR-mutated lung adenocarcinoma through a putative paracrine loop. *Biochim Biophys Acta Mol Basis Dis*. (2023) 1869:166684. doi: 10.1016/j.bbadis.2023.166684

53. Liu W, Maben Z, Wang C, Lindquist KC, Li M, Rayannavar V, et al. Structural delineation and phase-dependent activation of the costimulatory CD27:CD70 complex. *J Biol Chem*. (2021) 297:101102. doi: 10.1016/j.jbc.2021.101102

54. Liu H, Lv R, Song F, Yang Y, Zhang F, Xin L, et al. A near-IR ratiometric fluorescent probe for the precise tracking of senescence: a multidimensional sensing assay of biomarkers in cell senescence pathways. *Chem Sci*. (2024) 15:5681–93. doi: 10.1039/D4SC00595C

55. Huang W, Hickson LJ, Eirin A, Kirkland JL, Lerman LO. Cellular senescence: the good, the bad and the unknown. *Nat Rev Nephrol*. (2022) 18:611–27. doi: 10.1038/s41581-022-00601-z

56. Basisty N, Kale A, Jeon OH, Kuehnemann C, Payne T, Rao C, et al. A proteomic atlas of senescence-associated secretomes for aging biomarker development. *PloS Biol*. (2020) 18:e3000599. doi: 10.1371/journal.pbio.3000599

57. Hurd MW, Ralph MR. The significance of circadian organization for longevity in the golden hamster. *J Biol Rhythms*. (1998) 13:430–6. doi: 10.1177/074873098129000255

58. Yang WS, Stockwell BR. Inhibition of casein kinase 1-epsilon induces cancer-cell-selective, PERIOD2-dependent growth arrest. *Genome Biol*. (2008) 9:R92. doi: 10.1186/gb-2008-9-6-r92

59. Rodriguez N, Yang J, Hasselblatt K, Liu S, Zhou Y, Rauh-Hain JA, et al. Casein kinase I epsilon interacts with mitochondrial proteins for the growth and survival of human ovarian cancer cells. *EMBO Mol Med*. (2012) 4:952–63. doi: 10.1002/emmm.201101094

60. Yang J, Jiang Q, Liu L, Peng H, Wang Y, Li S, et al. Identification of prognostic aging-related genes associated with immunosuppression and inflammation in head and neck squamous cell carcinoma. *Aging*. (2020) 12:25778–804. doi: 10.18632/aging.v12i24

61. Park BJ, Kang JW, Lee SW, Choi SJ, Shin YK, Ahn YH, et al. The haploinsufficient tumor suppressor p18 upregulates p53 via interactions with ATM/ATR. *Cell*. (2005) 120:209–21. doi: 10.1016/j.cell.2004.11.054

62. Hassan MK, Kumar D, Naik M, Dixit M. The expression profile and prognostic significance of eukaryotic translation elongation factors in different cancers. *PloS One*. (2018) 13:e0191377. doi: 10.1371/journal.pone.0191377

63. Oh YS, Kim DG, Kim G, Choi EC, Kennedy BK, Suh Y, et al. Downregulation of lamin A by tumor suppressor AIMP3/p18 leads to a progeroid phenotype in mice. *Aging Cell*. (2010) 9:810–22. doi: 10.1111/j.1474-9726.2010.00614.x

64. Sirover MA. Pleiotropic effects of moonlighting glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in cancer progression, invasiveness, and metastases. *Cancer Metastasis Rev*. (2018) 37:665–76. doi: 10.1007/s10555-018-9764-7

65. Nicholls C, Li H, Liu JP. GAPDH: a common enzyme with uncommon functions. *Clin Exp Pharmacol Physiol*. (2012) 39:674–9. doi: 10.1111/j.1440-1681.2011.05599.x

66. Gamez-Belmonte R, Mahapatro M, Erkert L, Gonzalez-Acera M, Naschberger E, Yu Y, et al. Epithelial presenilin-1 drives colorectal tumour growth by controlling EGFR-COX2 signalling. *Gut*. (2023) 72:1155–66. doi: 10.1136/gutjnl-2022-327323

67. Shen J, Bronson RT, Chen DF, Xia W, Selkoe DJ, Tonegawa S. Skeletal and CNS defects in Presenilin-1-deficient mice. *Cell*. (1997) 89:629–39. doi: 10.1016/S0092-8674(00)80244-5

68. Yaffe MB, Rittinger K, Volinia S, Caron PR, Aitken A, Leffers H, et al. The structural basis for 14-3-3:phosphopeptide binding specificity. *Cell*. (1997) 91:961–71. doi: 10.1016/S0092-8674(00)80487-0

69. Ogihara T, Isobe T, Ichimura T, Taoka M, Funaki M, Sakoda H, et al. 14-3-3 protein binds to insulin receptor substrate-1, one of the binding sites of which is in the phosphotyrosine binding domain. *J Biol Chem*. (1997) 272:25267–74. doi: 10.1074/jbc.272.40.25267

70. Tatar M, Bartke A, Antebi A. The endocrine regulation of aging by insulin-like signals. *Sci (New York NY)*. (2003) 299:1346–51. doi: 10.1126/science.1081447

71. Chen CH, Chuang SM, Yang MF, Liao JW, Yu SL, Chen JJ. A novel function of YWHAZ/β-catenin axis in promoting epithelial-mesenchymal transition and lung cancer metastasis. *Mol Cancer Res*. (2012) 10:1319–31. doi: 10.1158/1541-7786.MCR-12-0189

72. Orjalo AV, Bhaumik D, Gengler BK, Scott GK, Campisi J. Cell surface-bound IL-1alpha is an upstream regulator of the senescence-associated IL-6/IL-8 cytokine network. *Proc Natl Acad Sci United States America*. (2009) 106:17031–6. doi: 10.1073/pnas.0905299106

73. Leon KE, Buj R, Lesko E, Dahl ES, Chen CW, Tangudu NK, et al. DOT1L modulates the senescence-associated secretory phenotype through epigenetic regulation of IL1A. *J Cell Biol*. (2021) 220:e202008101. doi: 10.1083/jcb.202008101

74. Garlanda C, Dinarello CA, Mantovani A. The interleukin-1 family: back to the future. *Immunity*. (2013) 39:1003–18. doi: 10.1016/j.immuni.2013.11.010

75. Wang YA, Sun Y, Palmer J, Solomides C, Huang LC, Shyr Y, et al. IGFBP3 modulates lung tumorigenesis and cell growth through IGF1 signaling. *Mol Cancer Res*. (2017) 15:896–904. doi: 10.1158/1541-7786.MCR-16-0390

76. Donoho G, Brenneman MA, Cui TX, Donoviel D, Vogel H, Goodwin EH, et al. Deletion of Brca2 exon 27 causes hypersensitivity to DNA crosslinks, chromosomal instability, and reduced life span in mice. *Genes Chromosomes Cancer*. (2003) 36:317–31. doi: 10.1002/gcc.10148

77. Vogel H, Lim DS, Karsenty G, Finegold M, Hasty P. Deletion of Ku86 causes early onset of senescence in mice. *Proc Natl Acad Sci United States America*. (1999) 96:10770–5. doi: 10.1073/pnas.96.19.10770

78. Fan Y, Gao Z, Li X, Wei S, Yuan K. Gene expression and prognosis of x-ray repair cross-complementing family members in non-small cell lung cancer. *Bioengineered*. (2021) 12:6210–28. doi: 10.1080/21655979.2021.1964193

79. Satyananda V, Oshi M, Endo I, Takabe K. High BRCA2 gene expression is associated with aggressive and highly proliferative breast cancer. *Ann Surg Oncol*. (2021) 28:7356–65. doi: 10.1245/s10434-021-10063-5

80. Yan B, Xie B, Huang M, Guo J, Sun J, Chen J, et al. Mutations and expressions of breast cancer 1/2 in lung cancer. *Thorac Cancer*. (2023) 14:1753–63. doi: 10.1111/1759-7714.14920

81. López-Otín C, Pietrocola F, Roiz-Valle D, Galluzzi L, Kroemer G. Meta-hallmarks of aging and cancer. *Cell Metab*. (2023) 35:12–35. doi: 10.1016/j.cmet.2022.11.001

82. Bao H, Cao J, Chen M, Chen M, Chen W, Chen X, et al. Biomarkers of aging. *Sci China Life Sci*. (2023) 66:893–1066. doi: 10.1007/s11427-023-2305-0

83. Lian I, Kim J, Okazawa H, Zhao J, Zhao B, Yu J, et al. The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. *Genes Dev*. (2010) 24:1106–18. doi: 10.1101/gad.1903310

84. Pan D. The hippo signaling pathway in development and cancer. *Dev Cell*. (2010) 19:491–505. doi: 10.1016/j.devcel.2010.09.011

85. Wang CX, Yan J, Lin S, Ding Y, Qin YR. Mutant-allele dispersion correlates with prognosis risk in patients with advanced non-small cell lung cancer. *J Cancer Res Clin Oncol*. (2023) 149:8545–55. doi: 10.1007/s00432-023-04801-3

86. Hernández Borrero LJ, El-Deiry WS. Tumor suppressor p53: Biology, signaling pathways, and therapeutic targeting. *Biochim Biophys Acta Rev Cancer*. (2021) 1876:188556. doi: 10.1016/j.bbcan.2021.188556

87. Qin C, Fan X, Sai X, Yin B, Zhou S, Addeo A, et al. Development and validation of a DNA damage repair-related gene-based prediction model for the prognosis of lung adenocarcinoma. *J Thorac Dis*. (2023) 15:6928–45. doi: 10.21037/jtd

88. Zhao Y, Qing B, Xu C, Zhao J, Liao Y, Cui P, et al. DNA damage response gene-based subtypes associated with clinical outcomes in early-stage lung adenocarcinoma. *Front Mol Biosci*. (2022) 9:901829. doi: 10.3389/fmolb.2022.901829

89. Chen HA, Ho YJ, Mezzadra R, Adrover JM, Smolkin R, Zhu C, et al. Senescence rewires microenvironment sensing to facilitate antitumor immunity. *Cancer Discovery*. (2023) 13:432–53. doi: 10.1158/2159-8290.CD-22-0528

90. Marin I, Boix O, Garcia-Garijo A, Sirois I, Caballe A, Zarzuela E, et al. Cellular senescence is immunogenic and promotes antitumor immunity. *Cancer Discovery*. (2023) 13:410–31. doi: 10.1158/2159-8290.CD-22-0523

91. Shahbandi A, Chiu FY, Ungerleider NA, Kvadas R, Mheidly Z, Sun MJS, et al. Breast cancer cells survive chemotherapy by activating targetable immune-modulatory programs characterized by PD-L1 or CD80. *Nat Cancer*. (2022) 3:1513–33. doi: 10.1038/s43018-022-00466-y

92. Prieto LI, Sturmlechner I, Goronzy JJ, Baker DJ. Senescent cells as thermostats of antitumor immunity. *Sci Trans Med*. (2023) 15:eadg7291. doi: 10.1126/scitranslmed.adg7291

93. Holze C, Michaudel C, Mackowiak C, Haas DA, Benda C, Hubel P, et al. Oxeiptosis, a ROS-induced caspase-independent apoptosis-like cell-death pathway. *Nat Immunol*. (2018) 19:130–40. doi: 10.1038/s41590-017-0013-y

94. Chen Z, Han F, Du Y, Shi H, Zhou W. Hypoxic microenvironment in cancer: molecular mechanisms and therapeutic interventions. *Signal Transduct Target Ther*. (2023) 8:70. doi: 10.1038/s41392-023-01332-8

Frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 19 September 2024
DOI 10.3389/fimmu.2024.1398685

OPEN ACCESS

EDITED BY
Petar Ozretić,
Rudjer Boskovic Institute, Croatia

REVIEWED BY
Somayah Abdullah Albaradei,
King Abdullah University of Science and Technology, Saudi Arabia
Arkady Bedzhanyan,
Petrovsky National Research Center of Surgery, Russia

*CORRESPONDENCE
Liang Yang
✉ ncu_yangliang@163.com

†These authors have contributed equally to this work

RECEIVED 10 March 2024
ACCEPTED 29 August 2024
PUBLISHED 19 September 2024

CITATION
Yang K, Wu J, Xu T, Zhou Y, Liu W and Yang L (2024) Machine learning to predict distant metastasis and prognostic analysis of moderately differentiated gastric adenocarcinoma patients: a novel focus on lymph node indicators.
Front. Immunol. 15:1398685.
doi: 10.3389/fimmu.2024.1398685

Machine learning to predict distant metastasis and prognostic analysis of moderately differentiated gastric adenocarcinoma patients: a novel focus on lymph node indicators

Kangping Yang[1]†, Jiaqiang Wu[2]†, Tian Xu[3]†, Yuepeng Zhou[1], Wenchun Liu[4] and Liang Yang[1]*

[1]Department of Gastroenterological Surgery, The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, Jiangxi, China, [2]Department of General Surgery, First Medical Center of the Chinese People's Liberation Army General Hospital, Beijing, China, [3]Department of Gastroenterological Surgery, Jiangxi Hospital of Integrated Traditional Chinese and Western Medicine, Nanchang, Jiangxi, China, [4]The Second Department of Internal Medicine, Anfu People's Hospital, Anfu, Jiangxi, China

Background: Moderately differentiated gastric adenocarcinoma (MDGA) has a high risk of metastasis and individual variation, which strongly affects patient prognosis. Using large-scale datasets and machine learning algorithms for prediction can improve individualized treatment. The specific efficacy of several lymph node indicators in predicting distant metastasis (DM) and patient prognosis in MDGA remains obscure.

Methods: We collected data from MDGA patients from the SEER database from 2010 to 2019. Additionally, we collected data from MDGA patients in China. We used nine machine learning algorithms to predict DM. Subsequently, we used Cox regression analysis to determine the risk factors affecting overall survival (OS) and cancer-specific survival (CSS) in DM patients and constructed nomograms. Furthermore, we used logistic regression and Cox regression analyses to assess the specific impact of six lymph node indicators on DM incidence and patient prognosis.

Results: We collected data from 5,377 MDGA patients from the SEER database and 109 MDGC patients from hospitals. T stage, N stage, tumor size, primary site, number of positive lymph nodes, and chemotherapy were identified as independent risk factors for DM. The random forest prediction model had the best overall predictive performance (AUC = 0.919). T stage, primary site, chemotherapy, and the number of regional lymph nodes were identified as prognostic factors for OS. Moreover, T stage, number of regional lymph nodes, primary site, and chemotherapy were also influential factors for CSS. The nomograms showed good predictive value and stability in predicting the 1-, 3-, and 5-year OS and CSS in DM patients. Additionally, the log odds of a metastatic lymph node and the number of negative lymph nodes may be risk factors for DM, while the regional lymph node ratio and the number of regional lymph nodes are prognostic factors for OS.

Conclusion: The random forest prediction model accurately identified high-risk populations, and we established OS and CSS survival prediction models for MDGA patients with DM. Our hospital samples demonstrated different characteristics of lymph node indicators in terms of distant metastasis and prognosis.

KEYWORDS

moderately differentiated gastric adenocarcinoma, prognosis, nomogram, lymph node indicators, distant metastasis, machine learning

1 Introduction

Gastric cancer, a very prevalent gastrointestinal tumor, is the fifth most prevalent tumor worldwide (1). In 2020, there were more than one million additional cases of gastric cancer (2). The histologic type of gastric cancer is predominantly adenocarcinoma, and the pathologic grade includes highly, moderately, and poorly differentiated and undifferentiated (3, 4). Although progressive gastric cancer is predominantly poorly differentiated, some moderately differentiated gastric adenocarcinomas (MDGAs) still have a high risk of metastasis and individual differences, which have been reported in animal models and clinical studies (5–7). There is no doubt that the occurrence of distant metastasis (DM) directly affects patient prognosis (8). According to the latest eighth revision of the UICC/AJCC TNM classification for gastric cancer, once DM occurs, the disease has already entered stage IV, at which time the patient's survival chances are extremely poor (9). A retrospective study showed that the median overall survival (OS) time for patients with liver metastases from gastric cancer was 7 months and that for patients with lung and brain metastases (10) was only 5 months. Timely and accurate determination of the distant metastasis status of gastric cancer patients has important positive implications for avoiding missing opportunities for early and effective interventions and improving patient survival.

Currently, tests to clarify the occurrence of DM mainly rely on multidetector computed tomography (CT), positron emission tomography-CT (PET/CT), and other imaging methods (11, 12). However, all of these methods have the problem of insufficient sensitivity in practical applications (13). For example, in PET/CT, some poorly differentiated carcinomas, mucinous carcinomas, and indolent cell carcinomas usually have low ^{18}F-FDG uptake, which often results in false-negative results and delayed therapy (14). Therefore, there is an urgent need for an accurate, convenient, yet affordable method for DM diagnosis and prediction. The use of emerging machine learning (ML) algorithms and large-scale datasets to construct predictive models is currently a popular solution (15–17). ML algorithms are able to accurately process raw data originating from databases, analyze the relationships between important data, and ultimately build and filter the best predictive models (18–21). This prediction model, which integrates clinical manifestations and imaging data to form a comprehensive assessment tool, can be used to diagnose the presence or absence of DM early and accurately and can better guide subsequent clinical diagnosis and treatment.

For patients with already occurring DM, the median OS after performing conventional chemotherapy is approximately 12 months (22). With regard to cancer-specific survival (CSS), the 1- and 3-year CSS rates for the younger group (≤60 years of age) were 29.0% and 6.2%, respectively, compared with 22.8% and 4.8% for the older group (>60 years of age), respectively (23). These findings suggest that there are many factors that can influence DM patient prognosis, and clarifying the effects of these factors and applying them in a targeted manner are important ways to improve patient prognosis. Many studies have demonstrated that factors such as age, tumor size, sex, degree of differentiation, and primary site are directly associated with DM patient prognosis (24–26). Moreover, recent studies have demonstrated a strong association between various lymph node indicators and DM and the prognosis of moderately differentiated gastric adenocarcinoma patients. For example, lymph node-specific indicators include the number of positive lymph nodes (PLNs), the lymph node ratio (LNR), and the log odds of metastatic LNs (LODDS) (27–29). However, the specific efficacy of these lymph node indicators in predicting DM and patient prognosis is unclear (30–33). This study explored these prognostic factors in DM patients in the MDGA to provide strong theoretical support for individualized treatment in this population. Afterward, the above factors were combined to construct OS and CSS prognostic nomograms at 1, 3, and 5 years for DM patients with MDGA, which is a simplified visualization model for statistical prediction in combination with independent factors.

Our goal was to formulate models for predicting DM in MDGA patients and to ensure the stability and accuracy of these models through both database validation and external validation. A prognostic analysis of DM patients was then performed to plot OS and CSS prognostic nomograms for MDGA patients. Importantly, we focused on exploring the relationships between various lymph node indicators whose efficacy is still unclear and between DM and prognosis to further promote the application of lymph node indicators in the clinical practice of stomach cancer diagnosis.

2 Materials and methods

2.1 Sources of data and sample selection

The primary training dataset was obtained by collecting all 2010–2019 gastric cancer patient data from the Surveillance, Epidemiology, and End Results (SEER) database. The SEER database is the most detailed publicly available cancer database. Moreover, we collected the clinical data of MDGA patients treated at the Second Affiliated Hospital of Nanchang University between 2008 and 2010 as an external validation dataset. The inclusion criteria were as follows: 1) had a diagnosis of MDGA, 2) did not receive preoperative radiotherapy or immunotherapy, and 3) had comprehensive and searchable prognostic data. The exclusion criteria were as follows: 1) patients whose primary tumor was not gastric cancer, 2) patients whose tumor and lymph node status were not clear, and 3) patients whose other basic information was incomplete. The specific data selection steps are illustrated below in Figure 1.

2.2 Variable selection

Variables in the present study included age, TNM stage, primary site, tumor size, sex (male or female), and two therapeutic variables (chemotherapy and radiation) obtained from the diagnostic information, as well as several lymph node indicators. Multiple lymph node indicators included the number of Reg LNs, number of all LNs, number of Reg LNs, number of Neg LNs, gross LN metastasis, LN positivity rate, log odds of metastatic LNs, and lymph node ratio (number of metastatic LNs to total number of LNs examined).

OS and CSS are the main outcomes for predicting the prognosis of patients with DM. In OS, deaths due to any cause will be counted, while in CSS analysis, only deaths due to MDGA will be considered events, and deaths due to other factors as well as survival will be excluded.

2.3 Statistical methods

The research procedure is illustrated in Figure 2. Heatmapping was first developed to correlate the proposed study variables. We use regression analysis and machine learning for dual validation of risk factors; regression analysis is performed using the full SEER data, and machine learning uses the training set, the test set, and the external validation set to construct predictive models. Independent risk factors influencing DM in moderately differentiated gastric adenocarcinoma patients were screened by logistic regression analysis. The outcomes are expressed as hazard ratios (HRs) and 95% confidence intervals (CIs). The patient data screened from the SEER dataset were randomized 7:3 into a training set and a test set. Then, the training set will be utilized to build the predictive model. The constructed predictive models are then tested and evaluated using the test set data. We constructed nine ML algorithms in the training set, including RF (random forest), LR (logistic regression), LASSO (least absolute shrinkage and selection operator), SVM

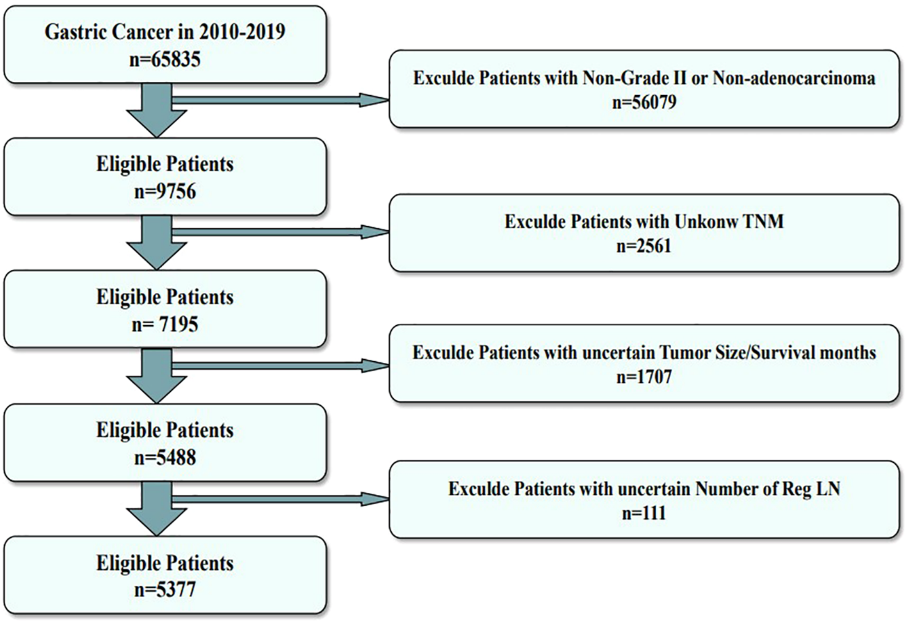


FIGURE 1
Flowchart of the data screening process. The figure shows the process of filtering eligible patient data from the SEER database.

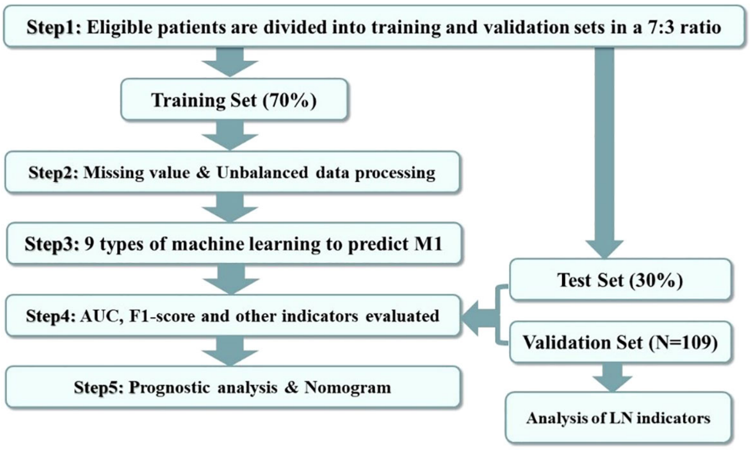


FIGURE 2
Data analysis guide. The figure shows the procedure of this study for processing and analyzing the screened data.

(support vector machine), KNN (K-nearest neighbor), NBC (naive Bayes classifier), and ANN (artificial neural network). The receiver operating characteristic (ROC) curve, the area under the ROC curve (AUC), sensitivity, specificity, F1 score, and accuracy were used to compare the performance of the models. Additionally, the predictive models were evaluated and validated using test set data. Self-collected hospital patient data were used as an external validation set to validate the best predictive model that assessed the generalization ability of the model.

We used several R packages in R for data analysis and visualization. The createDataPartition function of the caret package was used for grouping the training and validation sets. The imp function of randomForest package was used to construct the importance scores of RF. The coords function of the pROC package was used to construct the confusion matrix. The randomForest package, the MASS package, the rms package, the glmnet package, the e1071 package, the xgboost package, the adabag package, and the neuralnet package were all used for machine learning model construction. The MASS package, rms package, glmnet package, e1071 package, xgboost package, adabag package, and neuralnet package were used for the construction of machine learning models. The ggplot package and pROC package were used for the visualization of ROC curves and importance scores.

For survival prognostic analyses, single-variable Cox regression analysis was first adopted to screen the relevant variables that could influence the prognosis ($P < 0.05$), and then multifactorial analyses were carried out on the screened variables. Moreover, we used the Kaplan–Meier curves to assess the differences in survival prognosis among patients stratified by different variables and compared the results by means of the log-rank test. The independent risk factors identified through Cox regression analysis were used to construct the nomogram. Moreover, using multifactor Cox regression analysis, the regression coefficients β (coe β) for each variable were normalized and are displayed as risk scores on the nomograms. The accuracy and discriminatory power of the generated nomograms were assessed with the AUC, calibration curves, and consistency index (C-index). In addition, we evaluated the clinical value of the nomograms by using decision curve analysis (DCA). This is a commonly used measure to assess model validity by quantitatively estimating the net effectiveness under the exposure threshold.

Finally, the impact of multiple nuanced tumor-associated LN indicators on the development of DM in MDGA patients was investigated by logistic regression analysis of patient data collected at our institution, as well as factors affecting patient prognosis. For descriptive statistics, categorical variables were compared using the chi-square test or Fisher's exact probability method. P <0.05 indicated statistical significance.

2.4 Ethics approval

The use of patient data in this research has been authorized. The approval document from the Ethics Committee is shown in the attachment. Patients from the SEER database provided consent for open research in any scientific study worldwide.

3 Results

3.1 Basic features and patient subgroups

Altogether, 5,377 patients from the database were included in this study; 749 (13.93%) had DM, and 4,628 (86.07%) had no DM. The local patient dataset, which served as an external validation set, included a total of 109 patients, of whom 15 (13.76%) had DM and 94 (86.24%) had no DM. The patient data screened from the SEER dataset were randomized 7:3 into training and testing sets. The

TABLE 1 Comparison of the general features of the training and test sets.

Variable	Training set (%) *N* = 3,764	Test set (%) *N* = 1,613	*P*	Validation set (%) *N* = 109
Age (years)			0.800	
<40	42 (1.1%)	23 (1.4%)		4
40–60	675 (17.9%)	283 (17.5%)		35
60–80	2,215 (58.8%)	948 (58.8%)		67
>80	832 (22.1%)	359 (22.3%)		3
Sex			0.443	
Male	2,673 (71.0%)	1,128 (69.9%)		83
Female	1,087 (29.0%)	489 (30.1%)		26
T stage			0.612	
1	1,364 (36.2%)	573 (35.5%)		23
2	561 (14.9%)	230 (14.3%)		17
3	1,334 (35.4%)	573 (35.5%)		55
4	505 (13.4%)	237 (14.7%)		14
N stage			0.997	
0	2,012 (53.5%)	865 (53.6%)		45
1	976 (25.9%)	417 (25.9%)		40
2	471 (12.5%)	199 (12.3%)		12
3	305 (8.1%)	132 (8.2%)		12
M stage			0.919	
0	3,238 (86.0%)	1,390 (86.2%)		94
1	526 (14.0%)	223 (13.8%)		15
Primary site			0.419	
Body	303 (8.0%)	143 (8.9%)		7
Cardia	1,509 (40.1%)	650 (40.3%)		40
Fundus	134 (3.6%)	59 (3.7%)		2
Gastric antrum	817 (21.7%)	378 (23.4%)		30
Greater curvature	141 (3.7%)	50 (3.1%)		2
Lesser curvature	342 (9.1%)	131 (8.1%)		8
Overlapping lesion	182 (4.8%)	77 (4.8%)		12
Pylorus	136 (3.6%)	43 (2.7%)		4
Stomach	200 (5.3%)	82 (5.1%)		4
Tumor size (cm)			0.280	
<2	971 (25.8%)	457 (28.3%)		24
2 to 5	1792 (47.6%)	744 (46.1%)		51
5 to 8	766 (20.4%)	312 (19.3%)		23
>8	235 (6.2%)	100 (6.2%)		11
Number of Reg LN			0.997	
None	1,204 (32.0%)	515 (31.9%)		47
1 to 3	167 (4.4%)	71 (4.4%)		22
4 or more	2,393 (63.6%)	1,027 (63.7%)		40

(Continued)

TABLE 1 Continued

results of the analyses, as shown in Table 1, show the comprehensive demographic and clinical characteristics of the three groups of MDGA patients. Additionally, there were no statistically significant differences ($P > 0.05$) in any of the clinical characteristics analyzed, such as tumor size, primary site, TNM stage, or number of Reg LNs, between the patients in the training and testing sets.

3.2 Comparison and analysis of model variables

Pearson correlation analysis was used to examine the relationship between each variable (Figure 3A). Moreover, with the comprehensive consideration of the type of data, distribution characteristics, and other factors, all the variables are independent and well-distributed and can be included in the following statistical analysis. By multifactorial logistic regression analysis, this study revealed that six variables were statistically significant in predicting the occurrence of DM in patients with MDGA (Table 2). These included the T and N stages, but the M stage seemed to be not significantly different. Other variables included primary site, tumor size, number of Reg LNs, and chemotherapy. In addition, the importance scores from the random forest model indicated variable significance (as displayed in Figure 3B). The number of Reg LNs, N stage, T stage, chemotherapy, age, tumor size, and primary site were positively related to the occurrence of DM in MDGA patients. Specifically, the outcome was the same as the findings of the former correlation analyses, except for age. With distant metastasis as the outcome variable, we conducted single- and multiple-factor logistic regression analyses on eight factors: primary site, tumor size, age, sex, T stage, N stage, number of positive LNs, and chemotherapy. Multiple factor regression was performed, and step-forward analysis revealed that the *P*-values for T stage, N stage, primary site, number of positive LNs, tumor size, and chemotherapy were less than 0.05 and were considered statistically significant independent risk factors. The results of forward regression analysis indicated the meaningful impact of six variables on distant metastasis: sex, T stage, N stage, primary

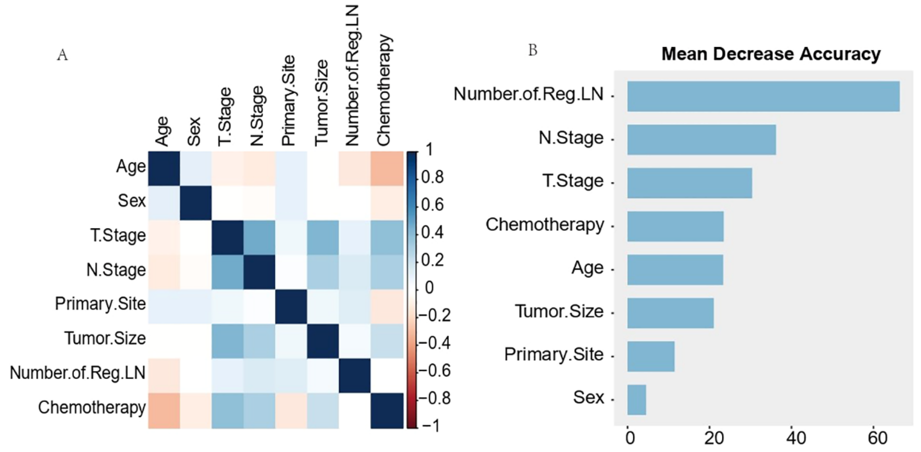


FIGURE 3
Results of Pearson correlation analysis for each variable **(A)** and ranking of importance of predictive model characteristics **(B)**. The results of Pearson correlation analysis for each variable **(A)** showed that all variables existed independently of each other. The predictive model characteristics **(B)** were the number of Reg LNs, N stage, T stage, chemotherapy, age, tumor size, and primary site, in order of importance.

site, tumor size, and number of positive LNs (the detailed results are presented in the Supplementary Material).

3.3 Establishment of predictive models for distant metastasis

This research used nine distinct machine learning models individually to construct a distant metastasis prediction model for MDGA patients. The built models were trained with data from the training set. The symptoms were finely tuned to stabilize the models and prevent them from overfitting.

Table 3 and Figure 4A present the evaluation standards for each model comparison, including ROC curves, specificity, sensitivity, accuracy, recall, and F1 score. Based on the comparison results, it is concluded that the random forest model has the highest predictive value. Its AUC (0.913), specificity (0.891), and accuracy (0.880) were the best among the nine models. The results in the testing set verified this point again. The AUC of the ROC curve for the RF model was 0.848 (Figure 4B), which was noticeably superior to those of the other

TABLE 2 Multifactorial analysis of moderately differentiated distant metastatic gastric adenocarcinoma.

Variables	Beta	S.E	*Z*	OR (95% CI)	*P*	aBeta	aS.E	a*Z*	aOR (95% CI)	a*P*
Age										
<40				1.00 (reference)						
40–60	−0.26	0.37	−0.71	0.77 (0.37–1.59)	0.478					
60–80	−0.60	0.36	−1.64	0.55 (0.27–1.12)	0.100					
>80	−0.62	0.37	−1.67	0.54 (0.26–1.11)	0.094					
Sex										
Male				1.00 (reference)						
Female	−0.16	0.11	−1.53	0.85 (0.69–1.05)	0.125					
T stage										
1				1.00 (reference)					1.00 (reference)	
2	−0.55	0.19	−2.90	0.58 (0.40–0.84)	0.004	−0.69	0.21	−3.27	0.50 (0.33–0.76)	0.001
3	0.25	0.12	2.12	1.28 (1.02–1.61)	0.034	−0.19	0.15	−1.25	0.82 (0.61–1.12)	0.210
4	1.25	0.13	9.72	3.49 (2.71–4.49)	<0.001	0.68	0.17	3.94	1.97 (1.41–2.76)	<0.001
N stage										
0				1.00 (reference)					1.00 (reference)	
1	1.12	0.11	10.19	3.08 (2.48–3.82)	<0.001	1.00	0.14	7.16	2.73 (2.07–3.60)	<0.001

(Continued)

TABLE 2 Continued

Variables	Beta	S.E	*Z*	OR (95% CI)	*P*	aBeta	aS.E	a*Z*	aOR (95% CI)	a*P*
N stage										
2	0.59	0.16	3.75	1.80 (1.32–2.44)	<0.001	0.89	0.20	4.52	2.45 (1.66–3.60)	<0.001
3	1.23	0.16	7.83	3.42 (2.52–4.66)	<0.001	1.84	0.21	8.76	6.29 (4.17–9.49)	<0.001
Primary site										
Cardia				1.00 (reference)					1.00 (reference)	
Gastric antrum	−0.16	0.13	−1.22	0.85 (0.66–1.10)	0.222	0.58	0.17	3.49	1.78 (1.29–2.47)	<0.001
Lesser curvature	−0.29	0.19	−1.57	0.75 (0.52–1.07)	0.116	0.21	0.22	0.98	1.24 (0.81–1.89)	0.326
Pylorus	−0.85	0.35	−2.41	0.43 (0.21–0.85)	0.016	−0.11	0.40	−0.29	0.89 (0.41–1.94)	0.772
Body	0.14	0.17	0.84	1.15 (0.83–1.60)	0.400	0.58	0.20	2.88	1.79 (1.20–2.66)	0.004
Greater curvature	0.08	0.24	0.31	1.08 (0.67–1.74)	0.754	0.36	0.29	1.22	1.43 (0.81–2.52)	0.222
Stomach	0.30	0.20	1.54	1.35 (0.92–1.99)	0.124	0.83	0.24	3.52	2.29 (1.44–3.63)	<0.001
Overlapping lesion	0.11	0.22	0.51	1.12 (0.73–1.70)	0.611	0.08	0.26	0.32	1.09 (0.65–1.81)	0.746
Fundus	0.13	0.23	0.56	1.14 (0.72–1.80)	0.579	0.13	0.28	0.45	1.13 (0.66–1.95)	0.649
Tumor size										
<2				1.00 (reference)					1.00 (reference)	
2 to 5	1.11	0.15	7.18	3.02 (2.23–4.09)	<0.001	0.85	0.17	4.93	2.35 (1.67–3.30)	<0.001
5 to 8	1.55	0.17	9.35	4.69 (3.39–6.49)	<0.001	1.24	0.19	6.38	3.44 (2.36–5.04)	<0.001
>8	1.75	0.20	8.68	5.75 (3.87–8.54)	<0.001	1.15	0.24	4.78	3.15 (1.97–5.04)	<0.001
Number of Reg LN										
None				1.00 (reference)					1.00 (reference)	
1 to 3	−1.18	0.24	−4.88	0.31 (0.19–0.49)	<0.001	−1.38	0.27	−5.22	0.25 (0.15–0.42)	<0.001
4 or more	−2.08	0.11	−18.95	0.12 (0.10–0.15)	<0.001	−2.75	0.14	−19.25	0.06 (0.05–0.08)	<0.001
Chemotherapy										
No/unknown				1.00 (reference)					1.00 (reference)	
Yes	0.66	0.10	6.90	1.94 (1.61–2.35)	<0.001	0.36	0.12	2.92	1.43 (1.13–1.82)	0.003

eight models. Finally, the RF models were externally validated using our 109 hospital patients (AUC = 0.728) (Figure 4C). We also made an aggregation of the previous ROC curves (Figure 4D). In summary, we trained eight machine learning prediction models with data from the training set. Through the experimental results of the test set and validation set, it was determined that the RF model has a relatively accurate ability to predict the risk of DM in MDGA patients and has high clinical value.

4 Prognostic analysis and prediction of MDGA patients with established DM

4.1 Patient baseline characteristics

The 749 eligible MDGA patients with DM were randomized into two groups in the same 7:3 split. The training set included 524 patients, while the testing set included 225 patients. In terms of clinical characteristics, 40–60 years of age was the most common age for distant metastasis according to the MDGA (55.41% according to the SEER data), and the highest proportion of distant metastases according to the MDGA originated in cardia (42.86% according to the SEER data). Descriptions of other clinical characteristics are summarized in the accompanying table (Table 4). The results suggested that no statistically significant differences were found between the basic information of the two datasets.

4.2 Analysis of prognosis-related factors

Using OS and CSS as prognostic endpoints, we performed univariate and multivariate Cox regression analyses on data from eligible patients screened from the SEER database. Nine variables were included in the univariate analysis, and the detailed results are

TABLE 3 Comparison of the predictive performance values of nine forecasting models in the training set.

Models	AUC	Specificity	Sensitivity	Accuracy	Precision	Recall	F1 score
RF	0.913	0.891	0.811	0.880	0.548	0.811	0.654
LR	0.848	0.791	0.766	0.787	0.372	0.766	0.501
LASSO	0.848	0.791	0.766	0.787	0.372	0.766	0.501
SVM	0.872	0.834	0.760	0.823	0.425	0.760	0.545
XGBoost	0.989	0.792	0.836	0.798	0.394	0.836	0.536
KNN	0.885	0.740	0.853	0.756	0.348	0.853	0.494
NBC	0.825	0.641	0.870	0.673	0.282	0.870	0.426
AdaBoost	0.900	0.811	0.823	0.813	0.414	0.823	0.551
ANN	0.850	0.749	0.792	0.755	0.338	0.792	0.474

RF, random forest; LR, logistic regression; LASSO, least absolute shrinkage and selection operator; SVM, support vector machine; XGBoost, eXtreme Gradient Boosting; KNN, K-nearest neighbor; NBC, naive Bayesian model; AdaBoost, adaptive boosting; ANN, artificial neural network.

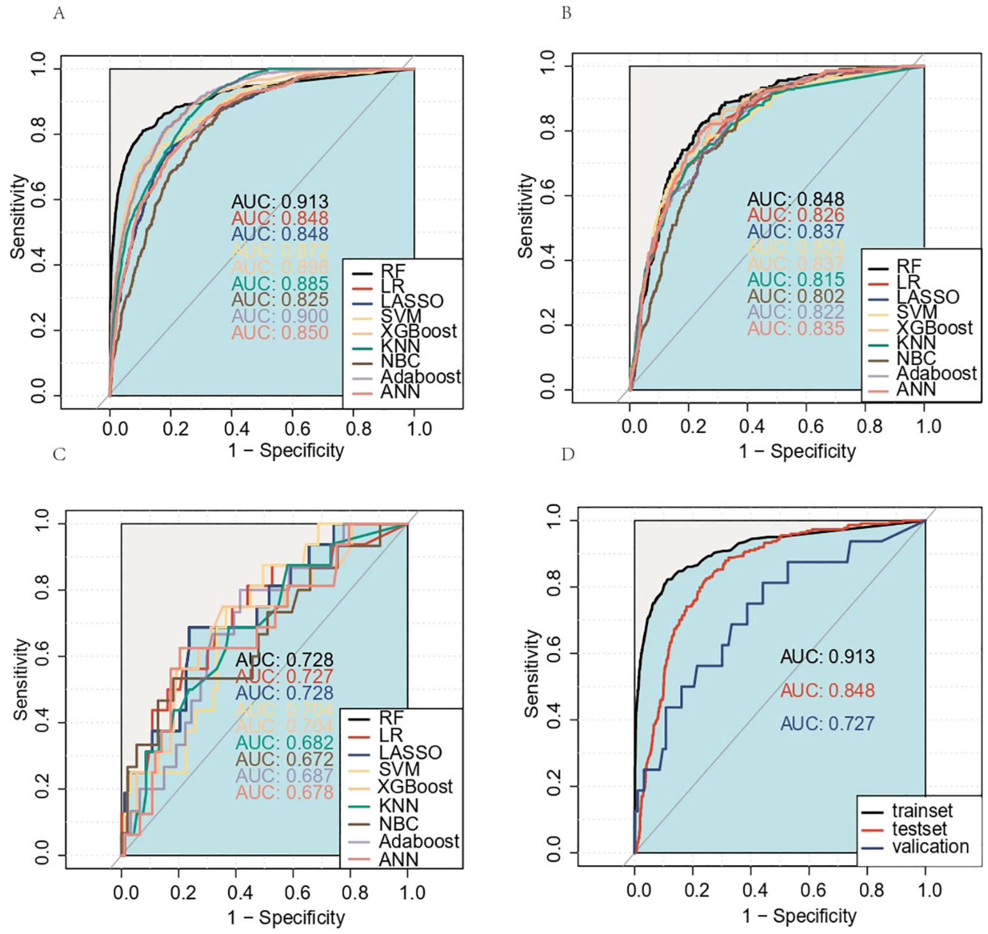


FIGURE 4
Receiver operating characteristic (ROC) curves for the training set, test set, and external validation set prediction models. **(A)** Training set; **(B)** test set; **(C)** external validation set. The aggregation of the previous ROC curves for the RF model **(D)**. AUC, area under the ROC curve; RF, random forest; LR, logistic regression; LASSO, least absolute shrinkage and selection operator; SVM, support vector machine; KNN, K-nearest neighbor; NBC, naive Bayes classifier; ANN, artificial neural network.

TABLE 4 Basic information about MDGA patients with DM.

Variable	Total (n = 749)	Train_set (n = 524)	Valid_set (n = 225)	Statistic	P
Survival months, M (Q_1, Q_3)	8.00 (3.00–18.00)	9.00 (3.00–19.00)	7.00 (2.00–16.00)	Z = 1.771	0.077
Age, n (%)				χ^2 = 4.450	0.217
<40	16 (2.14)	14 (2.67)	2 (0.89)		
40–60	169 (22.56)	123 (23.47)	46 (20.44)		
60–80	415 (55.41)	280 (53.44)	135 (60.00)		
>80	149 (19.89)	107 (20.42)	42 (18.67)		
Sex, n (%)				χ^2 = 0.022	0.882
Male	552 (73.7)	387 (73.85)	165 (73.33)		
Female	197 (26.3)	137 (26.15)	60 (26.67)		
T stage, n (%)				χ^2 = 0.592	0.898
1	214 (28.57)	152 (29.01)	62 (27.56)		
2	59 (7.88)	43 (8.21)	16 (7.11)		
3	257 (34.31)	179 (34.16)	78 (34.67)		
4	219 (29.24)	150 (28.63)	69 (30.67)		
N stage, n (%)				χ^2 = 3.476	0.324
0	249 (33.24)	179 (34.16)	70 (31.11)		
1	295 (39.39)	212 (40.46)	83 (36.89)		
2	96 (12.82)	62 (11.83)	34 (15.11)		
3	109 (14.55)	71 (13.55)	38 (16.89)		
Primary site, n (%)				χ^2 = 6.002	0.647
Body	69 (9.21)	46 (8.78)	23 (10.22)		
Cardia	321 (42.86)	221 (42.18)	100 (44.44)		
Fundus	31 (4.14)	25 (4.77)	6 (2.67)		
Gastric antrum	145 (19.36)	100 (19.08)	45 (20.00)		
Greater curvature	27 (3.6)	20 (3.82)	7 (3.11)		
Lesser curvature	51 (6.81)	34 (6.49)	17 (7.56)		
Overlapping lesion	46 (6.14)	35 (6.68)	11 (4.89)		
Pylorus	11 (1.47)	6 (1.15)	5 (2.22)		
Stomach	48 (6.41)	37 (7.06)	11 (4.89)		
Tumor size, n (%)				χ^2 = 2.719	0.437
<2	78 (10.41)	51 (9.73)	27 (12.00)		
2 to 5	374 (49.93)	271 (51.72)	103 (45.78)		
5 to 8	216 (28.84)	145 (27.67)	71 (31.56)		
>8	81 (10.81)	57 (10.88)	24 (10.67)		
Number of Reg LN, n (%)				χ^2 = 0.619	0.734
None	528 (70.49)	372 (70.99)	156 (69.33)		
1 to 3	27 (3.6)	20 (3.82)	7 (3.11)		
4 or more	194 (25.9)	132 (25.19)	62 (27.56)		

(Continued)

TABLE 4 Continued

Variable	Total (n = 749)	Train_set (n = 524)	Valid_set (n = 225)	Statistic	P
Chemotherapy, n (%)				χ^2 = 0.584	0.445
No/unknown	294 (39.25)	201 (38.36)	93 (41.33)		
Yes	455 (60.75)	323 (61.64)	132 (58.67)		
Radiation, n (%)				χ^2 = 0.496	0.481
None/unknown	575 (76.77)	406 (77.48)	169 (75.11)		
Beam radiation	174 (23.23)	118 (22.52)	56 (24.89)		
Cause, n (%)				χ^2 = 0.291	0.590
Alive or dead of other cause	141 (18.83)	96 (18.32)	45 (20.00)		
Dead (attributable to this cancer dx)	608 (81.17)	428 (81.68)	180 (80.00)		
Status, n (%)				χ^2 = 0.079	0.778
Alive	70 (9.35)	50 (9.54)	20 (8.89)		
Dead	679 (90.65)	474 (90.46)	205 (91.11)		
Survival months, M (Q_1, Q_3)	8.00 (3.00–18.00)	9.00 (3.00–19.00)	7.00 (2.00–16.00)	Z = 1.771	0.077

shown in the left half of Tables 5 and 6. Afterward, according to the outcome, statistically significant variables were included in the multivariate analyses.

The Cox regression results for OS are shown in Table 5. The detailed outcomes suggested that age, T stage, primary site, chemotherapy, radiation, and the number of Reg LNs were correlated with OS in MDGA patients. Multifactorial analysis for OS revealed that only T stage (2 and 3), primary site, chemotherapy, and number of Reg LNs were statistically significant independent risk factors for MDGA patients with established DM. Moreover, patients with higher T stages (T3 and 4) and without chemotherapy had significantly shorter survival and worse outcomes. Patients with superficial primary sites (gastric antrum and greater curvature) and a greater number of Reg LNs could have improved outcomes. More comprehensive OS analysis information, such as the analytical CIs and P-values for each variable, is collated and displayed in Table 5.

The outcome of the Cox regression analysis using CSS as the endpoint is presented in Table 6. The primary site, number of Reg LNs, age, T stage, chemotherapy, and radiotherapy variables were integrated into the multifactorial analysis. The analysis indicated that the number of Reg LNs, T stage, primary site, and chemotherapy were considered statistically significant independent risk factors for CSS. The CIs and the corresponding P-values are summarized in Table 6.

4.3 Nomogram

According to the outcomes obtained in the previous steps, this study developed a visual nomogram to predict the survival time of MDGA patients with established DM. The nomogram, derived from the prognostically relevant risk factors that have been identified, provides a score based on the patient's current condition. Physicians can assess a patient's probability of 1-, 3-, and 5-year OS/CSS based on this nomogram (Figure 5). According to the OS nomogram (Figure 5A), of the five independent risk factors, chemotherapy had the greatest impact on survival, followed by the primary tumor site, while T stage had the least impact. According to the CSS nomogram (Figure 5B), the presence or absence of chemotherapy was considered to be the most influential factor for survival, followed by the lymph node positivity rate.

A simple example of how to use a nomogram is given below. Suppose a 60-year-old patient with distant metastases from MDGA has received conventional chemotherapy but no radiotherapy. At the same time, the pathological findings suggest that the tumor originated in the greater curvature, the current T stage is 3, and the number of regional LNs reaches more than four. At this point, an approximate score can be calculated based on the nomogram (age, 17.5 points; T stage 3, 2 points; primary site, 17 points; number of regional LNs, 11 points; received chemotherapy, 0; not received radiotherapy, 18 points). This hypothetical patient would therefore have a total score of 65.5, and this score was plotted against a scale of total scores. By plotting vertically on a straight line of survival probability, one can derive the probability that the overall survival available for reference is approximately 0.78, 0.55, and 0.45 for 1, 3, and 5 years, respectively. Similarly, the corresponding CSS for this patient can be calculated using the same steps as above.

4.4 Evaluation and validation of the nomograms

The predictive results and clinical value of the nomograms were assessed and verified by the C-index, AUC, calibration curve, and

TABLE 5 Cox regression analysis of OS in the SEER cohort.

Variables	Beta	S.E	*Z*	*P*	HR (95% CI)	m_Beta	m_S.E	m_*Z*	a*P*	aHR (95% CI)
Age										
<40					Ref					Ref
40–60	0.19	0.29	0.66	0.506	1.21 (0.69–2.14)	0.13	0.29	0.45	0.655	1.14 (0.64–2.03)
60–80	0.36	0.28	1.27	0.204	1.43 (0.82–2.49)	0.18	0.29	0.63	0.526	1.20 (0.68–2.12)
>80	0.91	0.29	3.12	0.002	2.48 (1.40–4.38)	0.53	0.30	1.77	0.077	1.70 (0.94–3.07)
Sex										
Male					Ref					
Female	−0.00	0.09	−0.05	0.960	1.00 (0.84–1.18)					
T stage										
4					Ref					Ref
2	−0.20	0.15	−1.32	0.187	0.82 (0.61–1.10)	−0.33	0.16	−2.11	0.035	0.72 (0.53–0.98)
1	0.27	0.10	2.68	0.007	1.31 (1.08–1.60)	−0.16	0.11	−1.41	0.157	0.85 (0.69–1.06)
3	−0.19	0.10	−1.98	0.048	0.82 (0.68–0.99)	−0.29	0.10	−2.90	0.004	0.75 (0.61–0.91)
N stage										
3					Ref					
1	0.08	0.12	0.64	0.523	1.08 (0.85–1.37)					
0	0.21	0.12	1.70	0.089	1.23 (0.97–1.56)					
2	0.08	0.15	0.55	0.586	1.08 (0.81–1.45)					
Primary site										
Lesser curvature					Ref					Ref
Cardia	0.04	0.16	0.25	0.800	1.04 (0.76–1.42)	0.16	0.17	0.93	0.351	1.17 (0.84–1.62)
Fundus	0.65	0.24	2.78	0.005	1.92 (1.21–3.05)	0.50	0.24	2.10	0.036	1.65 (1.03–2.63)
Stomach	−0.14	0.22	−0.65	0.516	0.87 (0.56–1.33)	−0.44	0.22	−2.00	0.046	0.64 (0.41–0.99)
Gastric antrum	−0.00	0.17	−0.00	0.998	1.00 (0.71–1.40)	−0.04	0.17	−0.25	0.805	0.96 (0.68–1.35)
Overlapping lesion	0.06	0.21	0.26	0.797	1.06 (0.69–1.61)	−0.09	0.22	−0.40	0.691	0.92 (0.60–1.41)
Body	0.07	0.19	0.37	0.710	1.08 (0.73–1.57)	0.10	0.20	0.49	0.626	1.10 (0.75–1.63)
Greater curvature	−0.04	0.25	−0.15	0.879	0.96 (0.59–1.58)	−0.46	0.26	−1.78	0.075	0.63 (0.38–1.05)
Pylorus	−0.37	0.38	−0.97	0.334	0.69 (0.33–1.46)	−0.02	0.39	−0.05	0.959	0.98 (0.46–2.10)
Tumor size										
<2					Ref					
>8	0.02	0.17	0.10	0.923	1.02 (0.73–1.41)					
2 to 5	0.20	0.13	1.51	0.131	1.22 (0.94–1.57)					
5 to 8	0.11	0.14	0.80	0.426	1.12 (0.85–1.46)					
Number of Reg LN										
None					Ref					Ref
1 to 3	−0.24	0.21	−1.13	0.257	0.79 (0.52–1.19)	−0.60	0.22	−2.74	0.006	0.55 (0.35–0.84)
4 or more	−0.59	0.09	−6.49	<0.001	0.55 (0.46–0.66)	−0.78	0.10	−7.48	<0.001	0.46 (0.37–0.56)

(Continued)

TABLE 5 Continued

Variables	Beta	S.E	*Z*	*P*	HR (95% CI)	m_Beta	m_S.E	m_*Z*	a*P*	aHR (95% CI)
Chemotherapy										
No/unknown					Ref					Ref
Yes	−1.05	0.08	−13.00	<0.001	0.35 (0.30–0.41)	−1.20	0.09	−12.76	<0.001	0.30 (0.25–0.36)
Radiation										
None/unknown					Ref					Ref
Beam radiation	−0.23	0.09	−2.53	0.011	0.79 (0.66–0.95)	−0.17	0.10	−1.82	0.068	0.84 (0.70–1.01)

DCA. In the training set, the AUC values for predicting 1-, 3-, and 5-year OS were 0.797, 0.807, and 0.737, respectively (Figure 6A), while in the validation set, they were 0.757, 0.737, and 0.718, respectively (Figure 6B). The C-index of the nomograms was 0.726 (95% CI, 0.703–0.748) in the training set for OS and 0.703 (95% CI, 0.663–0.744) in the validation set. The fit of the 1-, 3-, and 5-year calibration curves for predicting OS was satisfactory (Figures 6C–H). In the calibration curves of the nomograms in the training and validation cohorts, the red curve fit line matches the gray diagonal line (representing the predicted probability of the ideal state) to a high degree, suggesting that the predicted probability of survival via the nomograms remains generally consistent with the observed probability of survival, with no excessive overestimation or underestimation of risk. The DCA curve presented graphically in Figures 6I–N suggests that this nomogram for OS has excellent net clinical efficacy.

Similarly, the results for evaluating the CSS nomograms show a positive applicability. The C-index was 0.727 (95% CI, 0.703–0.751) for the training set and 0.705 (95% CI, 0.663–0.748) for the validation set. In addition, the AUCs of the nomograms were 0.747, 0.737, and 0.699 for 1-, 3-, and 5-year CSS, respectively, in the training cohort (Figure 7A), and in the validation cohort, the

TABLE 6 Cox regression analysis of CSS according to the SEER data.

Variables	Beta	S.E	*Z*	*P*	HR (95% CI)	m_Beta	m_S.E	m_*Z*	a*P*	aHR (95% CI)
Age										
<40					Ref					Ref
40–60	0.10	0.29	0.36	0.721	1.11 (0.63–1.96)	0.04	0.30	0.15	0.884	1.04 (0.58–1.86)
60–80	0.22	0.28	0.79	0.432	1.25 (0.72–2.18)	0.06	0.29	0.22	0.828	1.06 (0.60–1.88)
>80	0.75	0.29	2.55	0.011	2.11 (1.19–3.74)	0.39	0.30	1.28	0.201	1.48 (0.81–2.68)
Sex										
Male					Ref					
Female	0.00	0.09	0.04	0.965	1.00 (0.84–1.20)					
T stage										
4					Ref					Ref
2	−0.22	0.16	−1.37	0.172	0.80 (0.58–1.10)	−0.38	0.17	−2.28	0.023	0.68 (0.49–0.95)
1	0.29	0.11	2.70	0.007	1.33 (1.08–1.65)	−0.17	0.12	−1.44	0.150	0.84 (0.67–1.06)
3	−0.14	0.10	−1.39	0.166	0.87 (0.71–1.06)	−0.25	0.11	−2.29	0.022	0.78 (0.63–0.96)
N stage										
3					Ref					
1	0.08	0.13	0.65	0.519	1.09 (0.85–1.39)					
0	0.20	0.13	1.56	0.119	1.22 (0.95–1.58)					
2	0.13	0.15	0.83	0.407	1.14 (0.84–1.54)					
Primary site										
Lesser curvature					Ref					Ref

(Continued)

TABLE 6 Continued

Variables	Beta	S.E	*Z*	*P*	HR (95% CI)	m_Beta	m_S.E	m_*Z*	a*P*	aHR (95% CI)
Primary site										
Cardia	0.02	0.16	0.11	0.916	1.02 (0.74–1.40)	0.11	0.17	0.65	0.516	1.12 (0.80–1.57)
Fundus	0.61	0.25	2.49	0.013	1.84 (1.14–2.99)	0.45	0.25	1.79	0.073	1.56 (0.96–2.55)
Stomach	−0.18	0.23	−0.79	0.428	0.83 (0.53–1.31)	−0.47	0.23	−2.01	0.045	0.63 (0.40–0.99)
Gastric antrum	−0.09	0.18	−0.50	0.616	0.91 (0.64–1.30)	−0.12	0.18	−0.67	0.500	0.88 (0.62–1.27)
Overlapping lesion	0.07	0.22	0.33	0.744	1.07 (0.70–1.66)	−0.08	0.22	−0.36	0.717	0.92 (0.59–1.43)
Body	0.02	0.20	0.09	0.932	1.02 (0.68–1.52)	0.02	0.21	0.12	0.908	1.02 (0.68–1.54)
Greater curvature	−0.38	0.29	−1.31	0.192	0.68 (0.38–1.21)	−0.81	0.30	−2.74	0.006	0.44 (0.25–0.79)
Pylorus	−0.30	0.39	−0.77	0.439	0.74 (0.35–1.58)	0.07	0.39	0.18	0.854	1.07 (0.50–2.32)
Tumor size										
<2					Ref					
>8	0.06	0.18	0.35	0.725	1.07 (0.75–1.51)					
2 to 5	0.26	0.14	1.86	0.062	1.30 (0.99–1.70)					
5 to 8	0.12	0.15	0.81	0.419	1.13 (0.84–1.51)					
Number of Reg LN										
None					Ref					Ref
1 to 3	−0.23	0.22	−1.07	0.287	0.79 (0.52–1.22)	−0.61	0.23	−2.65	0.008	0.54 (0.35–0.85)
4 or more	−0.68	0.10	−6.83	<0.001	0.51 (0.42–0.62)	−0.88	0.11	−7.82	<0.001	0.42 (0.33–0.52)
Chemotherapy										
No/unknown					Ref					Ref
Yes	−1.04	0.09	−12.28	<0.001	0.35 (0.30–0.42)	−1.24	0.10	−12.46	<0.001	0.29 (0.24–0.35)
Radiation										
None/unknown					Ref					Ref
Beam radiation	−0.22	0.10	−2.33	0.020	0.80 (0.66–0.97)	−0.16	0.10	−1.61	0.108	0.85 (0.70–1.04)

AUCs were 0.661, 0.713, and 0.892, respectively (Figure 7B). Moreover, both the calibration curves and DCA curves used for the 1-, 3-, and 5-year CSS forecasts also exhibited satisfactory fits and net gains (Figures 7C–N). In summary, the nomograms produced to predict the prognosis of MDGA patients with DM had considerable discriminatory and calibrating power.

5 Analysis of the impact of more detailed LN indicators on the occurrence of DM and prognosis of MDGA

The above studies have suggested a strong association between multiple lymph node indices and DM and the prognosis of MDGA. Although good predictive efficacy can be achieved by categorizing the number of positive LNs (0, 1 to 3, 3+), 70% of patients in the database had a positive lymph node clearance of 0. This suggests that the existing lymph node indices may not describe a patient's prognosis specifically; thus, more diversified ways of evaluating the metastasis and immune mechanisms of patients are needed. Lymph node positivity, the specific number of negative/positive lymph nodes, and visualization of LN metastasis may be better indicators of DM risk and survival; therefore, we collected more detailed data from our institution and performed a logistic analysis to identify risk factors associated with DM.

We collected data from 109 patients with moderately differentiated gastric adenocarcinoma in our hospital. Data such as LODDS and the number of Neg LNs were analyzed and calculated, followed by logistic regression to explore the risk factors for distant metastasis in patients with MDGA and Cox regression to analyze the risk factors affecting the prognosis of patients with MDGA. Our univariate logistic regression results showed that the number of negative LNs and the LODDS were considered to be influential factors for the occurrence of DM in MDGA (Table 7). However, it is noteworthy that in regard to our multifactor logistic regression analysis of the variables of interest,

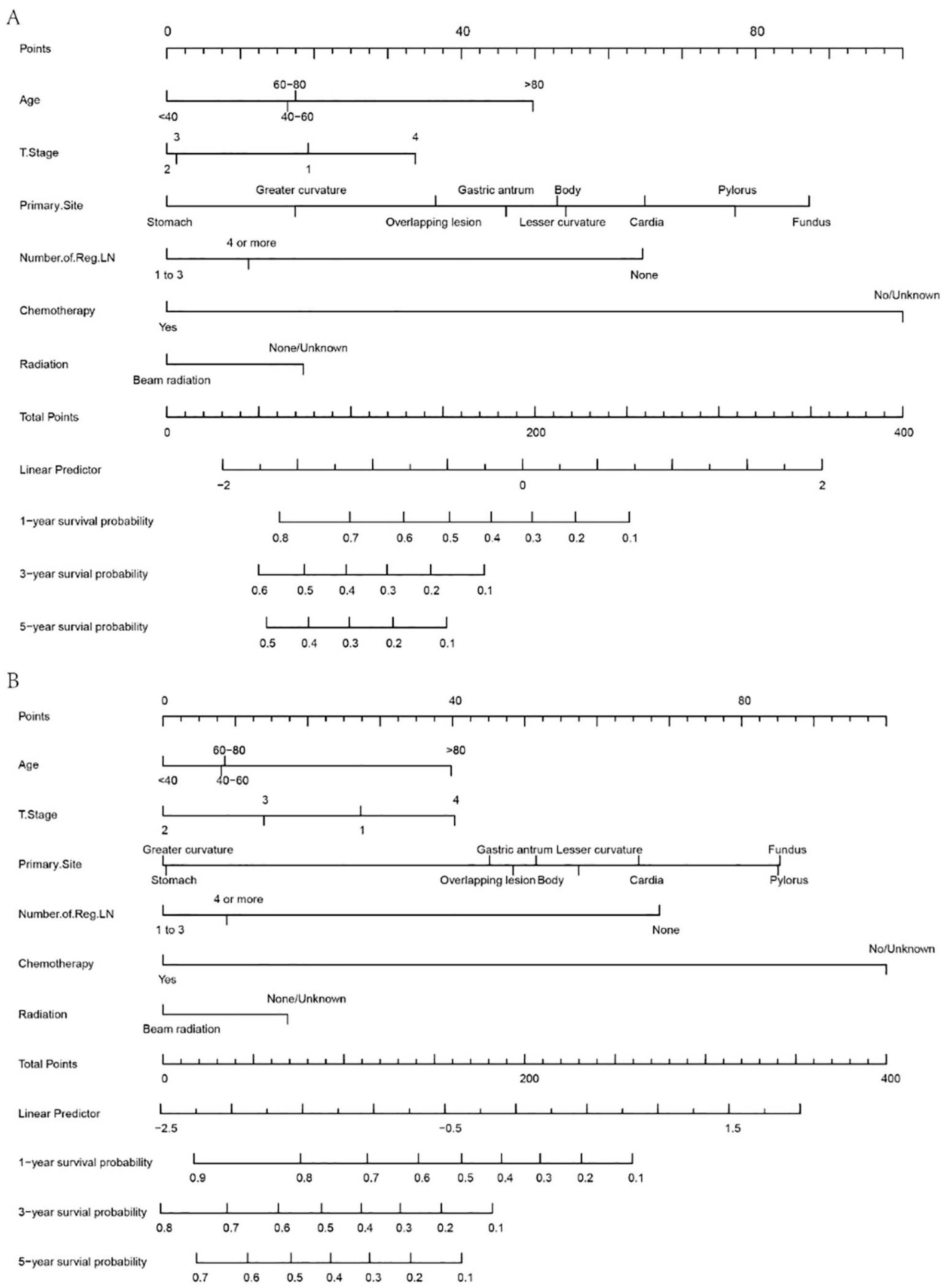


FIGURE 5
Nomograms for 1-, 3-, and 5-year OS **(A)** and CSS **(B)** in MDCA patients with DM.

our results lacked statistical significance. We conducted single- and multivariate Cox analyses of our patient data. As shown in Table 8, 15 variables were included. The results of the univariate analysis revealed that nine variables, including the number of Reg LNs, LNR, age >80 years, TNM stage, tumor size, gross LN metastasis, and number of Reg LNs, had an impact on the prognosis of MDGA patients ($P < 0.05$). These findings were subsequently incorporated into a multifactorial analysis, which indicated that the LNR, T stage (1 and 2), and gross LN metastasis 3 cm away from the tumor were independent risk factors, whereas the number of Reg LNs and the number of Reg LNs in groups 1–3 were considered protective factors. More specific data are shown in Tables 7 and 8.

6 Discussion

Moderately differentiated gastric adenocarcinoma is common in clinical practice and has a high risk of metastasis and individual variability (34). Once a patient develops DM, the prognosis becomes extremely poor (35, 36). The OS of MDGA patients

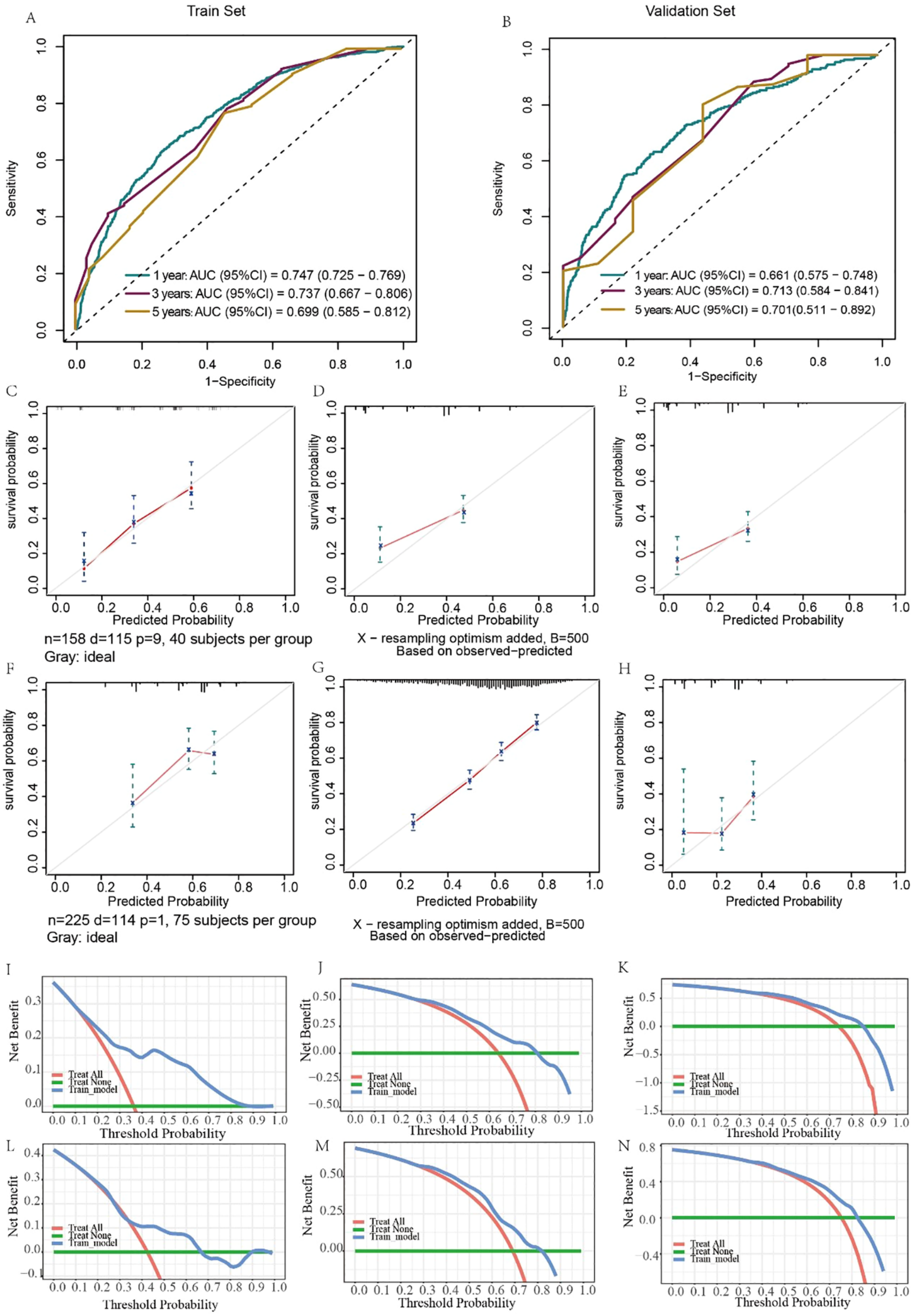


FIGURE 6
Evaluation of the ability of the nomogram to predict OS. ROC curves validating the OS prediction nomogram for 1-, 3-, and 5-year OS in the training set **(A)** and validation set **(B)**. Calibration curves validating the OS prediction nomograms for 1-, 3-, and 5-year OS in the training set **(C–E)** and validation set **(F–H)**. Decision curve analysis validating the OS prediction nomogram for 1-, 3-, and 5-year OS in the training set **(I–K)** and validation set **(L–N)**.

without DM is generally considered to be more than 22.3 months after surgical treatment (37). However, after the onset of DM, survival decreases in patients receiving conventional chemotherapy, with a median overall survival of just under 12 months (22, 38). Determining whether a patient has distant metastases is therefore particularly important and is vital for providing individualized prevention and treatment strategies in the clinic. In addition, the current prognostic method for patients with DM is relatively limited, and some DM-related indices, especially lymph node indices such as the LNR and LODDS, are

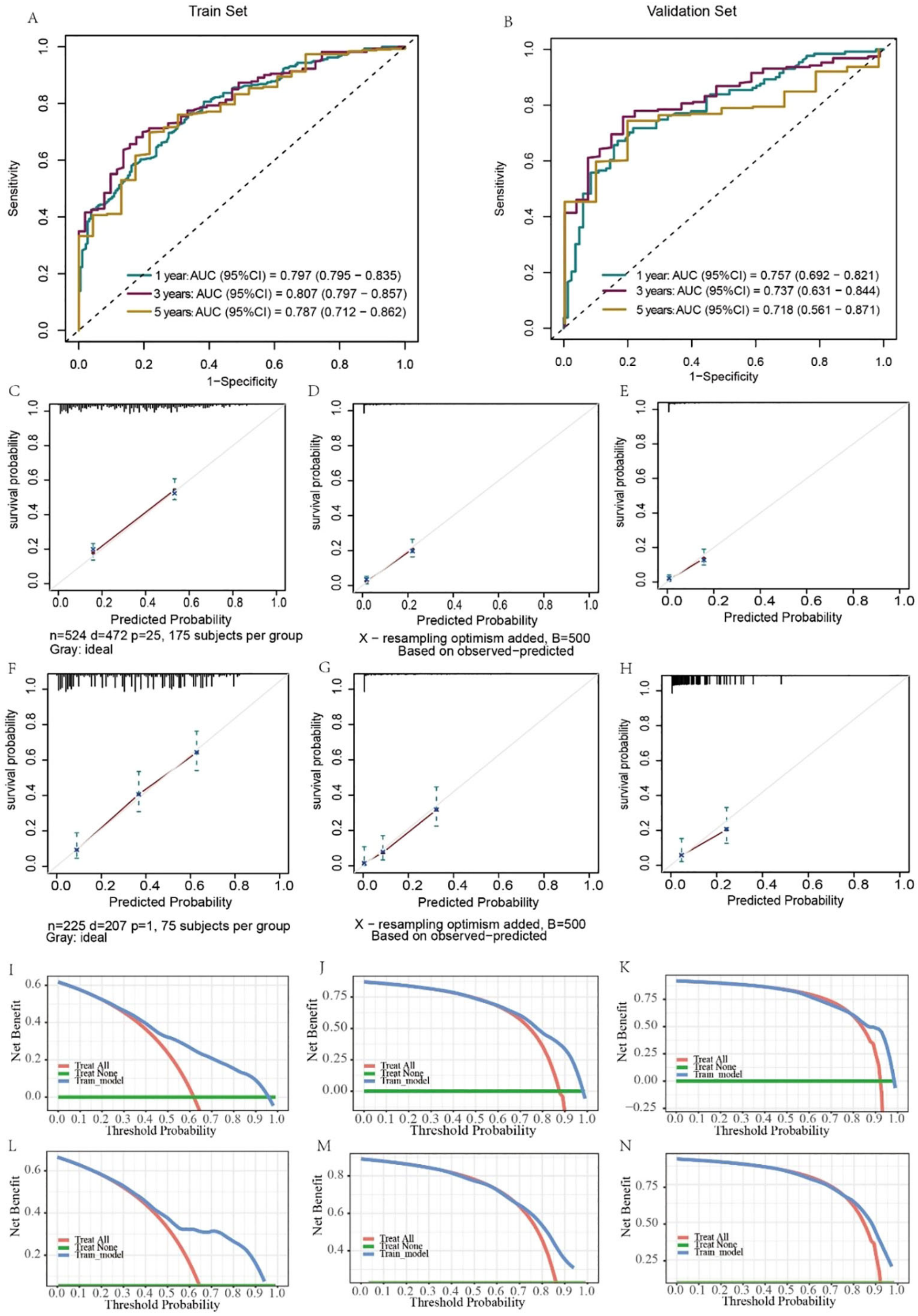


FIGURE 7
Evaluation of the ability of the nomogram to predict CSS. ROC curves validating the CSS prediction nomogram for 1-, 3-, and 5-year RFS in the training set **(A)** and validation set **(B)**. Calibration curves validating the CSS prediction nomograms for 1-, 3-, and 5-year survival in the training cohort **(C–E)** and validation cohort **(F–H)**. Decision curve analysis validating the CSS prediction nomogram for 1-, 3-, and 5-year RFS in the training set **(I–K)** and validation set **(L–N)**.

considered to be important indicators of prognosis (39, 40). However, its specific clinical effects have still not been extensively and comprehensively tested.

Our major objectives for the investigation were to develop a forecasting system to predict the development of DM in persons with MDGA and to analyze the risk factors influencing the prognosis of persons with DM. In addition, this study analyzed the specific ability of six lymph node indicators in our patients to predict DM and prognosis using logistic and Cox regression. Nine machine learning samples were utilized for predicting distant metastases, with the RF model considered the most effective. Multivariate Cox regression analysis for MDGA patients who

TABLE 7 The risk factors for developing DM in MDGA patients were analyzed by logistic regression based on our data.

Variables	Beta	S.E	*Z*	OR (95% CI)	*P*	aBeta	aS.E	a*Z*	aOR (95% CI)	a*P*
Number of all LNs	−0.03	0.02	−1.27	0.97 (0.93–1.02)	0.203					
Number of Reg LN	0.05	0.04	1.50	1.06 (0.98–1.13)	0.135					
Number of Neg LN	−0.06	0.03	−2.00	0.94 (0.89–0.99)	0.046	−0.04	0.03	−1.48	0.96 (0.90–1.01)	0.138
LODDS	−3.08	1.48	−2.09	0.05 (0.00–0.83)	0.037	−1.68	1.59	−1.05	0.19 (0.01–4.24)	0.292
LNR	0.94	1.18	0.80	2.57 (0.25–25.89)	0.423					
Number of Reg LN_group										
None				1.00 (reference)						
1 to 3	−0.10	0.74	−0.14	0.90 (0.21–3.88)	0.890					
4 or more	−0.20	0.63	−0.32	0.82 (0.24–2.80)	0.747					
Gross LN metastasis										
None				1.00 (reference)						
3 cm away from the tumor	0.69	0.68	1.02	2.00 (0.52–7.62)	0.310					
Within 3 cm of the tumor	−1.11	0.81	−1.38	0.33 (0.07–1.60)	0.169					
Age										
<40				1.00 (reference)						
40–60	−0.95	1.27	−0.75	0.39 (0.03–4.67)	0.455					
60–80	−0.76	1.21	−0.63	0.47 (0.04–4.98)	0.527					
>80	0.41	1.68	0.24	1.50 (0.06–40.63)	0.810					
Sex										
Male				1.00 (Reference)						
Female	−0.26	0.69	−0.38	0.77 (0.20–2.98)	0.707					
T stage										
1				1.00 (reference)					1.00 (reference)	
2	0.27	1.45	0.19	1.31 (0.08–22.62)	0.851	0.24	1.48	0.16	1.27 (0.07–23.15)	0.872
3	1.25	1.09	1.15	3.50 (0.41–29.78)	0.251	1.27	1.12	1.13	3.56 (0.39–32.31)	0.259
4	2.46	1.17	2.11	11.67 (1.19–114.57)	0.035	2.34	1.21	1.93	10.40 (0.97–111.50)	0.053
N stage										
0				1.00 (reference)						
1	1.06	0.73	1.45	2.88 (0.69–12.00)	0.146					
2	1.54	0.90	1.72	4.67 (0.81–26.98)	0.085					
3	1.13	0.98	1.15	3.11 (0.45–21.40)	0.249					
Primary site										
Body				1.00 (reference)						
Cardia	16.83	2,465.33	0.01	20,408,610.53 (0.00–Inf)	0.995					
Gastric antrum	15.93	2,465.33	0.01	8,260,628.07 (0.00–Inf)	0.995					
Lesser curvature	18.06	2,465.33	0.01	69,389,275.80 (0.00–Inf)	0.994					

(Continued)

TABLE 7 Continued

Variables	Beta	S.E	*Z*	OR (95% CI)	*P*	aBeta	aS.E	a*Z*	aOR (95% CI)	a*P*
Primary site										
Greater curvature	18.57	2,465.33	0.01	115,648,792.99 (0.00–Inf)	0.994					
Overlapping lesion	16.17	2,465.33	0.01	10,513,526.64 (0.00–Inf)	0.995					
Stomach	17.47	2,465.33	0.01	38,549,597.66 (0.00–Inf)	0.994					
Pylorus	0.00	4,088.28	0.00	1.00 (0.00–Inf)	1.000					
Fundus	18.57	2,465.33	0.01	115,648,792.99 (0.00–Inf)	0.994					
Tumor size										
<2				1.00 (reference)						
2 to 5	−0.04	0.90	−0.05	0.96 (0.16–5.63)	0.961					
5 to 8	1.06	0.89	1.19	2.89 (0.50–16.67)	0.234					
>8	1.84	0.97	1.90	6.29 (0.94–41.96)	0.058					
Chemotherapy										
No/unknown				1.00 (reference)						
Yes	0.39	0.56	0.69	1.47 (0.49–4.42)	0.488					

LODDS, log odds of metastatic lymph nodes; LNR, lymph node ratio.

TABLE 8 Cox regression analysis of risk factors affecting patient OS based on our data.

Variables	Beta	S.E	*Z*	*P*	HR (95% CI)	m_Beta	m_S.E	m_*Z*	a*P*	aHR (95% CI)
Number of all LNs	−0.02	0.01	−1.35	0.179	0.98 (0.96–1.01)					
Number of Reg LN	0.05	0.02	2.95	0.003	1.05 (1.02–1.08)	−0.12	0.05	−2.68	0.007	0.89 (0.81–0.97)
Number of Neg LN	−0.06	1.02	−0.06	0.950	0.94 (0.13–6.91)					
LODDS	0.25	0.33	0.78	0.436	1.29 (0.68–2.44)					
LNR	1.94	0.52	3.76	<0.001	6.93 (2.53–19.01)	5.09	1.67	3.04	0.002	162.60 (6.12–4,318.06)
Age										
<40					Ref					Ref
40–60	0.24	1.04	0.23	0.816	1.28 (0.16–9.88)	−1.36	1.27	−1.07	0.285	0.26 (0.02–3.10)
60–80	0.64	1.02	0.63	0.528	1.90 (0.26–13.92)	−0.48	1.23	−0.39	0.698	0.62 (0.06–6.93)
>80	2.46	1.16	2.12	0.034	11.71 (1.20–114.06)	1.59	1.47	1.08	0.279	4.92 (0.28–88.06)
Sex										
Male					Ref					
Female	−0.04	0.35	−0.11	0.910	0.96 (0.49–1.89)					
T stage										
1					Ref					Ref
2	−1.49	1.10	−1.36	0.173	0.22 (0.03–1.92)	−2.21	1.17	−1.90	0.058	0.11 (0.01–1.07)
3	1.15	0.48	2.38	0.017	3.15 (1.22–8.13)	−0.08	0.60	−0.14	0.890	0.92 (0.28–3.00)
4	1.36	0.56	2.43	0.015	3.89 (1.30–11.66)	−0.41	0.66	−0.63	0.532	0.66 (0.18–2.42)

(Continued)

TABLE 8 Continued

Variables	Beta	S.E	*Z*	*P*	HR (95% CI)	m_Beta	m_S.E	m_*Z*	a*P*	aHR (95% CI)
N stage										
0					Ref					Ref
1	1.23	0.38	3.20	0.001	3.42 (1.61–7.26)	1.72	0.74	2.31	0.021	5.56 (1.30–23.90)
2	2.11	0.48	4.41	<0.001	8.25 (3.23–21.08)	2.13	0.88	2.42	0.016	8.42 (1.50–47.32)
3	1.00	0.55	1.82	0.069	2.72 (0.93–7.99)	1.37	1.03	1.33	0.183	3.94 (0.52–29.61)
M stage										
0					Ref					Ref
1	1.04	0.35	2.99	0.003	2.82 (1.43–5.58)	0.79	0.50	1.56	0.120	2.19 (0.82–5.90)
Primary site										
Body					Ref					
Cardia	0.16	0.62	0.27	0.791	1.18 (0.35–3.97)					
Gastric antrum	−0.16	0.65	−0.24	0.809	0.86 (0.24–3.03)					
Lesser curvature	0.12	0.76	0.16	0.875	1.13 (0.25–5.05)					
Greater curvature	1.33	0.92	1.45	0.148	3.78 (0.62–22.84)					
Overlapping lesion	−1.02	0.91	−1.11	0.266	0.36 (0.06–2.17)					
Stomach	0.11	0.91	0.12	0.901	1.12 (0.19–6.72)					
Pylorus	−0.64	1.16	−0.55	0.582	0.53 (0.05–5.10)					
Fundus	−16.30	3,293.13	−0.00	0.996	0.00 (0.00–Inf)					
Tumor size										
<2					Ref					Ref
2 to 5	1.83	0.74	2.48	0.013	6.25 (1.47–26.61)	1.60	0.92	1.75	0.081	4.94 (0.82–29.73)
5 to 8	2.33	0.76	3.08	0.002	10.30 (2.34–45.40)	1.43	0.91	1.57	0.116	4.19 (0.70–24.97)
>8	2.38	0.79	3.01	0.003	10.78 (2.29–50.84)	1.62	0.92	1.77	0.077	5.08 (0.84–30.76)
Chemotherapy										
No/unknown					Ref					
Yes	−0.10	0.31	−0.32	0.747	0.91 (0.50–1.65)					
Gross LN metastasis										
None					Ref					Ref
3 cm away from the tumor	1.40	0.47	2.99	0.003	4.04 (1.62–10.09)	1.96	0.66	2.97	0.003	7.11 (1.95–25.99)
Within 3 cm of the tumor	0.72	0.47	1.54	0.124	2.05 (0.82–5.13)	1.21	0.64	1.88	0.060	3.36 (0.95–11.90)
Number of Reg LN group										
None					Ref					Ref
1 to 3	0.98	0.40	2.43	0.015	2.66 (1.21–5.87)	−1.67	0.80	−2.08	0.038	0.19 (0.04–0.91)
4 or more	1.06	0.36	2.99	0.003	2.90 (1.44–5.82)	−1.45	0.88	−1.65	0.098	0.23 (0.04–1.31)

already had DM indicated that higher T stage (2 and 3), primary site, chemotherapy, and number of Reg LNs were independent risk factors for prognosis. Moreover, specialized nomograms created from our analysis results were evaluated and tested to show convincing prognostic discrimination and calibration capabilities.

Notably, the categorization of the number of positive LNs (0, 1 to 3, 4 or more) in the SSER database may achieve good predictive efficacy. However, nearly 70% of the patients in the database had a positive lymph node clearance of 0. This suggests that our single reliance on lymph node clearance results may not be effective in characterizing the prognosis of patients, and more diverse classifications and metrics are needed to evaluate a patient's metastatic and prognostic condition. In this study, based on our patient data, we revealed that the LNR, gross LN metastasis, and the number of Reg LNs were found to be independent factors influencing the prognosis of MDGA patients with DM.

It is worth noting that the AUC of the validation set in this paper is generally slightly lower than that of the training set, which is a relatively common phenomenon, and the possible reasons are that the training set adopts the U.S. population samples from the SEER database, and extrapolation is not strong enough in the Chinese population, or the sample size of the validation set is not sufficient. In future research, we will further consider the extrapolation of the population and the adequacy of the samples to deepen and improve the prediction ability of the validation set.

Nevertheless, it must be noted that the research has been limited by its retrospective nature. Although the SEER database is very detailed and reliable, there are some more exhaustive data that it is unable to provide (41). For example, data on some noteworthy laboratory tests were not included, and some of the more nuanced pictures of the lymph nodes, as previously mentioned, were lacking. Furthermore, for the practical application of the nomogram, additional clinical information must be considered, including the ethnicity of the patient, their geographical location, and other pertinent factors. These data, which are absent from the database and not included in the study, have an impact on the results, and more information is required to enhance the nomogram. For our data, because of the sample size and other reasons, it is not as effective as it should be in carrying out some statistical studies, and in the future, it is necessary to collect more case and patient information for more in-depth analysis and studies.

7 Conclusion

In conclusion, this research investigated the variables linked to the development of DM in MDGA, including T stage, N stage, primary site, tumor size, number of positive LNs, and chemotherapy. Then, we investigated the prognostic factors, including T stage, primary location, chemotherapy, and number of Reg LNs, in MDGA patients with DM. Additionally, based on the prognostic analysis, separate nomograms of OS and CSS were produced for relevant influencing factors. Finally, the effect of multiple lymph node indicators on the metastasis and prognosis of MDGA patients was investigated. This study provides a reference for subsequent clinical studies and further suggests the importance of lymph node indicators.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by The Ethics Committee of the Second Affiliated Hospital of Nanchang University (examination and approval no. review (2019) no. (105)). The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from primarily isolated samples as part of your previous study for which ethical approval was obtained. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

KY: Writing – original draft. JW: Writing – original draft. TX: Writing – original draft. YZ: Writing – original draft. WL: Writing – original draft. LY: Writing – review & editing.

Funding

The present research is funded by grants from the National Natural Science Foundation of China (81960103 to LY).

Acknowledgments

The graphical abstracts were created with BioRender software (BioRender.com).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1398685/full#supplementary-material

References

1. Lin Y, Zheng Y, Wang HL, Wu J. Global patterns and trends in gastric cancer incidence rates (1988-2012) and predictions to 2030. *Gastroenterology*. (2021) 161:116–27.e8. doi: 10.1053/j.gastro.2021.03.023

2. 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: Cancer J Clin*. (2021) 71:209–49. doi: 10.3322/caac.21660

3. López Sala P, Leturia Etxeberria M, Inchausti Iguíñiz E, Astiazaran Rodríguez A, Aguirre Oteiza MI, Zubizarreta Etxaniz M. Gastric adenocarcinoma: A review of the TNM classification system and ways of spreading. *Radiologia*. (2023) 65:66–80. doi: 10.1016/j.rxeng.2022.10.011

4. Nagtegaal ID, Odze RD, Klimstra D, Paradis V, Rugge M, Schirmacher P, et al. The 2019 WHO classification of tumours of the digestive system. *Histopathology*. (2020) 76:182–8. doi: 10.1111/his.13975

5. Xiang Y, Yao L. Retrospective analysis of diagnosis and treatment of gastric cancer at Huzhou central hospital. *Altern Ther Health Med*. (2023) 29:302–9.

6. Liao JB, Lee HP, Fu HT, Lee HS. Assessment of EGFR and ERBB2 (HER2) in gastric and gastroesophageal carcinomas: EGFR amplification is associated with a worse prognosis in early stage and well to moderately differentiated carcinoma. *Appl Immunohistochem Mol Morphol*. (2018) 26:374–82. doi: 10.1097/PAI.0000000000000437

7. Li H, Zhang YC, Tsuchihashi Y. Invasion and metastasis of SY86B human gastric carcinoma cells in nude mice. *Jpn J Cancer Res*. (1988) 79:750–6. doi: 10.1111/j.1349-7006.1988.tb02232.x

8. Lord AC, D'Souza N, Shaw A, Rokan Z, Moran B, Abulafi M, et al. MRI-diagnosed tumor deposits and EMVI status have superior prognostic accuracy to current clinical TNM staging in rectal cancer. *Ann Surg*. (2022) 276:334–44. doi: 10.1097/SLA.0000000000004499

9. Guan WL, He Y, Xu RH. Gastric cancer treatment: recent progress and future perspectives. *J Hematol Oncol*. (2023) 16:57. doi: 10.1186/s13045-023-01451-3

10. Lin Z, Wang R, Zhou Y, Wang Q, Yang CY, Hao BC, et al. Prediction of distant metastasis and survival prediction of gastric cancer patients with metastasis to the liver, lung, bone, and brain: research based on the SEER database. *Ann Trans Med*. (2022) 10:16. doi: 10.21037/atm-21-6295

11. Young JJ, Pahwa A, Patel M, Jude CM, Nguyen M, Deshmukh M, et al. Ligaments and lymphatic pathways in gastric adenocarcinoma. *Radiographics: Rev Publ Radiological Soc North America Inc*. (2019) 39:668–89. doi: 10.1148/rg.2019180113

12. Lim JS, Yun MJ, Kim MJ, Hyung WJ, Park MS, Choi JY, et al. CT and PET in stomach cancer: preoperative staging and monitoring of response to therapy. *Radiographics: Rev Publ Radiological Soc North America Inc*. (2006) 26:143–56. doi: 10.1148/rg.261055078

13. Kwee RM, Kwee TC. Modern imaging techniques for preoperative detection of distant metastases in gastric cancer. *World J Gastroenterol*. (2015) 21:10502–9. doi: 10.3748/wjg.v21.i37.10502

14. Lordick F, Carneiro F, Cascinu S, Fleitas T, Haustermans K, Piessen G, et al. Gastric cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann oncology: Off J Eur Soc Med Oncol*. (2022) 33:1005–20. doi: 10.1016/j.annonc.2022.07.004

15. Xiang R, Song W, Ren J, Wu J, Fu J, Fu T. Identification of stem cell-related subtypes and risk scoring for gastric cancer based on stem genomic profiling. *Stem Cell Res Ther*. (2021) 12:563. doi: 10.1186/s13287-021-02633-x

16. Pai RK, Hartman D, Schaeffer DF, Rosty C, Shivji S, Kirsch R, et al. Development and initial validation of a deep learning algorithm to quantify histological features in colorectal carcinoma including tumour budding/poorly differentiated clusters. *Histopathology*. (2021) 79:391–405. doi: 10.1111/his.14353

17. Liu X, Zhang D, Liu Z, Li Z, Xie P, Sun K, et al. Deep learning radiomics-based prediction of distant metastasis in patients with locally advanced rectal cancer after neoadjuvant chemoradiotherapy: A multicentre study. *EBioMedicine*. (2021) 69:103442. doi: 10.1016/j.ebiom.2021.103442

18. Tian H, Liu Z, Liu J, Zong Z, Chen Y, Zhang Z, et al. Application of machine learning algorithm in predicting distant metastasis of T1 gastric cancer. *Sci Rep*. (2023) 13:5741. doi: 10.1038/s41598-023-31880-6

19. Allesøe RL, Thompson WK, Bybjerg-Grauholm J, Hougaard DM, Nordentoft M, Werge T, et al. Deep learning for cross-diagnostic prediction of mental disorder diagnosis and prognosis using danish nationwide register and genetic data. *JAMA Psychiatry*. (2023) 80:146–55. doi: 10.1001/jamapsychiatry.2022.4076

20. van der Vliet R, Selles RW, Andrinopoulou ER, Nijland R, Ribbers GM, Frens MA, et al. Predicting upper limb motor impairment recovery after stroke: A mixture model. *Ann Neurol*. (2020) 87:383–93. doi: 10.1002/ana.25679

21. Poldrack RA, Huckins G, Varoquaux G. Establishment of best practices for evidence for prediction: A review. *JAMA Psychiatry*. (2020) 77:534–40. doi: 10.1001/jamapsychiatry.2019.3671

22. Shitara K. Chemotherapy for advanced gastric cancer: future perspective in Japan. *Gastric cancer: Off J Int Gastric Cancer Assoc Japanese Gastric Cancer Assoc*. (2017) 20:102–10. doi: 10.1007/s10120-016-0648-7

23. Li X, Wang W, Ruan C, Wang Y, Wang H, Liang X, et al. Age-specific impact on the survival of gastric cancer patients with distant metastasis: an analysis of SEER database. *Oncotarget*. (2017) 8:97090–100. doi: 10.18632/oncotarget.21350

24. Yeoh KG, Tan P. Mapping the genomic diaspora of gastric cancer. *Nat Rev Cancer*. (2022) 22:71–84. doi: 10.1038/s41568-021-00412-7

25. Li H, Wang C, Lan L, Behrens A, Tomaschko M, Ruiz J, et al. High expression of vinculin predicts poor prognosis and distant metastasis and associates with influencing tumor-associated NK cell infiltration and epithelial-mesenchymal transition in gastric cancer. *Aging (Albany NY)*. (2021) 13:5197–225. doi: 10.18632/aging.202440

26. Guo S, Shang MY, Dong Z, Zhang J, Wang Y, Zheng ZC, et al. Clinicopathological features and prognostic analysis of signet ring cell gastric carcinoma: a population-based study. *Transl Cancer Res*. (2019) 8:1918–30. doi: 10.21037/tcr.2019.09.06

27. Zhu YF, Liu K, Zhang WH, Song XH, Peng BQ, Liao XL, et al. Is no. 12a lymph node dissection compliance necessary in patients who undergo D2 gastrectomy for gastric adenocarcinomas? A population-based retrospective propensity score matching study. *Cancers (Basel)*. (2023) 15(3):749. doi: 10.3390/cancers15030749

28. Matsushima J, Sato T, Yoshimura Y, Mizutani H, Koto S, Matsusaka K, et al. Clinical utility of artificial intelligence assistance in histopathologic review of lymph node metastasis for gastric adenocarcinoma. *Int J Clin Oncol*. (2023) 28:1033–42. doi: 10.1007/s10147-023-02356-4

29. Zhou R, Zhang J, Sun H, Liao Y, Liao W. Comparison of three lymph node classifications for survival prediction in distant metastatic gastric cancer. *Int J Surg*. (2016) 35:165–71. doi: 10.1016/j.ijsu.2016.09.096

30. Feng X, Hong T, Liu W, Xu C, Li W, Yang B, et al. Development and validation of a machine learning model to predict the risk of lymph node metastasis in renal carcinoma. *Front Endocrinol (Lausanne)*. (2022) 13:1054358. doi: 10.3389/fendo.2022.1054358

31. Wang X, Chen Y, Gao Y, Zhang H, Guan Z, Dong Z, et al. Predicting gastric cancer outcome from resected lymph node histopathology images using deep learning. *Nat Commun*. (2021) 12:1637. doi: 10.1038/s41467-021-21674-7

32. Sun Y, Wu X, Lin H, Lu X, Huang Y, Chi P. Lymph node regression to neoadjuvant chemoradiotherapy in patients with locally advanced rectal cancer: prognostic implication and a predictive model. *J Gastrointest Surg*. (2021) 25:1019–28. doi: 10.1007/s11605-020-04566-x

33. Gao P, Zhu T, Gao J, Li H, Liu X, Zhang X. Impact of examined lymph node count and lymph node density on overall survival of penile cancer. *Front Oncol*. (2021) 11:706531. doi: 10.3389/fonc.2021.706531

34. Feng F, Liu J, Wang F, Zheng G, Wang Q, Liu S, et al. Prognostic value of differentiation status in gastric cancer. *BMC Cancer*. (2018) 18:865. doi: 10.1186/s12885-018-4780-0

35. Willmann J, Vlaskou Badra E, Adilovic S, Christ SM, Ahmadsei M, Mayinger M, et al. Distant metastasis velocity as a novel prognostic score for overall survival after disease progression following stereotactic body radiation therapy for oligometastatic disease. *Int J Radiat Oncol Biol Phys*. (2022) 114:871–82. doi: 10.1016/j.ijrobp.2022.06.064

36. Chen J, Wu L, Zhang Z, Zheng S, Lin Y, Ding N, et al. A clinical model to predict distant metastasis in patients with superficial gastric cancer with negative lymph node metastasis and a survival analysis for patients with metastasis. *Cancer Med*. (2021) 10:944–55. doi: 10.1002/cam4.v10.3

37. Lazăr D, Tăban S, Sporea I, Dema A, Cornianu M, Lazăr E, et al. Gastric cancer: correlation between clinicopathological factors and survival of patients. *II. Rom J Morphol Embryol*. (2009) 50:185–94.

38. Sato Y, Okamoto K, Kawano Y, Kasai A, Kawaguchi T, Sagawa T, et al. Novel biomarkers of gastric cancer: current research and future perspectives. *J Clin Med*. (2023) 12(14):4646. doi: 10.3390/jcm12144646

39. Spolverato G, Capelli G, Mari V, Lorenzoni G, Gregori D, Poultsides G, et al. Very early recurrence after curative-intent surgery for gastric adenocarcinoma. *Ann Surg Oncol*. (2022) 29:8653–61. doi: 10.1245/s10434-022-12434-y

40. Kim J, Park J, Park H, Choi MS, Jang HW, Kim TH, et al. Metastatic lymph node ratio for predicting recurrence in medullary thyroid cancer. *Cancers (Basel)*. (2021) 13 (22):5842. doi: 10.3390/cancers13225842

41. Charlton ME, Kahl AR, McDowell BD, Miller RS, Komatsoulis G, Koskimaki JE, et al. Cancer registry data linkage of electronic health record data from ASCO's cancerLinQ: evaluation of advantages, limitations, and lessons learned. *JCO Clin Cancer Inform*. (2022) 6:e2100149. doi: 10.1200/CCI.21.00149

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 22 November 2024
DOI 10.3389/fimmu.2024.1491508

Check for updates

OPEN ACCESS

EDITED BY
Petar Ozretić,
Rudjer Boskovic Institute, Croatia

REVIEWED BY
Antonio Giovanni Solimando,
University of Bari Aldo Moro, Italy
Sreeram Vallabhaneni,
Harvard Medical School, United States

*CORRESPONDENCE
Runhong Mu
✉ murunhong@126.com
Tao Wang
✉ wangtaoGPPH@gzu.edu.cn

†These authors have contributed equally to this work

RECEIVED 05 September 2024
ACCEPTED 06 November 2024
PUBLISHED 22 November 2024

CITATION
Guo X, Xing J, Cao Y, Yang W, Shi X, Mu R and Wang T (2024) Machine learning based anoikis signature predicts personalized treatment strategy of breast cancer.
Front. Immunol. 15:1491508.
doi: 10.3389/fimmu.2024.1491508

Machine learning based anoikis signature predicts personalized treatment strategy of breast cancer

Xiao Guo[1†], Jiaying Xing[1†], Yuyan Cao[1], Wenchuang Yang[1], Xinlin Shi[1], Runhong Mu[2*] and Tao Wang[3*]

[1]School of Pharmacy, Beihua University, Jilin, China, [2]School of Basic Medical Sciences, Beihua University, Jilin, China, [3]Research Laboratory Center, Guizhou Provincial People's Hospital, Guiyang, Guizhou, China

Background: Breast cancer remains a leading cause of mortality among women worldwide, emphasizing the urgent need for innovative prognostic tools to improve treatment strategies. Anoikis, a form of programmed cell death critical in preventing metastasis, plays a pivotal role in breast cancer progression.

Methods: This study introduces the Artificial Intelligence-Derived Anoikis Signature (AIDAS), a novel machine learning-based prognostic tool that identifies key anoikis-related gene patterns in breast cancer. AIDAS was developed using multi-cohort transcriptomic data and validated through immunohistochemistry assays on clinical samples to ensure robustness and broad applicability.

Results: AIDAS outperformed existing prognostic models in accurately predicting breast cancer outcomes, providing a reliable tool for personalized treatment. Patients with low AIDAS levels were found to be more responsive to immunotherapies, including PD-1/PD-L1 inhibitors, while high-AIDAS patients demonstrated greater susceptibility to specific chemotherapeutic agents, such as methotrexate.

Conclusions: These findings highlight the critical role of anoikis in breast cancer prognosis and underscore AIDAS's potential to guide individualized treatment strategies. By integrating machine learning with biological insights, AIDAS offers a promising approach for advancing personalized oncology. Its detailed understanding of the anoikis landscape paves the way for the development of targeted therapies, promising significant improvements in patient outcomes.

KEYWORDS

breast cancer, anoikis, personalized treatment, PD-L1, methotrexate

Introduction

Breast cancer (BC) is the most common malignant tumor in women in the world, and its incidence rate has gradually increased in recent years (1). The diagnosis, treatment and prognosis of BC have a great impact on the health, lifestyle and work of individuals as well as their family life (2). With the continuous development of biomedical technology, the research on the prognosis of BC has also made much progress, and people's awareness of personalized treatment is increasing (3). Multiple sets of data have been integrated to predict the prognosis of BC. For example, prediction models based on genomics, transcriptomics and proteomics data can be used to predict the survival rate and recurrence of BC patients (4). In recent years, artificial intelligence technology has also been widely used in predicting the prognosis of BC, and the prediction model based on machine learning can more accurately evaluate the prognosis by integrating a large number of clinical data and bioinformatics data (5).

Anoikis is a specialized form of programmed cell death triggered by the loss of cellular attachment to the extracellular matrix and neighboring cells, playing a pivotal role in tumor development and metastasis (6). While anoikis is crucial in tumor invasion and infiltration, there are limited studies systematically evaluating and predicting BC prognosis based on anoikis.

We conducted a comprehensive analysis to elucidate the importance of anoikis. Leveraging bulk and single-cell sequencing techniques, we evaluated anoikis activity across various cell types. Machine learning algorithms were employed to identify anoikis genes associated with BC prognosis, allowing us to construct predictive models. These models demonstrated the efficacy of anoikis in predicting BC patient outcomes, immune status, responsiveness to immune checkpoint inhibitors (ICIs) and chemotherapy, as well as in identifying potential therapeutic targets and drugs. Through rigorous evaluations, anoikis emerged as a promising tool for precise prognostication and treatment stratification in BC patients.

Methods

Data acquisition

We retrospectively collected data from 12 distinct breast cancer cohorts derived from The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), and Metabric (7). These cohorts included samples with comprehensive survival information, enabling thorough analysis. Our study encompassed a total of 11,033 patients across the 12 cohorts for prognostic evaluation. The patient distribution was as follows: TCGA-BRCA (n = 1076), GSE202203 (n = 3206), GSE96058 (n = 3409), GSE20685 (n = 327), GSE58812 (n = 107), GSE21653 (n = 244), GSE7390 (n = 198), GSE11121 (n = 200), GSE86166 (n = 330), GSE88770 (n = 108), GE48391 (n = 81), and Metabric (n = 1747). Genes implicated in the anoikis process were obtained from the Molecular Signature Database on the GSEA website (8).

Machine learning derived anoikis signature

To develop a breast cancer-specific anoikis signature, we employed the methodology established in our previous research (9). Our approach involved utilizing ten diverse computational Survival algorithms: Random Survival Forest (RSF), Least Absolute Shrinkage and Selection Operator (LASSO), Gradient Boosting Machine (GBM), Survival Support Vector Machine (Survival-SVM), Supervised Principal Component (SuperPC), Ridge Regression, Partial Least Squares Cox Regression (plsRcox), CoxBoost, Stepwise Cox regression, and Elastic Net (Enet). Among these, RSF, LASSO, CoxBoost, and Stepwise Cox were chosen for their ability to reduce dimensionality and identify relevant variables. These techniques were combined into 108 unique configurations to construct a predictive signature. By evaluating all cohorts, including TCGA and other datasets, we identified the most robust prognostic model through the calculation of the average Concordance index (C-index). This iterative process culminated in the creation of an anoikis-specific signature designed to predict outcomes in breast cancer.

Genomic alteration analysis

To elucidate genetic disparities between the two AIDAS groups, we analyzed genetic mutation levels and Copy Number Alterations (CNA) using the TCGA-BRCA database. The Tumor Mutation Burden (TMB) for both high- and low-AIDAS breast cancer patients was derived from the raw mutation data. Utilizing the maftools landscape, we depicted the most frequently mutated genes (mutation rate > 5%). Patient-specific mutational signatures were identified using the deconstructSigs package (10), emphasizing four prominent mutational signatures (SBS3, SBS1, SB12, SBS11) that exhibited elevated mutation frequencies in the TCGA-BRCA dataset. We identified the five most common regions of amplification and deletion, with a specific focus on the four predominant genes in chromosomal regions 3q26.32 and 5q21.3.

Single-cell data processing

We applied Seurat (v4.0) to process the single-cell data from GSE161529 (11). This involved filtering out genes with zero expression and retaining those with nonzero expression levels. The expression matrix was normalized using Seurat's "SCTransform" function. Dimensionality reduction was performed using principal component analysis (PCA) and UMAP techniques. To identify distinct cellular groupings, we employed Seurat's "FindNeighbors" and "FindClusters" functions. To ensure dataset integrity, the DoubletFinder package was used to eliminate potential doublets (12). Cells failing to meet quality standards, such as those with mitochondrial gene content exceeding 15% or fewer than 500 genes, were excluded. Following stringent quality control measures, 64,308 cells were retained for analysis. Cell types were determined by manual annotation based on established marker genes.

Inference of regulons and their activity

We utilized the Single-Cell rEgulatory Network Inference (SCENIC) approach to construct gene regulatory networks (GRNs) from single-cell RNA sequencing data. SCENIC involves a three-step process: first, it identifies co-expression modules between transcription factors (TFs) and their potential target genes. Next, it identifies the direct target genes for each module, prioritizing those enriched with the motif of the associated TF, thereby defining a regulon comprising a TF and its direct targets. Finally, the regulatory activity score (RAS) is computed for each cell by evaluating the area under the recovery curve.

To address the conventional SCENIC protocol's challenges with scalability for extensive datasets and its susceptibility to sequencing depth variations, we modified it to enhance both scalability and robustness. This involved partitioning the data into metacells before applying SCENIC to these gene expression profiles (13). This adjustment significantly improved data quality and reduced computational demands, representing a notable advancement in the application of SCENIC to single-cell RNA-seq data analysis.

Regulon clustering

We employed a robust computational method to dissect the regulatory relationships between transcription factors (TFs) and their target genes, with a focus on TF clustering. The process began by filtering TF-target interaction data to isolate pairs exceeding a significance threshold (>1), prioritizing the most critical regulatory interactions. We then identified key regulatory TFs by assessing their influence on target gene regulation, highlighting them as central nodes in the regulatory network for detailed analysis.

To visualize the intricate network of TF-target interactions, we constructed a graph model. A force-directed algorithm was used to refine the spatial layout of the graph, intuitively representing the network's structure and the interplay between TFs and their targets. For an enhanced understanding of the network's architecture, the Leiden algorithm was applied for community detection. This revealed the modular organization of TFs based on their regulatory connections, assigning each TF to a specific cluster. This approach allowed for a detailed analysis of the regulatory landscape, providing insights into the functional organization of TFs within the network.

Cell-cell communication analysis

Using the "CellChat" R package, we generated CellChat objects from the UMI count matrices for each group (14). The "CellChatDB.human" database was used as the reference for ligand-receptor interactions. Intercellular communication was interpreted using the default settings of the package. To compare interaction counts and intensities, we merged CellChat objects from each group with the "mergeCellChat" function. Differences in interaction numbers and intensities among specific cell types were visualized using the "netVisual_diffInteraction" function. Changes in signaling pathways were identified using the "rankNet" function, and the distribution of signaling gene expression among groups was displayed with the "netVisual_bubble" and "netVisual_aggregate" functions.

Additionally, we employed the NicheNet package to analyze intercellular communication from the perspective of ligand activity and the expression patterns of specific downstream targets regulated by these key ligands (15). This method provided a detailed understanding of the signaling processes underlying cell-cell interactions, using ligand-target relationships to infer communication pathways within the cellular microenvironment.

Evaluation of TME disparities and immunotherapy response

To comprehensively and accurately assess immune cell infiltration levels, we analyzed adverse infiltrated immune cells using multiple algorithms, including MCPcounter, EPIC, xCell, CIBERSORT, quanTIseq, and TIMER, among patients stratified by the AIDAS (16–22). Additionally, to precisely depict the immune landscape and architecture within the tumor microenvironment (TME), we evaluated the ESTIMATE and TIDE indices. These metrics provide critical insights into the potential for immunotherapy and offer prognostic implications for breast cancer patients.

Moreover, we quantified immune checkpoints, which serve as indicators of the immune state and offer preliminary predictions of patient responsiveness to ICI therapy. This comprehensive evaluation of the immune profile within the TME is crucial for advancing personalized medicine and refining treatment strategies for breast cancer patients.

Determination of therapeutic targets and drugs for high AIDAS patients

We identified therapeutic targets and drugs for high-AIDAS patients from the Drug Repurposing Hub and dropped out duplicate compounds, resulting in a refined list of 6,125 compounds. We established the selection of therapeutic targets associated with breast cancer outcomes through Spearman correlation analysis. Specifically, we assessed the relationship between the AIDAS and gene expression levels, selecting genes with a correlation coefficient greater than 0.3 and a P-value less than 0.05. Additionally, genes with a correlation coefficient below -0.3 and a P-value below 0.05 were identified as linked to poor prognosis. The significance of these genes was further evaluated by examining the relationship between CERES scores from the Cancer Cell Line Encyclopedia (CCLE) and model value (23).

To enhance predictions regarding drug responsiveness, we utilized data from the Cancer Therapeutics Response Portal (CTRP) and the PRISM project, both of which offer extensive drug screening and molecular data across diverse cancer cell lines. Differential expression analysis was conducted between bulk samples and cell lines. Subsequently, we employed the pRRophetic package to implement a ridge regression model for predicting drug response. This model, trained using expression data and drug response metrics from solid Cancer Cell Lines (CCLs), demonstrated excellent predictive accuracy, validated through 10-fold cross-validation (24).

Furthermore, to identify the most promising therapeutic drugs for breast cancer, we performed a Connectivity Map (CMap) analysis. This entailed comparing gene expression profiles across different risk subgroups and submitting the top 300 genes (150 up-regulated and 150 down-regulated) to the CMap website. A negative CMap score indicated a higher therapeutic potential against breast cancer, suggesting an inverse relationship between the CMap score and a compound's effectiveness as a potential treatment.

Patient stratification

To evaluate gene expression in breast cancer specimens, RNA extraction was conducted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). This was followed by cDNA synthesis and quantitative reverse transcription PCR (qRT-PCR) using GoScript reverse transcriptase and Master Mix (Promega), adhering to the manufacturer's instructions. Data acquisition was performed with the CFX96 Touch Real-Time PCR Detection System (BioRad, Hercules, CA, USA). Gene expression levels were quantified using the $2^{-\Delta\Delta Cq}$ method, with GAPDH serving as the normalization control. Patients were subsequently categorized based on their gene expression profiles using a predefined formula derived from the AIDAS. This stratification was crucial in identifying patients with distinct risk profiles, thus facilitating tailored therapeutic interventions.

Immunohistochemistry experiment

Tissue samples were collected from 30 breast cancer patients undergoing surgery at Guizhou Provincial People's Hospital. These samples were subjected to Hematoxylin and Eosin (H&E) staining following established protocols (25, 26), with diagnoses independently confirmed by two pathologists.

For immunohistochemistry (IHC) analysis, paraffin-embedded samples were processed according to procedures outlined in previous studies. Protein expression levels were evaluated independently by two pathologists, adhering to standardized protocols and scoring systems consistent with methodologies from prior research (26).

Results

Construction of an anoikis model using artificial intelligence

The comprehensive evaluation of the anoikis model was conducted using a combination of 108 machine learning algorithms with ten-fold cross-validation (Figure 1A). The performance of the models was assessed by calculating the average C-index across various cohorts, with the Random Survival Forest (RSF) algorithm demonstrating the highest average C-index (0.632). The key anoikis genes were identified based on the point with the lowest error rate of RSF in 1000 tests (Figures 1B, C). These genes underwent univariate Cox regression analysis to calculate the hazard ratio (HR) across nine enrolled cohorts (Figure 1D). Finally, four genes (PTK2, coef = 0.278; NOTCH1, coef = 0.145; PKD4, coef = -0.169; BCL2, coef = -0.236) were selected to construct an artificial intelligence-derived anoikis signature (AIDAS) (Figure 1E). The evaluation of AIDAS across the nine cohorts revealed that the binary classification model effectively classified patients into high and low-AIDAS groups (Supplementary Figure S1).

Assessment of AIDAS with 83 published models

We further conducted both univariate and multivariate Cox analysis to assess the independence of AIDAS and other clinical indices (Supplementary Figure S2A). Three significant indices, namely AIDAS, stage, and age, were chosen to develop a nomogram aimed at predicting patients' survival rates in clinical practice (Supplementary Figure S2B). The overall survival (OS) of breast cancer patients with different conditions was predicted, and the OS curve demonstrated a good fit with the standard curve, indicating the model's accuracy (Supplementary Figures S2C, D). Through comparisons with other factors, it was observed that AIDAS could provide more accurate predictions of patients' conditions (Supplementary Figure S2F).

The stability of the predictive model of the AIDAS was evaluated by collecting and assessing 83 published signatures in BC across 9 independent cohorts. It was demonstrated that only the AIDAS exhibited consistent statistical significance across all cohorts (Figure 2A). The predictive power of the AIDAS was compared with the 83 models across the 9 cohorts using the C-index (Figure 2B). The AIDAS showed significantly better accuracy than the others in almost all cohorts, ranking first in seven cohorts, fifth in one cohort, and seventh in one cohort, thereby revealing the stability of our model (Figure 2B).

Multi-omics analysis of genomic alterations based on AIDAS

Gene variations between the AIDAS groups were analyzed using multi-omics integration analysis. We observed a significant increase in TMB in high-AIDAS patients, accompanied by multigene mutation characteristics (Figures 3A, C). When considering 10 oncogenic signaling pathways together, classic tumor suppressor genes like TP53, RB1, and AXIN1/2 were found to mutate more frequently in the high-AIDAS group, while oncogenic genes such as RET, PIK3CA/B, and RPTOR mutated less

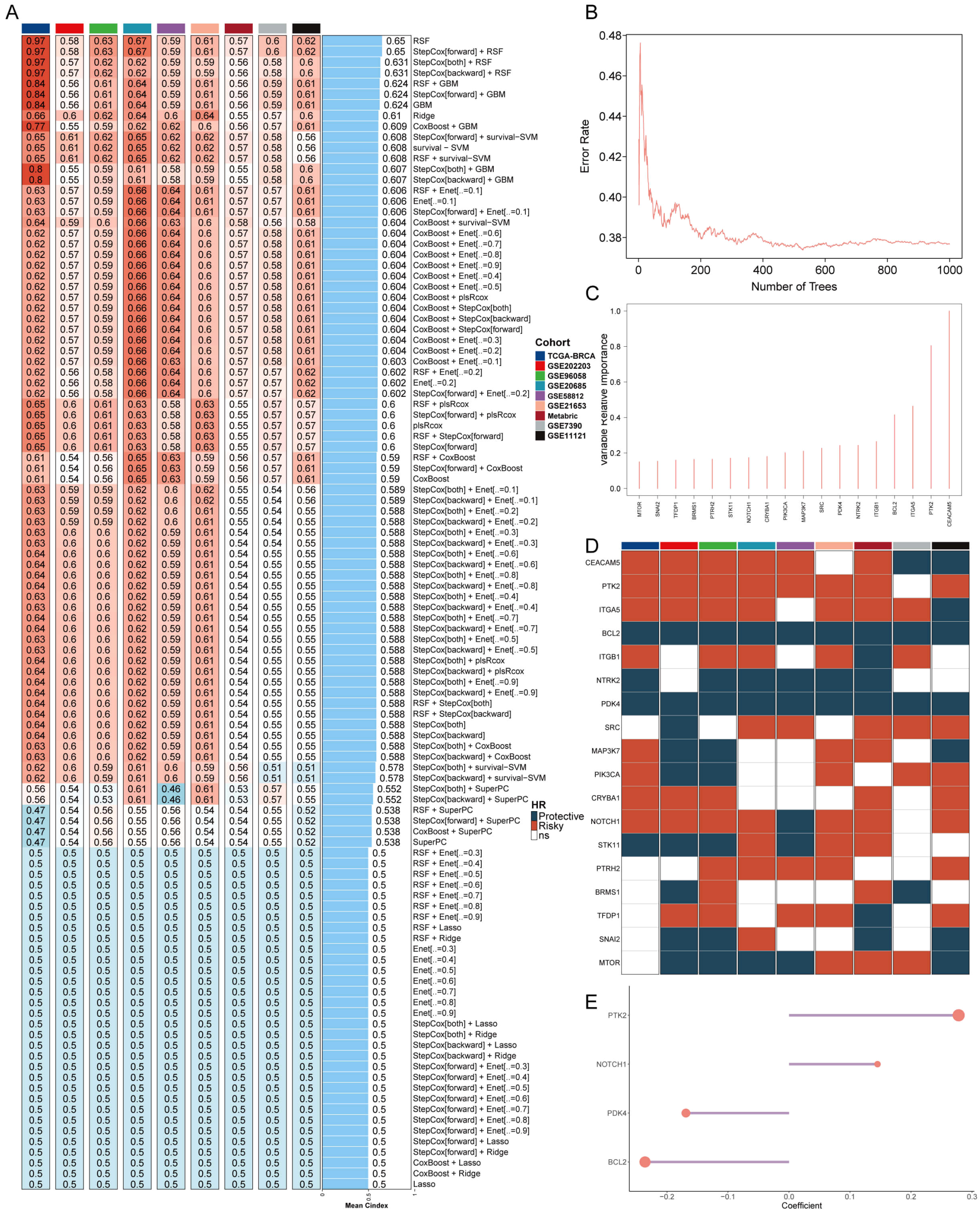


FIGURE 1
Construction of an anoikis model using artificial intelligence. **(A)** C-indexes of the 108 machine learning algorithm combinations in the nine cohorts. **(B)** Error rate of RSF after 1000 tests. **(C)** Key anoikis genes selected by RSF. **(D)** Prognostic value of key genes in nine BC cohorts. **(E)** Final selection of 4 anoikis genes based on an exhaustive search, with patient risk scores calculated according to the expression levels of these genes and their regression coefficients.

(Figures 3A, B). Further analysis of CNV between these subgroups revealed that amplifications and deletions at the level of chromosome arms were more pronounced in the high-AIDAS group, including amplifications of 3q26.32, 6p23, 6q21, 8q24.21, and 10p15.1, as well as deletions of 5q11.2, 5q21.3, 14q24.1, 14q32.12, and 19p13.3 (Figures 3A, D). These results suggest that the poor prognosis for high-AIDAS patients may be related to significant increases in the amplification of 3q26.32 and multiple oncogenes genes (ASAP1, PVT1, TMEM75, and MYC), as well as deletions of multiple tumor suppressor genes of 5q21.3 (GPBP1, RAB3C, DDX4, and ITGA1) (Figure 3A).

Deciphering the AIDAS at the single-cell level

The expression characteristics of different immune infiltrating cells were revealed at the single-cell level. The distribution of cells from 8 BC

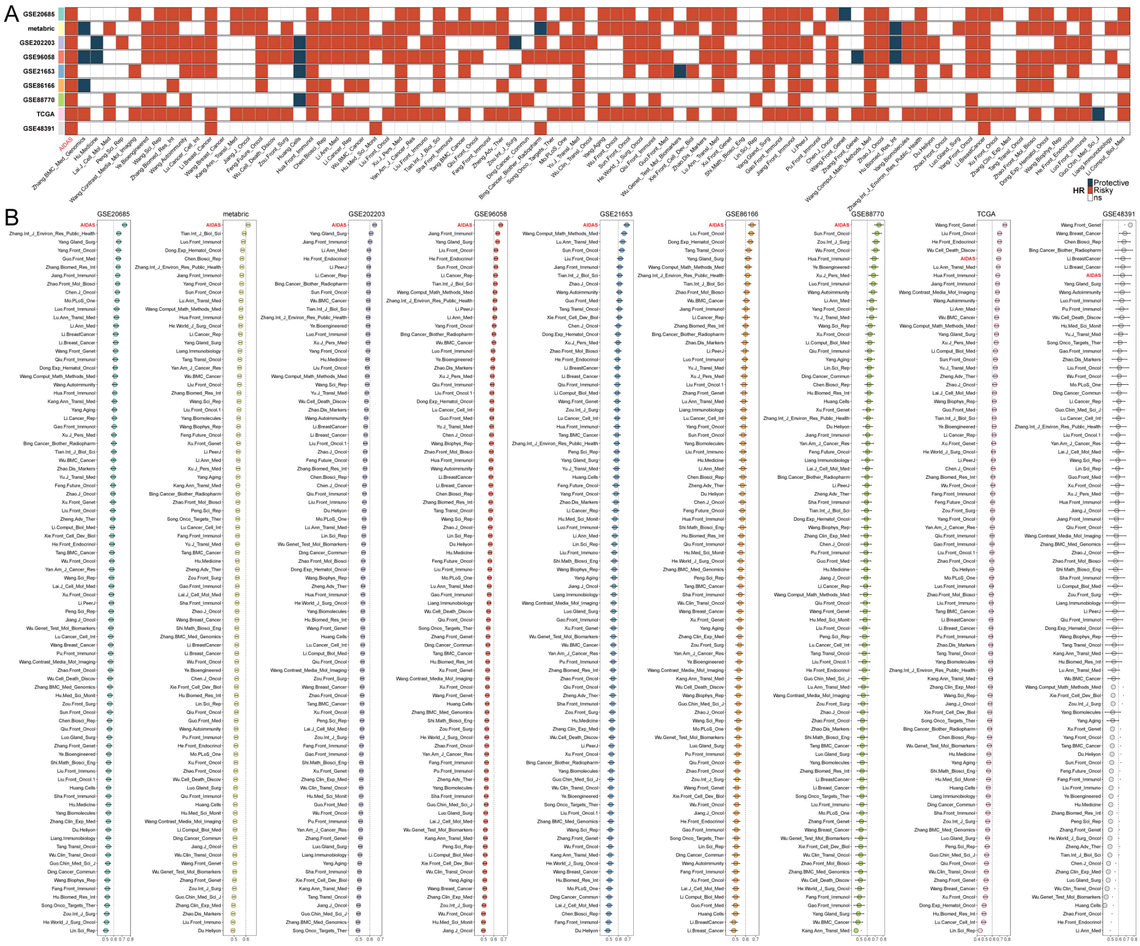


FIGURE 2
Assessment of AIDAS with 83 published models. **(A)** The stability of AIDAS was compared with 83 published models. **(B)** C-index values of AIDAS and 83 published models in 9 different datasets.

patients was analyzed, and the distribution of tumor and normal tissues (Supplementary Figures S3A, B), 17 cell clusters were identified and divided into 6 cell types (Figures 4A, B). The number of cells in these 6 types was statistically analyzed, and then their proportion in the bodies of these 8 tumor patients was calculated (Supplementary Figures S3C, D). The representative markers in these 6 types of cells, as well as their actual distribution in the cells (Figure 4C; Supplementary Figure S3E) were observed. The results showed that epithelial cells and macrophages accounted for a larger proportion of the tumor tissue, while fibroblasts, T cells, Pericytes, and endothelial cells accounted for a larger proportion in the normal tissue (Figure 4D).

Next, the AIDAS was incorporated into the single-cell distribution map (Figure 4E). All cells were divided into low- and high-AIDAS groups based on their peak of epithelial cells (Figure 4F), and then differential gene expression analysis and functional clustering were performed to elucidate potential functional pathways (Supplementary Figures S3F, G). Subsequently, copyKat analysis was performed to observe the CNV for distinguishing the normal cells and tumor cells (Figure 4G). We observed a higher AIDAS score in tumor-aneuploid than in tumor-diploid, implying the significance of AIDAS in breast cancer progression (Figure 4H).

Specific regulons for AIDAS and cell recognition

To comprehensively construct a GRNs of AIDAS, a SCENIC pipeline was applied to analyze single-cell RNA seq data with cis-regulatory sequence information (Figures 5A, B). PCA and variance analyses were performed on different cell types and AIDAS. PCA1 explained the specific transcription factors of different cell types, while PCA2 was associated with the regulons of AIDAS (Figures 5C, D). The key transcription factors for cell recognition were identified, and the regulon specificity score (RSS) of these specific transcription factors in different types of cells was evaluated (Figure 5E). The regulatory factors with higher RSS scores were selected from these six types of cells, and GATA3, SPDEF, and PITX1 were identified as the most relevant specific regulators to epithelial cells (Figure 5F). Similarly, the most relevant specific regulators to the other five types of cells were analyzed (Supplementary Figure S4A).

Understanding that TFs often collaborate to modulate gene expression, we systematically explored the combinatory patterns of these regulatory elements. Based on the Leiden algorithm, the similarity of RAS scores for each TF was compared, and the cluster

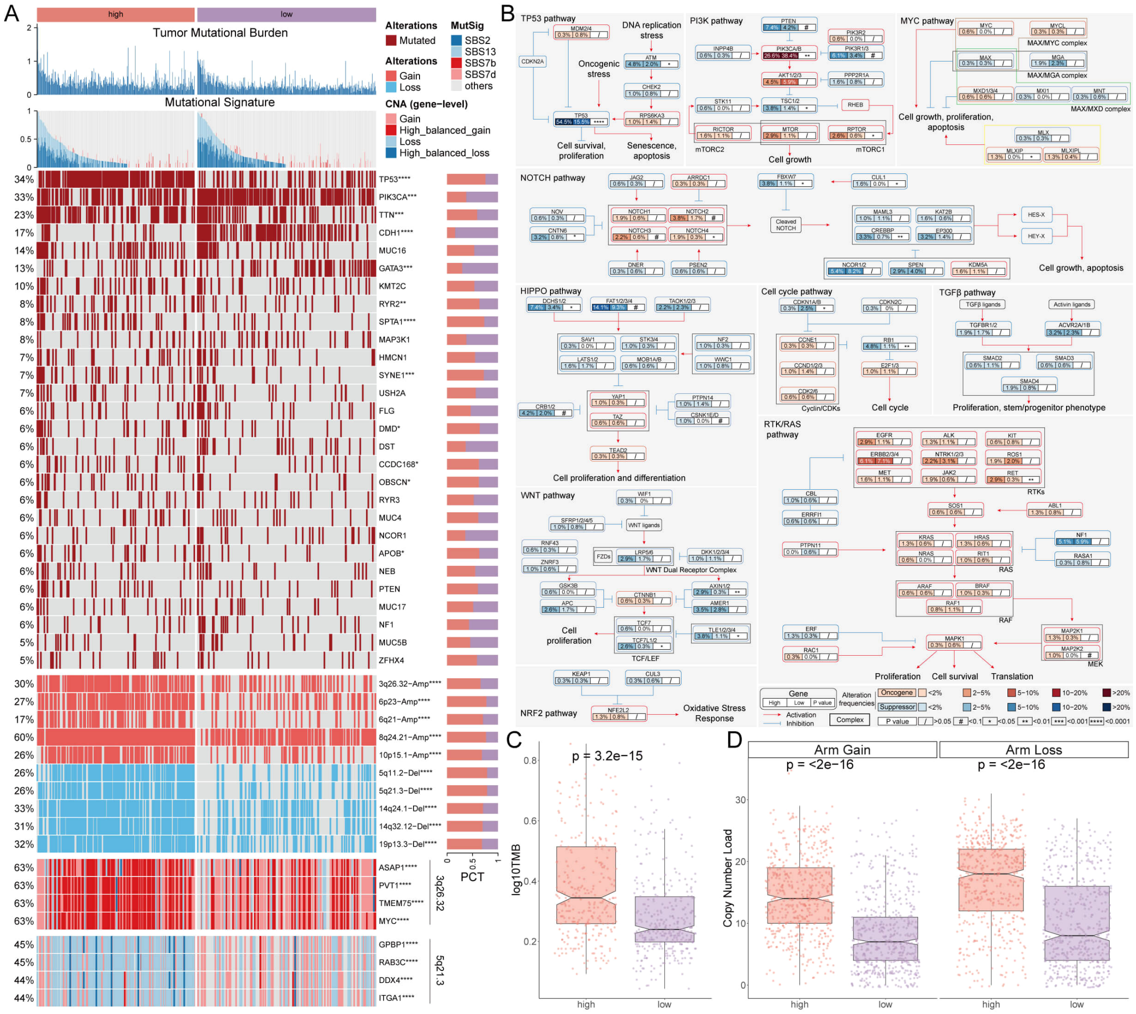


FIGURE 3
Multi-omics analysis of genomic alterations based on AIDAS. **(A)** Overview of genomic variations based on AIDAS. **(B)** Mutation atlas of 10 oncogenic pathways. **(C)** Difference of TMB values. **(D)** Comparison of copy number load between the two AIDAS groups. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

analysis of TFs was conducted to find 12 clusters of transcription factors, of which the contribution rate of transcription factor sets B and E to the development of AIDAS was greater than that of the other 11 clusters, so only transcription factors B and E were displayed (Figures 5G, H, Supplementary Figure S4B). We next focused on the exact TFs that drive epithelial cells' transcriptomic changes by AIDAS. Multiple pathways were identified by GSEA analysis. For example, collagen degradation was activated in epithelial cells in the high-AIDAS cells, while interference alpha beta signaling was inhibited (Figures 5I, J). Transcription factors contributing to these pathways were identified by further analysis (Figure 5K). The network diagrams of regulatory relationships among transcription factors were shown (Figure 5L).

Intercellular communications for AIDAS

Intercellular communication among six cell types was evaluated by CellChat analysis. We observed that the number and intensity of cell-cell interactions were stronger in the low-AIDAS cells, and the intercellular communication between epithelial cells and endothelial cells was elevated (Figures 6A, B). Some signaling pathways involved in intercellular communication were analyzed, and the results showed that most of them had stronger intercellular communications in the low-AIDAS cells (Figure 6C). By comparing changes in outgoing and incoming signals among different cells, it was found that incoming interactions of epithelial cells were stronger in the low-AIDAS cells, indicating that incoming interactions of epithelial cells in the low-AIDAS group may be enhanced after they receive signals from other cells (Figure 6D).

Potential ligands of epithelial cells in the different groups were speculated using nichenetr analysis. We inferred potential ligands that may regulate epithelial cells from other cells based on the AIDAS group. The potential ligand-receptor pairs were further evaluated (Figure 6F). A high degree of interaction between THBS1-SDC4 and CNN1-SDC4 was observed, indicating that fibroblasts are the main sending cells affecting changes in the epithelial cell pathway (Figure 6G). THBS1 ligand and CNN1 ligand could reach

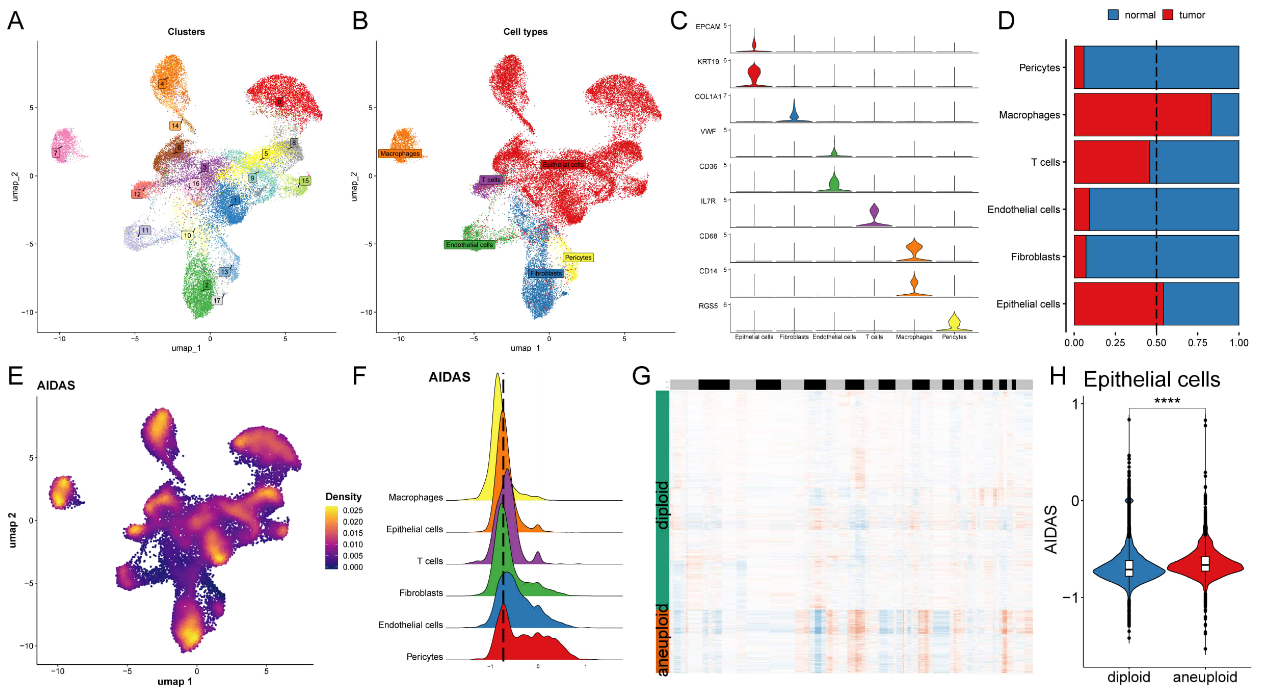


FIGURE 4
Deciphering the AIDAS at the single-cell level. **(A)** Distribution of 20 cell clusters. **(B)** Seven cell types identified by the established marker genes. **(C)** Representative markers of each cell type. **(D)** Proportion of seven types of cells between tumor patients and normal tissues; **(E)** Distribution of AIDAS value. **(F)** AIDAS value across cell types. **(G)** CNV evaluation using copyKat algorithm. **(H)** Comparison of the AIDAS value between diploid and aneuploid cells within the epithelial cells. ****P<0.0001.

the SDC4 through other receptors or transcription factors, in which high mutation rates of transcription factors such as TP53, MYC, and RAC1 in high-AIDAS (Figure 6H).

Personalized immunotherapy for low-AIDAS patients

Immune microenvironment is involved in breast cancer progression, six algorithms were applied to evaluate the immune infiltration of different AIDAS patients. A higher proportion of memory T cells, Tregs, M1 macrophages, and $CD8^+$ T cells were observed in the high-AIDAS patients (Figure 7A), and some ICIs were also overexpressed, such as PD-L1, CTLA4, and LAG3 (Figure 7B). IHC was performed to support the above results using the representative cell markers and clinical ICIs (Figure 7C).

Further analysis revealed that TIDE and Dysfunction values in the low-AIDAS group were higher than those in the high-AIDAS group, and there was no significant difference in the Exclusion value between the two groups (Figure 8A). There was a longer survival time in patients with a low-AIDAS and high-TIDE than in other combinations (Figure 8B). The correlation of AIDAS with the immune cycle and signaling showed that the anti-tumor immune activity of low-AIDAS patients was higher than that of high-AIDAS patients (Figure 8C).

ICIs have emerged as a transformative approach in cancer immunotherapy over the past several decades, yet their effectiveness in solid tumors, including breast cancer, remains limited. We sought to explore the predictive capability of AIDAS levels regarding the efficacy of immune checkpoint blockade therapies in the IMvigor210 (anti-PD-L1) and GSE78220 (anti-PD-1) cohorts.

Patients from low-AIDAS presented remarkable clinical benefits and better survival rates than the high-AIDAS in anti-PD-L1 response (Figures 8D–G). Prior benefits for low-AIDAS patients were also observed in anti-PD1 response (Figures 8H–L). Utilizing SubMap algorithms, we confirmed the response to immunotherapy, which was significantly more likely to benefit from treatments with anti-PD-L1 and CTLA4 treatments (Figure 8M). Based on the above research results, patients with the low-AIDAS can achieve better results in the treatment with ICIs.

Identification of therapeutic drugs for high-AIDAS patients

Chemotherapy is the standard treatment for anti-cancer, and data from multiple datasets have been used to develop potential drugs for BC patients with high-AIDAS. Seven therapeutic targets were identified using Spearman correlation analysis, and the results showed that high-AIDAS patients were positively correlated with the abundance of seven genes (MDH2, LIMK1, S100A2, TYRO3, COX7B, and ESRRA), and significantly negatively correlated with their CERES scores, suggesting that these seven genes can serve as a potential therapeutic target (Figure 9A). Potential drug targets were further analyzed based on drug sensitivity ratios, and it was revealed that these 7 genes had a high sensitivity to the drugs, so they were

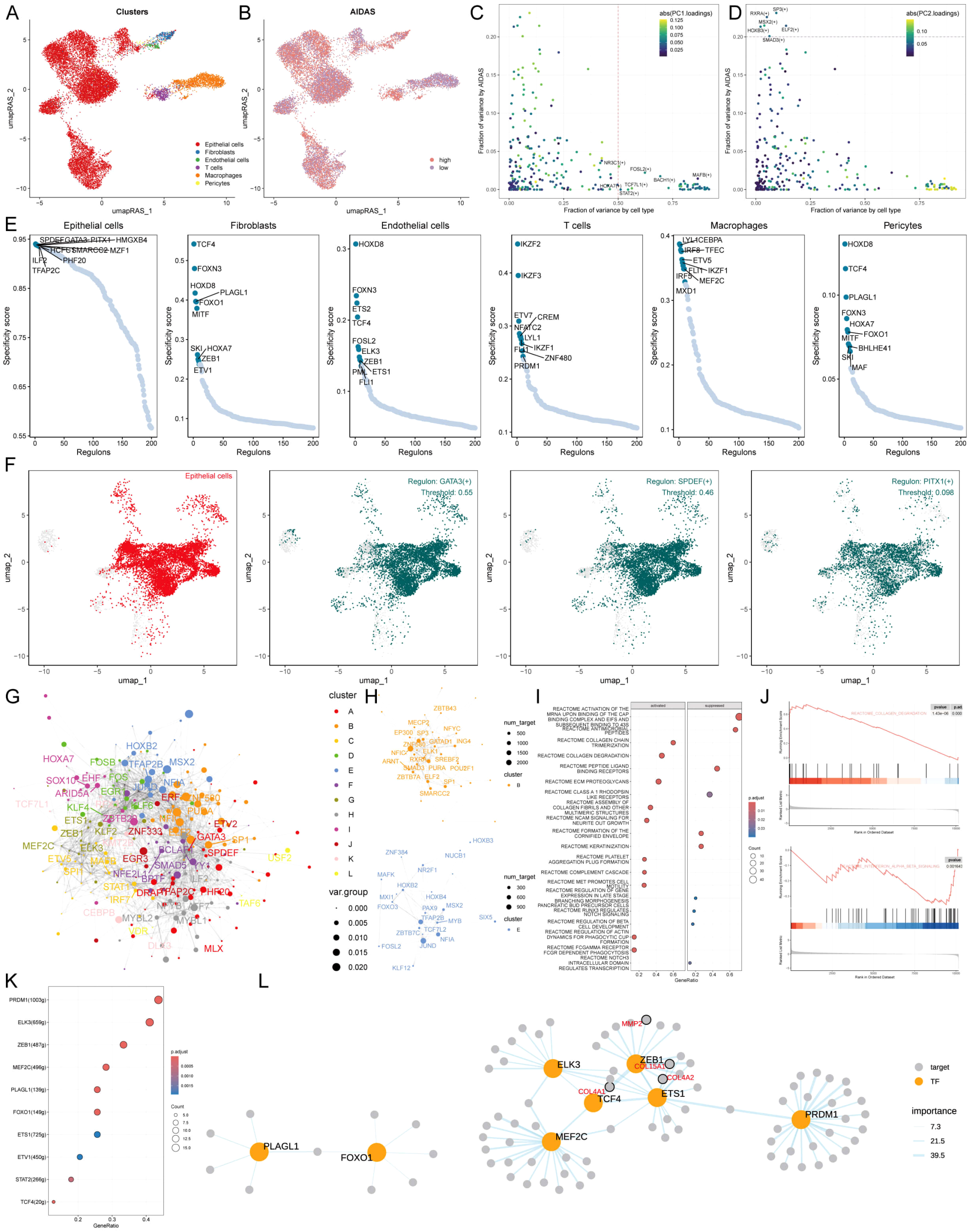


FIGURE 5
Specific regulons for AIDAS and cell recognition. **(A)** Distribution of cell types based on RAS. **(B)** Distribution of AIDAS value based on RAS. **(C)** Variance analysis highlights the impact of cell types and AIDAS on transcription factor activity, using color mapping to PC1 to emphasize the primary variance influenced by these factors. **(D)** Variance analysis with color mapped to PC2. **(E)** Scores of specific transcription factors in different types of cells. **(F)** Specific distribution of the most relevant specific regulators in epithelial cells. **(G)** Network of each transcription factor based on Leiden algorithm. **(H)** Specific transcription factor groups with higher scores in AIDAS. **(I)** GSEA identifies pathway variations linked to AIDAS in epithelial cells. **(K)** Transcription factors that could contribute to the collagen degradation. **(L)** Network of the regulatory relationship related to collagen degradation.

considered the key therapeutic targets for high-AIDAS patients (Figure 9B). Thirteen compounds were screened out from CTPR (CR-1-31B, SB-743921, BI2536, GSK461364, methotrexate, vincristine, paclitaxel, and leptomycin B) and PRISM datasets (docetaxel, vincristine, ispinesib, gemcitabine, and LY2606368), for evaluating candidate therapeutic drugs. The AUC values of the different compounds in the two groups were compared, and the results showed that lower AUC values were identified in high-AIDAS patients, indicating that these compounds may be suitable for the drug treatment of high-AIDAS patients (Figures 9C, D). The promising therapeutic agents were identified by CMap analysis, in which methotrexate, with a CMAP value of -99.82, was ultimately

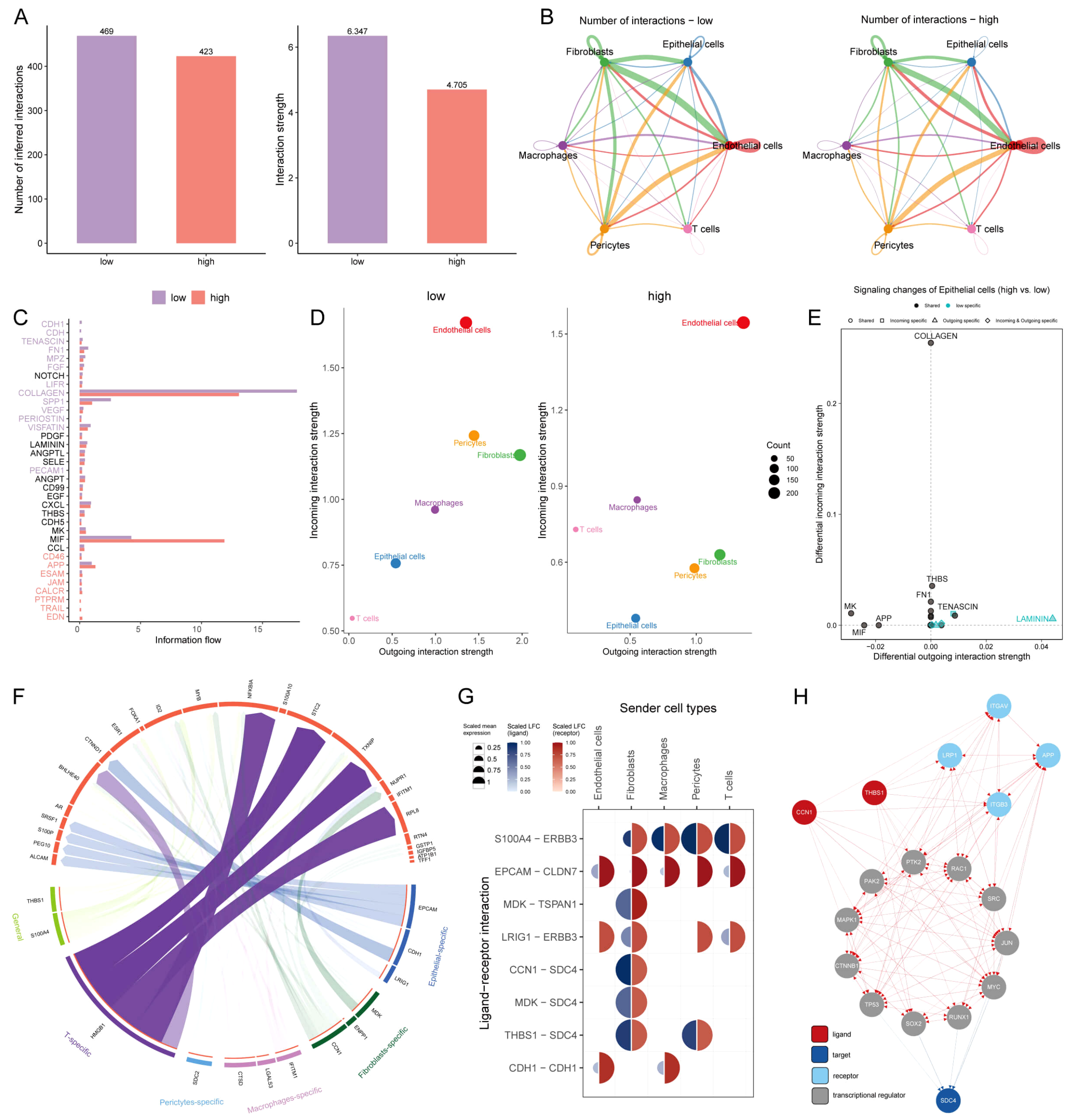


FIGURE 6
Intercellular communications for AIDAS. **(A)** Comparison of the number and intensity of cell interactions between two AIDAS groups. **(B)** Detail of cell communications among each cell type. **(C)** Differences of signaling pathways involved in the intercellular communication. **(D)** Intensity of incoming and outgoing interactions among different cells. **(E)** Specificity of incoming and outgoing signals of different signaling pathways. **(F)** Specific regulatory of ligands and receptors in cells. **(G)** Expression levels of ligands and receptors in different cells. **(H)** Route diagram of reaching target receptor SDC4 of CCN1 and THBS1 ligands through other receptors or transcription factors.

identified as the best potential therapeutic drug for high-AIDAS patients (Figure 9E).

Discussion

Considering the unique clinical characteristics of BC patients, it is necessary to customize specialized prognostic plans for these patients, and it is crucial to develop an accurate prognostic model. Anoikis is a specific form of programmed apoptosis caused by the disruption of cell-cell or cell-extracellular matrix attachment, and eliminating displaced or displaced cells can help maintain the dynamic balance of tissues (27), Anoikis is a term that describes the process of apoptosis that triggered by the detachment of cells from the extracellular matrix (28). It has been confirmed that anoikis is the first line of defense against cancer cell metastasis

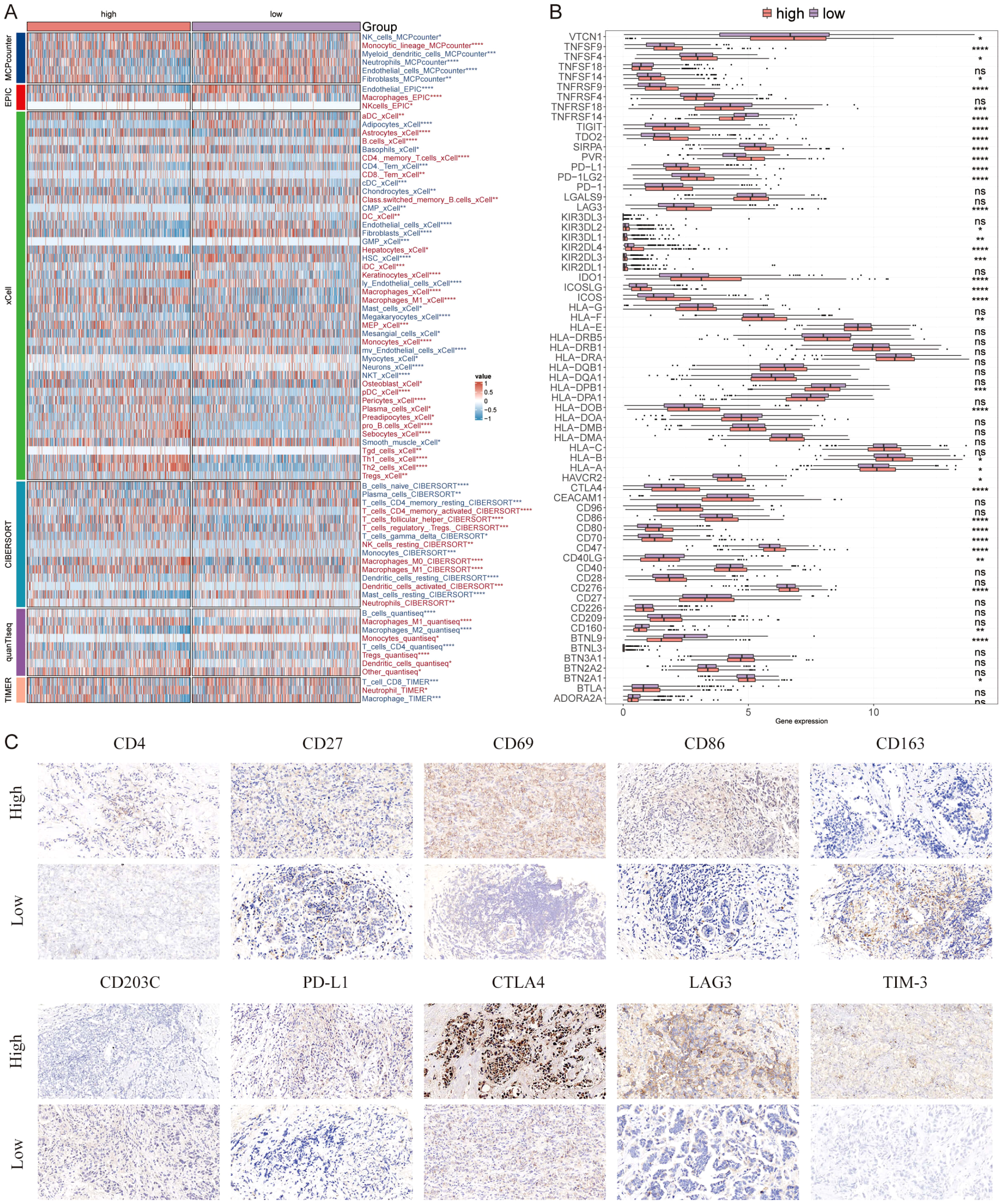


FIGURE 7
Differential expression and immunohistochemical analysis of immune markers in tumor microenvironments between AIDAS subgroups. **(A)** Heatmap provides a comparative view of immune cell infiltration in tumor samples with low and high AIDAS, utilizing various computational algorithms for quantification. Each row represents a different type of immune cell, with the color intensity reflecting the level of infiltration. **(B)** Box plots illustrate the distribution of gene expression levels for ICIs across low *vs*. high AIDAS conditions, with statistical significance denoted by ns for not significant; *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001. **(C)** Representative immunohistochemistry images showcase the staining intensity of various immune markers between high and low expression conditions, visually depicting the differential expression of these markers in correlation with AIDAS levels.

and an early intervention measure for preventing cancer metastasis (29). However, there is a limited prognostic model based on anoikis for predicting the prognosis and personalized treatment of BC.

By focusing on the process of anoikis—programmed cell death triggered by cellular detachment—AIDAS provides novel insights into how resistance to anoikis is linked to cancer aggressiveness and metastasis. Here, we discuss the clinical implications, biological rationale, and limitations of AIDAS, and outline directions for future research that could further enhance its utility as a personalized medicine tool.

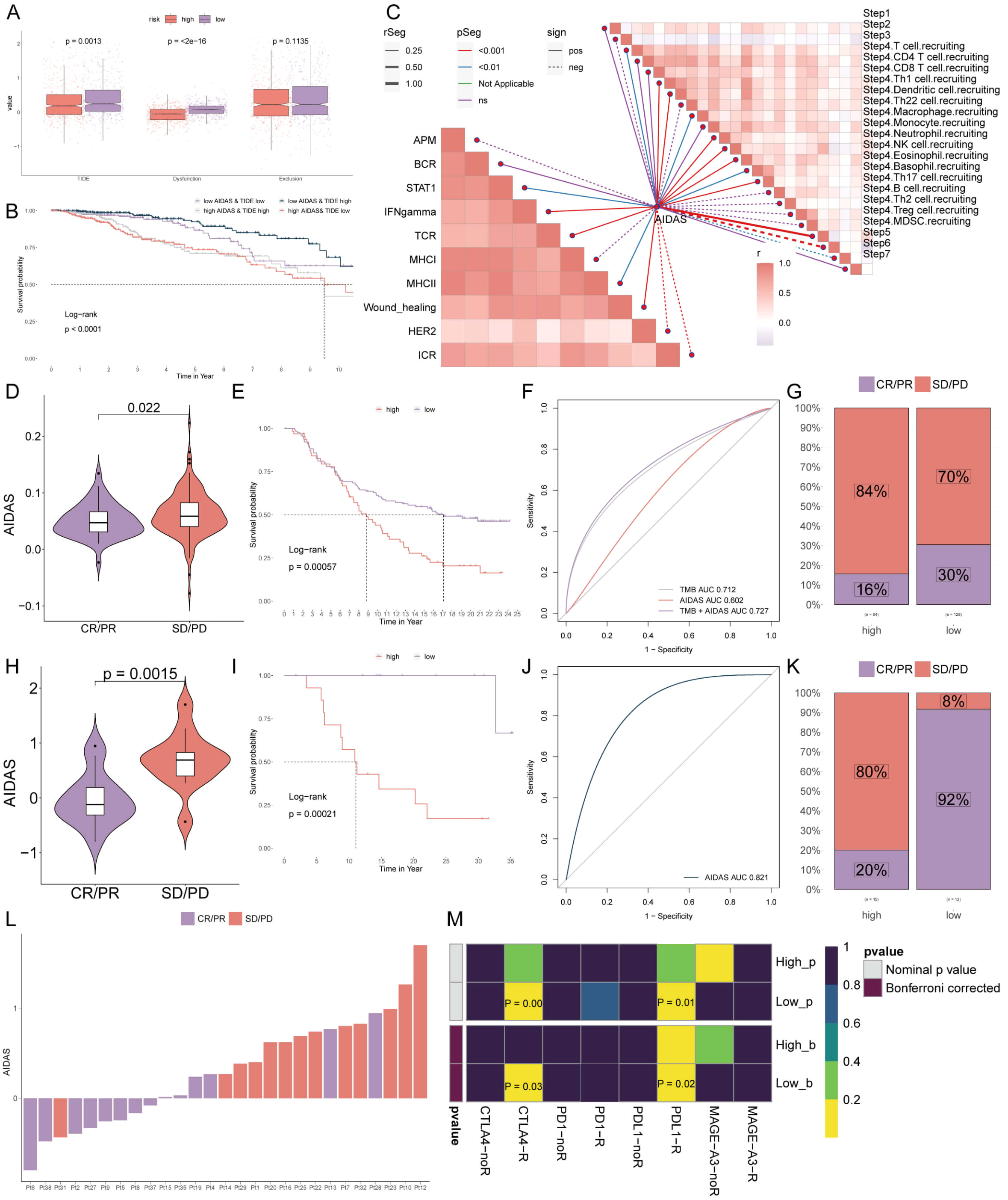
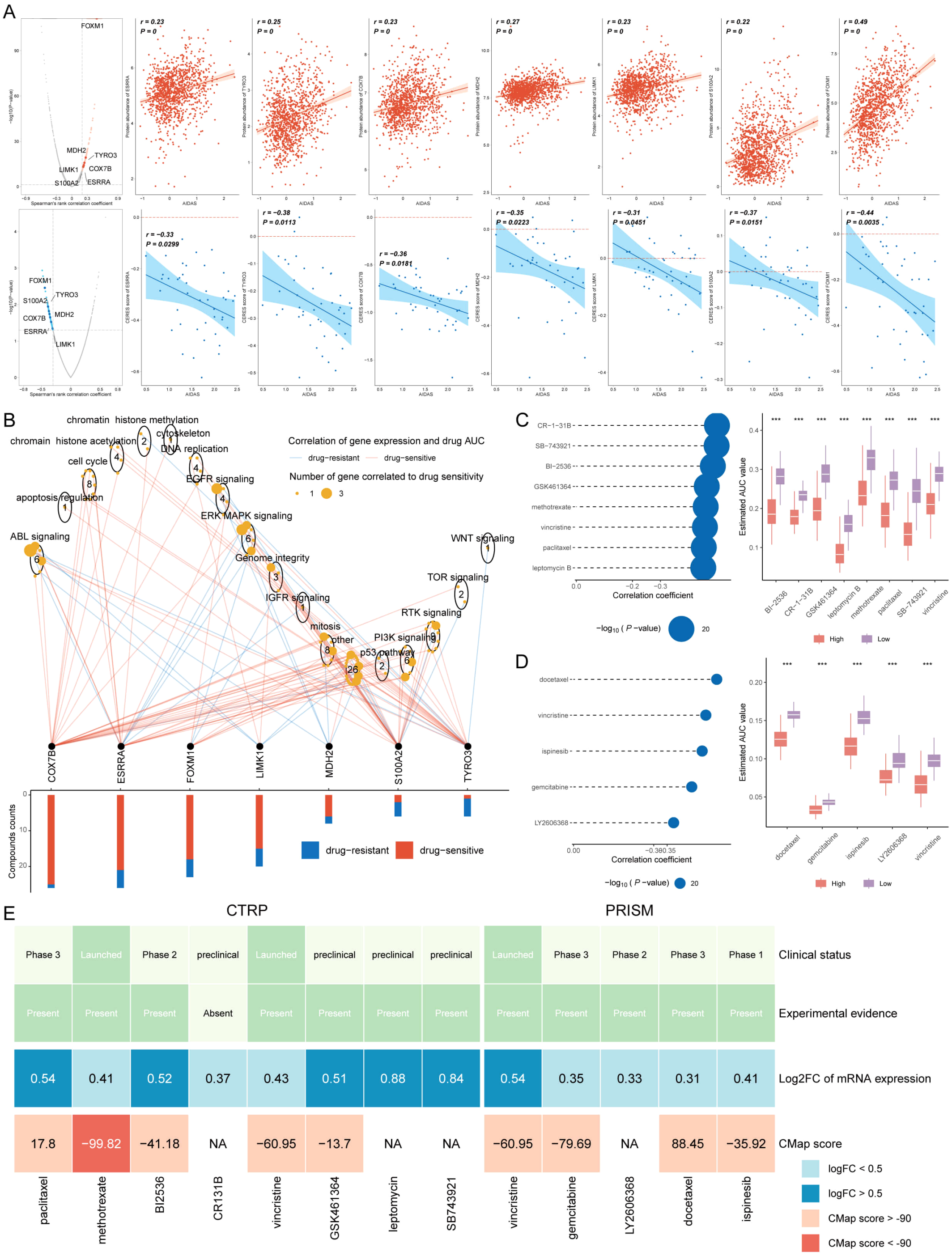


FIGURE 8
Personalized immunotherapy for low-AIDAS patients. **(A)** Differences in TIDE, Dysfunction, and Exclusion between patients in the low- and high-AIDAS groups. **(B)** Comparison of the survival probability of four combinations. **(C)** Correlation analysis of AIDAS with tumor immune cycle and ten immune pathways. **(D)** Correlation analysis of AIDAS value with anti-PD-L1 response. **(E)** KM survival curves of AIDAS after anti-PD-L1 treatment. **(F)** Accuracy of AIDAS and TMB in anti-PD-L1 treatment. **(G)** Proportion of CR/PR and SD/PD of anti-PD-L1 in ADIAS subgroups. **(H)** Correlation analysis of AIDAS value with anti-PD-1 response. **(I)** KM survival curves of AIDAS after anti-PD-1 treatment. **(J)** Accuracy of AIDAS and TMB in anti-PD-1 treatment. **(K)** Proportion of CR/PR and SD/PD of anti-PD-1 in ADIAS subgroups. **(L)** Distribution of ADIAS score of different patients after anti-PD-1 treatment. **(M)** Heatmap demonstrating the predictive power of ADIAS for responsiveness to different ICIs treatment.

By focusing on the process of anoikis—programmed cell death triggered by cellular detachment—AIDAS provides novel insights into how resistance to anoikis is linked to cancer aggressiveness and metastasis. Here, we discuss the clinical implications, biological rationale, and limitations of AIDAS, and outline directions for future research that could further enhance its utility as a personalized medicine tool. AIDAS leverages machine learning to capture complex interactions among anoikis-related genes, enabling



paclitaxel	methotrexate	BI2536	CR131B	vincristine	GSK461364	leptomycin	SB743921	vincristine	gemcitabine	LY2606368	docetaxel	ispinesib	
Phase 3	Launched	Phase 2	preclinical	Launched	preclinical	preclinical	preclinical	Launched	Phase 3	Phase 2	Phase 3	Phase 1	Clinical status
Present	Present	Present	Absent	Present	Present	Present	Present	Present	Present	Present	Present	Present	Experimental evidence
0.54	0.41	0.52	0.37	0.43	0.51	0.88	0.84	0.54	0.35	0.33	0.31	0.41	Log2FC of mRNA expression
17.8	−99.82	−41.18	NA	−60.95	−13.7	NA	NA	−60.95	−79.69	NA	88.45	−35.92	CMap score

FIGURE 9

Identification of therapeutic drugs for high-AIDAS patients. **(A)** Spearman correlation of 7 potential therapeutic targets, where red and blue represent positive and negative correlations, respectively. **(B)** Network analysis highlighting the connections between the 7 therapeutic targets and their related drug action pathways. **(C)** AUC values of identified compounds from CTRP database. **(D)** AUC values of identified compounds from PRISM database. **(E)** Analysis from multiple perspectives based on the clinical status, experimental evidence, mRNA expression, and CMap score of 13 compounds. ***P < 0.001.

us to explore how gene expression patterns associated with anoikis resistance influence breast cancer prognosis. Anoikis resistance is a critical step in metastasis, and understanding its molecular underpinnings could provide pathways for intervention in cancer progression. By identifying gene clusters and pathways linked to anoikis resistance, AIDAS deepens our understanding of this biological process and its role in breast cancer outcomes, highlighting potential targets for future therapeutic strategies that could re-sensitize tumor cells to anoikis. This mechanistic insight underscores the value of combining molecular biology with advanced computational techniques to address complex questions in cancer biology.

Immunotherapy is found to be more beneficial for low-AIDAS patients by studying the immune cell infiltration score and immune checkpoint count of patients in two AIDAS subgroups. To effectively examine which patient would be more sensitive to immunotherapy, multiple analyses were utilized, and it was concluded that low-AIDAS populations have greater advantages in the treatment of ICIs, especially in response to PD-1, PD-L1, and CTLA4 drugs. For aggressive subtypes like triple-negative breast cancer (TNBC), which frequently exhibit poor responses to chemotherapy, AIDAS could be a valuable tool for tailoring immunotherapy. By stratifying TNBC patients based on AIDAS and PD-L1 expression, clinicians may be able to identify those more likely to benefit from PD-1-targeted therapies, potentially improving outcomes in this difficult-to-treat population. Enhancing the patient's immune response to tumors by blocking the inhibitory signals of the human anti-tumor response is recognized as the most promising new cancer immunotherapy currently. CTLA-4 and PD-1 are considered two important checkpoints of the immune system, playing a negative regulatory role in the immune response of T cells. *In vivo* mouse experiments indicate that CTLA-4-dependent antibodies bind to Fc receptors rather than blocking the action of CTLA-4/B7, demonstrating the anti-tumor effect of CTLA-4 antibodies (30). Nikhil Joshi stated that PD-1 plays a crucial role in preventing T cells from attacking normal tissues in healthy individuals, and this finding may help look for a way to reduce or prevent the side effects of immunotherapy (31). Our study observed that patients in the low-AIDAS group tend to have lower PD-L1 expression, correlating with a less immunosuppressive tumor microenvironment. This reduced immune suppression may explain their improved responses to PD-1/PD-L1 inhibitors, as these therapies rely on reactivating the immune system to recognize and target cancer cells. Beyond PD-L1 expression levels, differences in the immune cell landscape and functional activity within the tumor microenvironment likely contribute to these divergent responses. Studies have shown that functional characteristics, such as T-cell activation and the presence of regulatory T-cells, can significantly impact immunotherapy effectiveness (32). Techniques such as leukosome isolation and single-cell profiling could further elucidate the immune cell dynamics within AIDAS groups, providing deeper insights into how these functional immune variations drive therapeutic responses.

Chemotherapy plays an important role in the treatment of tumors in the clinic. To study the chemotherapy efficacy among different patients, therapeutic targets and drugs were screened. After a series of analyses, it was found that BC patients with high-AIDAS are more suitable for chemotherapy. Finally, seven therapeutic targets and one drug were identified to improve the prognosis. These studies have demonstrated the effectiveness of methotrexate. For example, Methotrexate chemotherapy can induce the dysregulation of three types of glial cells, which forms the basis for chemotherapy-related cognitive impairment (33). Shen Y et al. reported that patients showed a good prognosis after they received four courses of methotrexate chemotherapy (34). Thomas S et al. believe that methotrexate is a promising drug for treating myeloproliferative tumors (35). Overall, the therapeutic potential of methotrexate has been repeatedly verified.

The genomic alterations identified in high-AIDAS tumors provide a biologically plausible explanation for the poorer prognosis associated with high anoikis resistance. High-AIDAS tumors frequently exhibit amplification of known oncogenes, such as MYC, and deletions in tumor suppressor genes, linking AIDAS with oncogenic pathways that drive tumor progression and therapeutic resistance. These findings add credibility to AIDAS as a prognostic tool, as they align with established mechanisms of cancer progression. Further exploration of these genetic drivers, within the context of AIDAS, could yield new insights into specific molecular targets, particularly for therapies aimed at reversing anoikis resistance.

Compared to other prognostic models, AIDAS offers a unique focus on anoikis-related gene expression patterns, which are particularly relevant in the context of metastasis and therapeutic resistance. Existing models tend to emphasize overall survival predictors or molecular subtypes without specifically addressing the role of anoikis and immune markers in treatment selection. AIDAS fills this gap by providing actionable insights that could directly influence treatment planning, such as recommending chemotherapy for high-AIDAS patients and immunotherapy for low-AIDAS patients. This targeted approach enhances the individualization of breast cancer treatment, which could improve outcomes by reducing unnecessary treatments and optimizing therapeutic choices based on tumor biology.

Despite the potential of AIDAS, several limitations need to be addressed. Firstly, the study's retrospective and observational design restricts our findings to associations, without the ability to infer causality. Prospective studies with standardized, long-term follow-up would be essential to confirm AIDAS's clinical relevance over time. Additionally, our IHC validation was conducted on a limited sample size of 30 tissue samples, which, although consistent with broader dataset findings, may not fully capture population-level heterogeneity. Expanding IHC validation to larger, multi-cohort studies would strengthen the generalizability of our results.

Our study also integrated data from multiple cohorts, each with potential variations in sample processing. Although we applied normalization and batch correction, residual technical variability may influence the findings. Future studies with harmonized, single-cohort data could provide a more uniform validation. Finally, while our bioinformatics analysis identified potential therapeutic targets through in silico drug screening, wet lab validation is essential to confirm these findings. Future research should incorporate *in vitro* and *in vivo* experiments to validate AIDAS-predicted drug responses and explore the efficacy of novel anoikis-targeting therapies.

Furthermore, the integration of AIDAS with PD-L1 expression and other immune markers offers a promising approach for precision oncology. For instance, stratifying TNBC patients by AIDAS and PD-L1 levels could help personalize immunotherapy choices, optimizing patient selection for anti-PD-1/L1 treatments. By combining molecular and immune landscape data, AIDAS represents a step towards fully personalized breast cancer management, offering a comprehensive molecular profile to guide treatment.

AIDAS exemplifies the potential of combining mechanistic understanding with machine learning to advance personalized medicine. By linking anoikis resistance with breast cancer prognosis and therapy response, AIDAS provides an actionable framework for individualized treatment selection in clinical settings. Future studies integrating multi-omics data, single-cell immune profiling, and *in vivo* validation will be crucial to refine AIDAS and maximize its clinical impact. These steps could ultimately lead to new therapeutic avenues, including anoikis-targeting agents and immunotherapy combinations, further expanding the clinical utility of AIDAS in breast cancer care.

Conclusion

In conclusion, this study advocates for a more nuanced understanding of the TME, suggesting that the interrelationships and functional states of different immune components can significantly influence the efficacy of immunotherapy. It underscores the potential of integrating comprehensive immune profiling into clinical decision-making to tailor immunotherapeutic strategies more precisely. The differential response to immunotherapy in breast cancer groups highlights the importance of considering qualitative and functional aspects of immune cells, beyond their numerical abundance. This approach could lead to more personalized and effective therapeutic interventions, particularly in the realm of immunotherapy.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

The studies involving humans were approved by Ethics Committee of Guizhou Provincial People's Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

XG: Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Writing – original draft. JX: Data curation, Formal analysis, Resources, Software, Writing – original draft. YC: Data curation, Methodology, Software, Writing – original draft. WY: Data curation, Formal analysis, Writing – original draft. XS: Investigation, Methodology, Writing – original draft. RM: Conceptualization, Methodology, Validation, Writing – review & editing. TW: Conceptualization, Funding acquisition, Investigation, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This word was supported by the Talent Fund of Guizhou Provincial People's Hospital ((2022)-33), the science and technology development plan of Jilin Province (YDZJ202401432ZYTS) and the Beihua University graduate innovation program ((2024)073).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1491508/full#supplementary-material

References

1. Ling Y, Liang G, Lin Q, Fang X, Luo Q, Cen Y-h, et al. circCDYL2 promotes trastuzumab resistance via sustaining HER2 downstream signaling in breast cancer. *Mol Cancer*. (2022) 21:8. doi: 10.1186/s12943-021-01476-7

2. Paraskevi T. Quality of life outcomes in patients with breast cancer. *Oncol Rev*. (2012) 6:e2. doi: 10.4081/oncol.2012.e2

3. Boughey JC, Hartmann LC, Anderson SS, Degnim AC, Vierkant RA, Reynolds CA, et al. Evaluation of the tyrer-cuzick (International breast cancer intervention study) model for breast cancer risk prediction in women with atypical hyperplasia. *J Clin Oncol*. (2010) 28:3591–6. doi: 10.1200/jco.2010.28.0784

4. Akcakanat A, Zheng X, Cruz Pico CX, Kim T, Chen K, Korkut A, et al. Genomic, transcriptomic, and proteomic profiling of metastatic breast cancer. *Clin Cancer Res*. (2021) 27:3243–3252. doi: 10.1158/1078-0432.CCR-20-4048

5. Mazo C, Aura C, Rahman A, Gallagher WM, Mooney C. Application of artificial intelligence techniques to predict risk of recurrence of breast cancer: A systematic review. *J Personalized Med*. (2022) 12:1496. doi: 10.3390/jpm12091496

6. Kakavandi E, Shahbahrami R, Goudarzi H, Eslami G, Faghihloo E. Anoikis resistance and oncoviruses. *J Cell Biochem*. (2018) 119:2484–91. doi: 10.1002/jcb.v119.3

7. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. (2012) 486:346–52. doi: 10.1038/nature10983

8. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. (2005) 102:15545–50. doi: 10.1073/pnas.0506580102

9. Wang S, Li Z, Hou J, Li X, Ni Q, Wang T. Integrating PANoptosis insights to enhance breast cancer prognosis and therapeutic decision-making. *Front Immunol*. (2024) 15:1359204. doi: 10.3389/fimmu.2024.1359204

10. Wang L, Liu Z, Liang R, Wang W, Zhu R, Li J, et al. Comprehensive machine-learning survival framework develops a consensus model in large-scale multicenter cohorts for pancreatic cancer. *Elife*. (2022) 11:e80150. doi: 10.7554/eLife.80150

11. Pal B, Chen Y, Vaillant F, Capaldo BD, Joyce R, Song X, et al. A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. *EMBO J*. (2021) 40:e107333. doi: 10.15252/embj.2020107333

12. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst*. (2019) 8:329–337.e4. doi: 10.1016/j.cels.2019.03.003

13. Baran Y, Bercovich A, Sebe-Pedros A, Lubling Y, Giladi A, Chomsky E, et al. MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol*. (2019) 20:206. doi: 10.1186/s13059-019-1812-2

14. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*. (2021) 12:1088. doi: 10.1038/s41467-021-21246-9

15. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods*. (2020) 17:159–62. doi: 10.1038/s41592-019-0667-5

16. Zeng D, Ye Z, Shen R, Yu G, Wu J, Xiong Y, et al. IOBR: multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures. *Front Immunol*. (2021) 12:687975. doi: 10.3389/fimmu.2021.687975

17. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. (2016) 17:218. doi: 10.1186/s13059-016-1070-5

18. Racle J, Gfeller D. EPIC: A tool to estimate the proportions of different cell types from bulk gene expression data. *Methods Mol Biol (Clifton NJ)*. (2020) 2120:233–48. doi: 10.1007/978-1-0716-0327-7_17

19. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. (2017) 18:220. doi: 10.1186/s13059-017-1349-1

20. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. (2015) 12:453–7. doi: 10.1038/nmeth.3337

21. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med*. (2019) 11:34. doi: 10.1186/s13073-019-0638-6

22. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res*. (2017) 77: e108–10. doi: 10.1158/0008-5472.Can-17-0307

23. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet*. (2017) 49:1779–84. doi: 10.1038/ng.3984

24. Yang C, Huang X, Li Y, Chen J, Lv Y, Dai S. Prognosis and personalized treatment prediction in TP53-mutant hepatocellular carcinoma: an in silico strategy towards precision oncology. *Brief Bioinform*. (2021) 22:bbaa164. doi: 10.1093/bib/bbaa164

25. Wang T, Li T, Li B, Zhao J, Li Z, Sun M, et al. Immunogenomic landscape in breast cancer reveals immunotherapeutically relevant gene signatures. *Front Immunol*. (2022) 13:805184. doi: 10.3389/fimmu.2022.805184

26. Wang T, Ba X, Zhang X, Zhang N, Wang G, Bai B, et al. Nuclear import of PTPN18 inhibits breast cancer metastasis mediated by MVP and importin β2. *Cell Death Dis*. (2022) 13:720. doi: 10.1038/s41419-022-05167-z

27. Han HJ, Sung JY, Kim S-H, Yun U-J, Kim H, Jang E-J, et al. Fibronectin regulates anoikis resistance via cell aggregate formation. *Cancer Lett*. (2021) 28(508):59-72. doi: 10.1016/j.canlet.2021.03.011

28. Rosner G. Anoikis–a specific form of programmed cell death. *Harefuah*. (2003) 142:857–61.

29. Buchheit CL, Weigel KJ, Schafer ZT. Cancer cell survival during detachment from the ECM: multiple barriers to tumour progression. *Nat Rev Cancer*. (2014) 14:632–41. doi: 10.1038/nrc3789

30. Ingram JR, Blomberg OS, Rashidian M, Ali L, Garforth S, Fedorov E, et al. Anti-CTLA-4 therapy requires an Fc domain for efficacy. *Proc Natl Acad Sci U.S.A*. (2018) 115:3912–7. doi: 10.1073/pnas.1801524115

31. Damo M, Hornick NI, Venkat A, William I, Clulo K, Venkatesan S, et al. PD-1 maintains CD8 T cell tolerance towards cutaneous neoantigens. *Nature*. (2023) 619:151–9. doi: 10.1038/s41586-023-06217-y

32. Shadbad MA, Safaei S, Brunetti O, Derakhshani A, Lotfinejad P, Mokhtarzadeh A, et al. A systematic review on the therapeutic potentiality of PD-L1-inhibiting microRNAs for triple-negative breast cancer: toward single-cell sequencing-guided biomimetic delivery. *Genes (Basel)*. (2021) 12:1206. doi: 10.3390/genes12081206

33. Gibson EM, Nagaraja S, Ocampo A, Tam LT, Wood LS, Pallegar PN, et al. Methotrexate chemotherapy induces persistent tri-glial dysregulation that underlies chemotherapy-related cognitive impairment. *Cell*. (2019) 176:43–55.e13. doi: 10.1016/j.cell.2018.10.049

34. Shen Y, Wan X, Xie X. A metastatic invasive mole arising from iatrogenic uterus perforation. *BMC Cancer*. (2017) 17:876. doi: 10.1186/s12885-017-3904-2

35. Thomas S, Fisher KH, Snowden JA, Danson SJ, Brown S, Zeidler MP. Effect of methotrexate on JAK/STAT pathway activation in myeloproliferative neoplasms. *Lancet*. (2015) 385:385. doi: 10.1016/S0140-6736(15)60413-5

Frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 22 November 2024
DOI 10.3389/fimmu.2024.1489171

OPEN ACCESS

EDITED BY
Tinka Vidovic,
University of Zagreb, Croatia

REVIEWED BY
Nahum Puebla-Osorio,
University of Texas MD Anderson Cancer Center, United States
Irma Olarte,
Hospital General de México Dr. Eduardo Liceaga, Mexico

*CORRESPONDENCE
Xiangyan Feng
fengxiangyan2014@163.com
Yucui Dong
yucuidong@bzmc.edu.cn
Xiao Zhu
xzhu@ytu.edu.cn

RECEIVED 31 August 2024
ACCEPTED 04 November 2024
PUBLISHED 22 November 2024

CITATION
Zhu C, Feng X, Tong L, Mu P, Wang F, Quan W, Dong Y and Zhu X (2024) Prediction of acute myeloid leukemia prognosis based on autophagy features and characterization of its immune microenvironment. *Front. Immunol.* 15:1489171. doi: 10.3389/fimmu.2024.1489171

Prediction of acute myeloid leukemia prognosis based on autophagy features and characterization of its immune microenvironment

Chaoqun Zhu[1], Xiangyan Feng[2]*, Lanxin Tong[3], Peizheng Mu[1], Fei Wang[1], Wei Quan[1], Yucui Dong[4]* and Xiao Zhu[1]*

[1]School of Computer and Control Engineering, Yantai University, Yantai, Shandong, China, [2]Department of Hematology, Yantai Yuhuangding Hospital Affiliated to Qingdao University, Yantai, Shandong, China, [3]Guangzhou Dublin International College of Life Sciences and Technology, South China Agricultural University, Guangzhou, Guangdong, China, [4]Department of Immunology, Binzhou Medical University, Yantai, Shandong, China

Background: Autophagy promotes the survival of acute myeloid leukemia (AML) cells by removing damaged organelles and proteins and protecting them from stress-induced apoptosis. Although many studies have identified candidate autophagy genes associated with AML prognosis, there are still great challenges in predicting the survival prognosis of AML patients. Therefore, it is necessary to identify more novel autophagy gene markers to improve the prognosis of AML by utilizing information at the molecular level.

Methods: In this study, the Random Forest, SVM and XGBoost algorithms were utilized to identify autophagy genes linked to prognosis, respectively. Subsequently, six autophagy genes (TSC2, CALCOCO2, BAG3, UBQLN4, ULK1 and DAPK1) that were significantly associated with patients' overall survival (OS) were identified using Lasso-Cox regression analysis. A prediction model incorporating these autophagy genes was then developed. In addition, the immunological microenvironment analysis of autophagy genes was performed in this study.

Results: The experimental results showed that the predictive model had good predictive ability. After adjusting for clinicopathologic parameters, this feature proved an independent prognostic predictor and was validated in an external AML sample set. Analysis of differentially expressed genes in patients in the high-risk and low-risk groups showed that these genes were enriched in immune-related pathways such as humoral immune response, T cell differentiation in thymus and lymphocyte differentiation. Then immune infiltration analysis of autophagy genes in patients showed that the cellular abundance of T cells CD4+ memory activated, NK cells activated and T cells CD4+ in the high-risk group was significantly lower than that in the low-risk group.

Conclusion: This study systematically analyzed autophagy-related genes (ARGs) and developed prognostic predictors related to OS for patients with AML, thus more accurately assessing the prognosis of AML patients. This not only helps to improve the prognostic assessment and therapeutic outcome of patients, but may also provide new help for future research and clinical applications.

KEYWORDS

autophagy gene, immune infiltration, random forest, acute myeloid leukemia, prognosis

Introduction

Acute myeloid leukemia (AML) is a complex and diverse blood cancer triggered by abnormal proliferation and immature differentiation of hematopoietic stem cells in the bone marrow (1–3). Despite previous studies on the role of autophagy in AML (4–6), the specific functions of ARGs and their interaction with immune infiltration have not been thoroughly explored. This gap not only makes the biological functions of ARGs unclear, but also limits their potential application in AML therapy. Therefore, this study aimed to reveal the key autophagy genes associated with the prognosis of AML and their role in relation to the immune microenvironment through comprehensive bioinformatics analysis, providing new targets and strategies for AML treatment.

Autophagy is an important cellular self-regulatory mechanism that maintains cellular and organismal homeostasis (7) and adapts to changes in the external environment by disassembling and removing damaged proteins and organelles inside the cell. Autophagy gene (ARG) mutations linked to cancer and other diseases (8). For example, autophagy levels are strongly associated with the prognosis of ovarian cancer patients (9). Recent studies have indicated that autophagy is closely linked to progression of leukemia, including AML. However, the exact mechanism of autophagy in AML and the expression and function of autophagy genes in AML are still limited.

Certain immune cells play an immunoregulatory role in the tumor microenvironment (TME) and are linked to the immune escape of tumor cells, thereby influencing tumor progression (10). Bansal et al. showed that the number of regulatory T cells was significantly higher in patients with AML than in the healthy population, and that the increased number of Tregs may be strongly associated with poor prognosis (11). Wan et al. further noted that Tregs contribute to immune escape of AML cells in the tumor microenvironment by enhancing the inhibitory effect on effector T cells (12). The study by Romee et al. demonstrates the potential of using cytokines to induce memory-like NK cells for immunotherapy in AML patients (13). Bioinformatics analysis of immune infiltration is a powerful approach that allows in-depth study of immune cell infiltration in TME and its relationship with tumor development by integrating multi-omics data. Although there have been several studies on immune infiltration in AML, the interaction between ARGs and immune infiltration has not been thoroughly investigated.

In this research, AML transcriptome data obtained from the GEO database was used to screen for AML-related ARGs (14–16). Then functional enrichment analyses were conducted to obtain the biological meaning and functional features of these ARGs. In addition, the autophagy genes obtained in this experiment were analyzed by protein–protein interaction (PPI) network to obtain the interactions between these autophagy genes and their regulatory mechanisms inside the cell. After that, Random Forest (17), SVM-RFE (18) and XGBoost were used in combination to identify key autophagy genes associated with AML. Lasso-Cox analysis was then conducted to identify six autophagy-related genes and construct a survival prediction model. After that, AML high and low risk groups divided according to the survival prediction model and differential expression analysis was performed. The genes obtained with significant differences were then analyzed for functional enrichment. The results indicated that these ARGs were primarily enriched in immune-related pathways such as T cell differentiation in thymus and lymphocyte differentiation. Accordingly, the autophagy genes were analyzed for immune infiltration. Moreover, the link between ARGs and immune infiltration was investigated. This study reveals the critical role of autophagy genes in acute myeloid leukemia and their interaction with the immune microenvironment, which is of great clinical significance. By constructing a survival prediction model, it can provide AML patients with prognostic assessment and personalized treatment plans. In addition, autophagy genes are expected to be used as potential targets for novel therapeutic strategies, especially showing great potential in combination with immunotherapy. The basic flow of this experiment is shown in Figure 1.

Methods

Data set acquisition

In this study, the original microarray dataset of GSE37642 (19) was downloaded from the GEO database, including transcriptome

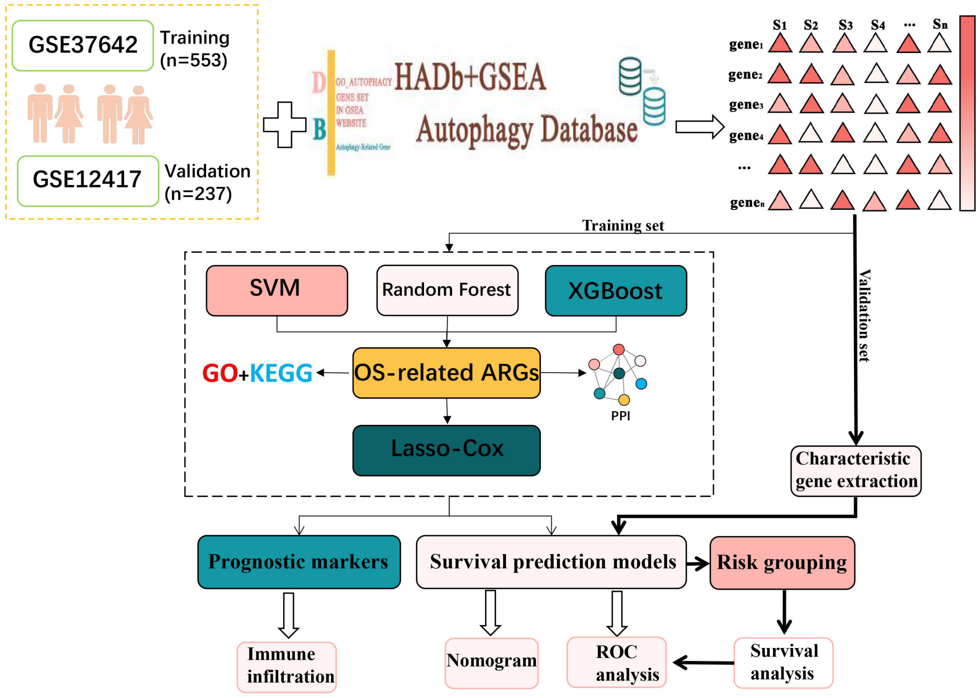


FIGURE 1
The general analytical flow of this experimental design.

data of GPL96 and GPL570 platforms. The data were first quality checked for missing values, outliers and distribution. Subsequently, the data were normalized using the robust multi-array average (RMA) algorithm in the affy package to normalize gene expression levels across arrays. To eliminate the batch effect due to different platforms, the Combat algorithm from the sva package was used for correction (20). Clinical information was then collated and integrated to remove samples lacking relevant clinical information, resulting in 553 usable acute myeloid leukemia samples. Clinical information on these samples is presented in Supplementary Table S1. The dataset GSE12417 was processed in the same way, resulting in a total of 237 samples with clinical information. The available clinical information for the samples used was shown in Supplementary Table S2. AML RNA-seq datasets were downloaded from the UCSC Xena database (https://xenabrowser.net/datapages/). Available clinical information for the samples used in this study is shown in Supplementary Table S3.

Acquisition of autophagy genes

Autophagy-related genes were obtained from the Human Autophagy Database (HADB, http://www.autophagy.lu/index.html) and from the GO_AUTOPHAGY gene set in GSEA website (http://software.broadinstitute.org/gsea/index.jsp). The Human Autophagy Database (HADb) is an authoritative database dedicated to autophagy-related genes, covering a large number of experimentally validated autophagy genes, which ensures the breadth and comprehensiveness of the data. The collection of GO_AUTOPHAGY genes on the GSEA website is based on the autophagy biological process as defined by Gene Ontology (GO), and these genes are strictly classified according to the GO classification criteria division, ensuring consistency in biological function and annotation. This enables the autophagy genes selected during the study to have a clear functional orientation and ensures their relevance to the autophagy process. The two obtained autophagy gene sets were combined to obtain 531 related ARGs (Supplementary Table S4). 392 ARGs were screened from GSE37642.

Random forest identifies overall survival-related ARGs

In this study, survival time and survival state information were extracted from AML patient data, and a random forest model with 1000 decision trees was constructed to predict patient survival. The model used multiple samples, each containing feature genes and their corresponding survival information. To build the decision trees, the random forest employed the log-rank split rule, which assessed the survival differences between two subsets. At each candidate split point, the log-rank statistic was calculated to measure the difference between two survival curves, using the formula $X^2 = \sum_{i=0}^{m} \frac{(O_i - E_i)}{E_i}$, where O_i was the observed number of events at time point i, E_i was the expected number of events at i, and m was the total number of time points. The split point with the highest log-rank statistic is selected as the optimal point, as it maximizes the distinction between the survival curves of the resulting subsets. This process continues recursively, splitting the data at the best split points until a stopping condition is met.

SVM identifies ARGs

In this study, a SVM model was used to identify the most important features for the classification task. The importance of each feature was determined by looking at how much influence it had on the model's decisions. This was done by calculating the absolute value of the product of the feature's weight and the corresponding support vector. In simpler terms, the importance of a feature depends on how much its weight, when multiplied by the support vector, affects the classification. After calculating these importance scores, they were sorted from highest to lowest to identify the most important features.

XGBoost Identifies ARGs

Firstly, the training data were preprocessed, including extracting the autophagy gene expression data from the transcriptome data, and also collecting clinical information such as the survival time and survival status of the patients. Handle missing and abnormal values to ensure complete autophagy gene expression data for each sample and remove abnormal or incomplete samples. Generate labels by combining survival time and survival status. Next, the data are converted to DMatrix format for XGBoost and the model parameters are set, where the objective function is Cox proportional risk model and the evaluation metric is negative log likelihood. The objective function of Cox proportional risk model (21) is defined as:

$$\boldsymbol{logL(\beta)} = \sum\nolimits_{i\in E}(\boldsymbol{x_i\beta} - \mathbf{log}(\sum\nolimits_{j\in R(T_i)}\mathbf{exp}(\boldsymbol{x_j\beta}))) \tag{1}$$

where E denotes the set of events, i.e., all samples of observed deaths, $R(T_i)$ denotes the set of samples at risk at time T_i. x_i dentes the eigenvector of sample i, and β is a parameter of the model. The model is trained through 100 rounds of iterations, setting the learning rate to 0.1, and recording the negative log-likelihood value and training error for each round as a function of the number of iterations. The model is then used to calculate the importance of the features. Feature importance (22) (Gain) indicates the contribution of each feature to the model with the following formula:

$$\boldsymbol{Gain(j)} = \sum\nolimits_{t\epsilon T_j}\boldsymbol{\Delta G_t} \tag{2}$$

where ΔG_t denotes the gain of feature j in tree t and T_j denotes the set of all trees in which feature j appears.

Permutation test

To further assess the impact of the identified ARGs on survival, a permutation test was conducted. This test aims to verify the reliability of the model's predictions by randomly shuffling the survival labels, as described below.

1. Randomly disrupt survival state labels to generate a new set of labels.
2. Retrain the Cox regression model using the disrupted data and record the C-index of the model each time.
3. Repeat the above process a certain number of times to generate a C-index replacement distribution.
4. The C-index of the original model was compared with the C-index distribution of the replacement and the p-value was calculated to assess the significance of the original model.

Functional enrichment analysis and PPI molecular interactions

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses of key autophagy genes were performed using clusterProfiler (version 3.14.3) to reveal the primary functions of these genes. We will apply the Benjamini & Hochberg correction method and use a corrected P value of less than 0.05 as the criterion for statistical significance.

To study the interactions between these key ARGs, a PPI will be constructed using the STRING database. Subsequently, the MCODE plugin was used in Cytoscape (v3.10.0) (23) to extract densely connected modules with default parameters "degree cutoff = 2", "node score cutoff = 0.2", "K-core = 2", and "Maximum depth = 100" to extract densely connected modules.

Construction and validation of survival prediction models

To avoid overfitting of prognostic risk features, we performed the following steps on the training set to construct survival prediction models.

1. A Cox regression method based on the least absolute shrinkage and selection operator (LASSO) was applied to the training dataset to identify significant features of ARGs associated with OS.
2. Subsequently, we performed multivariate Cox proportional risk regression on these candidate genes and stepwise variable selection using the Akaike information criterion.
3. Ultimately, risk scores for optimized prognostic markers were calculated.

$$\boldsymbol{Risk\ score} = \sum\nolimits_i^n \boldsymbol{Coef_i} \times \boldsymbol{A_i} \tag{3}$$

where $Coef_i$ represents the regression coefficient of the i gene, indicating the degree of influence of the expression level of the gene on the risk. A_i denotes the expression level of the i gene, and n denotes the total number of genes selected for characterization.

Differences in patient OS were assessed by Kaplan-Meier analysis and log-rank tests. The predictive power of ARG-based characteristics was assessed using time-dependent ROC curves (24).

To test the accuracy of the survival prediction model, external validation was performed using the GSE12417 (*n*=242) dataset and AML cohorts-TCGA-LAML (*n*=129). First, the risk scores of patients in each external validation dataset were calculated using the survival prediction model from the training set. Then, patients were categorized into high-risk and low-risk groups based on their risk scores. Next, the survival distribution of the model in the high- and low-risk groups was assessed using Kaplan-Meier curves, and the survival differences were compared to validate the predictive performance of the model.

Identification of differentially expressed genes

Differential expression analysis was performed on samples from the high-risk and low-risk groups using the limma package, setting the criteria of |log2FC| > 2 and a P-value < 0.05 to screen for DEGs. Next, volcano maps of DEGs were plotted using the EnhancedVolcano (25) function in the EnhancedVolcano package.

Immune infiltration analysis

The analysis of 22 immune cell types is of great importance during the progression of AML. These immune cells, including T cells, B cells, NK cells, T cells gamma delta and macrophages, are known to play a key immunomodulatory role in the tumor microenvironment (26). Ge Jiang et al. demonstrated that a significant elevation in the abundance of NK cells and macrophage infiltration was strongly associated with a poor prognosis in AML (27). Another study by Moore et al. demonstrated that macrophage reduction promoted AML cell growth *in vivo* (28).

To further investigate the relationship between immune cell infiltration and AML, the CIBERSORT algorithm was used to calculate the infiltration abundance of 22 immune cell types in gene expression data from AML patients. Subsequently, the association between hub genes and the abundance of 22 immune cells was detected and then visualized using the software package "ggcorrplot", and gene-immune cell correlations greater than 0.28 were considered significant.

Results

Using machine learning to select OS-related ARGs

Three hundred and ninety-two ARGs were screened from the gene expression matrix and screened for autophagy genes associated with survival prognosis using Random Forest, Support Vector Machine (SVM) and XGBoost (29) algorithms, respectively.

First, in the random forest model, 1000 decision trees were constructed and the variables were partitioned using the log-rank rule. The model assessed the relationship between gene expression and survival prognosis by calculating the importance of each variable and the proximity of the samples (30). The OBB error plot of the model showed a gradual decrease in error and improved performance as the number of trees increased (Figure 2A). The variable importance plot showed the importance of each gene (Figure 2B), and 146 genes with significant effects on survival analysis were screened (Supplementary Table S5). Meanwhile, the performance of the model was assessed by the C-index (consistency index), and a C-index value of 0.88 was obtained, indicating that the model was predicted relatively well.

Next, the XGBoost algorithm was employed for survival analysis. XGBoost used the Cox proportional risk model as the objective function and evaluated the model by optimizing the Cox negative log-likelihood ratio (cox-nloglik). Survival states and survival times were converted into a labelled format suitable for the Cox model, and the number of iterations of the model was set to 100 with a learning rate of 0.1. Figure 2C shows the trend of Cox negative log-likelihood value during the training process. From the figure, it can be seen that the model gradually converges and the performance of the model gradually improves as the number of iterations increases. The top 180 genes that had a significant effect on survival analysis were screened by feature importance analysis (Supplementary Table S6) and the performance of the model was assessed with a C-index of 0.99. The top 10 ranked important features are shown in Table 1. Figure 2D visualizes the top 10 ranked genes and their corresponding importance scores. These features had the highest importance scores in the model and significantly influenced survival prediction.

In addition, a support vector machine (SVM) was used for survival analysis, and a linear kernel function (31) and epsilon regression type were used for model training. The coefficients and support vectors of the model were used to calculate the importance scores of each feature, and the top 180 feature genes that had a significant effect on survival prediction were filtered out (Supplementary Table S7), and the top 10 features with the highest importance scores were visualized by bar graphs to show the importance scores of these feature genes (Figure 2E). The model has a C-index of 0.17. Table 2 demonstrates the top 10 significant feature genes and their importance scores. A total of 45 overlapping genes common to all three algorithms were screened by the above algorithm (32) (Figure 2F).

By comparing the importance scores of the top 10 genes screened by the three algorithms (Supplementary Figure 1), it was found that genes such as ITGB1, ANXA7, and ULK1 scored higher across all algorithms, suggesting a significant association of these genes with survival prognosis in AML. In model performance comparisons, XGBoost showed the best performance, while SVM performed relatively poorly. However, although XGBoost leads in prediction, it is too dependent on parameter tuning in the case of

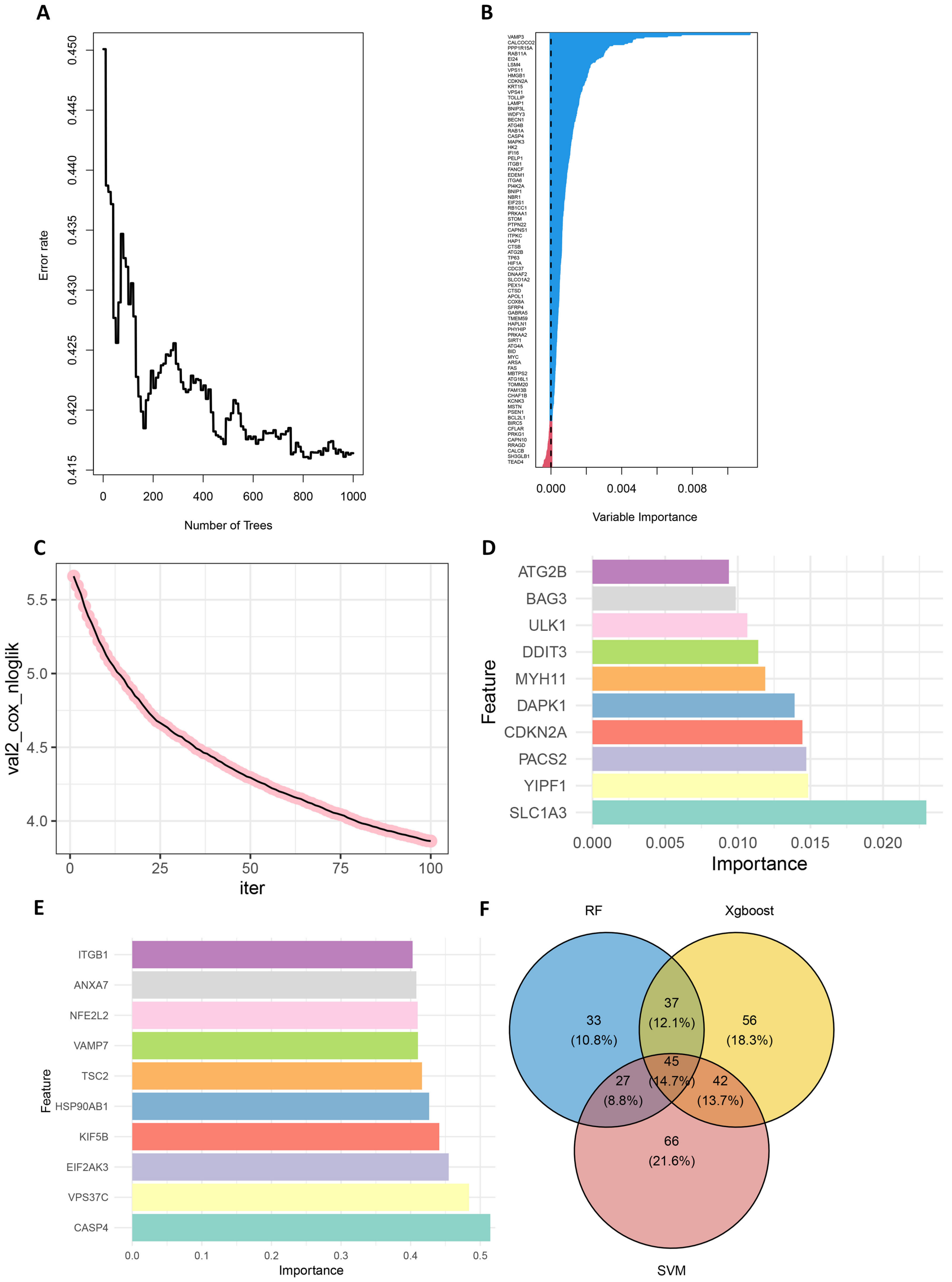


FIGURE 2
Screening for prognostically relevant autophagy genes using machine learning methods. **(A)** The OBB error plot of the random forest algorithm is used to estimate the generalization performance of the model. The graph shows that as the number of trees increases, the errors of model become smaller. **(B)** VIMP plot showing the importance scores of each variable to help identify the most important feature genes. **(C)** Plot of the number of iterations of the training process of the XGBoost algorithm versus the Cox negative log-likelihood value. **(D)** Bar chart of the top 10 genes and their corresponding importance scores screened by the XGBoost algorithm **(E)** Bar chart of the top 10 genes and their corresponding importance scores screened by the SVM algorithm. **(F)** Venn plots of overlapping genes shared by SVM, Random Forest and XGBoost.

small samples and is prone to overfitting if the parameters are not adjusted properly. SVM, on the other hand, is more suitable for handling high-dimensional data with small samples, and although its overall performance is not as good as that of XGBoost, it has a unique advantage in handling data dimensions.

In order to improve the stability and consistency of the screened genes, we adopted a combination strategy of multiple algorithms. By using SVM, Random Forest and XGBoost algorithms to identify prognostic genes from different angles, we further screened the overlapping genes that showed significance in all three algorithms. Finally, we screened 45 overlapping genes in total (Figure 2F).

To further validate the impact of the screened ARGs on the survival prognosis of AML patients, we used the replacement test to

TABLE 1 The top 10 important genes of XGBoost and their importance scores.

ID	Feature	Gain
1	SLC1A3	0.022978235
2	YIPF1	0.014842549
3	PACS2	0.014716427
4	CDKN2A	0.014447571
5	DAPK1	0.013909943
6	MYH11	0.011887605
7	DDIT3	0.011406339
8	ULK1	0.010654985
9	BAG3	0.00985662
10	ATG2B	0.009385788

assess their statistical significance. The results showed that the C-index of the original model was significantly higher than that of most of the replacement models, and was located at the rightmost end of the replacement distribution (Supplementary Figure 2). By comparing the C-index of the original model with that of the replacement models, a p-value of 0.0429 was calculated, indicating that the original model was statistically significant in predicting the survival of AML patients, further confirming the importance of the screened ARGs in survival prediction.

Enrichment analysis of ARGs

In order to better study the biological features in the autophagy gene data so as to understand the functions and regulatory mechanisms of the biological systems, GO and KEGG analyses were conducted. For GO enrichment analysis of autophagy genes, the genes related to total survival were analyzed in terms of biological processes (BP), cellular components (CC), and molecular functions (MF), respectively. BP analysis revealed that these genes were primarily associated with cytolytic metabolic processes, autophagy, and the regulation of processes that utilize autophagic mechanisms (Figure 3A). CC analysis indicated that these ARGs were predominantly distributed in cellular components of vesicle, cytoplasmic vesicle and bounding membrane of organelle (Figure 3B). MF analysis showed that most of these genes act together on a protein and enzyme with catalytic effects (Figure 3C). KEGG revealed that these ARGs were primarily enriched in the pathways of autophagy animal, AMPK signaling and longevity regulation in animals (Figure 3D). To gain insight into the interactions between these autophagy genes associated with overall survival, STRING (33) was utilized to construct the PPI network and identify two important modules: the HSP90AB1 module and the BECN1 module (Figure 3E). The BECN1 module contains 12 nodes and 29 edges, while the CASP3 module consists of 4 nodes and 6 edges. HSPA5, VDAC1, and BAG3 are the other 3 nodes of the CASP3 module. These ARGs may be important for the pathogenesis of AML.

TABLE 2 SVM top 10 significant genes and their importance scores.

ID	Feature	Gain
1	CASP4	0.514479265
2	VPS37C	0.48390987
3	EIF2AK3	0.454698135
4	KIF5B	0.441155029
5	HSP90AB1	0.426529294
6	TSC2	0.416280361
7	VAMP7	0.410471634
8	NFE2L2	0.410017974
9	ANXA7	0.40822591
10	ITGB1	0.40279245

Modelling survival predictions

In this study, survival data were systematically analyzed, and feature genes significantly associated with survival were screened by Lasso-Cox regression and used for modelling. First, the optimal lambda value (lambda.1se) of 0.09393562 was selected by 10-fold cross-validation, and the lambda plot and LASSO regression were plotted (Figures 4A, B). Next, the non-zero coefficients were extracted and the six characterized autophagy genes and their regression coefficients selected by LASSO were saved. Cox stepwise regression (34) analysis was then conducted to optimize the selection of feature genes (Table 3). The resulting risk score model for the patients was as follows:

$$\begin{aligned}\boldsymbol{Risk\ score} &= (\mathbf{0.14747} \times \mathbf{BAG3}) - (\mathbf{0.14437} \times \boldsymbol{TSC2}) \\ &\quad - (\mathbf{0.32652} \times \mathbf{CALCOCO2}) + (\mathbf{0.32410} \\ &\quad \times \mathbf{UBQLN4}) - (\mathbf{0.24254} \times \mathbf{ULK1}) + (\mathbf{0.23913} \\ &\quad \times \mathbf{DAPK1})\end{aligned} \tag{4}$$

Risk scores were subsequently calculated for each sample, and the samples were divided into high and low risk groups based on the median risk score. An increase in the risk score was correlated with a higher number of patient deaths (Figures 4C, D). Among the characterized genes screened, DAPK1, UBQLN4, and BAG3 were highly expressed in high risk, and ULK1, ALCOCO2, and TSC2 were highly expressed in low risk (Figure 4E).

To assess the difference in survival time, the Kaplan–Meier survival curves were used (35). The results showed that patients in the high-risk group had a shorter OS than those in the low-risk group (P< 0.0001, Figure 5A). The accuracy of the constructed survival prediction model was evaluated, and the results showed that the AUCs of 1-year, 3-year, and 5-year OS were 0.660, 0.733,

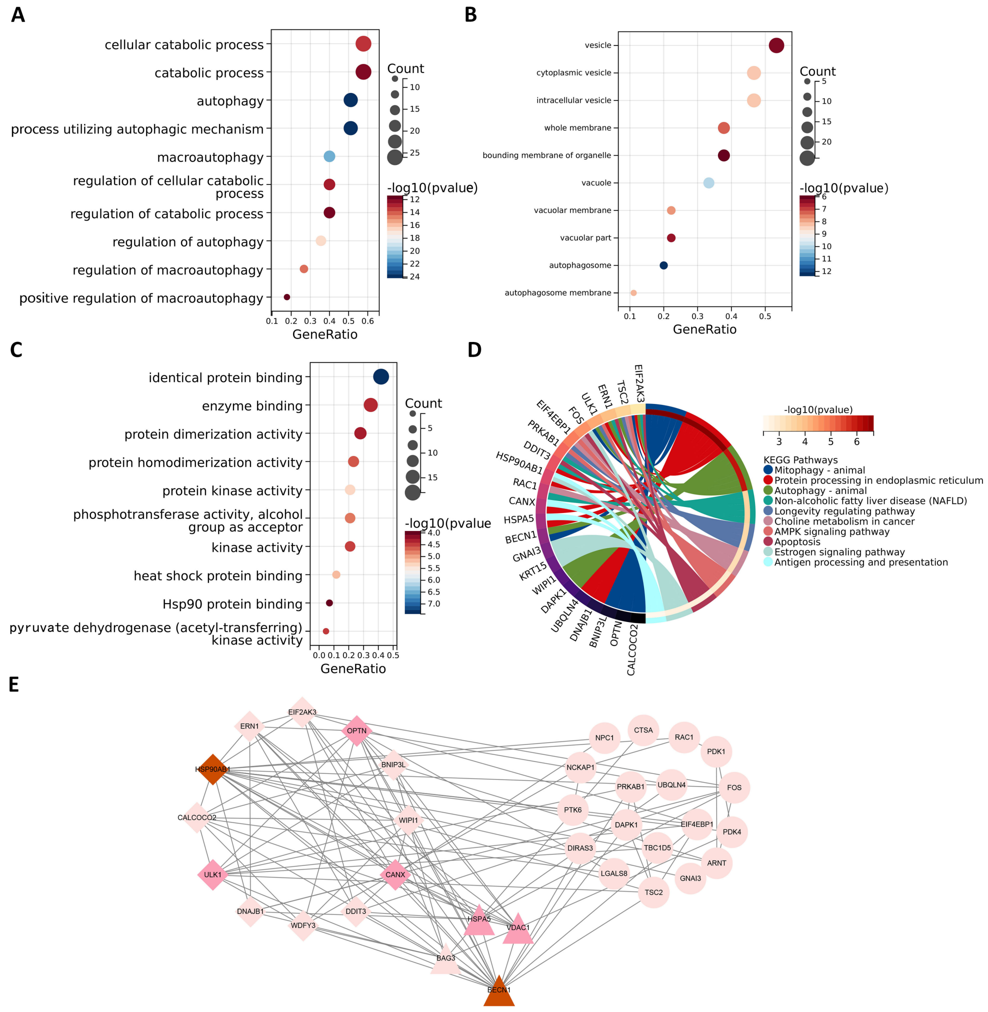


FIGURE 3

Functional enrichment analysis and PPI analysis of ARGs associated with survival. The gene functional enrichment analysis of key modules **(A)** BP; **(B)** CC; **(C)** MF; **(D)** KEGG pathways. **(E)** HSP90AB1 and BECN1 modules were identified by PPI analysis of ARGs. Darker node colors represent greater node degrees.

and 0.739, respectively (Figure 5B), which indicated that the survival prediction model constructed by using the prognostic genes screened in this experiment had high predictive ability.

To quantify the relative importance of the screened autophagy genes in the survival prediction model, we employed a game theory-based SHAP value (SHapley Additive exPlanations) technique. By using the SHAP values calculated by the iml package, we provided a quantitative relative importance score for each gene. Analysis of each gene in the model by SHAP value visually demonstrates the contribution of these genes to the prediction of AML survival (Supplementary Figure 3). The average contribution of each gene in the model to the prediction is summarized in Table 4. As shown in Table 4, these autophagy genes have high contribution values in the model, further supporting their key role in AML survival prognosis.

Univariate Cox regression analysis (UCRA) and multivariate Cox regression analysis (MCRA) were conducted to validate the independence of prognosis-related autophagy gene survival prediction. UCRA revealed that age, runx1 mutation, and risk score were significantly associated with patients' OS (Figure 5C). MCRA indicated that age and risk score were independent predictors for AML patients, respectively (Figure 5D).

To more precisely evaluate the survival prediction model's effectiveness, nomogram plot integrating risk scores and other survival information was constructed. (Figure 5E) The calibration curves demonstrated accurate predictions OS in AML patients (Figures 5F–H). This suggests that that integrating our risk score with clinical information can enhance the prediction of OS.

External validation set validation of survival prediction models

This study evaluated the diagnostic performance of the models in two external independent validation groups, GSE12417 and TCGA-LAML. Comparison of OS using Kaplan-Meier curves (36) and the log-rank test revealed that in the GSE12417 group, patients in the high-risk group had significantly shorter OS

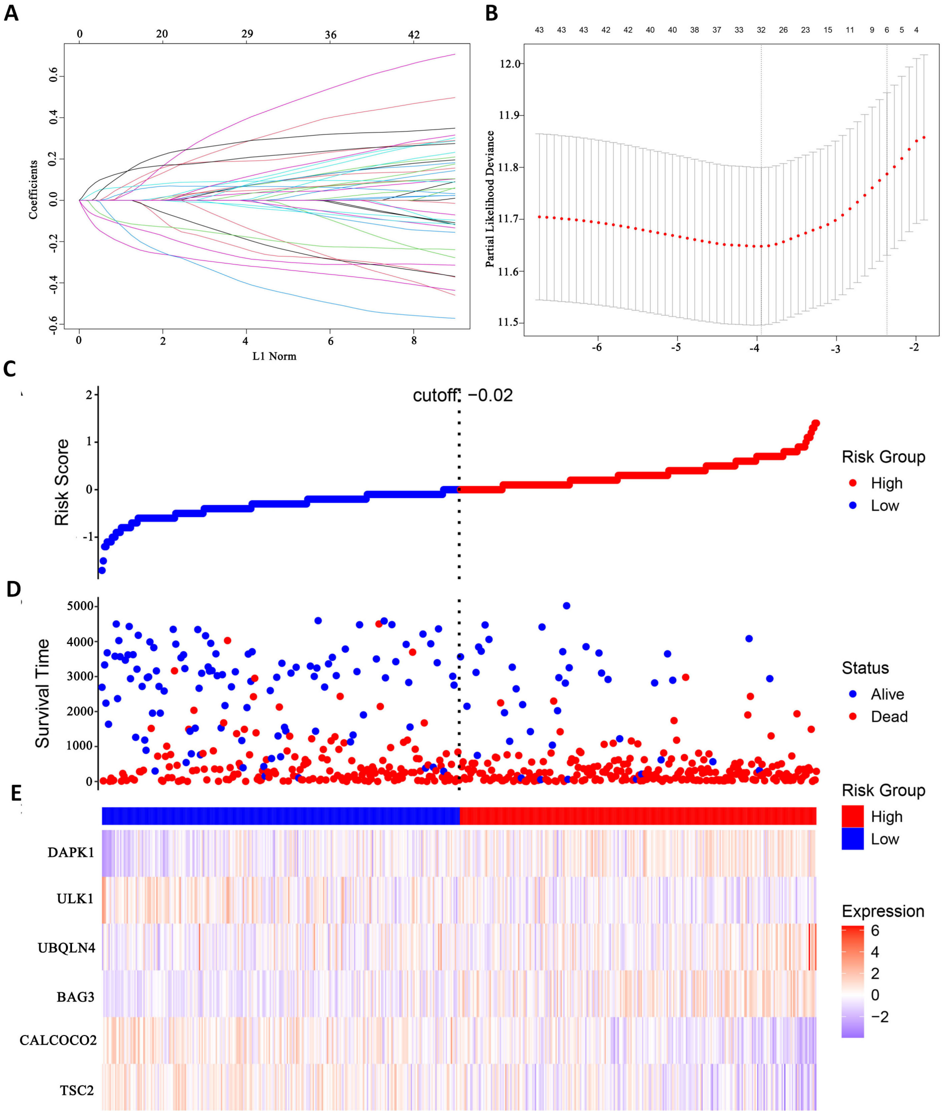


FIGURE 4
Identify autophagy genes associated with survival. **(A)** Path diagram of LASSO coefficients. **(B)** Cross-validation curve for LASSO regression analysis. **(C)** Change curves of patient risk scores. **(D)** The number of patients corresponding to different survival times. **(E)** Expression of six model genes.

compared to those in the low-risk group (P<0.0001, Figure 6A). Similarly, in the TCGA-LAML group, the prognosis of patients in the high-risk group was significantly worse than that in the low-risk group (P=0.015, Figure 6B).

To further evaluate the classification performance of the model for patient survival on different datasets, ROC curves for patient survival were plotted based on the model risk score. In the GSE12417 group, the area under the curve (AUC) for 1-year and 3-year OS was 0.633 and 0.651, respectively (Figure 6C). In the TCGA-LAML group, the AUC values for 1-, 3- and 5-year OS were 0.632, 0.612 and 0.704, respectively (Figure 6D). These results demonstrated the strong predictive power of the model in predicting survival in AML patients. Additionally, this study analyzed the distribution of patients' risk scores and OS, and found that the mortality rate in the high-risk group was higher than that in the low-risk group. In terms of gene expression, the validation group showed that DAPK1, UBQLN4, and BAG3 were significantly up-regulated in the high-risk group, whereas ULK1,

TABLE 3 Survival prediction models for acute myeloid leukemia.

Gene	Coefficients	Exp(coef)	P-value
TSC2	-0.14437	0.86556	0.128973
CALCOCO2	-0.32652	0.72143	0.000299
BAG3	0.14747	1.15890	3.70e-05
UBQLN4	0.32410	1.38278	0.009121
ULK1	-0.24254	0.78463	0.020660
DAPK1	0.23913	1.27015	3.26e-06

Coefficients was the regression coefficient for each variable, indicating the direction and magnitude of the variable's effect on survival time. Se(coef) was the standard error of the regression coefficient for each variable, indicating the uncertainty in the estimation.

ALCOCO2, and TSC2 were significantly down-regulated in the low-risk group (Figures 6E, F), which was consistent with the risk score calculation. Overall, the validation results indicated that the proportional risk model has reasonable accuracy and discriminative ability for independently predicting OS in AML patients.

Identification and enrichment of DEGs

Differential expression analysis of transcriptome data from patients in the high- and low-risk groups using the limma package identified 63 DEGs, including 47 up-regulated genes and 16 down-regulated genes (Figure 7A). The expression patterns of the differential genes are shown in Figure 7B. GO enrichment analysis revealed that these DEGs were mainly associated with BP such as T cell differentiation in thymus and lymphocyte differentiation. In terms of cellular components, these genes are predominantly found in the tertiary granule lumen, actin filament bundle, and platelet alpha granule. They are involved in molecular functions such as chemokine activity and cytokine receptor binding (Figure 7C).

KEGG pathway analysis indicated that these DEGs were primarily enriched in the IL-17 signaling pathway and Th1 and Th2 cell differentiation (Figure 7D). High and low risk group differential genes enriched in lymphocyte differentiation, humoral

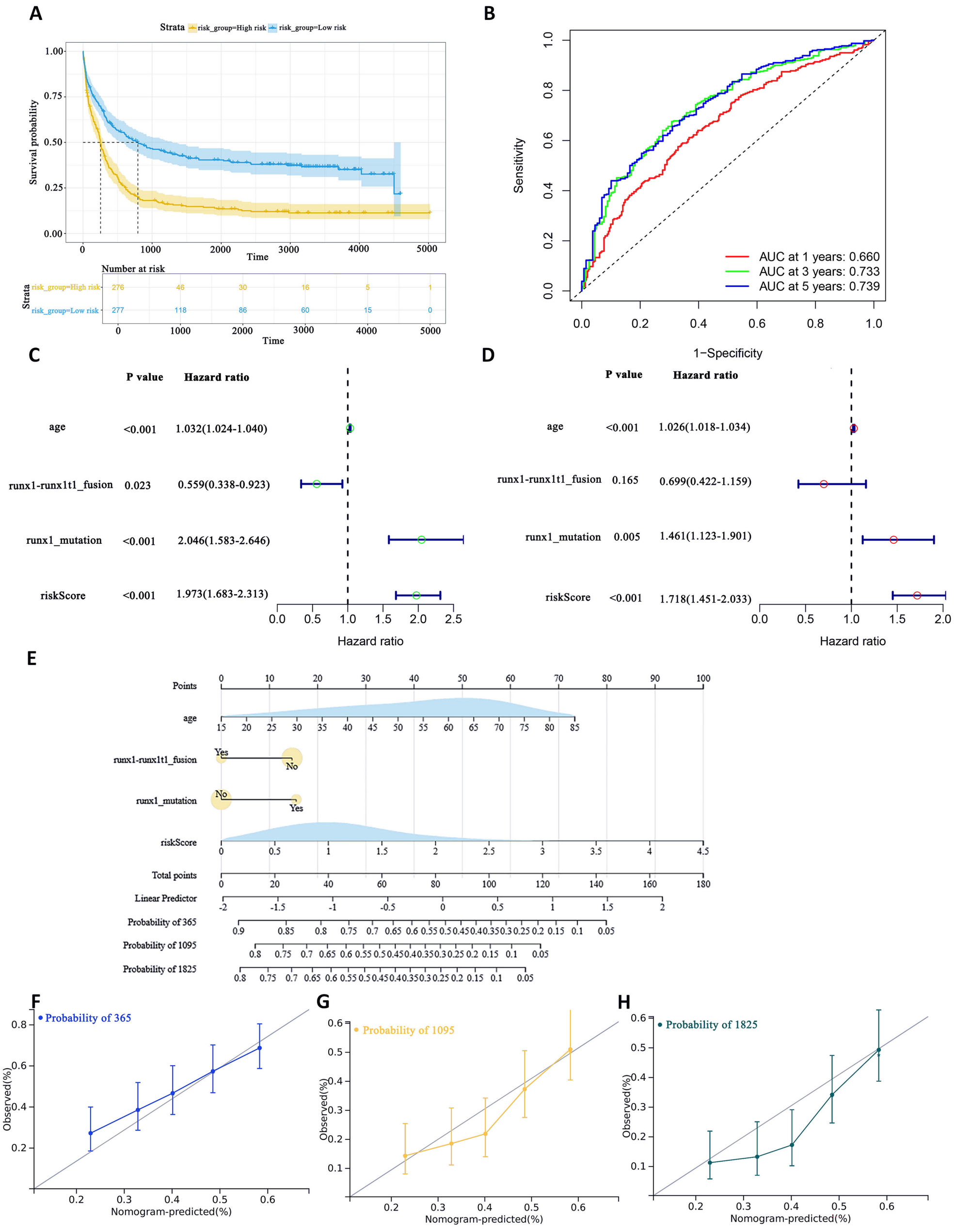


FIGURE 5
To assess the predictive accuracy of survival prediction models for patient OS. **(A)** Kaplan–Meier curves visualizing the difference in survival time. **(B)** AUC curves for prognostic markers. **(C)** UCRA **(D)** MCRA **(E)** Development of autophagic clinicopathological nomograms for the prediction of OS in AML patients by combining risk scores and clinical information. **(F–H)** Calibration curves-predicting 1-, 3-, and 5-year survival in AML patients. Solid lines indicate ideal performance.

TABLE 4 Relative importance ranking of autophagy genes based on SHAP values.

ID	Feature	Mean(|SHAP|)
1	BAG3	0.23646
2	TSC2	0.12995
3	UBQLN4	0.04845
4	DAPK1	0.04821
5	ULK1	0.01969
6	CALCOCO2	0.01892

Mean(|SHAP|) denotes the mean of the absolute value of the SHAP value for the gene or trait, i.e., the mean of the gene's contribution to the significance predicted by the model.

immune response and T cell differentiation in thymus associated with immune GO terms were TPD52, ZFP36L1 and GATA3 (Figure 7E). Figure 7F shows DEGs enriched in KEGG pathways such as IL-17 signaling pathway and so on.

In addition to these genes such as BAG3, DAPK1 and GATA3 are enriched in multiple other GO pathways (Supplementary Figure 4), and genes such as CXCL2, CXCL3 and CYP1B1 are also present in multiple other KEGG pathways (Supplementary Figure 5). This suggests that these genes play important roles in biological processes. In addition, by analyzing the relationships between the enriched pathways, the GO term network relationship map showed significant correlations between chemokine receptors and term such as activity, humoral immune response, and myeloid

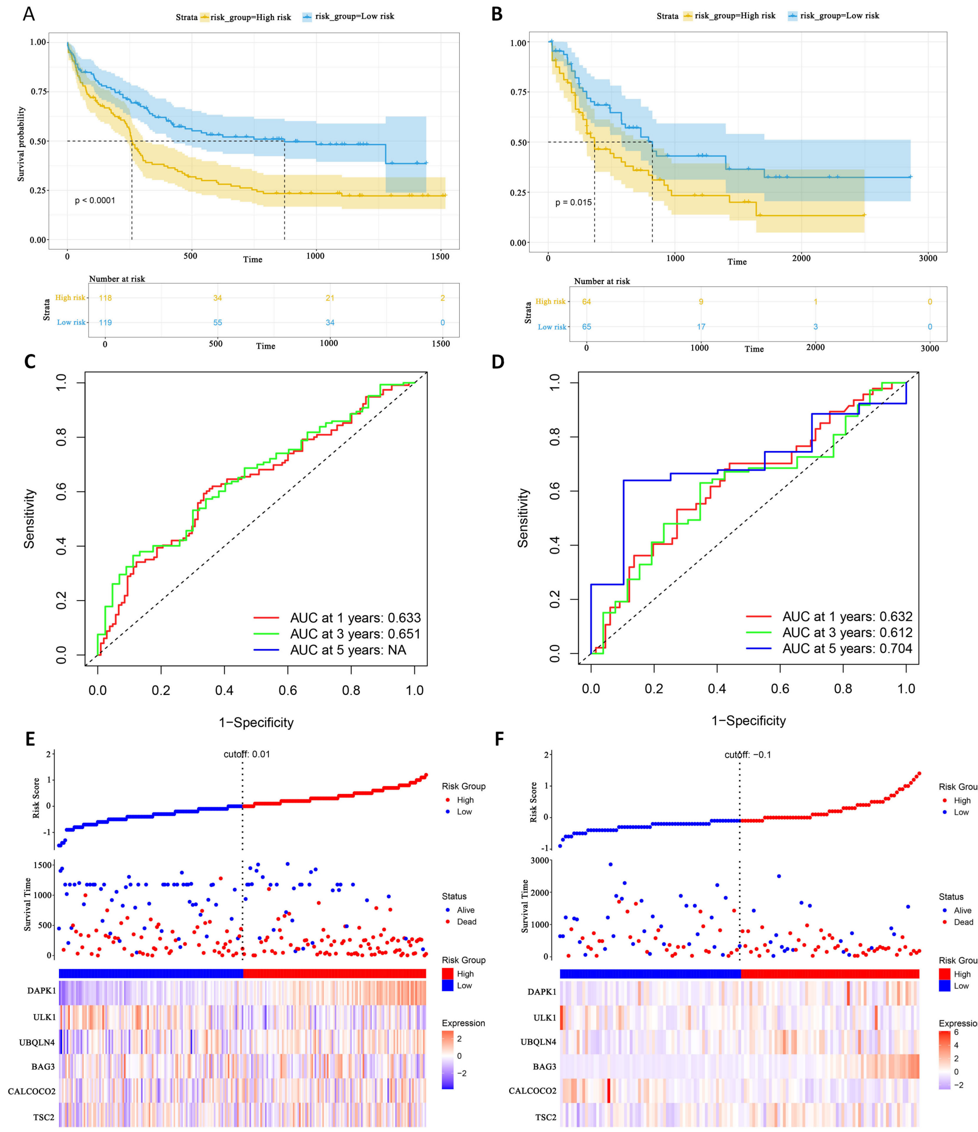


FIGURE 6
External gene set validation of survival prediction models. **(A, B)** Kaplan-Meier curves of prognostic genes in validation sets GSE12417 and TCGA-LAML. **(C, D)** AUC curve of the GSE12417 validation sets and TCGA-LAML. **(E, F)** Risk score distribution, survival status and 6 prognostic genes expression heatmap in the GSE12417 and TCGA-LAML cohorts.

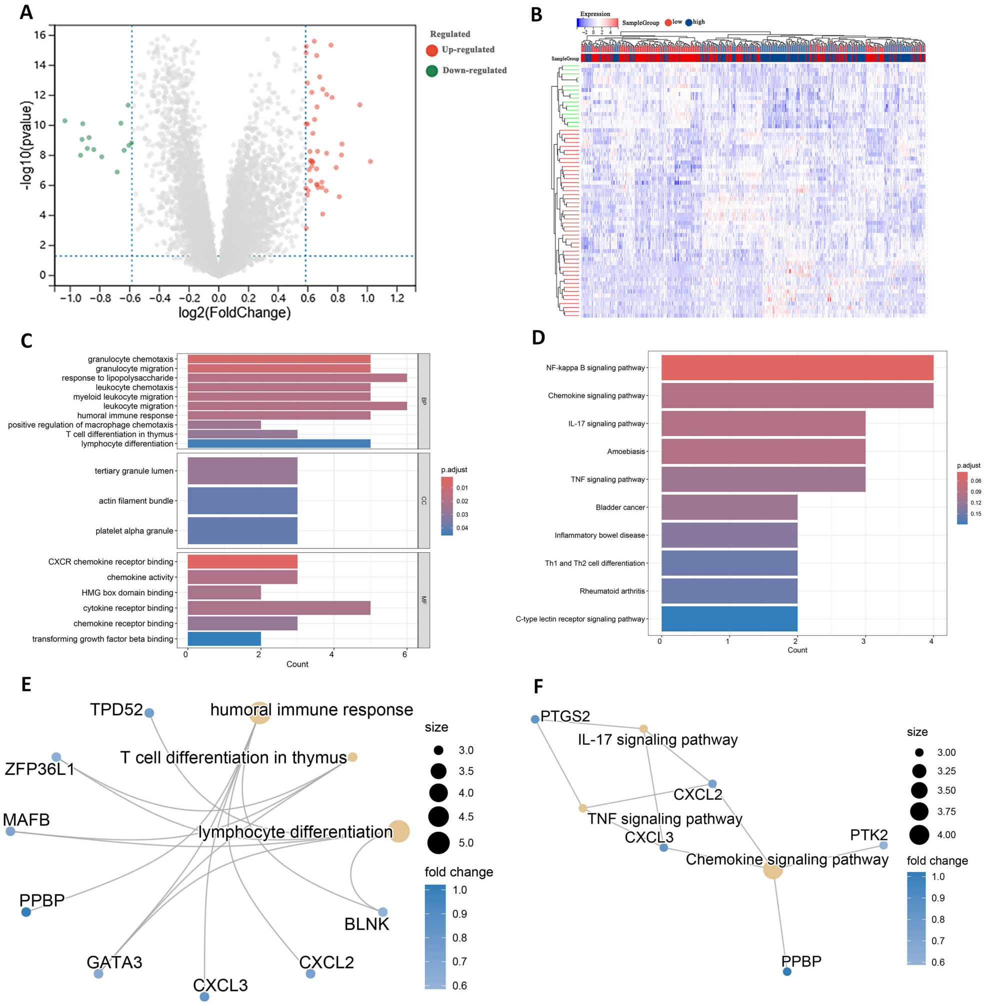


FIGURE 7
Identification of DEGs in high and low risk groups. **(A)** Volcano plot of DEGs. **(B)** Heatmap of DEGs. **(C)** GO terms analysis of differential genes. **(D)** KEGG pathway enrichment analysis. **(E)** Network diagram of GO terms enrichment with differential genes. **(F)** Network diagram of KEGG pathway enrichment of differential genes.

leukocyte migration (Supplementary Figure 6). The KEGG pathway showed that the IL-17 signaling pathway, Chemokine signaling pathway, and TNF signaling pathway also interacted with multiple other pathways (Supplementary Figure 7). The enrichment analysis results suggest that these DEGs may play a role in the prognosis and immune response in AML.

Immune infiltration and immune interactions

There are complex interactions and associations between leukemia and immune infiltration. The immune system was crucial in regulating the development of leukemia. The experiment used the CIBERSORT (37) algorithm to identify 22 subtypes of immune infiltrating cells in AML samples and investigated the interactions of different immune cell subpopulations in AML patients (Figure 8A).

Supplementary Figure 8 shows the ratio of each type of immune cell in AML patients, from which it can be seen that immune cells such as Mast cells activated and Macrophages M0 have a higher ratio in AML. The immune infiltration results indicated that the abundance of immune cells, including T cells CD4$^+$ memory activated, NK cells activated and T cells CD4$^+$ naive was higher in patients in the low-risk group of AMLs than in the high-risk group (Figure 8B). Additionally, the relationship between six key ARGs and immune infiltration was investigated in this experiment. The results showed that these six key ARGs were associated with T cells CD4$^+$ naive, T cells CD8$^+$, and Macrophages M1, respectively, and immune cells, and changes in the abundance of these immune cells may influence the pathogenesis of AML (Figure 8C). The above results suggest that key autophagy genes may affect the abundance of immune cells in AML patients, thereby attenuating the control of leukemia by the immune system and consequently affecting leukemia survival.

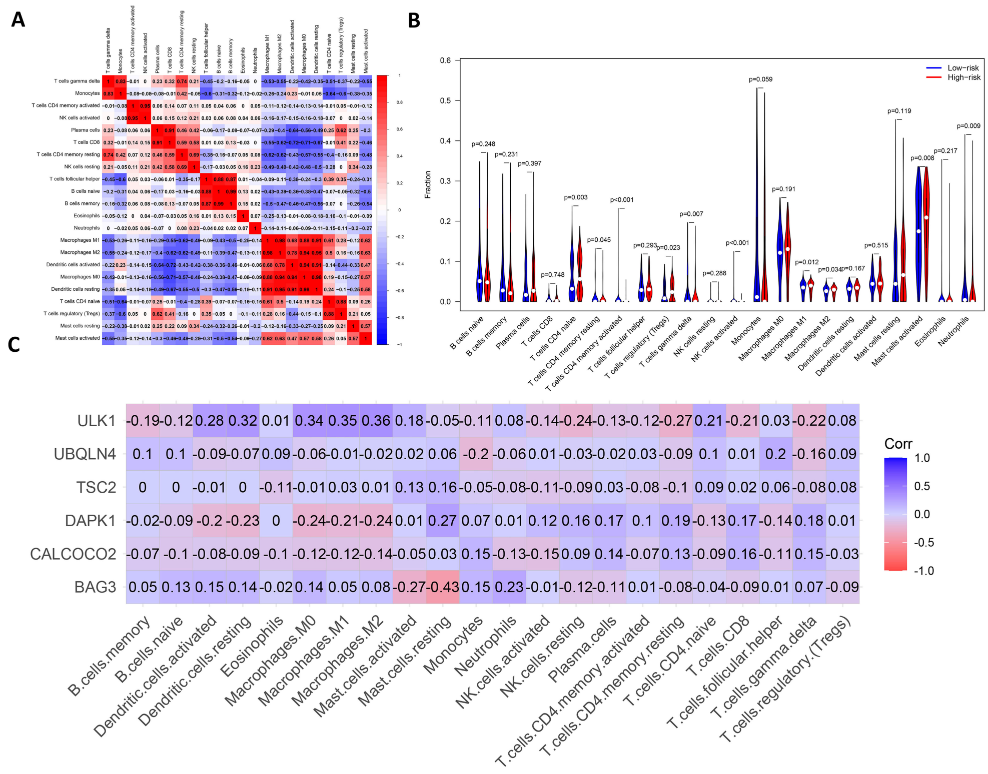


FIGURE 8

Analysis of leukemia autophagy gene immune infiltration and correlation with hub gene. **(A)** Heatmap of correlation of abundance of different immune cells. **(B)** Violin plots of immune cell abundance. Red represents the high-risk group and dark blue represents the low-risk group. **(C)** Heatmap of the correlation between six prognosis-related genes and immune cell.

Discussion

Despite significant progress in recent years in the study of prognostic markers for acute myeloid leukemia, the role of autophagy genes in AML is still understudied (38). In this experiment, machine learning methods such as XGBoost, Random Forest and SVM were used to identify potential prognostic markers associated with overall survival in AML (39). Lasso-Cox was then used to further screen for prognostic markers and a survival prediction model consisting of six genes was constructed. The model can predict the overall survival of patients with some generalization ability. In addition, an immune infiltration analysis of autophagy genes in transcriptomic data from AML patients was performed using the CIBERSORT algorithm, and the relationship between identified prognostic markers and immune cell infiltration was analyzed. These analyses have deepened our understanding of TME in AML patients and its impact on disease progression and prognosis.

AML is a severe blood cancer triggered by abnormal proliferation and differentiation of hematopoietic stem cells in the bone marrow. The role of autophagy genes in AML remains under-explored, despite significant progress in the study of AML prognostic markers in recent years. Since AML is a highly heterogeneous disease with multiple molecular features and significant biological differences between patients, it is difficult for a single prognostic marker to accurately predict the prognosis of all patients. Existing studies have mostly focused on common genetic prognostic markers, while studies on autophagy genes are more limited. However, the key role of autophagy in cell survival and death suggests that it may be an important factor influencing AML progression.

The role of autophagy in AML is dual, on the one hand helping AML cells to survive in a hostile environment by removing damaged organelles and proteins (7). On the other hand, autophagy can promote apoptosis in AML cells under certain circumstances (40). In recent years, some studies have begun to explore the role of autophagy in AML. For example, the study by Nan et al. demonstrated that FAT1 inhibited AML cell proliferation by reducing autophagy levels (41), but the study did not delve into the mechanism of the role of specific autophagy genes in AML prognosis. In contrast, Fu et al. used univariate Cox regression to initially screen autophagy genes associated with AML overall survival and further constructed a survival prediction model by Lasso-Cox regression (42). However, univariate Cox regression has limited predictive power when dealing with the complex effects of multivariate on survival. In this study, we conducted a more detailed molecular-level analysis of the relationship between autophagy genes and AML prognosis. We used a screening approach combining three machine learning algorithms, SVM, XGBoost, and Random Forest, which are capable of dealing with complex interactions of multivariate variables and generally provide higher prediction accuracy.

In this study, 45 autophagy genes associated with OS in patients with acute myeloid leukemia were screened using three machine algorithms, SVM, XGBoost and Random Forest algorithm. Among these genes, PDK4 regulates glucose metabolism by inhibiting pyruvate dehydrogenase complex and promotes glycolytic metabolism in tumor cells, an altered metabolism that is a typical feature of cancer cells. Low expression or mutation of BECN1 is closely associated with tumorigenesis. ULK1 plays a crucial role in regulating autophagy in cancer cells. These genes are clinically important as prognostic markers or potential therapeutic targets in cancers such as AML. These autophagy genes were subjected to PPI analysis and then the PPI network was further analyzed using MCODE. As a result, two important modules were identified, namely the HSP90AB1 module and the BECN1 module. It was shown that these ARGs modules have an important impact on OS in AML patients. For example, low expression of BECN1 was associated with poor prognosis in AML patients (43). High expression of ULK1 is associated with better prognosis and it may inhibit tumor growth by promoting autophagy to remove abnormal proteins and damaged organelles from AML cells (4). In addition, genes such as HSP90AB1, CALCOCO2, DNAJB1, and WDFY3 have not yet been extensively studied in the regulation of autophagy in AML. However, these genes are correlated in other cancers (44–46), and they may serve as important prognostic markers in AML. Further pathway enrichment analysis showed that these autophagy genes were mainly enriched in the AMPK signaling pathway, animal autophagy and longevity. It was shown that the activation of AMPK could inhibit the mTOR signaling pathway, promote autophagy and maintain cellular energy homeostasis. By inhibiting lipid and protein synthesis (47), AMPK can limit AML cell proliferation. In terms of GO term these genes are mainly associated with cytolytic metabolic processes, autophagy and the regulation of processes that utilize the autophagic machinery. Decreased cytolytic function is thought to correlate with immunosuppressive status and poor prognosis in AML. For example, Coles et al. showed that upregulation of the immunosuppressive glycoprotein CD200 significantly inhibited the cytolytic capacity of natural killer (NK) cells in AML patients, and that this inhibition reduced the efficiency of the immune system in the clearance of tumor cells, thereby worsening patient prognosis (48). In addition, autophagy, as a key metabolic regulatory mechanism, is closely related to drug resistance in AML cells. a study by Chen et al. indicated that autophagy not only helps leukemia cells to obtain energy and nutrients for metabolism, but also slows down the damage of drugs on AML cells by maintaining intracellular homeostasis under chemotherapeutic stress conditions through metabolic reprogramming (49). Therefore, over-activation of autophagy may make AML cells more resistant to drugs, which in turn affects the prognostic outcome of patients. In summary, cytolysis and autophagy regulation play key roles in the pathogenesis and prognosis of AML, providing a new entry point for the development of future targeted therapeutic strategies.

After Lasso-Cox regression analysis of 45 potential prognostic genes, 6 potential prognostic markers independently affecting AML survival were further screened. Kaplan-Meier analysis showed that the survival rate of the low-risk group was significantly better than that of the high-risk group on both the training. To assess the robustness of the model on different datasets, we validated the constructed survival prediction model using the GSE12417 dataset TCGA-LAML dataset combined with patients' survival information, respectively. The AUC values for 1-year and 3-year were 0.633 and 0.651, respectively, in the validation set GSE12417.In the validation set TCGA-LAML, the AUC values for 1-year, 3-year, and 5-year OS were 0.632, 0.612, and 0.704, respectively. These results indicate the robustness of the model. Differential expression analysis of patients in the high-risk and low-risk groups showed that these DEGs were mainly enriched in terms such as humoral immune response, T cell differentiation in thymus and lymphocyte differentiation. To investigate the relationship between immune cell abundance and autophagy genes and AML prognosis, an immune infiltration analysis of AML autophagy genes was performed using the CIBERSORT algorithm. The results showed that the abundance of T cells $CD4^+$ memory activated, NK cells activated and T cells $CD4^+$ naive was higher in patients in the AML low-risk group compared with the high-risk group. This suggests that alterations in the immune microenvironment may make the high-risk group less able to fight cancer. Further investigation of the relationship between these prognostic markers and immune cell abundance showed that ULK1 was positively associated with macrophage subtypes, whereas BAG3 was significantly negatively associated with Mast cells resting, and DAPK1 was negatively associated with multiple immune cell subtypes. DAPK1 was negatively associated with multiple immune cell subtypes. The results suggest that these autophagy genes may regulate AML progression by influencing immune cell infiltration. The underlying mechanisms may involve the central role of autophagy in regulating immune function, with ULK1 promoting anti-tumor immune responses by enhancing macrophage phagocytic activity, BAG3 inhibiting mast cell activity to weaken the immune response, and DAPK1 down-regulation inhibiting the activity of a variety of immune cells, resulting in difficulties for the immune system to recognize and destroy AML cells, which in turn drives tumor progression. In terms of clinical treatment, by targeting the autophagy pathway, it is possible to enhance the activity of specific immune cell subtypes or inhibit the autophagy escape mechanism of cancer cells. For example, activation of ULK1 may enhance the anti-tumor effect of macrophages, whereas by inhibiting BAG3, the control of AML by immune cells may be enhanced. In addition, DAPK1-associated negative regulatory effects could also serve as potential therapeutic targets aimed at restoring the immune system's ability to recognize and kill AML cells.

In this study, these gene-enriched pathways revealed the critical roles of autophagy and metabolic regulation in the pathogenesis of AML. Autophagy not only helps leukemia cells to meet their metabolic demands, but may also enable AML cells to better adapt to environmental stresses through inter-regulation with, for example, the AMPK signaling pathway. Therefore, targeting these aberrant pathways may provide new strategies for the treatment of AML. Survival prediction models constructed on the basis of these autophagy genes provide more comprehensive and precise prognostic information for personalized treatment of AML patients, helping clinicians to better assess the prognosis of patients and develop personalized treatment plans. In addition, autophagy genes play a key role in regulating immune cell infiltration and its prognostic impact on AML, which provides a research direction to further explore the complex relationship between autophagy and the immune microenvironment. Overall, the study of these pathways is important for an in-depth understanding of the prognostic mechanisms of AML and provides new targets for clinical treatment.

Although this study constructed a prognostic prediction model for AML based on autophagy genes, there are still some limitations. Firstly, although the joint screening of prognosis-related genes by three algorithms, SVM, Random Forest and XGBoost, can combine their respective advantages and improve the stability and consistency of the screening results, XGBoost and Random Forest are susceptible to overfitting when the sample sizes are small, especially when the parameters are not precisely adjusted. In addition, although SVM usually performs better on small sample data, the risk of overfitting may be further amplified when combining these three algorithms. Therefore, special attention needs to be paid to model tuning and validation when applying this combination strategy, especially when dealing with small-sample data, in order to reduce the potential overfitting problem. Secondly, this study mainly relied on transcriptomics data and did not address protein expression or functional status, thus some key biological processes may be missed. Although the model performed well in the validation set, further functional validation and experimental evaluation are needed for its clinical application prospects. Compared with other AML prognostic models, such as Guo et al. (19). who constructed models with common genetic markers or mutation information, our model, although incorporating the specific mechanism of autophagy genes, is slightly deficient in predictive ability, especially the low AUC value in the independent validation set, suggesting that the model's predictive performance needs to be further improved. In addition, the relatively small sample size of the 2 external validation datasets used in the study may not cover the diversity of AML patients. This limits the ability of the model to generalize to a wider patient population.

Therefore, future studies should further validate the robustness and accuracy of the model in larger and more diverse AML patient cohorts. Meanwhile, in addition to traditional transcriptomics data, multi-omics data such as proteomics and metabolomics can be integrated to provide a more comprehensive biological perspective and avoid missing biological processes that may play a key role in disease development. In addition, functional experiments should be performed on the screened autophagy genes to delve into the specific mechanisms of these genes in AML and to assess their impact on disease progression. In order to gain a deeper understanding of the complexity of the AML tumor immune microenvironment, future studies should be expanded to cover the analysis of more types and subpopulations of immune cells. Finally, based on the importance of these key autophagy genes, precision therapeutic strategies targeting these genes or their associated pathways could be explored in the future, thus promoting further development of personalized treatment for AML patients.

Conclusion

In this study, we screened six potential autophagy gene prognostic markers for AML (TSC2, CALCOCO2, BAG3, UBQLN4, ULK1, and DAPK1) and constructed a survival prediction model of eight autophagy genes for predicting the survival of AML patients. The model was validated by two validation sets, and the results showed that the survival prediction model had strong validity. In addition, autophagy gene pathway enrichment analysis as well as immune infiltration and immune correlation analysis of ARGs were performed to investigate the biological functions of autophagy genes and the prognostic markers of ARGs in correlation with many immune cells. However, although potential prognostic markers and correlations can be identified from transcriptomic data, the biological significance and clinical application of these results have not been fully confirmed due to the lack of further clinical validation. More medically relevant experiments are needed in the future to validate the potential molecular mechanisms of these genes to better understand their role and application value in AML.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: The public datasets used in this study are the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/, GSE12417 and GSE37642) databases and TCGA databases (https://xenabrowser.net/datapages/).

Author contributions

CZ: Conceptualization, Formal analysis, Methodology, Resources, Validation, Visualization, Writing – original draft. XF: Resources, Supervision, Writing – review & editing. LT: Validation, Visualization, Writing – review & editing. PM: Validation, Writing – review & editing. FW: Funding acquisition, Validation, Writing – review & editing. WQ: Formal analysis, Funding acquisition, Writing – review & editing. YD: Formal analysis, Funding acquisition, Supervision, Validation, Writing – review & editing. XZ: Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was partially supported by the National Natural Science Foundation of China (61902094, 62102346 and 81402054), the Shandong Provincial Natural Science Foundation of China (ZR2021MH036 and ZR2021QF123), the Science and Technology Program of Binzhou Medical University (50012304325), the Doctoral Scientific Research Foundation of Yantai University (JS22B171 and JS22B16).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1489171/full#supplementary-material

References

1. Forsberg M, Konopleva M. AML treatment: conventional chemotherapy and emerging novel agents. *Trends Pharmacol Sci*. (2024) 45:430–48. doi: 10.1016/j.tips.2024.03.005

2. Gurnari C, Voso MT, Maciejewski JP, Visconte VJC. From bench to bedside and beyond: therapeutic scenario in acute myeloid leukemia. *Cancers*. (2020) 12:357. doi: 10.3390/cancers12020357

3. Zeng T, Cui L, Huang W, Liu Y, Si C, Qian T, et al. The establishment of a prognostic scoring model based on the new tumor immune microenvironment classification in acute myeloid leukemia. *BMC Med*. (2021) 19:1–11. doi: 10.1186/s12916-021-02047-9

4. Du W, Xu A, Huang Y, Cao J, Zhu H, Yang B, et al. The role of autophagy in targeted therapy for acute myeloid leukemia. *Autophagy*. (2021) 17:2665–79. doi: 10.1080/15548627.2020.1822628

5. Joffre C, Ducau C, Poillet-Perez L, Courdy C, Mansat-De Mas VJB. Autophagy a close relative of AML biology. *Biology*. (2021) 10:552. doi: 10.3390/biology10060552

6. Zalpoor H, Bakhtiyari M, Akbari A, Aziziyan F, Shapourian H, Liaghat M, et al. Potential role of autophagy induced by FLT3-ITD and acid ceramidase in acute myeloid leukemia chemo-resistance: new insights. *Cell Communication Signaling*. (2022) 20:172. doi: 10.1186/s12964-022-00956-7

7. Yamamoto H, Zhang S, Mizushima N. Autophagy genes in biology and disease. *Nat Rev Genet*. (2023) 24:382–400. doi: 10.1038/s41576-022-00562-w

8. Levine B, Kroemer G. Biological functions of autophagy genes: a disease perspective. *Cell*. (2019) 176:11–42. doi: 10.1016/j.cell.2018.09.048

9. Ding J, Wang C, Sun Y, Guo J, Liu S, Cheng Z. Identification of an autophagy-related signature for prognosis and Immunotherapy Response Prediction in Ovarian Cancer. *Biomolecules*. (2023) 13:339. doi: 10.3390/biom13020339

10. Rui R, Zhou L, He S. Cancer immunotherapies: advances and bottlenecks. *Front Immunol*. (2023) 14:1212476. doi: 10.3389/fimmu.2023.1212476

11. Bansal AK, Sharawat SK, Gupta R, Vishnubhatla S, Dhawan D, Bakhshi S. Regulatory T cells in pediatric AML are associated with disease load and their serial assessment suggests role in leukemogenesis. *Am J Blood Res*. (2020) 10:90.

12. Wan Y, Zhang C, Xu Y, Wang M, Rao Q, Xing H, et al. Hyperfunction of CD4 CD25 regulatory T cells in *de novo* acute myeloid leukemia. *BMC cancer*. (2020) 20:1–10. doi: 10.1186/s12885-020-06961-8

13. Romee R, Rosario M, Berrien-Elliott MM, Wagner JA, Jewell BA, Schappe T, et al. Cytokine-induced memory-like natural killer cells exhibit enhanced responses against myeloid leukemia. *Sci Trans Med*. (2016) 8:357ra123–357ra123. doi: 10.1126/scitranslmed.aaf2341

14. Xue L, Zhang W, Ju Y, Xu X, Bo H, Zhong X, et al. TNFSF10, an autophagy related gene, was a prognostic and immune infiltration marker in skin cutaneous melanoma. *J Cancer*. (2023) 14:2417. doi: 10.7150/jca.86735

15. Wang Y, Zhao W, Xiao Z, Guan G, Liu X, Zhuang M. A risk signature with four autophagy-related genes for predicting survival of glioblastoma multiforme. *J Cell Mol Med*. (2020) 24:3807–21. doi: 10.1111/jcmm.14938

16. Luo Z, Nong B, Ma Y, Fang D. Autophagy related long non-coding RNA and breast cancer prognosis analysis and prognostic risk model establishment. *Ann Trans Med*. (2022) 10:58. doi: 10.21037/atm-21-6251

17. Hu J, Szymczak S. A review on longitudinal data analysis with random forest. *Briefings Bioinf*. (2023) 24:bbad002. doi: 10.1093/bib/bbad002

18. Duan K-B, Rajapakse JC, Wang H, Azuaje F. Multiple SVM-RFE for gene selection in cancer classification with expression data. *IEEE Trans nanobioscience*. (2005) 4:228–34. doi: 10.1109/tnb.2005.853657

19. Guo C, Gao Y-y, Ju Q-q, Zhang C-x, Gong M, Li Z-l. The landscape of gene co-expression modules correlating with prognostic genetic abnormalities in AML. *J Trans Med*. (2021) 19:228. doi: 10.1186/s12967-021-02914-2

20. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JDJB. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. (2012) 28:882–3. doi: 10.1093/bioinformatics/bts034

21. Verweij PJ, Van Houwelingen HC. Penalized likelihood in Cox regression. *Stat Med*. (1994) 13:2427–36. doi: 10.1002/sim.4780132307

22. Chang W, Ji X, Xiao Y, Zhang Y, Chen B, Liu H, et al. Prediction of hypertension outcomes based on gain sequence forward tabu search feature selection and xgboost. *Diagnostics*. (2021) 11:792. doi: 10.3390/diagnostics11050792

23. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: network analysis and visualization of proteomics data. *J Proteome Res*. (2018) 18:623–32. doi: 10.1021/acs.jproteome.8b00702

24. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J anesthesiology*. (2022) 75:25–36. doi: 10.4097/kja.21209

25. Yılmaz H, Toy HI, Marquardt S, Karakülah G, Küçük C, Kontou PI, et al. In silico methods for the identification of diagnostic and favorable prognostic markers in acute myeloid leukemia. *Int J Mol Sci*. (2021) 22:9601. doi: 10.3390/ijms22179601

26. Yu S, Jiang J. Immune infiltration-related genes regulate the progression of AML by invading the bone marrow microenvironment. *Front Immunol*. (2024) 15:1409945. doi: 10.3389/fimmu.2024.1409945

27. Jiang G, Jin P, Xiao X, Shen J, Li R, Zhang Y, et al. Identification and validation of a novel CD8+ T cell-associated prognostic model based on ferroptosis in acute myeloid leukemia. *Front Immunol*. (2023) 14:1149513. doi: 10.3389/fimmu.2023.1149513

28. Moore JA, Mistry JJ, Hellmich C, Horton RH, Wojtowicz EE, Jibril A, et al. LC3-associated phagocytosis in bone marrow macrophages suppresses acute myeloid leukemia progression through STING activation. *J Clin Invest*. (2022) 132:e153157. doi: 10.1172/JCI153157

29. Zhang Y, Xu W, Yang P, Zhang AJ, Making D. Machine learning for the prediction of sepsis-related death: a systematic review and meta-analysis. *BMC Med Inform Decis Mak*. (2023) 23:283. doi: 10.1186/s12911-023-02383-1

30. Hong W, Lu Y, Zhou X, Jin S, Pan J, Lin Q, et al. Usefulness of random forest algorithm in predicting severe acute pancreatitis. *Front Cell Infect Microbiol*. (2022) 12:893294. doi: 10.3389/fcimb.2022.893294

31. Xiao ML, Wei Y, Zhang J, Jian JM, Song Y, Lin ZJ, et al. MRI texture analysis for preoperative prediction of lymph node metastasis in patients with nonsquamous cell cervical carcinoma. *Acad Radiol*. (2022) 29:1661–71. doi: 10.1016/j.acra.2022.01.005

32. Zheng Y, Wang J, Ling Z, Zhang J, Zeng Y, Wang K, et al. A diagnostic model for sepsis-induced acute lung injury using a consensus machine learning approach and its therapeutic implications. *J Trans Med*. (2023) 21:620. doi: 10.1186/s12967-023-04499-4

33. Mering Cv, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res*. (2003) 31:258–61. doi: 10.1093/nar/gkg034

34. Li D, Ding L, Luo J, Li Q-G. Prediction of mortality in pneumonia patients with connective tissue disease treated with glucocorticoids or/and immunosuppressants by

machine learning. *Front Immunol*. (2023) 14:1192369. doi: 10.3389/fimmu.2023.1192369

35. Ma Z, Yang S, Yang Y, Luo J, Zhou Y, Yang H. Development and validation of prediction models for the prognosis of colon cancer with lung metastases: a population-based cohort study. *Front Endocrinology*. (2023) 14:1073360. doi: 10.3389/fendo.2023.1073360

36. Gomes AP, Costa B, Marques R, Nunes V, Coelho C. Kaplan-meier survival analysis: practical insights for clinicians. *Acta Médica Portuguesa*. (2024) 37:280–5. doi: 10.20344/amp.21080

37. Wu W, Wang X, Le W, Lu C, Li H, Zhu Y, et al. Immune microenvironment infiltration landscape and immune-related subtypes in prostate cancer. *Front Immunol*. (2023) 13:1001297. doi: 10.3389/fimmu.2022.1001297

38. Metzeler KH, Herold T, Rothenberg-Thurley M, Amler S, Sauerland MC, Gorlich D, et al. The Journal of the American Society of Hematology. Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia. *Blood*. (2016) 128:686–98. doi: 10.1182/blood-2016-01-693879

39. Moncada-Torres A, van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. *Sci Rep*. (2021) 11:6968. doi: 10.1038/s41598-021-86327-7

40. Das S, Shukla N, Singh SS, Kushwaha S, Shrivastava R. Mechanism of interaction between autophagy and apoptosis in cancer. *Apoptosis*. (2021) 26:1–22. doi: 10.1007/s10495-021-01687-9

41. Zhang N, M-y S, Meng Q-l, Sun H-p, Fan F-y, Yi H, et al. FAT1 inhibits AML autophagy and proliferation via downregulating ATG4B expression. *Biochim Biophys Acta (BBA)-General Subjects*. (2024) 1868:130519. doi: 10.1016/j.bbagen.2023.130519

42. Fu D, Zhang B, Wu S, Zhang Y, Xie J, Ning W, et al. Prognosis and characterization of immune microenvironment in acute myeloid leukemia through identification of an autophagy-related signature. *Front Immunol*. (2021) 12:695865. doi: 10.3389/fimmu.2021.695865

43. Aasebo E, Berven FS, Bartaula-Brevik S, Stokowy T, Hovland R, Vaudel M, et al. Proteome and phosphoproteome changes associated with prognosis in acute myeloid leukemia. *Cancers*. (2020) 12:709. doi: 10.3390/cancers12030709

44. Zullow HJ, Sankar A, Ingram DR, Guerra DDS, DAvino AR, Collings CK, et al. The FUS:: DDIT3 fusion oncoprotein inhibits BAF complex targeting and activity in myxoid liposarcoma. *Mol Cell*. (2022) 82:1737–1750.e8. doi: 10.1016/j.molcel.2022.03.019

45. Li T, Fu J, Cheng J, Elfiky AA, Wei C, Fu J. New progresses on cell surface protein HSPA5/BiP/GRP78 in cancers and COVID-19. *Front Immunol*. (2023) 14:1166680. doi: 10.3389/fimmu.2023.1166680

46. Zhang Y-J, Yi D-H. CDK1-SRC interaction-dependent transcriptional activation of HSP90AB1 promotes antitumor immunity in hepatocellular carcinoma. *J Proteome Res*. (2023) 22:3714–29. doi: 10.1021/acs.jproteome

47. Hsu C-C, Peng D, Cai Z, Lin H-K. AMPK signaling and its targeting in cancer progression and treatment. *Semin Cancer Biol*. (2022) 85:52–68 p. doi: 10.1016/j.semcancer.2021.04.006

48. Coles S, Wang ECY, Man S, Hills RK, Burnett AK, Tonks A, et al. CD200 expression suppresses natural killer cell function and directly inhibits patient anti-tumor response in acute myeloid leukemia. *Leukemia*. (2011) 25:792–9. doi: 10.1038/leu.2011.1

49. Chen Y, Chen J, Zou Z, Xu L, Li J. Crosstalk between autophagy and metabolism: implications for cell survival in acute myeloid leukemia. *Cell Death Discovery*. (2024) 10:46. doi: 10.1038/s41420-024-01823-9

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 27 November 2024
DOI 10.3389/fonc.2024.1499794

OPEN ACCESS

EDITED BY
Mohsin Saleet Jafri,
George Mason University, United States

REVIEWED BY
Eric Munger,
United States Department of Veterans Affairs,
United States
Soukaina Amniouel,
National Center for Advancing Translational
Sciences (NIH), United States

*CORRESPONDENCE
Yifan Jiang
✉ jyf016023@163.com

RECEIVED 21 September 2024
ACCEPTED 05 November 2024
PUBLISHED 27 November 2024

CITATION
Liu X, Shu X, Zhou Y and Jiang Y (2024) Construction of a risk prediction model for postoperative deep vein thrombosis in colorectal cancer patients based on machine learning algorithms.
Front. Oncol. 14:1499794.
doi: 10.3389/fonc.2024.1499794

Construction of a risk prediction model for postoperative deep vein thrombosis in colorectal cancer patients based on machine learning algorithms

Xin Liu[1], Xingming Shu[1], Yejiang Zhou[2] and Yifan Jiang[2]*

[1]Department of Clinical Medicine, Southwest Medical University, Luzhou, China, [2]Department of Gastrointestinal Surgery, The Affiliated Hospital of Southwest Medical University, Luzhou, Sichuan, China

Background: Colorectal cancer is a prevalent malignancy of the digestive system, with an increasing incidence. Lower extremity deep vein thrombosis (DVT) is a frequent postoperative complication, occurring in up to 40% of cases.

Objective: This research aims to develop and validate a machine learning model (ML) to predict the risk of lower limb deep vein thrombosis in patients with colorectal cancer, facilitating preventive and therapeutic measures to enhance recovery and ensure safety.

Methods: In this retrospective cohort study, we collected data from 429 colorectal cancer patients from January 2021 to January 2024. The medical records included age, blood test results, body mass index, underlying diseases, clinical staging, histological typing, surgical methods, and postoperative complications. We employed the Synthetic Minority Oversampling Technique to address imbalanced data and split the dataset into training and validation sets in a 7:3 ratio. Feature selection was performed using Random Forest (RF), XGBoost, and Least Absolute Shrinkage and Selection Operator algorithms (LASSO). We then trained six machine learning models: Logistic Regression (LR), Naive Bayes (NB), Gaussian Process (GP), Random Forest, XGBoost, and Multilayer Perceptron (MLP). The model's performance was evaluated using metrics such as area under the Receiver Operating Characteristic curve, accuracy, sensitivity, specificity, F1 score, and confusion matrix. Additionally, SHAP and LIME were used to enhance the interpretability of the results.

Results: The study combined Random Forest, XGBoost algorithms, and LASSO regression with univariate regression analysis to identify significant predictive factors, including age, preoperative prealbumin, preoperative albumin, preoperative hemoglobin, operation time, PIKVA2, CEA, and preoperative neutrophil count. The XGBoost model outperformed other ML algorithms, achieving an AUC of 0.996, an accuracy of 0.9636, a specificity of 0.9778, and an F1 score of 0.9576. Moreover, the SHAP method identified age and preoperative prealbumin as the primary determinants influencing ML model predictions. Finally, the study employed LIME for more precise prediction and interpretation of individual predictions.

Conclusion: The machine learning algorithms effectively predicted postoperative lower limb deep vein thrombosis in colorectal cancer patients. The XGBoost model demonstrated strong potential for improving early detection and treatment in clinical settings.

KEYWORDS

colorectal cancer, venous thrombosis, machine learning, prediction model, postoperative complications

1 Introduction

Colorectal cancer is among the most prevalent malignant tumors of the digestive system globally, ranking third in both incidence and mortality rates among malignant tumors (1). Currently, surgical treatment is the primary approach for colorectal cancer. However, ostoperative lower limb deep vein thrombosis has consistently been an issue that cannot be overlooked. Literature reports that the incidence of lower limb deep vein thrombosis after abdominal surgery is 15%-19%. Alarmingly, the incidence in colorectal cancer patients post-surgery is 40% (2). Additionally, since only 50% of patients with lower limb deep vein thrombosis exhibit symptoms and signs such as swelling and tenderness, many cases are overlooked postoperatively (3). Without timely diagnosis and intervention, the clot may detach and move through the veins to the lungs, leading to a life-threatening pulmonary embolism (4). However, lower limb deep vein thrombosis can be prevented in advance. Research suggests that prophylactic anticoagulant treatment can be suitably applied to bedridden patients in the perioperative phase (5, 6). Currently, the Caprini risk assessment model is the most widely used model in surgery. However, all colorectal cancer patients stratified postoperatively according to the Caprini model are considered high risk. Therefore, the Caprini model may not be a completely accurate indicator for DVT occurrence and intervention in colorectal cancer patients (7).

Additionally, most existing studies utilize traditional statistical methods rather than advanced machine learning algorithms, which often limits the models' ability to handle nonlinear relationships and multivariable interactions, thereby affecting their predictive performance and applicability (8). The purpose of this study is to integrate these common high-risk factors using machine learning by selecting shared features through three different machine learning algorithms and constructing multiple models to identify the optimal deep vein thrombosis risk prediction model for colorectal cancer patients. This model will assist clinicians in more accurately identifying high-risk patients and providing personalized, precise guidance for the prevention and treatment of deep vein thrombosis.

2 Materials and methods

2.1 Study design

The aim of this research is to develop a machine learning-based model to predict the risk of lower limb deep vein thrombosis in postoperative colorectal cancer patients. A retrospective study was conducted, including 429 colorectal cancer patients who underwent surgical treatment. Data were extracted from the hospital's electronic medical record system, which included demographic details, medical history, treatment information, disease severity, blood test results, and postoperative complications. The SMOTE algorithm was employed to address the issue of class imbalance. LASSO regression, Xgboost, and random forest were applied for feature selection to identify the features most associated with the risk of lower limb deep vein thrombosis. Following this, a range of ML models, such as LR, RF, GB, MLP, XGB, and KNN, were developed and optimized using the 10-fold cross-validation approach. The performance of these models was assessed through a range of metrics, including accuracy, sensitivity, specificity, positive predictive value, negative predictive value, F1 score, Kappa score, AUC, calibration curve, clinical impact curve, and confusion matrix. To enhance the transparency and interpretability of the model, SHAP and LIME methods were used to explain the prediction results, clarifying the impact of each feature on the predictions and thereby offering useful references for clinicians. Figure 1 illustrates the overall workflow of the proposed system more clearly.

2.2 Study data

We retrospectively selected 429 colorectal cancer patients who visited the Department of Gastrointestinal Surgery at the First Affiliated Hospital of Southwest Medical University from January 2022 to January 2024. Exclusion criteria include: patients with a history of prolonged bed rest or restricted activity; patients with a history of venous thrombosis; patients with a history of coagulation disorders; patients using drugs affecting coagulation function;

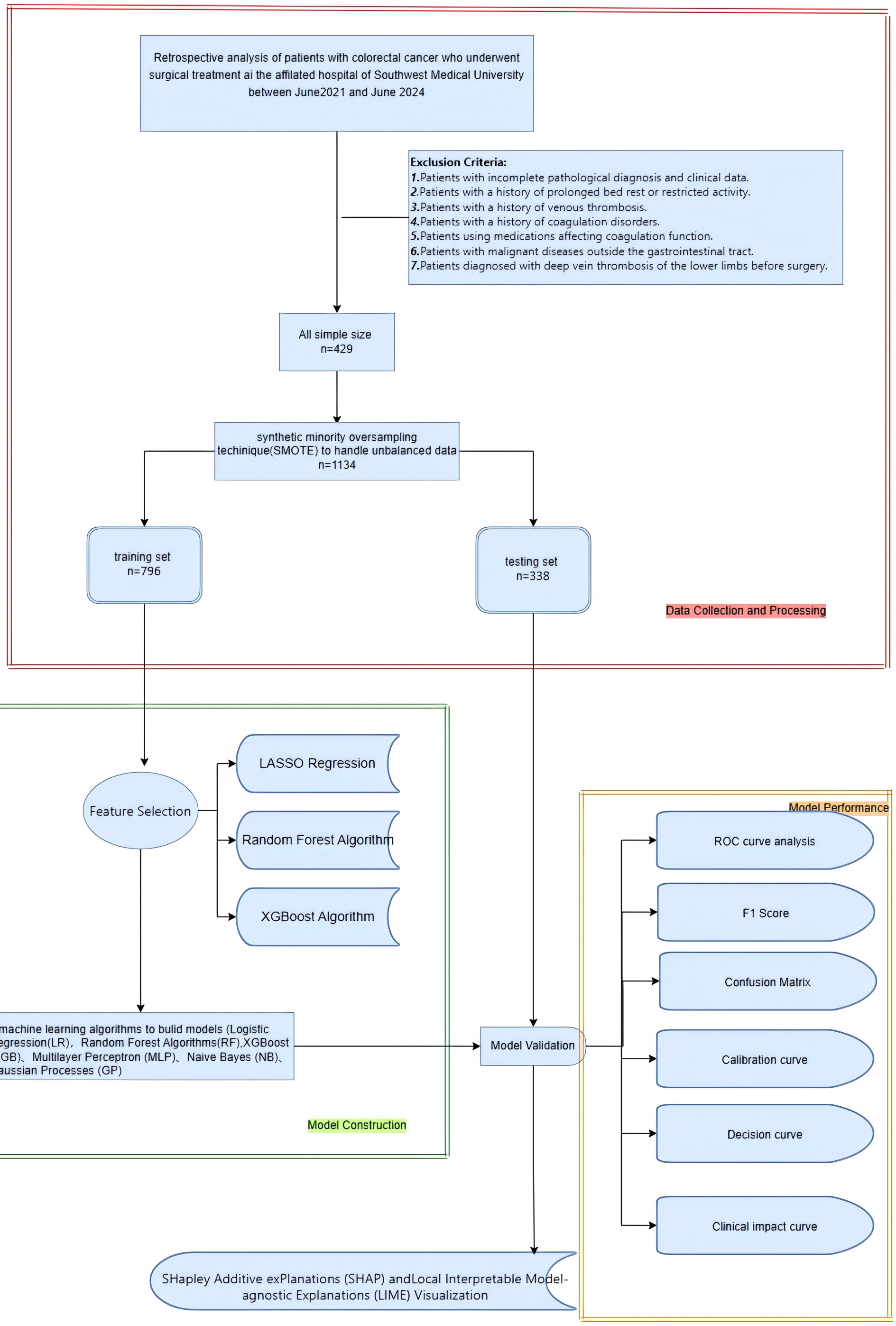


FIGURE 1
Research process.

patients with malignancies outside the gastrointestinal tract; and patients preoperatively diagnosed with lower extremity deep vein thrombosis. (Exclusion criteria are shown in Figure 1). As this study is retrospective, patients are exempt from providing informed consent according to the ethics review board's policy. The ethics committee has encrypted all personal information of patients involved in this study to prevent any leaks.

2.3 Study variables

The study includes 44 variables related to demographic factors (gender, age), medical history (history of diabetes, hypertension, coronary artery disease, chronic obstructive pulmonary disease), physical characteristics (BMI), disease severity (clinical stage, histological grade, presence of cancer embolus, nerve invasion, vascular invasion), treatment information (surgical method, surgery duration, use of specific cancer treatments), laboratory values (white blood cell count, neutrophil count, lymphocyte count, monocyte count, NLR, hemoglobin, prealbumin, albumin, creatinine clearance, platelet count, prothrombin time PT, fibrinogen, thrombin time TT, D-dimer), and postoperative complications (postoperative high fever, anastomotic leak). Venous blood samples were collected within 24 hours of admission.

2.4 Diagnosis

Patients were tested within 14 days postoperatively according to the diagnostic criteria for lower limb deep vein thrombosis. Specifically, color Doppler ultrasound showed an uneven echo solid mass in the lower limb, reduced or absent color blood flow and spectral signals, non-collapse of the venous lumen after compression, and venous incompressibility (9).

2.5 Data preprocessing

The structured database initially included 44 clinical variables. First, clinical variables with more than 30% missing data (n = 2) were excluded. The missing data were handled using 10-fold crossvalidation combined with the KNN imputation method. Subsequently, to prevent bias during later model training and improve interpretability, the Variance Inflation Factor (VIF) was employed to examine multicollinearity among the chosen features, ensuring all features' VIF values were less than 10. Additionally, we also removed variables with nearly zero variance to simplify the model and enhance its robustness. In the end, 39 clinical features of patients were chosen to construct the predictive model. The SMOTE algorithm was used to address the class imbalance issue, balancing the dataset and avoiding bias. Subsequently, patient data were randomly divided into two datasets: (1) a training dataset (70%) for feature selection and model training, and (2) a testing dataset (30%) for model performance evaluation.

2.6 Feature selection

For predicting postoperative DVT occurrence in colorectal cancer patients, features were selected using training group samples through three machine learning models: LASSO regression, random forest, and XGboost. The results showed that 29, 15, and 15 feature vectors were selected in the three models, Ultimately, we selected 8 common feature variables from the three models: age, preoperative prealbumin, preoperative albumin, preoperative hemoglobin, CEA, PIKVA2, surgery time, and preoperative white blood cell count.

2.7 Model development and evaluation

The machine learning task is to predict the probability distribution of patients developing lower extremity deep vein thrombosis based on these clinical variables. Model development involves experimenting with six machine learning algorithms: Logistic Regression (LR), Multilayer Perceptron (MLP), Extreme Gradient Boosting (XGBoost), Gaussian Process (GP), Random Forest (RF), and Naive Bayes (NB). During the training phase, we employed the 10-fold cross-validation method to train the models in order to achieve optimal predictive performance. To evaluate the predictive performance of each model, we primarily measured the Receiver Operating Characteristic (ROC) curve. In addition, we calculated sensitivity, specificity, accuracy, false positive (FP) rate, positive predictive value (PPV), negative predictive value (NPV), Brier score, F1 score, Decision Curve Analysis (DCA) curve, calibration curve, and Clinical Impact Curve (CIC) for a comprehensive assessment of the model's performance.

2.8 Statistical analysis

All data analyses in this study were carried out using SPSS (27.0) and R language (version 4.3.3). Preliminary analysis of the dataset used descriptive statistics. Data points that followed a normal distribution were represented by mean ± standard deviation, whereas data points deviating from a normal distribution were shown as median (interquartile range). Subsequently, an independent samples t-test was employed to compare two groups of normally distributed data. In contrast, the Mann-Whitney U test was used for comparing two groups of non-normally distributed data. We resolved the sample imbalance problem by oversampling the minority classes using the SMOTE function from the DMwR2 package in R. To build the predictive model, the dataset was randomly split into a training subset comprising 70% of the total data and a testing subset making up 30% of the total data. Subsequently, various machine learning methods were executed using R, including logistic regression (glm package), Gaussian model (e1071 package), random forest (randomForest package), XGBoost (XGBoost package), feedforward neural network (nnet package), and naive Bayes model (e1071 package). Models were trained using the training subset data with these six ML algorithms. During the model training, a 10-fold cross-validation method was adopted to optimize the model parameters, aiming to prevent overfitting. Statistical significance was defined at the level of $P<0.05$.

2.9 Feature interpretation

We used the Shapley Additive Explanations (SHAP) algorithm and the Local Interpretable ModelAgnostic Explanations (LIME) algorithm to interpret the main feature contributions after machine learning model training. In particular, the SHAP algorithm assesses the average contribution of each feature value by computing its Shapley value within all possible combinations of features. By taking the weighted average of each feature value's Shapley value, we can assess the impact of that feature on the overall prediction. Meanwhile, the LIME algorithm analyzes the model from a local perspective to explain the feature importance of specific predictions, providing an additional layer of interpretation and transparency. The combination of these two methods provides us with a multidimensional understanding of model interpretability.

3 Results

3.1 Characteristics of patients

This study encompassed 429 colorectal cancer patients who underwent surgical treatment. The median age of the patients was 67 years (range: 16-91), with 258 males (60.24%) and 171 females (39.76%). The original data from 429 cases includes 267 cases without lower extremity deep vein thrombosis (62.23%) and 162 cases with lower extremity deep vein thrombosis (37.77%). The baseline characteristics comparison of the two patient groups in the original data reveals that age, preoperative white blood cell count, preoperative lymphocyte count, preoperative hemoglobin, preoperative albumin, preoperative prealbumin count, preoperative glomerular filtration rate, gender, preoperative acute complete intestinal obstruction, and surgical method are all statistically significant (refer to Table 1).

3.2 Prediction factor screening

A total of 1134 patients with colorectal cancer receiving surgical treatment were involved after data imbalance. Patients were split into a training group with 796 cases and a test group with 338 cases in a 7:3 ratio. LASSO regression, as a shrinkage estimation method, achieves variable selection and complexity adjustment by formulating an optimization objective function with a penalty term. This study utilized LASSO regression to identify features including age, surgical procedure, acute intestinal obstruction, nerve invasion, preoperative lymphocyte count, preoperative fibrinogen, preoperative prothrombin time, coronary artery disease, and diabetes (Figure 2A). Random forest builds multiple decision trees through the random selection of data subsets and features. Each feature's importance score reflects its contribution to the model's predictions, allowing the extraction of the most predictive features and the identification of characteristic factors. Features including age, preoperative prealbumin, preoperative albumin, preoperative hemoglobin, CA724, CEA, and CA242 were selected (Figure 2B). Xgboost improves prediction performance by constructing multiple weak learners and using an additive model approach. The importance of features is assessed by calculating gain, coverage, and frequency for each one, identifying factors like age, preoperative prealbumin, preoperative white blood cell count, preoperative hemoglobin, preoperative glomerular filtration rate, BMI, and preoperative prothrombin time (Figure 2C). By comparing the selection results of LASSO regression, Xgboost algorithm, and random forest algorithm, we identified the common subset of features selected by these three methods. These selected features were eventually used to construct the model, including age, preoperative prealbumin, preoperative albumin, preoperative hemoglobin, operation time, PIKVA2, CEA, and preoperative neutrophil count (Figure 2D).

3.3 Model performance

In the training dataset, the RF model demonstrated excellent predictive performance with an AUC of 1.00, indicating very high prediction accuracy. In comparison, the AUC values for the remaining five models are as follows: XGB's AUC is 0.996 (95% CI [0.994, 0.999]), GP's AUC is 0.950 (95% CI [0.935, 0.966]), MLP's AUC is 0.938 (95% CI [0.918, 0.958]), NB's AUC is 0.882 (95% CI [0.859, 0.905]), and LR's AUC is 0.814 (95% CI [0.785, 0.844]) (Figure 3A). The F1 scores of these models are as follows: RF 1.0, XGB 0.976, GP 0.878, MLP 0.889, NB 0.740, LR 0.720. In the testing dataset, the AUC values for XGB, GP, MLP, NB, LR, and RF are 0.936 (95% CI [0.907, 0.966]), 0.919 (95% CI [0.890, 0.949]), 0.884 (95% CI [0.843, 0.925]), 0.826 (95% CI [0.781, 0.871]), 0.806 (95% CI [0.760, 0.853]), and 0.973 (95% CI [0.959, 0.986]), respectively (Figure 3B). The F1 scores for XGB, GP, MLP, NB, LR, and RF are respectively 0.853, 0.816, 0.825, 0.693, 0.696, and 0.881. In this research, the accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and kappa value of each model were computed and compared (Figures 3C, D). The RF model performed excellently in the training dataset. Due to concerns about potential overfitting, the XGB model was ultimately selected as the optimal model.

3.4 Model performance evaluation

In our study, we evaluated the predictive accuracy and calibration of the model by analyzing calibration curves for the training and test sets. The calibration curve results showed that the model in the training set had high predictive accuracy, with a Somers' D coefficient of 0.992 and an area under the ROC curve of 0.996, indicating good discriminatory power (Figure 4A). Additionally, the regression calibration slope of the training set model is 0.9934, close to the ideal value of 1.000, and the intercept is -0.0175, demonstrating excellent calibration ability. The Brier score is 0.038, reflecting the high reliability of the model's predictions. In contrast, the model's discriminatory power in the test set decreased but still maintained a high level, with an area under the ROC curve of 0.936 and a Somers' D coefficient of 0.873 (Figure 4B). Decision curves for the training set (Figure 4C) indicate that the model's net benefit is significantly above the baseline strategy. On the test set (Figure 4D), the model likewise exhibits good net benefit, particularly in the threshold probability range of 0.1 to 0.95, where it maintains a high level of net benefit. The confusion matrix results show the performance differences of the model across different datasets. In the training set (Figure 4E), the model correctly identified 440 true negatives and 327 true positives, with 10 false positives and 19 false negatives, the true positive rate is 85.0%, and the true negative rate is 89.7%.In the test set (Figure 4F), the model correctly identified 119 true negatives and 178 true positives, misidentifying 20 false positives and 21 false negatives, with a true positive rate of 85.0% and a true negative rate of 89.7%. During the model development process, we considered applying a penalty to the confusion matrix to reduce Type II errors (false negatives). Specifically, we explored methods such as

TABLE 1 Raw data in Three-Baseline table.

Variables	Total (n=429)	Missing Value (%)	Deep Vein Thrombosis Occurrence After Colorectal Surgery		P
			N0 (n=267)	YES (n=162)	
age	67 (57-73)	0	61 (54-71)	71 (65-74)	<0.01
prewbc	6.57 (5.25-8.08)	0	6.51 (5.34-7.58)	6.72 (5.12-8.89)	<0.01
prene	4.31 (3.27-5.77)	0	4.27 (3.28-5.35)	4.44 (3.26-6.38)	<0.01
prelym	1.37 (1.07-1.75)	0.7	1.39 (1.12-1.79)	1.32 (1.02-1.72)	0.06
premon	0.39 (0.3-0.51)	0.7	0.37 (0.3-0.5)	0.4 (0.31-0.53)	0.48
preNLR	68.3 (61.2-75.08)	0	67.4 (60.1-73.7)	70.25 (63.25-76.95)	0.08
prehb	126 (111-140)	0	131 (115-143)	120 (103-129)	<0.01
prepab	194.9 (156.3-230.2)	0.2	207.3 (176.32-242.95)	165.3(136.12-201.9)	<0.01
prealb	41.5 (38.6-44.1)	0	42.3 (39.95-44.9)	39.75 (37.23-42.5)	<0.01
precrci	93.05 (82.77-102.2)	0.2	95.9 (84.65-105.35)	90.8 (80.9-98)	<0.01
preplt	235 (194.75-304)	0.2	230 (191.5-299.75)	246 (198.25-318)	0.1
AFP	2.93 (2.14-3.96)	1.4	2.99 (2.21-4.08)	2.82 (2.02-3.66)	0.12
CEA	5.64 (3.08-15.25)	1.2	5.25 (3.07-13.88)	6.05 (3.18-19.52)	0.08
FER	64.18 (21.79129.28)	2.6	74.44 (26.35128.98)	45.53 (15.5-130.88)	0.91
CA50	8.92 (4.54-18.37)	2.8	8.92 (4.19-15.6)	8.91 (4.83-23.03)	0.13
CA242	5.66 (2.9-14.61)	2.8	5.66 (3.05-13.57)	5.67 (2.63-18.53)	0.3
CA724	2.83 (1.29-7.97)	2.8	2.82 (1.27-8.13)	2.96 (1.38-6.49)	0.46
CA199	12.03 (3.91-24.02)	2.1	11.5 (3.78-21.81)	14.68 (4.34-26.86)	0.03
PIVKA2	23.69 (18.37-31.35)	2.6	24.52 (18.72-31.68)	22.99 (18.26-30.52)	0.46
prept	11.3 (10.9-12)	1.4	11.3 (10.9-11.9)	11.4 (11-12.12)	0.14
prefib	3.69 (3.02-4.27)	1.4	3.58 (3-4.16)	3.83 (3.13-4.49)	<0.01
prett	17 (16.15-17.8)	1.4	17.1 (16.3-17.8)	16.8 (16.1-17.6)	0.08
pred2	0.4 (0.3-0.6)	53.4	0.37 (0.27-0.55)	0.46 (0.33-0.8)	0.95
time	225 (195-265)	0.7	220 (185.5-260)	234 (200-270)	0.01
Blood	50 (20-50)	4	50 (20-50)	50 (20-50)	0.08
BMI	22.77 (20.72-24.97)	4.9	22.77 (20.96-24.84)	22.83 (20.4-25.11)	0.94
Gender					< 0.01
Female	171 (39.86%)	0	86 (32.21%)	85 (52.47%)	
Male	258 (60.14%)		181 (67.79%)	77 (47.53%)	
Region					0.04
Ascending Colon	85 (19.81%)	0	47 (17.60%)	38 (23.46%)	
Transverse Colon	28 (6.53%)		11 (4.12%)	17 (10.49%)	
Descending and Sigmoid Colon	89 (20.75%)		60 (22.47%)	29 (17.90%)	
Upper-middle Rectum	154 (35.90%)		101 (37.83%)	53 (32.72%)	
Lower rectum	73 (17.02%)		48 (17.98%)	25 (15.43%)	

(Continued)

TABLE 1 Continued

Variables	Total (n=429)	Missing Value (%)	Deep Vein Thrombosis Occurrence After Colorectal Surgery		P
			N0 (n=267)	YES (n=162)	
Obstruction					<0.01
No	379 (88.34%)	0	248 (92.88%)	131 (80.86%)	
Yes	50 (11.66%)		19 (7.12%)	31 (19.14%)	
Specialreatment					1
No	312 (72.73%)	0	194 (72.66%)	118 (72.84%)	
Yes	117 (27.27%)		73 (27.34%)	44 (27.16%)	
stag					0.1
Stag0	11 (2.56%)	0	9 (3.37%)	2 (1.23%)	
StagI	30 (6.99%)		22 (8.24%)	8 (4.94%)	
StagII	174 (40.56%)		110 (41.20%)	64 (39.51%)	
StagIII	155 (36.13%)		97 (36.33%)	58 (35.80%)	
StagIV	59 (13.75%)		29 (10.86%)	30 (18.52%)	
tissue					0.82
Intramucosal Carcinoma	6 (1.40%)	0.2	5 (1.88%)	1 (0.62%)	
Highly Differentiated Adenocarcinoma	46 (10.75%)		29 (10.90%)	17 (10.49%)	
Moderately Differentiated Adenocarcinoma	293 (68.46%)		183 (68.80%)	110 (67.90%)	
Poorly Differentiated Adenocarcinoma	24 (5.61%)		14 (5.26%)	10 (6.17%)	
undifferentiated carcinoma	59 (13.79%)		35 (13.16%)	24 (14.81%)	
Tumor Embolus					0.71
No	349 (81.92%)	0.7	216 (81.20%)	133 (83.12%)	
Yes	77 (18.08%)		50 (18.80%)	27 (16.88%)	
Vascular Invasion					0.05
No	323 (75.64%)	0.5	211 (79.03%)	112 (70.00%)	
Yes	104 (24.36%)		56 (20.97%)	48 (30.00%)	
Perineural Invasion					0.76
No	325 (76.11%)	0.5	205 (76.78%)	120 (75.00%)	
Yes	102 (23.89%)		62 (23.22%)	40 (25.00%)	
Microsatellites					0.16
Stable	244 (93.13%)	38.9	156 (95.12%)	88 (89.80%)	
Unstable	18 (6.87%)		8 (4.88%)	10 (10.20%)	
Hypertension					0.23
No	322 (75.41%)	0.5	207 (77.53%)	115 (71.88%)	
Yes	105 (24.59%)		60 (22.47%)	45 (28.12%)	

(Continued)

TABLE 1 Continued

Variables	Total (n=429)	Missing Value (%)	Deep Vein Thrombosis Occurrence After Colorectal Surgery		P
			N0 (n=267)	YES (n=162)	
Diabetes					0.31
No	372 (87.32%)	0.7	237 (88.76%)	135 (84.91%)	
Yes	54 (12.68%)		30 (11.24%)	24 (15.09%)	
CAD					0.39
No	402 (94.59%)	0.9	255 (95.51%)	147 (93.04%)	
Yes	23 (5.41%)		12 (4.49%)	11 (6.96%)	
Pneumonia					0.09
No	387 (90.85%)	0.7	248 (92.88%)	139 (87.42%)	
Yes	39 (9.15%)		19 (7.12%)	20 (12.58%)	
Approach					<0.01
Laparotomy	78 (18.22%)	0.2	36 (13.48%)	42 (26.09%)	
Laparoscopic Surgery	350 (81.78%)		231 (86.52%)	119 (73.91%)	
Fever					0.87
No	336 (78.50%)	0.2	210 (78.95%)	126 (77.78%)	
Yes	92 (21.50%)		56 (21.05%)	36 (22.22%)	
Leak					1
No	290 (97.97)	31	180 (97.83)	110 (98.21)	
Yes	6 (2.03)		4 (2.17)	2 (1.79)	

adjusting the classification threshold and using weighted loss functions to impose a higher penalty on false negatives during model training. However, after several experiments, we found that while these adjustments could reduce false negatives, they also led to an increase in false positives, which in turn affected the overall performance metrics of the model (such as AUC and accuracy). Therefore, we ultimately decided not to apply such penalties to maintain the overall balanced performance of the model. Finally, we plotted Clinical Impact Curves (CICs) to evaluate the net benefit of the model with the highest diagnostic value in terms of clinical utility and applicability. Clinical Impact Curves (Figures 4G, H) offer insights into the model's capability to predict high-risk patients at various cost-benefit ratio thresholds. The test set's curve indicates that when prediction score probabilities exceed 65%, the model's predictions for postoperative colorectal cancer patients align closely with those who actually develop lower extremity deep vein thrombosis, confirming the model's high clinical efficacy.

3.5 Model-based interpretability analysis

This study evaluated the relative importance of various factors affecting the susceptibility of colorectal cancer patients to developing lower extremity deep vein thrombosis post-surgery. Figure 5A visually represents this ranking, with each point indicating a sample and the color gradient from purple to yellow indicates the magnitude of sample feature values. The vertical axis shows the importance ranking of features alongside the correlation and distribution of feature values with SHAP values. Figure 5B illustrates the hierarchical significance of features in the XGB model. The vertical axis shows individual features ranked in descending order of importance, and the horizontal axis represents the average SHAP values. The analysis shows that age, preoperative albumin, preoperative white blood cell count, surgery duration, and preoperative hemoglobin are the top five ranked features in terms of importance, indicating their critical impact on the occurrence of DVT. To better understand the model's decision-making process at the individual level, we performed detailed interpretability analyses using LIME on two representative samples(As illustrated in Figures 5C, D). Through model visualization, we can discern the impact of each feature on the model predictions for these specific instances.

4 Discussion

The migration of deep vein thrombosis from the lower extremities into the pulmonary artery through the circulatory system is a major trigger for fatal pulmonary embolism (10).

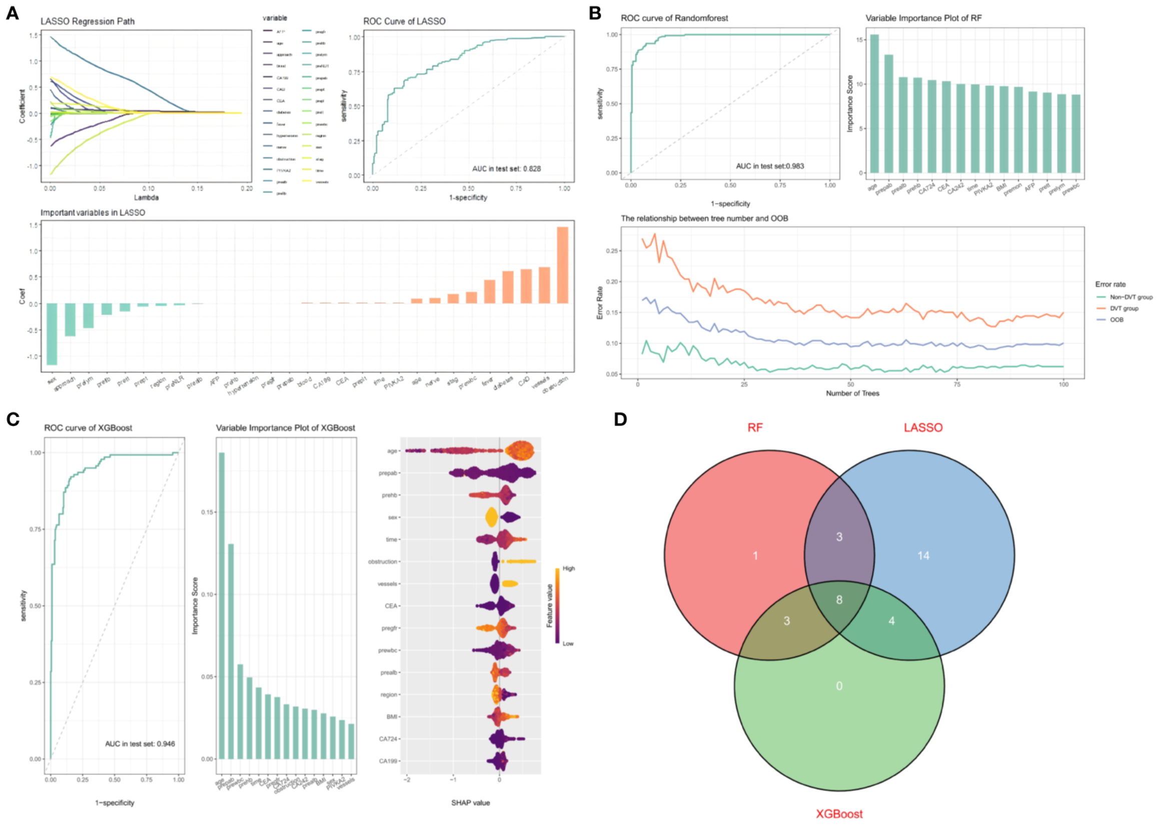


FIGURE 2
(A) AUC curve, path diagram, and importance ranking of selected feature variables from univariate combined with LASSO regression. 1. Penalization process of variables in LASSO. 2. Evaluation of predictive performance of LASSO model in testing set. 3. Feature importance ranking in LASSO model. **(B)** AUC curve, OOB plot, and importance ranking of selected feature variables from random forest. 1. Evaluation of predictive performance of RF model in testing set. 2.Feature importance ranking in RF model. 3. Relationship between number of trees and OOB (Out-of-Bag) error. **(C)** AUC curve, feature importance ranking, and SHAP visualization for XGBOOST model evaluation. 1. Evaluation of predictive performance of XGBOOST model in testing set. 2.Feature importance ranking in XGBOOST model. 3.SHAP value visualization for XGBOOST variables. **(D)** Eight common feature variables selected by three predictive models.

The differences in disease onset and progression characteristics across various specialties result in varying incidence rates of lower extremity DVT (11). Literature reports indicate that the incidence of lower extremity deep vein thrombosis in colorectal cancer patients post-surgery is 40% (2). At present, there is a lack of effective evidence-based research on the risk factors, clinical characteristics, and targeted prevention and treatment measures for lower extremity DVT following gastrointestinal surgery. The American College of Chest Physicians Guidelines define cancer surgery as a high-risk factor for venous thromboembolism and recommend the use of intermittent pneumatic compression and certain medications (such as low molecular weight heparin, low-dose unfractionated heparin, and Xa inhibitors) to prevent the occurrence of venous thromboembolism (7). Caprini, Geneva, and Rapt scores are commonly used tools for assessing DVT, but they are limited in their applicability to colorectal cancer patients. The Caprini assessment rates all colorectal cancer patients undergoing abdominal surgery as high-risk, therefore, current risk assessment models are insufficient to identify patients truly at risk of DVT post-surgery. Many studies have examined the risk factors for postoperative DVT in colorectal cancer patients, such as open surgery, age, D-dimer, pulmonary disease, hemoglobin, and more (12, 13). Although many risk factors have been identified, the available assessment systems are still limited and unable to accurately predict the occurrence of postoperative DVT.

With the continuous advancement of surgical techniques for colorectal cancer, the differences in intraoperative factors are becoming less apparent. Therefore, we aim to develop a preoperative risk assessment tool similar to the Caprini score to facilitate early diagnosis and prevention of postoperative DVT in colorectal cancer patients.

Traditional approaches to identifying risk factors usually depend on developing risk models through univariate or multivariate regression, yet these models often ignore the interactions among variables and nonlinear relationships. In contrast, machine learning models are flexible enough to handle nonlinear and complex data structures, and can effectively address the challenges of high dimensional data and missing values. By training models on large datasets and continuously optimizing their performance, they improve prediction and classification accuracy (14–18). The SHAP algorithm utilizes the Shapley value concept from game theory, calculating the average contribution of each feature to the prediction. This approach enables us to thoroughly quantify each feature's influence on the model's overall predictions,

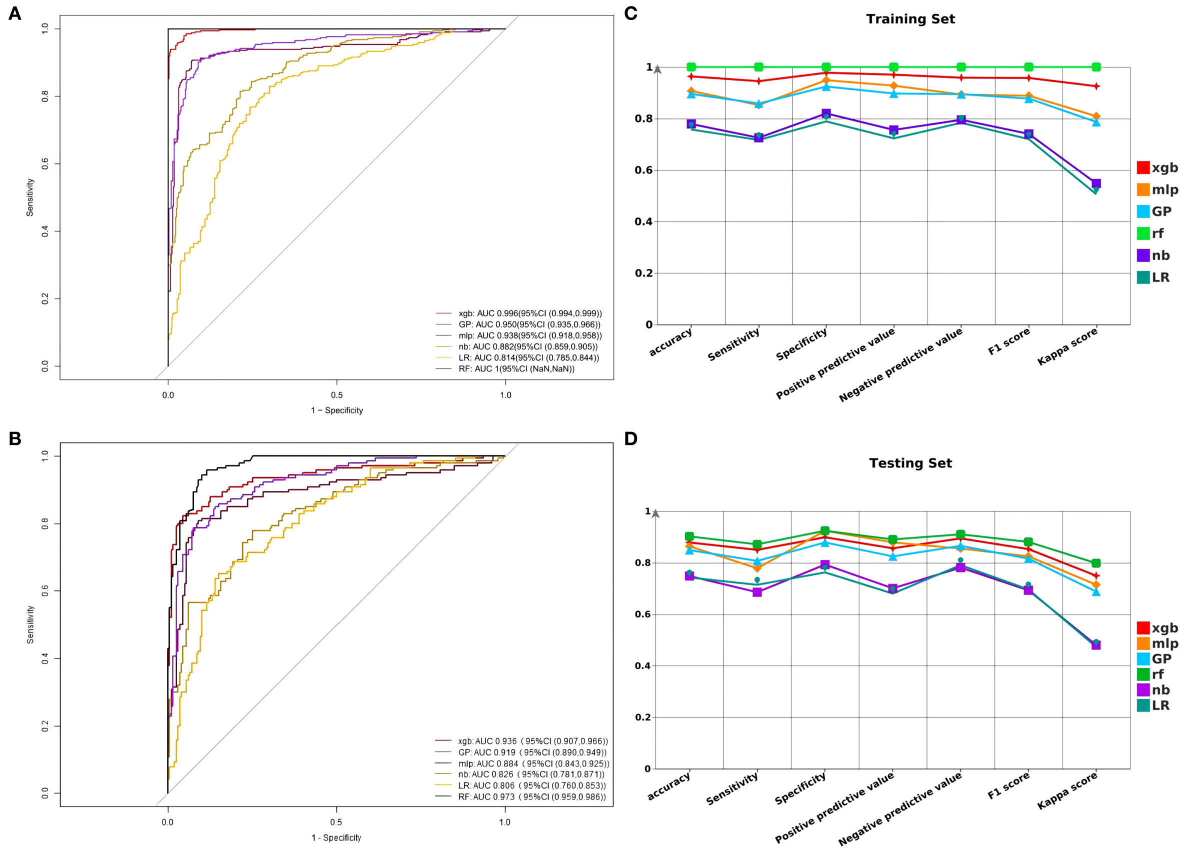


FIGURE 3
(A) Comparison of AUC models in the training set. **(B)** Comparison of AUC models in the testing set. **(C)** Comparison of F1 score, accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and kappa value in the training set. **(D)** Comparison of F1 score, accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and kappa value in the testing set.

thus providing a deeper understanding of the model's workings (19). On the other hand, the LIME algorithm provides localized and transparent explanations by analyzing the feature importance of individual predictions. This local interpretability allows us to understand the reasons behind specific predictions in detail (20). The combination of these two approaches provides us a multidimensional model interpretation framework, capable of capturing global feature impacts and providing thorough insights into specific predictions.

In this study, we first used three machine learning models to construct a prediction model for DVT in patients with gastrointestinal tumors among postoperative colorectal cancer patients. Lasso, Xgboost, and Random Forest each filtered out 29, 15, and 15 feature vectors, respectively. In the end, we selected 8 common feature variables among the three models. During the feature selection process, we adopted a model-based feature selection method. This approach selects the most relevant features by evaluating each feature's contribution to the model's performance. Specifically, we employed algorithms such as Lasso regression, Xgboost, and Random Forest, which effectively handle high-dimensional data and identify features that most significantly impact the prediction results. Existing studies have shown that feature selection plays an important role in cancer prediction models; for example, Sun Tao employed LASSO regression combined with the Boruta algorithm for feature selection, thereby enhancing the accuracy of predicting the risk of pulmonary infection in lung cancer patients post-chemotherapy (21). The ROC curve constructed from these feature vectors indicates that the AUC values for Xgboost and Decision Tree are both greater than 0.900, and the AUC value for Lasso regression is 0.823. The findings indicate that the Lasso, Xgboost, and Decision Tree models have high clinical value in predicting postoperative DVT occurrence in colorectal cancer patients. In contrast, in the research conducted by Xiuying L et al. (22) the DVT model developed through the Caprini Risk Assessment Model exhibited an AUC value of merely 0.701, with a sensitivity of 80.6% and specificity of 56.3%. These comparative results highlight the superiority of the machine learning models in this study, providing powerful tools for accurately predicting postoperative DVT in colorectal cancer patients, indicating that machine learning technology has high potential for application in clinical research. We utilized six machine learning models to build and compare prediction models, from which we selected the optimal model. Through comparison, we found that the XGBOOST model has extremely high prediction accuracy, with an area under the ROC curve larger than 0.99. Additionally, the internally validated DCA and calibration curve confirmed the model's consistency in net clinical benefit and prediction probability, indicating its high predictive value. Literature has shown that the Xgboost model has a higher predictive value for DVT prediction in gastrointestinal tumors, with an AUC value

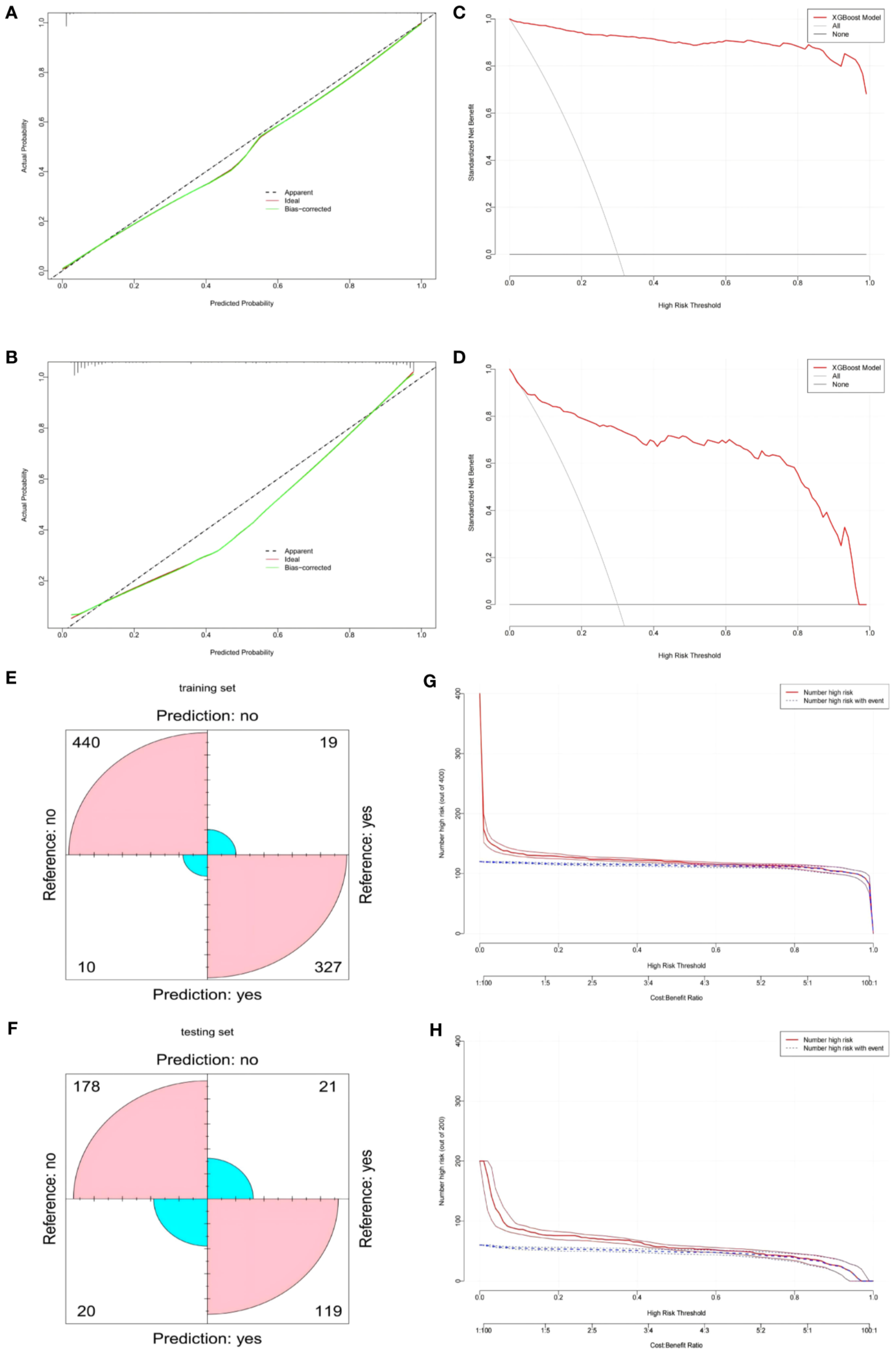


FIGURE 4
(A) XGBOOST model calibration curve in the training set. **(B)** XGBOOST model calibration curve in the testing set. **(C)** XGBOOST model clinical decision curve in the training set. **(D)** XGBOOST model clinical decision curve in the testing set. **(E)** XGBOOST model confusion matrix in the training set. **(F)** XGBOOST model confusion matrix in the testing set. **(G)** XGBOOST Model Clinical Impact Curve (CIC) in the training set. **(H)** XGBOOST model Clinical Impact Curve (CIC) in the testing set.

significantly higher than that of nomograms (23). Additionally, RuifengD et al. (24) constructed a model using the Xgboost model to predict early postoperative DVT in patients after hip surgery. In their study, the Xgboost model achieved an AUC of 0.991 ± 0.012 in the training cohort and an AUC of 0.982 in the validation cohort, with a sensitivity of 0.913 and a specificity of 0.998.The calibration and DCA curves in the validation cohort indicated good performance by the Xgboost model. Our study showed similar performance on these evaluation metrics, validating the model's effectiveness and reliability.

Consistent with some studies (25), advanced age is a predictive factor for VTE occurrence. In our predictive model, SHAP feature importance ranking shows that advanced age is the most important

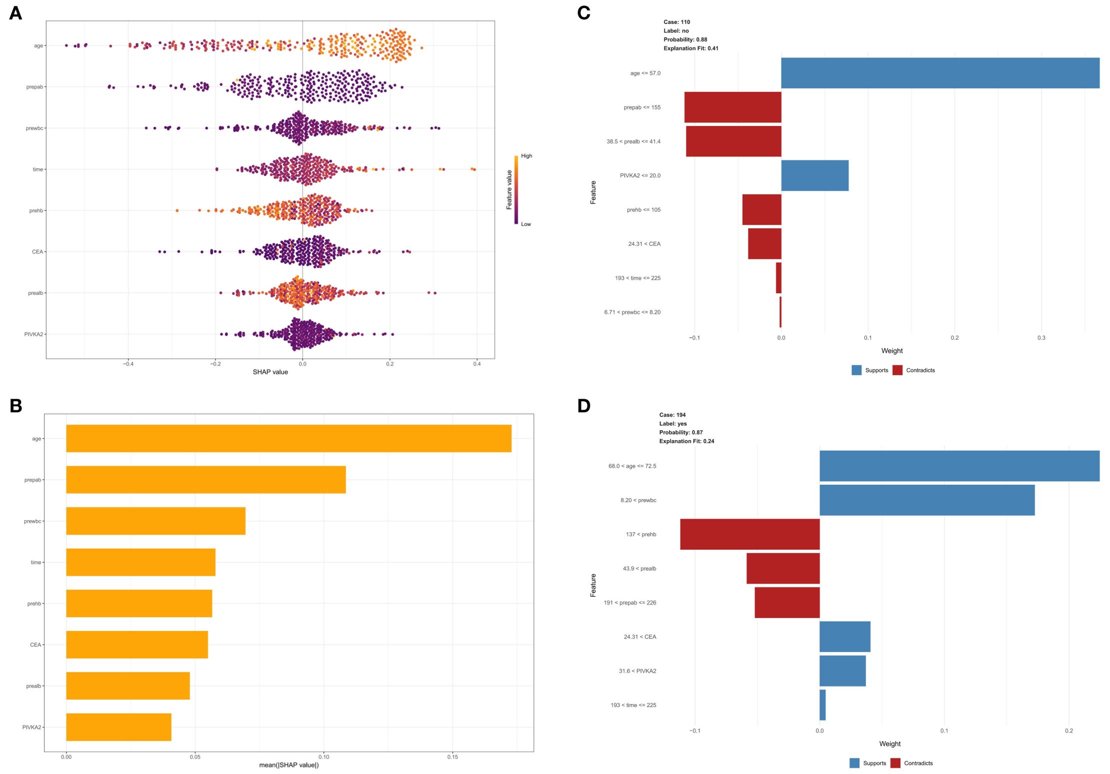


FIGURE 5
(A) SHAP interpretability analysis. The color gradient from purple to yellow represents the magnitude of the sample feature values. The vertical axis displays the importance ranking of features, along with the correlation and distribution of feature values with SHAP values. **(B)** Hierarchical importance ranking of features in the XGBOOST model. **(C, D)** Detailed interpretability analysis of two representative samples using LIME.

predictive factor. This indicates that age plays a crucial role in predicting the risk of VTE occurrence. As age increases, reduced vascular elasticity and changes in coagulation mechanisms can increase the risk of thrombosis. Additionally, reduced activity and the presence of multiple comorbidities in the elderly also increase the likelihood of VTE occurrence.

Prealbumin is a protein synthesized in the liver, commonly used to assess nutritional status and liver function. Its levels can reflect a person's nutritional state and inflammatory response (26, 27). Low levels of prealbumin are often associated with malnutrition, which may increase the risk of DVT (28). Malnutrition can lead to increased blood viscosity and endothelial dysfunction, thereby promoting thrombosis. Meanwhile, prealbumin levels decrease during acute inflammation or infection. The inflammatory response is a crucial mechanism in thrombosis as it can lead to endothelial damage and activation of coagulation factors (29, 30).

Studies have shown that there is a complex relationship between leukocyte activity and venous thrombosis, and the activity of inflammatory cells may play an important role in the natural history of thrombosis (31). Furthermore, research points out that when hematocrit is controlled, an increased white blood cell count (>12) is significantly correlated with the risk of thrombotic events (32). These discoveries highlight the significance of including white blood cell count as a factor in managing VTE, particularly among high-risk groups like surgical and cancer patients.

Our diagnostic tools encompass several additional features, including preoperative hemoglobin, preoperative albumin, CEA, and PIKVA2, all of which are essential preoperative laboratory checks. Additionally, we included surgery duration as a history-related feature. Some features in the tool have SHAP values that are inconsistent with clinical knowledge. However, it is important to consider that these features contribute differently to the overall model and should be viewed as a whole.

Our study has some limitations. Due to the limitations of retrospective studies, we were unable to include some highly valuable data that could be crucial and closely related to colorectal cancer. Despite extensive literature indicating that DD values might be closely linked to the occurrence of postoperative DVT (6, 33), unfortunately, due to a large number of missing values in preoperative DD, it was removed during preprocessing. We anticipate that with the advancements in genetics and bioinformatics, more predictive biomarkers will be identified and utilized, such as tumor genomic features in the Tic-ONCO model (34), among others. Additionally, due to limitations of the constraints of the data system, we could not perform extended observations on patients who were moved to rehabilitation facilities approximately 10 days after surgery. Finally, due to the lack of external validation, it is unclear whether our results are applicable to other populations, necessitating further research on more groups. In summary, these limitations hinder the clinical application of this predictive model, requiring further prospective studies with larger samples and meticulous design. As an initial exploration of this research theme, we hope this study offers some guidance for future prospective research.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by the Medical Ethics Committee of the Affiliated Hospital of Southwest Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because according to national legislation and institutional requirements, participants and their legal guardians/next of kin are not required to provide written informed consent.

Author contributions

XL: Conceptualization, Data curation, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. XS: Investigation, Project administration, Resources, Writing – original draft, Writing – review & editing. YZ: Conceptualization, Data curation, Formal analysis, Resources, Writing – original draft, Writing – review & editing. YJ: Conceptualization, Data curation, Formal analysis, Funding acquisition, Resources, Validation, Visualization, Writing – review & editing, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

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

Publisher's note

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA A Cancer J Clin*. (2021) 71:7–33. doi: 10.3322/caac.21654

2. Bikdeli B, Caraballo C, Trujillo-Santos J, Galanaud JP, di Micco P, Rosa V, et al. Clinical presentation and short- and long-term outcomes in patients with isolated distal deep vein thrombosis vs proximal deep vein thrombosis in the RIETE registry. *JAMA Cardiol*. (2022) 7:857–65. doi: 10.1001/jamacardio.2022.1988

3. Behrendt CA, Twerenbold R, Blankenberg S. The everlasting challenge to identify deep vein thrombosis in both clinical practice and research. *Eur Heart J*. (2022) 43:1882–3. doi: 10.1093/eurheartj/ehac164

4. Mrinalini Tadigiri M, Imam A J, Martins R, Daruwala F. A rare case of rectal carcinoma with pulmonary artery thrombosis. *Cureus*. (2024) 16:e56095. doi: 10.7759/cureus.56095

5. Lyman GH, Bohlke K, Falanga A. Venous thromboembolism prophylaxis and treatment in patients with cancer: American society of clinical oncology clinical practice guideline update. *JOP*. (2015) 11:e442–4. doi: 10.1200/JOP.2015.004473

6. Zhang W, Sun R, Hu X, Chen Z, Lai C. Caprini risk assessment model combined with D-dimer to predict the occurrence of deep vein thrombosis and guide intervention after laparoscopic radical resection of colorectal cancer. *World J Surg Onc*. (2023) 21:299. doi: 10.1186/s12957-023-03183-7

7. Gould MK, Garcia DA, Wren SM, Arcelus JI, Heit JA, Samama CM. Prevention of VTE in nonorthopedic surgical patients. *Chest*. (2012) 141:e227S–77S. doi: 10.1378/chest.11-2297

8. Rafique R, Islam SR, Kazi JU. Machine learning in the prediction of cancer therapy. *Comput Struct Biotechnol J*. (2021) 19:4003. doi: 10.1016/j.csbj.2021.07.003

9. Needleman L, Cronan JJ, Lilly MP, Merli GJ, Adhikari S, Hertzberg BS, et al. Ultrasound for lower extremity deep venous thrombosis: multidisciplinary recommendations from the society of radiologists in ultrasound consensus conference. *Circulation*. (2018) 137:1505–15. doi: 10.1161/circulationaha.117.030687

10. Barrosse-Antle ME, Patel KH, Kramer JA, Baston CM. Point-of-care ultrasound for bedside diagnosis of lower extremity DVT. *CHEST*. (2021) 160:1853–63. doi: 10.1016/j.chest.2021.07.010

11. Panpikoon T, Chuntaroj S, Treesit T, Chansanti O, Bua-ngam C. Lower-extremity venous ultrasound in DVT-unlikely patients with positive D-dimer test. *Acad Radiology*. (2022) 29:1058–64. doi: 10.1016/j.acra.2020.06.028

12. Wei Q, Wei ZQ, Jing CQ, Li YX, Zhou DB, Lin MB, et al. Incidence, prevention, risk factors, and prediction of venous thromboembolism in Chinese patients after colorectal cancer surgery: a prospective, multicenter cohort study. *Int J Surg*. (2023) 109:3003–12. doi: 10.1097/JS9.0000000000000553

13. Moghadamyeghaneh Z, Hanna MH, Carmichael JC, Nguyen NT, Stamos MJ. A nationwide analysis of postoperative deep vein thrombosis and pulmonary embolism in colon and rectal surgery. *J Gastrointestinal Surgery*. (2014) 18:2169–77. doi: 10.1007/s11605-014-2647-5

14. Chen K, Shiomi A, Kagawa H, Matsuda T, Tanaka Y, Yamamoto S, et al. Efficacy of a robotic stapler on symptomatic anastomotic leakage in robotic low anterior resection for rectal cancer. *Surg Today*. (2024) 54:1–10. doi: 10.1007/s00595-021-02313-6citeas

15. Mponponsuo K, Leal J, Spackman E, Somayaji R, Gregson D, Rennert-May E. Mathematical model of the cost-effectiveness of the BioFire FilmArray Blood Culture Identification (BCID) Panel molecular rapid diagnostic test compared with conventional methods for identification of *Escherichia coli* bloodstream infections. *J Antimicrobial Chemotherapy*. (2022) 77:507–16. doi: 10.1093/jac/dkab398

16. Johnson PM, Lin DJ, Zbontar J, Zitnick CL, Sriram A, Muckley M, et al. Deep learning reconstruction enables prospectively accelerated clinical knee MRI. *Radiology*. (2023) 307:e220425. doi: 10.1148/radiol.220425

17. Aromolaran O, Aromolaran D, Isewon I, Oyelade J. Machine learning approach to gene essentiality prediction: a review. *Briefings Bioinf*. (2021) 22:1–10. doi: 10.1093/bib/bbab128

18. Garriga R, Mas J, Abraha S, Harrison O, Tadros G, Matic A. Machine learning model to predict mental health crises from electronic health records. *Nat Med*. (2022) 28:1240–8. doi: 10.1038/s41591-022-01811-5

19. Sun J, Sun CK, Tang YX, Liu TC, Lu CJ. Application of SHAP for explainable machine learning on age-based subgrouping mammography questionnaire data for positive mammography prediction and risk factor identification. *Healthcare (Basel)*. (2023) 11:2000. doi: 10.3390/healthcare11142000

20. Mateussi N, Rogers MP, Grimsley EA, Read M, Parikh R, Pietrobon R, et al. Clinical applications of machine learning. *Ann Surg Open*. (2024) 5:e423. doi: 10.1097/AS9.0000000000000423

21. Sun T, Liu J, Yuan H, Li X, Yan H. Construction of a risk prediction model for lung infection after chemotherapy in lung cancer patients based on the machine learning algorithm. *Front Oncol*. (2024) 14:1403392. doi: 10.3389/fonc.2024.1403392

22. Lu X, Zeng W, Zhu L, Liu L, Du F, Yang Q. Application of the Caprini risk assessment model for deep vein thrombosis among patients undergoing laparoscopic surgery for colorectal cancer. *Med (Baltimore)*. (2021) 100:e24479. doi: 10.1097/MD.0000000000024479

23. Zhang Y, Ma Y, Wang J, Guan Q, Yu B. Construction and validation of a clinical prediction model for deep vein thrombosis in patients with digestive system tumors based on a machine learning. *Am J Cancer Res*. (2024) 14:155–68. doi: 10.62347/LNDL8700

24. Ding R, Ding Y, Zheng D, Huang X, Dai J, Jia H, et al. Machine learning-based screening of risk factors and prediction of deep vein thrombosis and pulmonary embolism after Hip arthroplasty. *Clin Appl Thromb Hemost*. (2023) 29:10760296231186145. doi: 10.1177/10760296231186145

25. Tan WJ, Chen L, Yang SJ, Zhang BY, Sun ML, Lin YB, et al. Development and validation of a prediction model for venous thrombus embolism (VTE) in patients with colorectal cancer. *Technol Cancer Res Treat*. (2023) 22:1–10. doi: 10.1177/15330338231186790

26. Wang QR, Long J, Wang CC, Hu JL, Lin N, Tang SH. Case report of atypical undernutrition of hypoproteinemia type. *Open Life Sci*. (2023) 18:20220766. doi: 10.1515/biol-2022-0766

27. Loftus TJ, Brown MP, Slish JH, Rosenthal MD. Serum levels of prealbumin and albumin for preoperative risk stratification. *Nutr Clin Pract*. (2019) 34:340–8. doi: 10.1002/ncp.10271

28. Truong A, Hanna MH, Moghadamyeghaneh Z, Stamos MJ. Implications of preoperative hypoalbuminemia in colorectal surgery. *World J Gastrointest Surg*. (2016) 8:353–62. doi: 10.4240/wjgs.v8.i5.353

29. López B, Castañón-Apilánez M, Molina-Gil J, Fernández-Gordón Sánchez S, González G, Reguera Acuña A, et al. Serum prealbumin levels on admission as a prognostic marker in stroke patients treated with mechanical thrombectomy. *Cerebrovasc Dis Extra*. (2022) 12:102–7. doi: 10.1159/000526354

30. Bittar LF, da Silva LQ, Orsi FL de A, Zapponi KCS, Mazetto de B M, de Paula EV, et al. Increased inflammation and endothelial markers in patients with late severe post-thrombotic syndrome. *PloS One*. (2020) 15:e0227150. doi: 10.1371/journal.pone.0227150

31. Saha P, Humphries J, Modarai B, Mattock K, Waltham M, Evans CE, et al. Leukocytes and the natural history of deep vein thrombosis: current concepts and future directions. *Arterioscler Thromb Vasc Biol*. (2011) 31:506–12. doi: 10.1161/ATVBAHA.110.213405

32. Cutsem EV, Mahé I, Felip E, Agnelli G, Awada A, Cohen A, et al. Treating cancer-associated venous thromboembolism: A practical approach. *Eur J Cancer*. (2024) 209:1–10. doi: 10.1016/j.ejca.2024.114263

33. Wu Y, Wang L, Yin Q, Deng L, Ma J, Tian X. Establishment and validation of a postoperative VTE prediction model in patients with colorectal cancer undergoing radical resection: CRSPOT nomogram. *Clin Appl Thromb Hemost*. (2023) 29:10760296231216966. doi: 10.1177/10760296231216966

34. Muñoz Martín AJ, Ortega I, Font C, Pachón V, Castellón V, Martínez-Marín V, et al. Multivariable clinical-genetic risk model for predicting venous thromboembolic events in patients with cancer. *Br J Cancer*. (2018) 118:1056–61. doi: 10.1038/s41416-018-0027-8

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 19 December 2024
DOI 10.3389/fimmu.2024.1493528

Exploration and validation of a novel reactive oxygen species–related signature for predicting the prognosis and chemotherapy response of patients with bladder cancer

Yulei Li[1†], Lulu Zhang[2†], Gang Xu[3†], Gang Xu[1], Jiajun Chen[1], Keyuan Zhao[1], Mengyao Li[4], Jing Jin[1], Chao Peng[1], Kaifang Wang[5], Shouhua Pan[1*] and Ke Zhu[3*]

[1]Department of Urology, Shaoxing People's Hospital, Zhejiang, Shaoxing, China, [2]Medical Research Center, Shaoxing People's Hospital, Zhejiang, Shaoxing, China, [3]Department of Urology, Nanchang People's Hospital, Nanchang, China, [4]Department of Pathology, Shaoxing People's Hospital, Zhejiang, Shaoxing, China, [5]Cancer Centre, Faculty of Health Sciences, University of Macau, Macau, Macao SAR, China

Background: Reactive Oxygen Species (ROS), a hallmark of cancer, is related to prognosis, tumor progression, and treatment response. Nevertheless, the correlation of ROS-based molecular signature with clinical outcome and immune cell infiltration has not been thoroughly studied in bladder cancer (BLCA). Accordingly, we aimed to thoroughly examine the role and prognostic value of ROS-related genes in BLCA.

Methods: We obtained RNA sequencing and clinical data from The Cancer Genome Atlas (TCGA) for bladder cancer (BLCA) patients and identified ROS-associated genes using the GeneCards and Molecular Signatures Database (MSigDB). We then analyzed differential gene expression between BLCA and normal tissues and explored the functions of these ROS-related genes through Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Protein-Protein Interaction (PPI) analysis. Prognostic ROS-related genes were identified using Univariate Cox regression (UCR) and LASSO analyses, which were further refined in a Multivariate Cox Regression (MCR) analysis to develop a Prognostic Signature (PS). This PS was validated in the GSE13507 cohort, assessing its predictive power with Kaplan-Meier survival and time-dependent ROC curves. To forecast BLCA outcomes, we constructed a nomogram integrating the PS with clinical variables. We also investigated the signature's molecular characteristics through Gene Set Enrichment Analysis (GSEA), Immune Cell Infiltration (ICI), and Tumor Mutational Burden (TMB) analyses. The Genomics of Drug Sensitivity in Cancer (GDSC) database was used to predict chemotherapy responses based on the PS. Additionally, we screened for Small-Molecule Drugs (SMDs) targeting ROS-related genes using the CMAP database. Finally, we validated our findings by checking protein levels of the signature genes in the Human Protein Atlas (HPA) and confirmed the role of Aldo–keto reductase family 1 member B1 (AKR1B1) through *in vitro* experiments.

EDITED BY
Tinka Vidovic,
University of Zagreb, Croatia

REVIEWED BY
Hadi Esmaili Gouvarchinghaleh,
Baqiyatallah University of Medical Sciences, Iran
Jiajia Huang,
Sun Yat-sen University Cancer Center (SYSUCC), China
Xin Yu,
Baylor College of Medicine, United States
Yu Yan,
University of North Carolina at Chapel Hill, United States

*CORRESPONDENCE
Ke Zhu
✉ 18357509340@163.com
Shouhua Pan
✉ 13606550587@163.com

†These authors have contributed equally to this work

RECEIVED 09 September 2024
ACCEPTED 03 December 2024
PUBLISHED 19 December 2024

CITATION
Li Y, Zhang L, Xu G, Xu G, Chen J, Zhao K, Li M, Jin J, Peng C, Wang K, Pan S and Zhu K (2024) Exploration and validation of a novel reactive oxygen species–related signature for predicting the prognosis and chemotherapy response of patients with bladder cancer. *Front. Immunol.* 15:1493528.
doi: 10.3389/fimmu.2024.1493528

Results: The constructed and validated PS that comprised 17 ROS-related genes exhibited good performance in predicting overall survival (OS), constituting an independent prognostic biomarker in BLCA patients. Additionally, we successfully established a nomogram with superior predictive capacity, as indicated by the calibration plots. The bioinformatics analysis findings showcased the implication of PS in several oncogenic pathways besides tumor ICI regulation. The PS was negatively associated with the TMB. The high-risk group patients had greater chemotherapy sensitivity in comparison to low-risk group patients. Further, 11 candidate SMDs were identified for treating BLCA. The majority of gene expression exhibited a correlation with the protein expression. In addition, the expression of most genes was consistent with protein expression. Furthermore, to test the gene reliability we constructed, AKR1B1, one of the seventeen genes identified, was used for in-depth validation. *In vitro* experiments indicate that siRNA-mediated AKR1B1 silencing impeded BLCA cell viability, migration, and proliferation.

Conclusions: We identified a PS based on 17 ROS-related genes that represented independent OS prognostic factors and 11 candidate SMDs for BLCA treatment, which may contribute to the development of effective individualized therapies for BLCA.

KEYWORDS

bladder cancer, reactive oxygen species, prognostic signature, chemotherapy response, overall survival, AKR1B1

1 Introduction

Bladder cancer (BLCA) has the sixth worldwide prevalence of new cases and the ninth-highest number of fatalities among male cancer patients globally. In 2020, there were nearly 573,000 new cases and nearly 213,000 deaths caused by BLCA (1, 2). Based on the depth of muscle invasion, BLCA can be mainly classified into non-muscle-invasive BLCA (NMIBC) and muscle-invasive BLCA (MIBC) (3). Despite remarkable advancements in treatments, including adjuvant chemotherapy, immune checkpoint inhibitor therapy, robot-assisted surgery systems, and targeted therapy, the overall survival (OS) of BLCA patients remains unfavorable (4–6). In addition, BLCA is a cancerous malignancy with notable and substantial heterogeneity, and conventional clinical predictive factors, including tumor grade and TNM stage, can be utilized for predicting BLCA patient prognosis accurately (7). Hence, identifying novel biomarkers for predicting the BLCA patient survival time is of crucial practical clinical significance.

Reactive oxygen species (ROS), characterized by molecules that contain oxygen with oxidizing properties, are the reduction products of oxidative metabolism and consist of nonradicals, mainly hydrogen peroxide (H2O2), hypochlorous acid (HOC1), and organoid hydroperoxides (ROOH), and free radicals, mainly hydroxyl and superoxide anion radicals (8). Mitochondria, peroxisomes, the endoplasmic reticulum (ER), metabolic enzymes, and the Warburg effect are the main endogenous sources of ROS (9). ROS can also be produced by physical agent exposure (ultraviolet rays and heat), chemotherapy, and radiotherapy (10, 11). ROS has been indicated to be crucial secondary messengers governing various cellular biological processes, including proliferation, angiogenesis, differentiation, metastasis, autophagy, drug resistance, immune response, and cancer stem cells (12). Moderate ROS levels are believed to be essential for cell growth and differentiation. Nevertheless, the excessive accumulation of ROS is involved in multiple diseases (13), particularly malignant tumors (14, 15). Recent studies have indicated that an imbalance in ROS is closely related to BLCA development and progression (16, 17). Therefore, comprehensively investigating the functions of ROS-related genes and identifying ROS-related biomarkers to accurately predict BLCA patients' OS is highly important.

The relationship between genes and reactive oxygen species (ROS) is multifaceted, including the regulation of ROS production and clearance by genes, and the influence of ROS on gene expression (18). Here are some key points that outline the interaction between genes and ROS: (1) Regulation of ROS by Genes. Genes such as p53 play a critical role in maintaining genomic integrity and orchestrating cellular responses to stress, including the modulation of ROS activity. ROS can act as signaling molecules to initiate p53 activation in response to DNA damage, leading to transcriptional regulation of genes involved in cell cycle arrest, DNA repair, and apoptosis (19, 20). (2) ROS Influence on Gene Expression. The Keap1-Nrf2-ARE signaling pathway is a well-studied regulatory system that preserves cellular redox homeostasis (21, 22). ROS act as central players in this mechanism, providing a dynamic balance between Nrf2 activation and its inhibition by Keap1. When cellular ROS levels rise, certain cysteine residues in Keap1 are oxidized, disrupting its ability to ubiquitinate Nrf2, leading to the accumulation of Nrf2 in the nucleus and the transcriptional activation of antioxidant and detoxification genes (22). (3) ROS and Chromatin. ROS influence the activity of epigenetic modulators, such as histone deacetylases (HDACs) or DNA methyltransferases (DNMTs), affecting the expression of target genes. They also oxidize DNA, particularly adenine and guanine, which can lead to mutations and contribute to tumorigenesis (23, 24). (4) ROS and Cancer. In cancer therapy, ROS can either activate or suppress NF-kB signaling involved in the control of cellular processes such as embryogenesis, cell proliferation and death, and responses to stress stimuli (21). Additionally, ROS can induce DNA hypermethylation, potentially affecting tumor phenotype when promoter regions of tumor suppressor genes are involved (25).

Our study comprehensively investigated the functions and prognostic values of ROS-associated genes in BLCA by accessing a public database via bioinformatics methods, aiming at constructing and validating a novel Prognostic Signature (PS) relying on ROS-related genes in BLCA through LASSO and Cox regression analyses. We also explored the associations between PS and Immune Cell Infiltration (ICI), Tumor Mutational Burden (TMB), and chemosensitivity. A nomogram was established by combining the Risk Scores (RSs) based on the seventeen prognostic ROS-associated genes and clinical characteristics. Additionally, we identified 11 candidate Small-Molecule Drugs (SMDs) for BLCA treatment. To verify the authenticity of the data, *in vitro* experiments revealed that siRNA-mediated AKR1B1 silencing impeded BLCA cell viability, migration, and proliferation, aligning with our expectations and demonstrating the constructed ROS-related gene reliability. We identified a PS based on 17 ROS-related genes that represented independent OS prognostic factors and 11 candidate SMDs for BLCA treatment, which may contribute to the development of effective individualized therapies for BLCA.

2 Methods and methods

2.1 Data acquisition

We first obtained ROS-related genes from the GeneCards database (https://www.genecards.org/) and Molecular Signature Database v7.1 (MSigDB; https://www.gsea-msigdb.org/gsea/msigdb). Then, we downloaded the level-three transcriptome RNA sequencing information and clinicopathological features of BLCA patients by accessing The Cancer Genome Atlas (TCGA) (https://gdc-portal.nci.nih.gov/). Further, we utilized the GSE13507 acquired from the Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/) as the validation set.

2.2 Identification of ROS-associated differentially expressed genes

Employing the R edge package (version R 4.0.5, https://bioconductor.org/packages/release/bioc/), the ROS-related DEGs between BLCA and normal bladder samples were screened, setting the cutoff criteria as a False Discovery Rate (FDR) < 0.05 and a |log2-fold change (FC)| > 1.

2.3 Enrichment analysis of ROS-related DEGs

Gene Ontology (GO) analysis that includes molecular function (MF), cell component (CC), and biological process (BP) analyses was implemented to explore the possible molecular mechanisms behind ROS-related DEGs via the clusterProfiler package of R, utilizing the same approach for Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis (26–28) and considering $P < 0.05$ as significant enrichment.

2.4 Protein-protein interactions

ROS-related DEGs were uploaded to the STRING database (http://www.string-db.org/) to obtain PPI information. The PPI network establishment and visualization were conducted via Cytoscape software, using the MCODE plug-in to screen the considerable PPI network modules.

2.5 Identification of potential small-molecule drugs

The Connectivity Map (CMAP) database (http://www.broadinstitute.org) could be beneficial for researchers in the identification of probable molecular drugs closely associated with diseases, including cancer. The enrichment scores were -1–1, with a negative score showing that BLCA patients could benefit from this drug.

Abbreviations: BLCA, Bladder Cancer; ROS, Reactive Oxygen Species; MSigDB, Molecular Signature Database; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; KEGG, Kyoto Encyclopedia of Genes and Genomes; FDR, False Discovery Rate; FC, Fold Change; GO, Gene Ontology; PPI, Protein-Protein Interaction; UCR, Univariate Cox Regression; MCR, Multivariate Cox Regression; PS, Prognostic Signature; GSEA, Gene Set Enrichment Analysis; ICI, Immune Cell Infiltration; TMB, Tumor Mutational Burden; GDSC, The Genomics of Drug Sensitivity in Cancer; SMDs, Small-Molecule Drugs; HPA, Human Protein Atlas; AKR1B1, Aldo–keto reductase family 1 member B1.

2.6 Construction and validation of the prognostic signature of ROS

The prognosis-associated ROS-related genes were identified via Univariate Cox regression (UCR) analysis (survival package) and least absolute shrinkage and selection operator (LASSO) regression analysis (glmnet and survival package) with $P < 0.05$ in the TCGA dataset, followed by incorporating the results into the Multivariate Cox Regression (MCR) analysis. Finally, a ROS-correlated gene signature related, to OS was constructed based on MCR analysis results. The Risk Score (RS) was generated by this formula: RS = (Coef1*expression mRNA1) + (Coef2*expression mRNA2) + (Coef n * expression mRNA n), where Coef represents the MCR model coefficient of relevant mRNA. Based on the RS mean, patients were classified into high-risk group (HRG) and low-risk group (LRG), employing the Kaplan-Meier (K-M) method to compare the survival outcomes between different groups. Our study deployed time-related ROC analysis to determine the predictive prognostic value of the PS. Both T-distributed stochastic neighbor embedding (t-SNE) analysis alongside principal component analysis (PCA) were implemented to examine the risk signature classification capacity with the R packages “Rtsne” and “ggplot2”, employing the same approach to calculate the RS and then validated the ROS-related gene signature in the GSE13507 dataset.

2.7 Development of a nomogram

We explored the relationships between the PS and clinical features (age, sex, T/N/TNM stages, and tumor grade) in the TCGA dataset via the chi-squared test. Then, stratified analysis was performed to further examine the PS reliability and stability of ROS in the prediction of BLCA patients’ OS. Additionally, we implemented UCR and MCR analyses to explore whether the RS was of independent prognostic value. Both RS and clinical features were incorporated to establish an OS-related nomogram, estimating the nomogram’s predictive capability by generating a calibration curve.

2.8 Gene set enrichment analysis and immune cell infiltration and tumor mutational burden analyses

GSEA was implemented to investigate the latent mechanisms among different groups based on GSEA software (version 4.1.0). Then, we acquired mutation information for BLCA patients by accessing the TCGA database, calculated the total mutation number for each sample, and analyzed the top mutational genes among different risk groups using the maftools package. The TIMER, CIBERSORT, CIBERSORT-ABS, XCELL, QUANTISEQ, EPIC, and MCP-counter methods were utilized for the analysis of the ICI levels of 22 distinct leukocyte subsets in both groups. $P < 0.05$ indicated statistically significant.

2.9 Chemotherapeutic response analysis

The Genomics of Drug Sensitivity in Cancer (GDSC, http://www.cancerrxgene.org) database was accessed to predict BLCA patients’ response in both groups to chemotherapy drugs. Eventually, we assessed chemosensitivity by calculating the half-maximal inhibitory concentration (IC50) through the R package pRRophetic, with $P < 0.05$ indicating statistical significance.

2.10 Patient sample

Between 2022 and 2024, 20 BLCA tissue and their corresponding non-tumor tissue specimens were collected from Shaoxing People’s Hospital for immunohistochemical staining (IHC) and western blot. No patient in this study had received radiation therapy, adjuvant therapy, or preoperative chemotherapy. The samples from Shaoxing People’s Hospital were collected with informed consent, and the use of the stored cancer specimens and clinical data was granted clearance by the Academic Ethical Committee of Shaoxing People’s Hospital (ethical approval number: 2022-K-Y-054-01). The study was executed in a way that aligned with the Declaration of Helsinki.

2.11 Immunohistochemistry

IHC images of key genes in BLCA and normal tissue samples were acquired through the Human Protein Atlas (HPA) database while evaluating the staining intensity following the HPA database standard (https://www.proteinatlas.org/). Use anti-AKR1B1 (1:1000; Proteintech,15439-1-AP). Rabbit monoclonal antibody was used for immunohistochemistry of paraffin-embedded human and nude mouse BLCA specimens. In short, samples were processed using dewaxing, hydration, antigen extraction, IHC labeling, and pathology scores.

2.12 Cell culture, treatments, and siRNA transfection

Human BLCA cells (T24 and 5637) were procured from Procell Life Science & Technology Company (Hubei, China). Herein, we grouped the logarithmic growth phase cells into the control, siAKR1B1-negative control (NC), and siAKR1B1 groups. The two cell lines were cultured in MCCOY’S 5A (Gibco, USA) and 1640 (Gibco, USA) medium supplemented with 10% fetal bovine serum (FBS, Gibco, USA) and 1% penicillin/streptomycin at 37°C and 5% CO_2. The cells went through treatment with 5 μl of siAKR1B1, using Lipofectamine 2000 to dilute the solution in Opti-MEM for 5 min. The solution was thereafter mixed and allowed to incubate at ambient temperature for a duration of 20 min, followed by introducing the composite into the cell culture plate. After a 48-h period of transfection, the cells were gathered for additional assessments.

2.13 Western blot analysis

Proteins were subjected to extraction using RIPA buffers and quantification by BCA kits. The protein eluate went through separation utilizing 10% SDS-PAGE and transferred to a PVDF membrane that was blocked and then incubated with primary and secondary antibodies. AKR1B1 (Proteintech, 15439-1-AP) and β-catenin (Abcam, ab32572) were detected by imaging with enhanced chemiluminescence reagents (Merck Millipore, Billerica, MA).

2.14 CCK8 assay

100 μL of suspension containing 5000 transfected cells was dispensed into each well of a 96-well plate along with 10 μL of CCK8 solution (MCE, HY-K0301). The plate was then placed in a cell culture incubator for 1 hour, following which absorbance readings were taken at 450 nm and recorded.

2.15 Colony formation assay

3600 BLCA cells were equally distributed into six-well plates and incubated at 37°C with 5% CO_2 for 14 days with regular medium changes. Post-incubation, the cells were fixed and stained using 4% paraformaldehyde and 0.1% crystal violet for 20 minutes each, after which images were captured and data documented.

2.16 Edu assay

5-ethynyl-2′-deoxyuridine (EdU) assay kit (MCE, China) was used as instructed by the manufacturer. In this experiment, BLCA cells were cultured in 96-well plates, with a seeding density of 4,000 cells per well, after incubation at 37°C for 72 hours. Next, BLCA cells were exposed to 10 μM EdU for 2 hours at 37°C. Subsequently, the cells were fixed using 4% paraformaldehyde and permeabilized with 0.5% Triton X-100 for 15 minutes at room temperature. After removing the fixatives, the cells were washed with PBS containing 1% BSA. Lastly, the cells were incubated in Click Additive Solution, protected from light, for 30 minutes and then stained with Hoechst to label the nucleus. Microscopic images were captured to observe the EdU detection samples. The proliferation of cells was further assessed by calculating the ratio of EdU-positive cells to the overall cell count.

2.17 Transwell assay

The transfected HOS and 143B cell lines were cultured with the serum-free DMEM and serum-free 1640, respectively, in a Transwell upper chamber. Corresponding culture medium containing 10% FBS was added to the lower chamber. The cells were incubated at 37°C with 5% CO2 for 48 hours, fixed with formaldehyde, stained with crystal violet, and visualized under a microscope for analysis.

2.18 Statistical analysis

Statistical analysis was conducted using the R software (version 4.0.5). The significance of differential gene expression was ascertained using adjusted p-value to correct for the multiple testing phenomenon, with a significance threshold set at p-value < 0.05. Another statistical analysis was conducted with SPSS Statistics software version 20. The values were compared by one-way ANOVA or independent-samples Student's t test. Statistical significance was determined at $^*p < 0.05$, $^{**}p < 0.01$, or $^{***}p < 0.001$. Values are presented as the mean ± SEM. Error bars indicate the SEM unless otherwise noted.

3 Results

3.1 Identification of ROS-related genes in BLCA

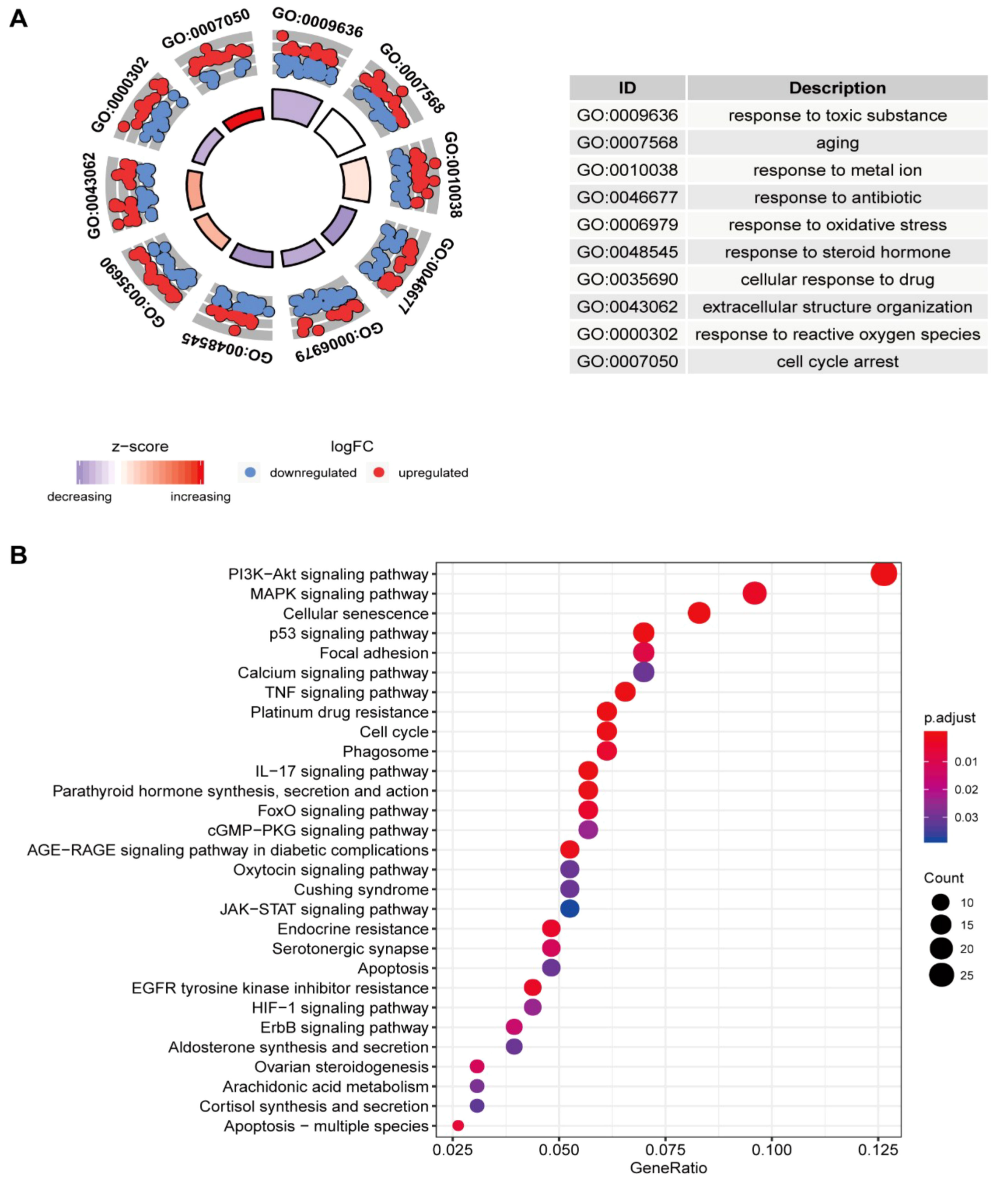
Supplementary Figure 1 illustrates the workflow diagram of this study. In total, we obtained 1749 ROS-related genes with relevance scores > 0.5 from the Gene Cards database and 70 ROS-related genes from the MSigDB database, acquiring 1767 genes after removing the overlapping genes. However, from the 1,767 genes, we eventually extracted the expression profiles of 1,719 ROS-associated genes identical to those in 412 and 19 BLCA and normal bladder tissue samples, respectively, in the TCGA dataset. By applying cutoff criteria of FDR < 0.05 and |log2 FC| > 1, 308 ROS-related Differentially Expressed Genes (DEGs) were identified; of them, 138 were downregulated, and 170 were upregulated. Moreover, GO, KEGG, and PPI analyses were deployed to explore the possible roles of ROS-associated genes. Both Univariate Cox regression (UCR) and LASSO analyses were performed to screen for prognostic ROS-related genes, and 71 genes were included in subsequent analyses (p<0.05). As a means to guarantee the clinical outcomes' stability and reliability based on the 71 genes, we conducted LASSO analysis to further screen for prognostic ROS-related genes, identifying 31 genes related to OS. The MCR analysis identified 17 ROS-related genes (JUN, CALR, P4HB, ELN, MYC, FASN, REV3L, VHL, NID1, SLC38A1, TFRC, AKR1B1, ITGA3, CGB5, HLA-G, FADS1, and ORM1) that were utilized to construct a PS, which was subsequently validated in the GSE13507 cohort. Both K–M survival and time-dependent receiver operating characteristic (ROC) curves were employed to evaluate the prognostic value of the PS. A nomogram was constructed, aiming at predicting the outcomes of BLCA patients in combination with the PS and clinical factors. GSEA, ICI, and TMB analysis were implemented for the exploration of the molecular characteristics of the PS. The GDSC database was accessed for the prediction of chemotherapy response according to the PS. Candidate SMDs targeting ROS-related genes were screened against the CMAP database. To verify the authenticity of the data, the PS protein expression levels were detected through the HPA. AKR1B1 was selected for *in vitro* experimental validation, demonstrating the reliability of the ROS-related genes we constructed.

3.2 Functional assays of the selected prognostic genes and Protein-protein interaction network construction

The GO analysis findings represented that ROS-associated genes were involved in multiple biological processes, including the response to toxic substances, aging, metal ions, oxidative stress, and ROS, besides cell cycle arrest and the cellular response to drugs (Figure 1A). The KEGG analysis findings showcased that these genes exhibited main involvement in multiple pathways, including the p53, platinum drug resistance, cell cycle, ErbB, PI3K-Akt, TNF, cellular senescence, IL-17, MAPK, HIF-1, and cGMP-PKG signaling pathways (Figure 1B) (26–28). For better comprehension of the involvements of ROS-associated genes in BLCA, a PPI network was established and visualized through the utilization of STRING database and Cytoscape software, which included 298 nodes and 2859 edges (Figure 2A). The MCODE plugin identified three crucial modules of target genes, and the critical modules consisted of 39 nodes and 321 edges, 29 nodes and 250 edges, and 31 nodes and 123 edges (Figures 2B–D).

3.3 Small-molecule drugs

Using the CMAP database, candidate SMDs for BLCA were identified based on ROS-related DEGs, identifying eleven SMDs



ID	Description
GO:0009636	response to toxic substance
GO:0007568	aging
GO:0010038	response to metal ion
GO:0046677	response to antibiotic
GO:0006979	response to oxidative stress
GO:0048545	response to steroid hormone
GO:0035690	cellular response to drug
GO:0043062	extracellular structure organization
GO:0000302	response to reactive oxygen species
GO:0007050	cell cycle arrest

FIGURE 1
Enrichment analysis of ROS-related differentially expressed genes (DEGs) and PPI. **(A)** GO and **(B)** KEGG analyses.

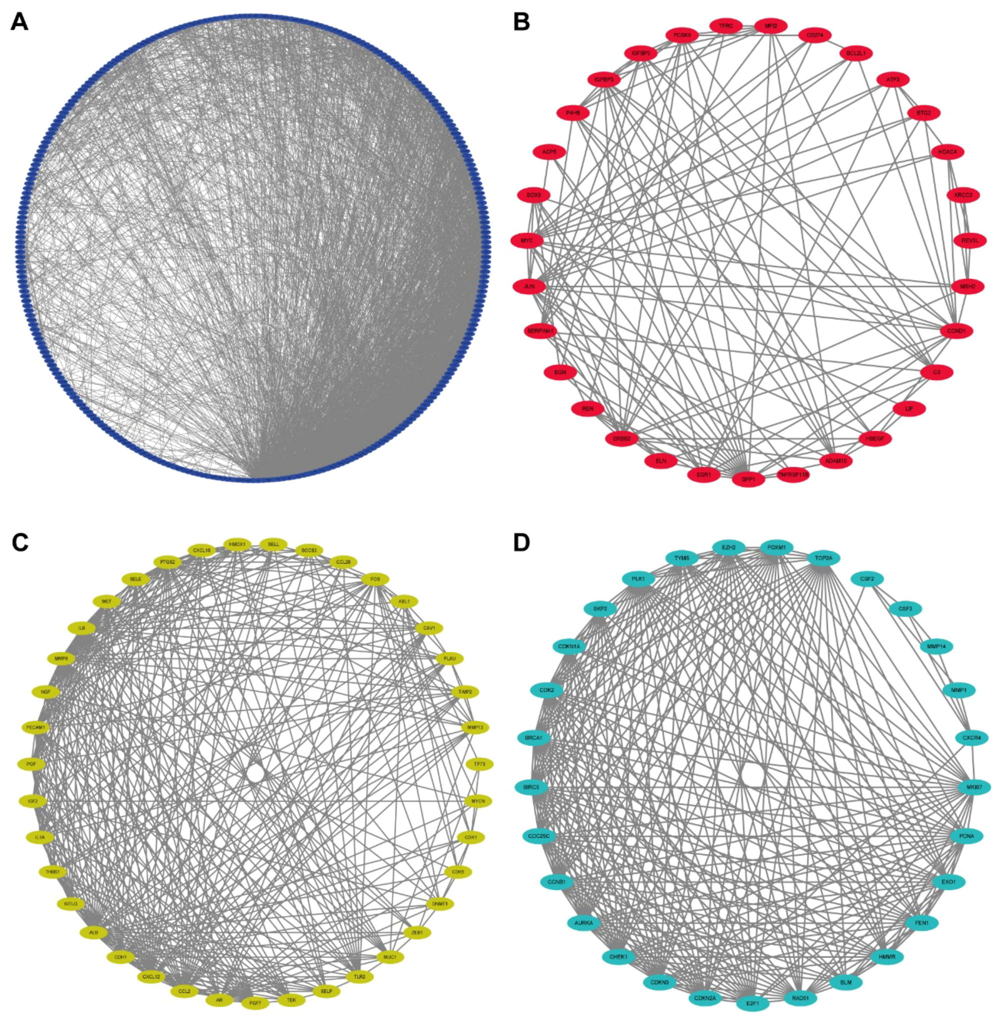


FIGURE 2
Protein-protein interaction (PPI) network. **(A)** Protein-protein interaction (PPI) network of differentially expressed ROS-related genes. **(B-D)** Key models of PPI networks.

(0297417-0002B, 5248896, puromycin, blebbistatin, anisomycin, STOCK1N-35215, methylergometrine, clofilium tosylate, verteporfin, withaferin A, and rottlerin) with anticancer functions in BLCA progression with enrichment scores < -0.8, $p < 0.01$, and n > 2 as the screening criteria (Table 1).

3.4 Construction and validation of the ROS-based prognostic signature

Our study conducted UCR analysis to identify ROS-related DEGs notably correlated with OS, and 71 genes were included in subsequent analyses (p<0.05) (Figure 3A). Aiming to ensure the clinical outcomes stability and reliability based on the 71 genes, LASSO analysis was conducted to further screen for prognostic ROS-related genes, and we identified 31 genes related to OS (Figures 3B, C). Multivariate Cox Regression (MCR) analysis identified 17 ROS-related genes (JUN, CALR, P4HB, ELN, MYC, FASN, REV3L, VHL, NID1, SLC38A1, TFRC, AKR1B1, ITGA3, CGB5, HLA-G, FADS1, and ORM1) that were used to construct a PS (Figure 3D). A ROS-based RS was established depending on the coefficient of 17 genes according to this formula: risk score = (0.3078 × FASN expression) + (0.305 × CALR expression) + (0.3832 × P4HB expression) + (0.1599 × ELN expression) + (0.2941 × MYC expression) + (0.3702 × REV3L expression) + (-0.4548 × VHL expression) +(0.147 × NID1 expression) + (-0.2213 × SLC38A1 expression) + (0.1687 × TFRC expression) +(0.123 × AKR1B1 expression) + (-0.1371 × ITGA3 expression) +(0.1762 × CGB5 expression) + (-0.1483 × HLA-G expression) +(0.1368 × FADS1 expression) + (-0.252 × ORM1 expression) + (0.1274 × JUN expression). Subsequently, we classified patients into HRG and LRG in accordance with the median RS. The LRG patients had longer OS than those in the HRG (p < 0.05) (Figures 4A, C). Time-dependent ROC analysis depicted that the signature AUC in the TCGA cohort was 0.78 at 5 years (Figure 4B). A heatmap was generated to show the differences in 17 ROS-related genes between the different groups (Figure 4D). PCA and t-SNE analyses indicated the signature's good classification ability (Figures 4E, F). Additionally, the prognostic

TABLE 1 The 11 small molecule drugs of CMP dataset analyses results.

cmap name	mean	n	enrichment	*p*-value	percent non-null
0297417-0002B	-0.779	3	-0.979	0.00004	100
puromycin	-0.765	4	-0.952	0	100
5248896	-0.668	2	-0.948	0.00594	100
blebbistatin	-0.679	2	-0.936	0.00861	100
anisomycin	-0.662	4	-0.933	0.00002	100
STOCK1N-35215	-0.691	3	-0.926	0.00062	100
methylergometrine	-0.64	4	-0.863	0.00064	100
verteporfin	-0.607	3	-0.844	0.00757	100
rottlerin	-0.68	3	-0.84	0.00817	100
withaferin A	-0.569	4	-0.832	0.00145	100
clofilium tosylate	-0.597	3	-0.832	0.00937	100

capacity of our constructed PS was validated in the GSE13507 dataset. The results in GSE13507 (Supplementary Figure 2) were consistent with previous results, which demonstrated the good performance of the PS in predicting OS.

3.5 Establishment and validation of the risk scores model

We first compared the ROS-based RSs among various subgroups classified by clinicopathological characteristics (TNM stage, sex, grade, age, T stage, and N stage). The RSs exhibited a significant correlation with clinicopathological factors and were markedly elevated in the following subgroups: >65 years of age, advanced T stage (T3/4), N stage (N1/2/3), pathological grade (High), and TNM stage (Stage III-IV) (Figure 5). Subsequently, stratification analysis was conducted relying upon the clinical characteristics (age, sex, grade, and TNM/T/N stages). Male or female sex, age (>65 years) or (<=65 years), T stage (T3-T4), N stage (N0), pathological grade (High), and TNM stage (Stage III–IV) were associated with inferior OS in the high-risk subgroup ($P < 0.05$) (Supplementary Figure 2), with no difference in OS in the T stage (T1/2), N stage (N1/2/3), or TNM stage (Stage I-II) subgroup (Figure 6). Furthermore, to evaluate whether the RS was an autocephalous prognostic indicator for BLCA patients, univariate and multivariate Cox proportional hazard models were implemented. According to UCR analysis results, age, TNM/T/N stages, and RS were related to unfavorable OS (Figure 7A). According to the multivariate analysis, age, N stage, and RS were still associated with unfavorable OS (Figure 7B). Multiparameter ROC curve analyses also revealed that the AUC of the RS was 0.769 (Figure 7C), indicating that compared with traditional clinical prognostic indicators, the ROS-based RS exhibited remarkable performance in predicting prognosis. Collectively, the ROS-based RS was an autocephalous prognostic factor. A nomogram including RS, age, and N stage we established to predict the outcomes of BLCA patients (Figure 8A), with the calibration curve elucidating the nomogram's good performance in predicting patient prognosis (Figures 8B, C).

3.6 GSEA

The GSEA results demonstrated that carcinogenic signaling pathways, such as calcium, focal adhesion, ECM receptor interaction, MAPK, BLCA, GAP junction, Wnt, Hedgehog, cancer, and TGF-β signaling pathways, exhibited main enrichment in the HRG (Figure 9). Several metabolism-associated signaling pathways, including autophagy regulation, peroxisomes, and oxidative phosphorylation, were highly enriched in the LRG.

3.7 Immune cell infiltration

A heatmap of the ICI data obtained via CIBERSORT, MCPcounter, XCELL, TIMER, CIBERSORT-ABS, QUANTISEQ, and EPIC analyses (Figure 10) suggested that the RS was correlated with ICI in BLCA. Additionally, significant differences were observed in the fractions of distinct leukocyte subsets between both groups. The proportions of naive B cells and M0/M2 macrophages were lower in the HRG, whereas the proportions of CD8+/CD4+T cells/were greater in the LRG.

3.8 Tumor mutational burden analysis

The mutation profile results among different risk groups in the TCGA dataset revealed somatic mutations in 93.53% (118) and 94.55% (191) of the BLCA patients in the HRG (Figure 11A) and LRG (Figure 11B), respectively. TP53, TTN, KMT2D, MUC16, ARID1A, KDM6A, PIK3CA, SYNE1, RB1, and KMT2C were the top 10 mutated genes in the HRG. TP53, TTN, KMT2D, MUC16, ARID1A, KDM6A, PIK3CA, SYNE1, RB1, and FGFR3 were the top 10 mutated genes in the LRG. Furthermore, the proportions of somatic mutations in

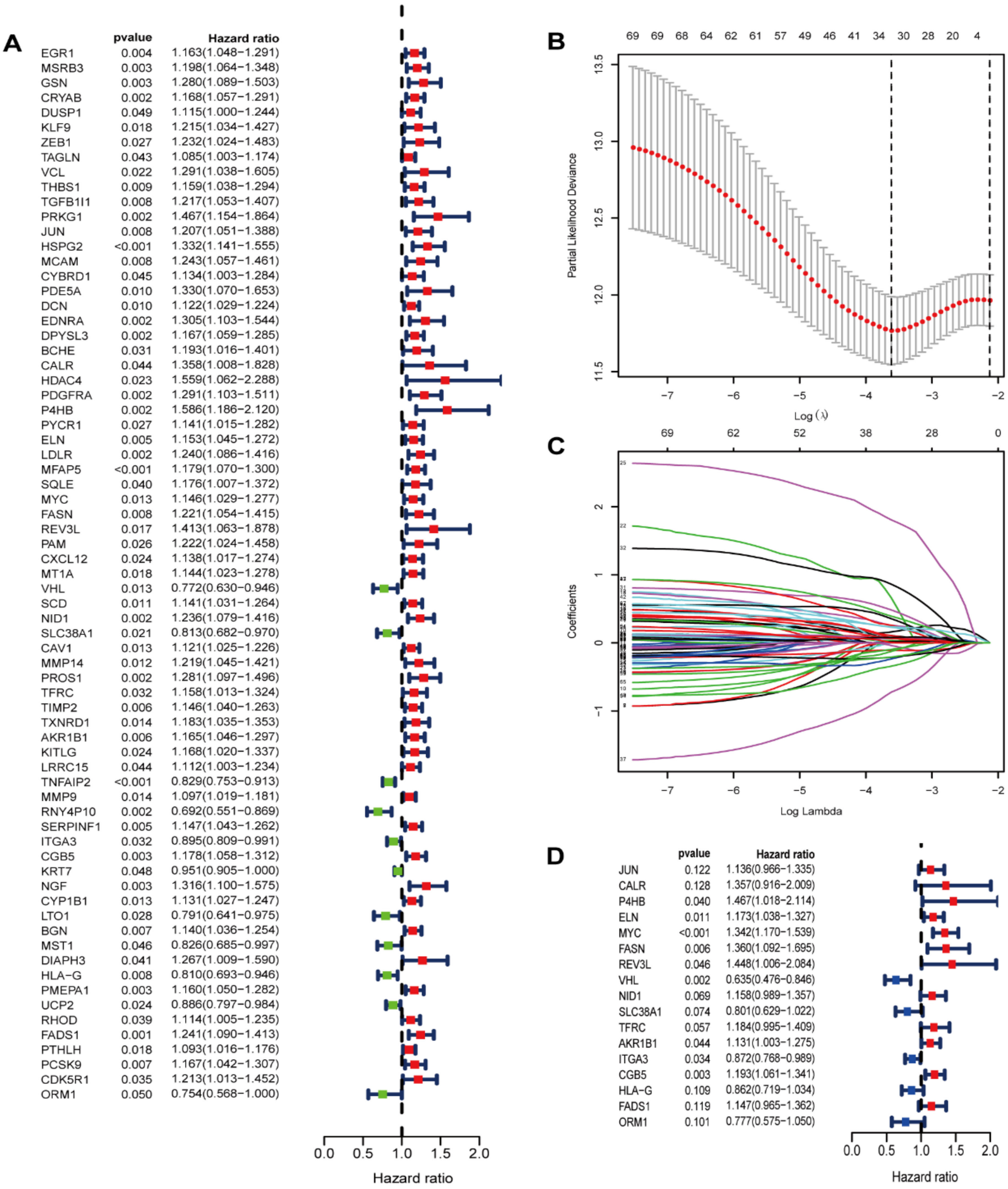


FIGURE 3
Identification of prognostic ROS-related genes in TCGA dataset. **(A)** Screening prognostic ROS-associated genes through univariate Cox regression analysis; **(B)** Incorporating the prognostic ROS-associated genes into the LASSO regression analysis; **(C)**The prognostic ROS-related genes were incorporated into the LASSO regression analysis. **(D)** Screening prognostic ROS-related genes through multivariate Cox regression analysis.

KDM6A and FGFR3 significantly differed between both groups. Additionally, the LRG patients had more mutation events than the HRG (Figure 11C). Patients having a high TMB appeared to possess a better prognosis than those with a low TMB (Figure 11D). Further, we investigated the collaborative interaction effect of the ROS-based RS and TMB on prognosis (Figure 11E). We found that the HRG patients having a high TMB had shorter OS than those in the LRG with a high TMB, and the LRG patients having a low TMB had longer OS than those in the HRG with a low TMB. Interestingly, patients having a high TMB exhibited better OS than those having a low TMB in the HRG, and patients having a low TMB displayed worse OS than those having a high TMB in the LRG. Patients possessing high TMB in the LRG had a greater OS than patients in the other three patient groups, and patients having low TMB in the HRG tended to have a significantly worse OS than patients in the other three patient groups. Collectively, the ROS-based RS might be a probable biomarker for predicting OS in BLCA patients.

3.9 Chemotherapeutic response analysis

The GDSC database analysis findings depicted that the IC50 values of chemotherapy drugs, including GSK269962A, BMS.536924, JNJ.26854165, docetaxel, temsirolimus, cisplatin, thapsigargin,

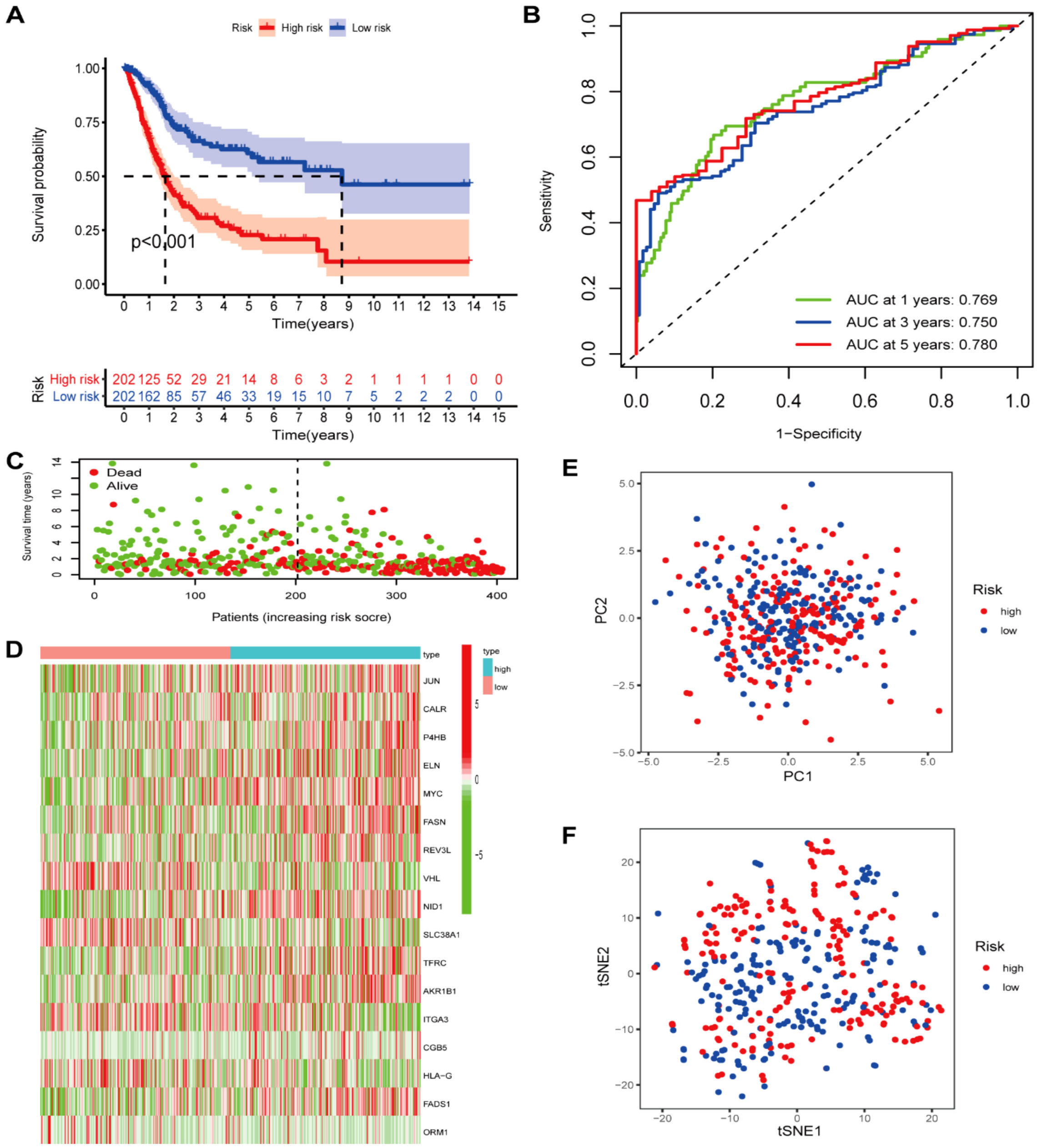


FIGURE 4
Prognostic ROS-based signature construction in TCGA dataset. **(A)** Kaplan-Meier survival analysis of BLCA patients between different groups; **(B)** Survival status distribution relying on the median risk score; **(C)**Time-independent ROC analysis of 5-year survival risk scores; **(D)** Heatmap showing the differences of 17 ROS-related genes between different groups. **(E)** PCA analysis; **(F)** t-SNE analysis.

sunitinib, rapamycin, and paclitaxel, were greater in LRG patients than in those the HRG. In comparison, the IC50 values of BIBW2992 and gefitinib were greater in HRG patients than in those at LRG (Figure 12).

3.10 Expression analysis of nine genes in the Human Protein Atlas database

IHC was utilized to additionally investigate the nine gene protein expression in the HPA database between BLCA and normal control tissues (Figure 13). In line with the RNA sequencing data, P4BH, FASN, AKR1B1, and CBG5 proteins, which have a high prognostic risk, were upregulated in tumor tissues, and MYC proteins, which have a low prognostic risk, were downregulated in tumor tissues compared with normal controls.

3.11 AKR1B1 affected BLCA cell viability, migration, and proliferation

While AKR1B1 has been documented in other types of cancer (29–31), its impact on BLCA remains unreported. Therefore, AKR1B1 was selected for further analysis. IHC and WB analyses elucidated AKR1B1

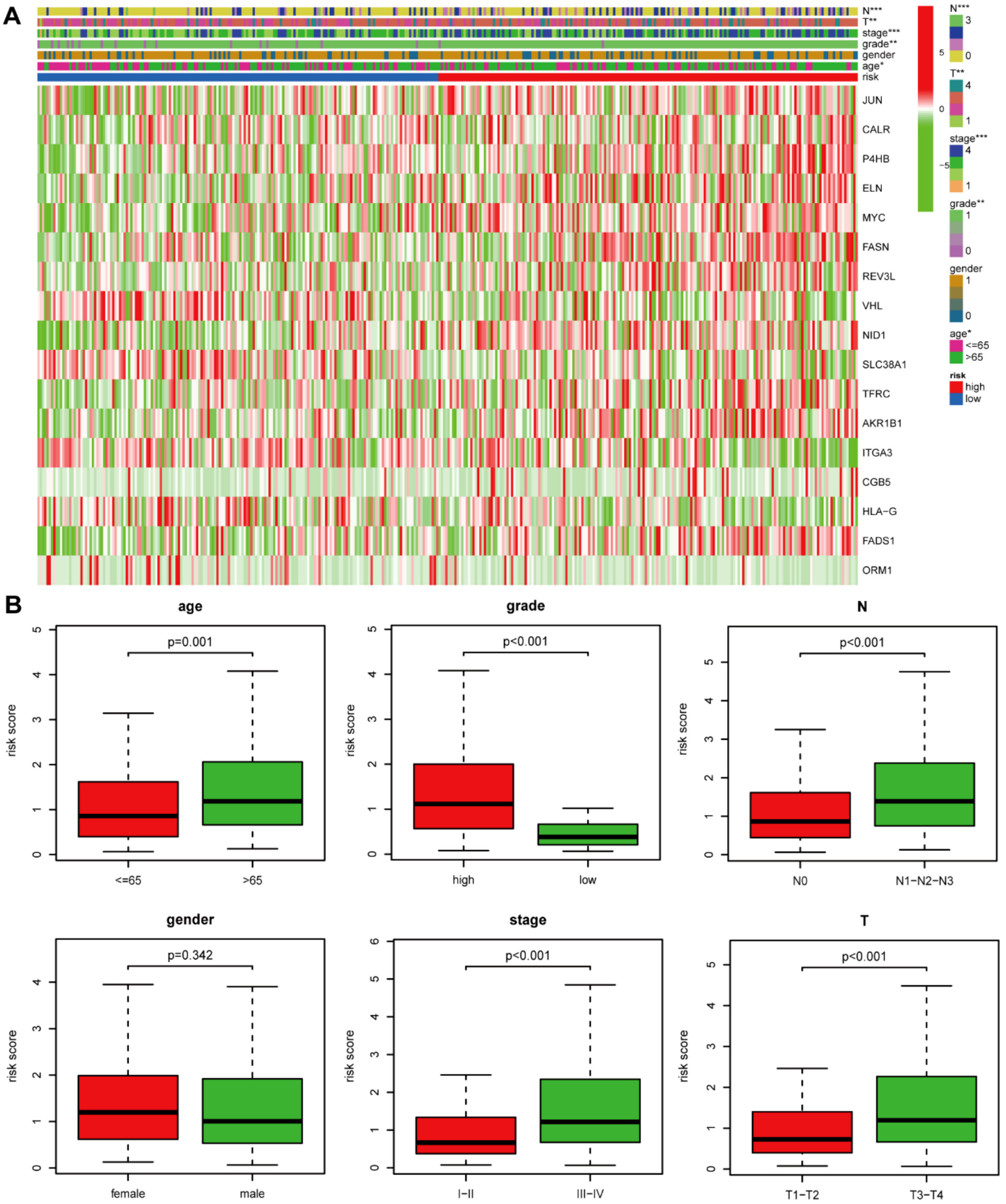


FIGURE 5
The risk score and clinicopathological factor correlation in the TCGA dataset. **(A)** The heatmap (*: 0.01<*P*<0.05; **: 0.001<*P*<0.01; ***: *P*<0.001) and **(B)** Boxplot show the risk score and clinicopathological factor correlation.

overexpression in BLCA tissues (Figure 14A). To understand the role of AKR1B1 in BLCA, we further investigated the effect of increased AKR1B1 levels in BLCA cell lines (T24 and 5637) via *in vitro* experiments. siRNA transfection successfully interfered with the mRNA expression of AKR1B1, which was confirmed by WB (Figure 14B). To further understand the effect of AKR1B1, colony formation analysis was also performed (Figure 14C), which showed that BLCA cell viability was significantly hindered after AKR1B1 was silenced by siRNA. The CCK-8 assay showcased that cell viability was impeded after the silencing of AKR1B1 expression (Figure 14D). In addition, an EdU proliferation assay showed that inhibiting AKR1B1 significantly lowered the percentage of EdU-positive BLCA cells (Figure 14E). To further test whether AKR1B1 affects BLCA cell metastasis, a Transwell assay was performed (Figure 14F), revealing that siRNA-mediated silencing of AKR1B1 inhibited BLCA cell migration and invasion.

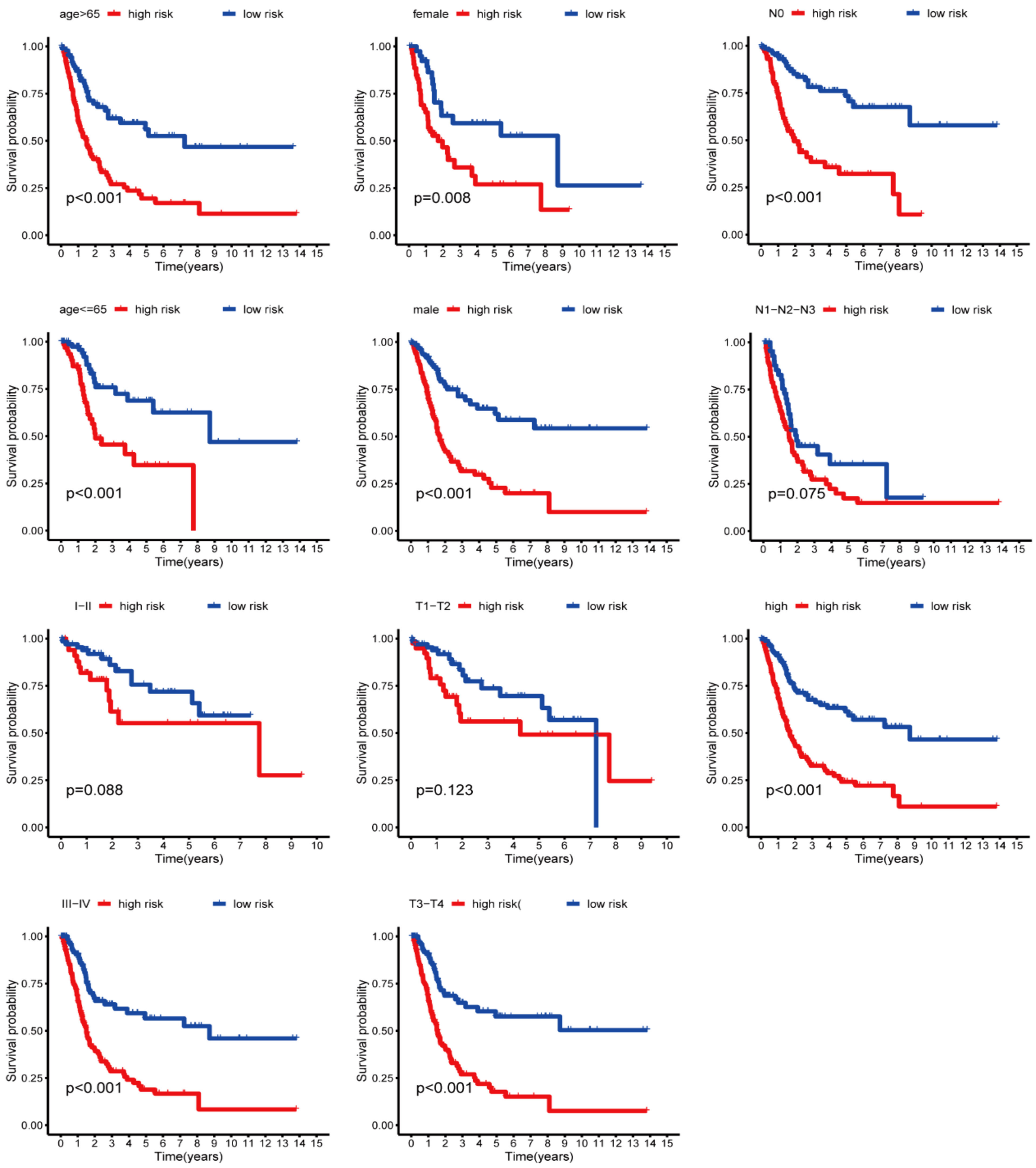


FIGURE 6
Kaplan-Meier curves stratification of OS by gender, age, grade, or N/T/TNM stages between both risk groups.

4 Discussion

Reactive Oxygen Species (ROS) include hydroxyl radicals (·OH), superoxide anions radicals (·O2-), and hydrogen peroxide (H2O2), are considered a double-edged sword (32). Physiologically, ROS play a crucial role in organisms. Excessive ROS can damage proteins and DNA through oxidative damage, causing many diseases, including cancer. ROS can cause cancer cells to die in high concentrations (33, 34). However, the possible mechanisms and prognostic value of ROS-associated genes in BLCA remain indefinite. Our study systematically explored the expression patterns and correlations of ROS-associated genes with outcomes in BLCA. Furthermore, we established a prognosis-correlated novel signature relying on 17 ROS-related genes. Here, the ROS-based signature was associated with CD8+ T cells and chemotherapy responses. Eleven drugs were screened for treating BLCA patients. Our results offer novel insights into ROS involvement in BLCA development and progression.

In BLCA, 308 ROS-related genes were identified as Differentially Expressed Genes (DEGs); of them, 138 and 170

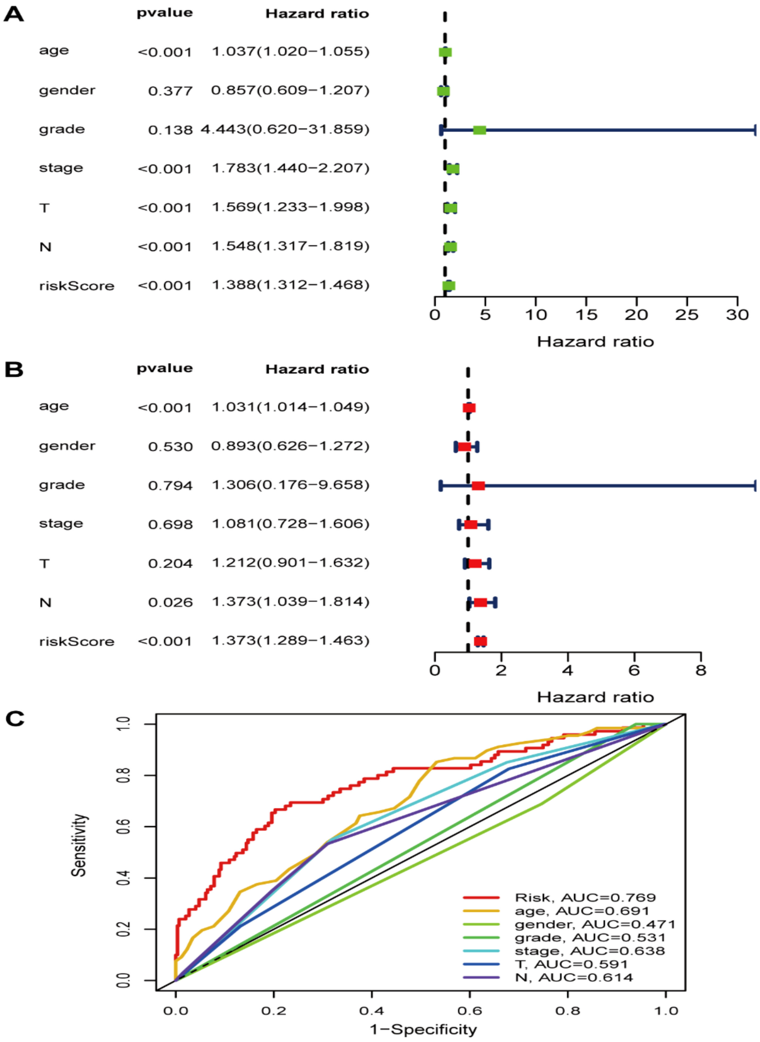


FIGURE 7
The risk signature as an independent BLCA prognostic factor in the TCGA dataset. **(A)** The OS risk score and clinicopathological factor correlations by univariate and **(B)** multivariate Cox regression analysis. **(C)** ROC curves of the clinical characteristics and risk score.

were downregulated and upregulated genes, respectively. Then, we explore the ROS-related DEGs' functions through GO and KEGG analyses. According to GO analysis, these genes exhibited main enrichment in response to toxic substances, aging, metal ions, oxidative stress, and ROS, besides cellular response to drugs. KEGG analysis showcased that these genes were closely associated with cancer-, immune-, and drug resistance-correlated pathways, such as the p53, platinum drug resistance, PI3K-Akt, TNF, IL-17, MAPK, HIF-1, and cGMP-PKG pathways, suggesting that ROS-related genes are involved in tumorigenesis. Subsequently, according to the results of differential expression analyses, a PS consisting of 17 ROS-related genes (JUN, CALR, P4HB, ELN, MYC, FASN, REV3L, VHL, NID1, SLC38A1, TFRC, AKR1B1, ITGA3, CGB5, HLA-G, FADS1, and ORM1) was constructed and validated via LASSO and Cox regression analyses.

Among the seventeen ROS-related genes in our established signature, calreticulin (CALR), an ER protein with high Ca2+-binding activity, is crucial in maintaining cell homeostasis and initiating the anticancer immune response to immunogenic cell death (35, 36). Elevated CALR was correlated with favorable prognosis in distinct tumor types (36–39). CALR overexpression was linked to worse OS in natural-killer T-cell lymphoma patients (40). CALR silencing suppressed BLCA cell proliferation, migration, and lung metastasis (41). FASN can serve as an oncogene by regulating AKT signaling pathways in BLCA (42, 43). Overexpression of P4HB was notably associated with inferior

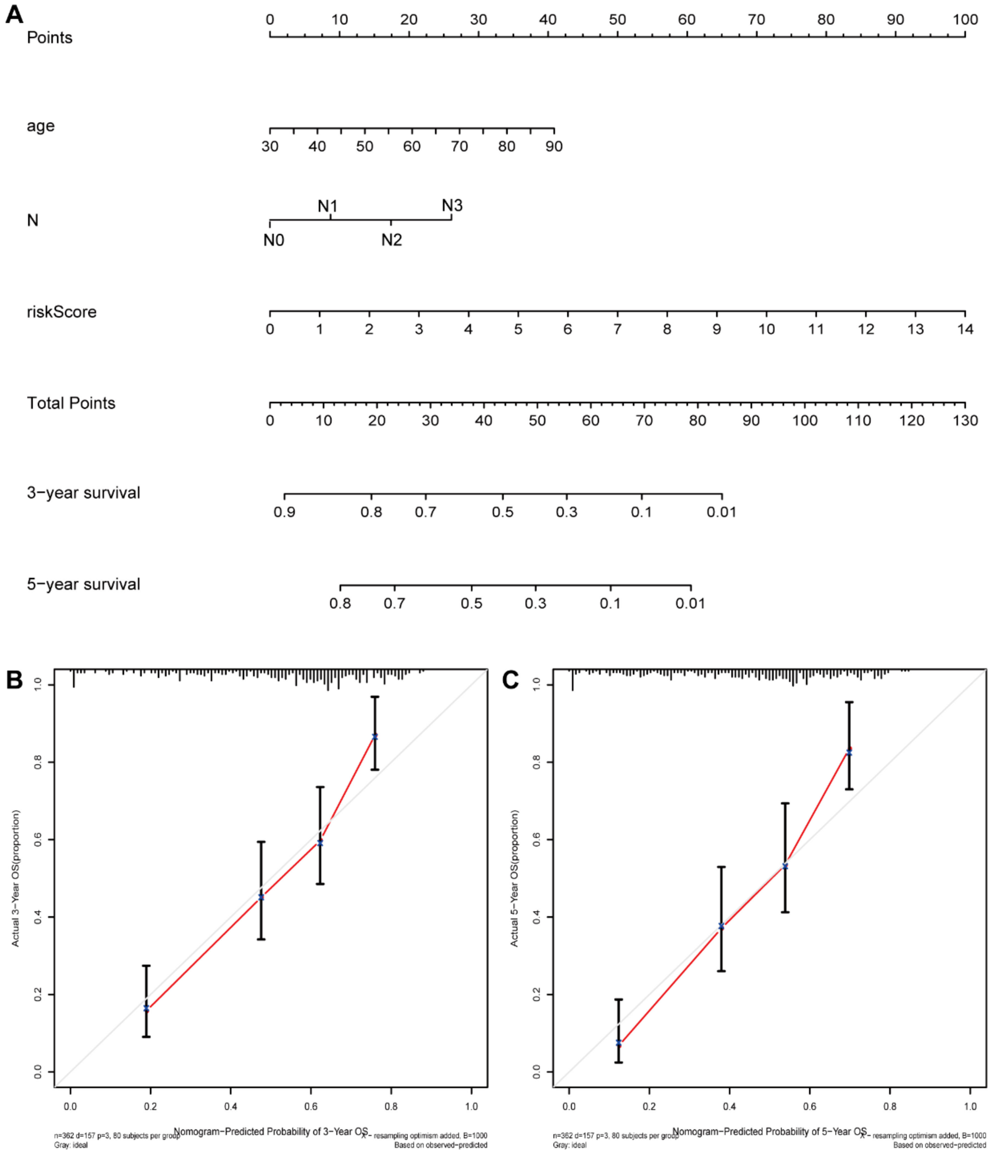


FIGURE 8
The nomogram construction. **(A)** Nomogram predicting 3- or 5-year OS. **(B)** Calibration plots predicting 3- and **(C)** 5-year OS.

outcomes, and knocking down P4HB impeded cell proliferation and enhanced GEM sensitivity via the PERK/eIF2α/ATF4/CHOP signaling pathways in BLCA (44). ITGA3 downregulation hindered tumor cell invasion and proliferation by regulating the FAK/PI3K/AKT pathway and epithelial-mesenchymal transition (45, 46). SLC38A1, a vital transporter of glutamine, has been implicated in tumorigenesis (47, 48). The expression of TFRC, a crucial member involved in ferroptosis, was significantly elevated in BLCA and promoted the tumorigenic phenotype of BLCA cells by inducing EMT (49). Aldo–keto reductase family 1 member B1 (AKR1B1) is closely implicated in cancer development and progression through various mechanisms, including EMT, ERK1/2, Ras, and PI3K-AKT signaling pathways (50). Additionally, ARK1B1 was also related to chemotherapeutic resistance and cancer stem cells (51, 52). REV3L is highly overexpressed in several cancers and facilitates cancer cell proliferation, metastasis, and insensitivity to cisplatin (53, 54). Elastin (ELN), a crucial member of the extracellular matrix family, has been documented to contribute to cancer cell invasion (55, 56). Orosomucoid 1 (ORM1), an essential immune system regulator in acute-phase reactions, might facilitate cancer cell immune evasion (57, 58). Nidogen1 (NID1), a vital component of the basement membrane,

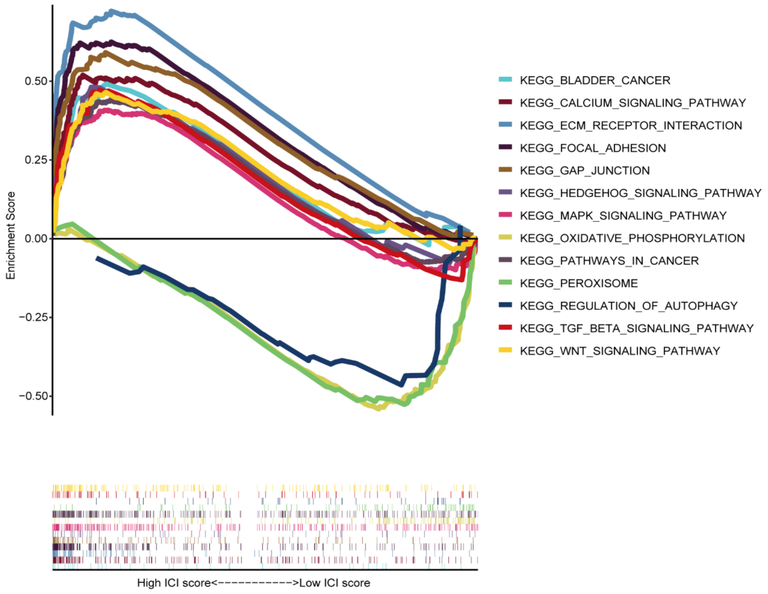


FIGURE 9
Gene set enrichment analysis among different groups.

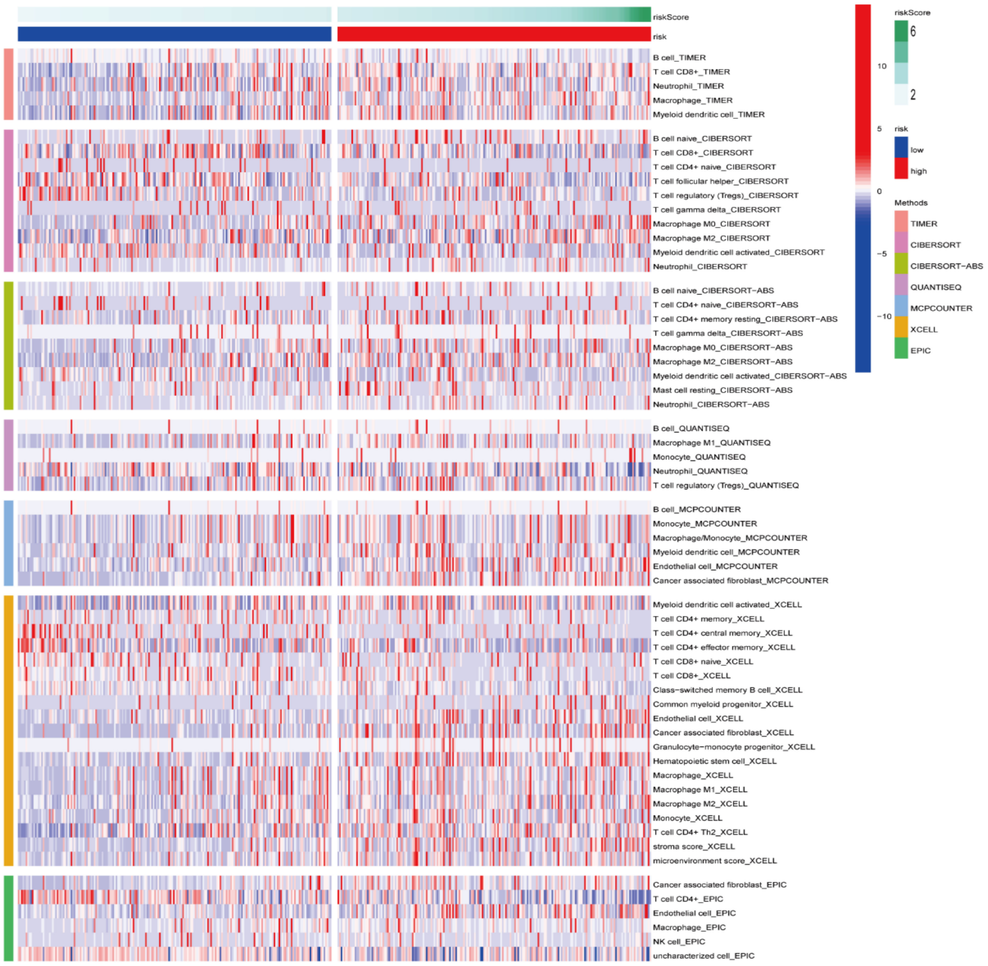


FIGURE 10
Immune cell infiltration between both risk groups.

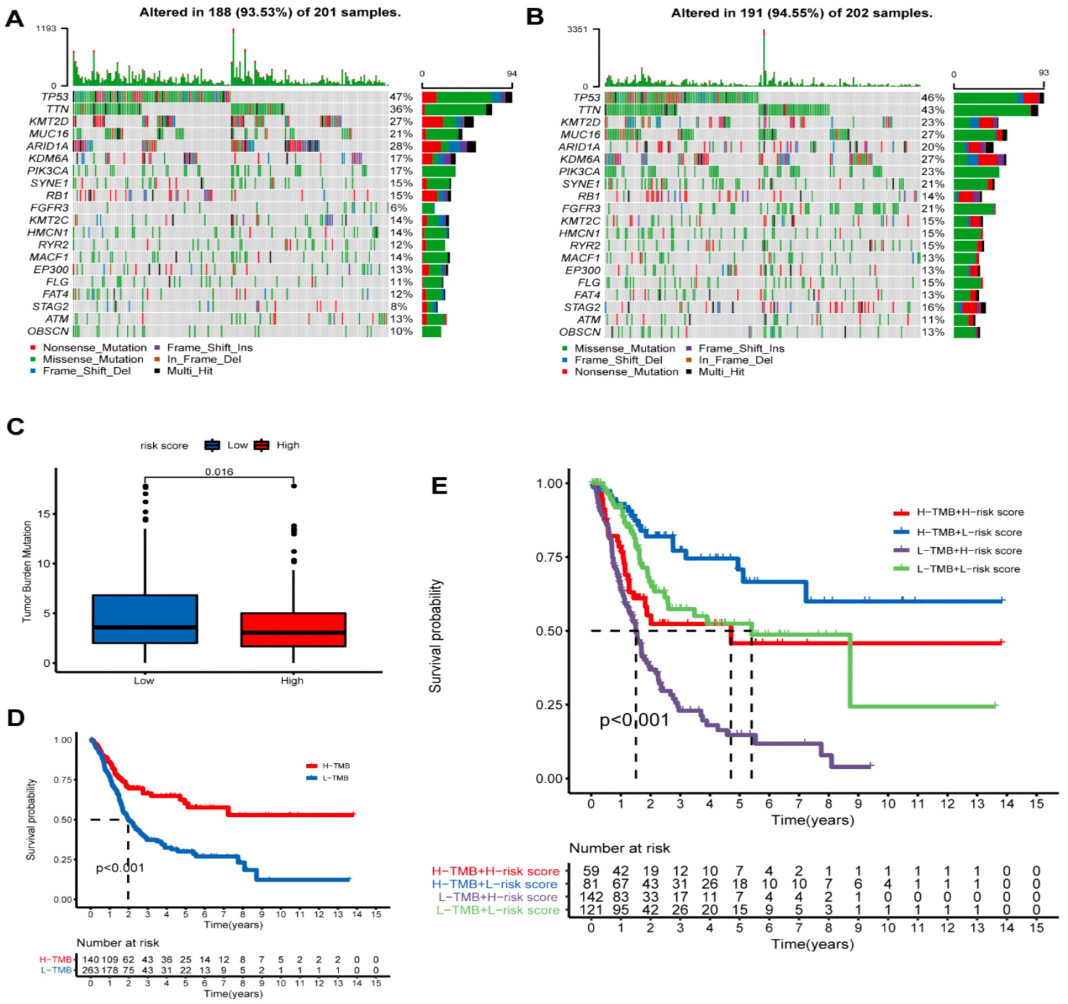


FIGURE 11
Tumor mutational burden (TMD) analysis. **(A)** Demonstrating the top 20 mutational genes within the high- and **(B)** low-risk groups. **(C)** TMB difference in both risk groups. **(D)** Kaplan-Meier (K-M) survival analysis of BLCA patients with high or low TMB. **(E)** K-M curve analysis stratification of OS by TMB and the prognostic signature.

serves as an oncogene in several tumors (59–62). Chorionic gonadotropin beta polypeptide 5 (CGB5) can accelerate cancer growth and vasculogenic mimicry formation by activating the LHR signaling pathway (63). Jun represents a critical transcription factor implicated in various biological processes, including autophagy, proliferation, apoptosis, metastasis, and inflammation (64, 65). FADS1 silencing reduced cell growth by arresting the cell cycle in the G1 phase (66).

Our Univariate Cox regression (UCR) and Multivariate Cox Regression (MCR) analyses results demonstrated that the RS was a negative prognostic factor of OS in BLCA patients. Further, ROC analysis suggested that the RS outbalanced the conventional clinical characteristics in OS prediction of BLCA patients. Herein, BLCA patients with advanced clinical features (III-IV stage, Grade high, T3/4 stage, and N1/2/3 stage) had elevated RSs in comparison with patients with early clinical features (I-II stage, Grade low, T1/2 stage, and N0 stage). The RS was also related to age. Stratification analyses revealed that the RS could effectively predict BLCA patient outcomes in most subgroups other than subgroups (N1/2/3 stage, T1/2 stage, and I-II stage). Finally, we constructed a ROS-related nomogram to evaluate 3- and 5-year OS comprehensively. The calibration curve results implied that the nomogram showed excellent performance in predicting BLCA patient prognosis.

To deeply understand the potential mechanisms behind ROS-mediated differential outcomes in BLCA patients, we implemented GSEA analyses for different groups relying upon the ROS-based PS.

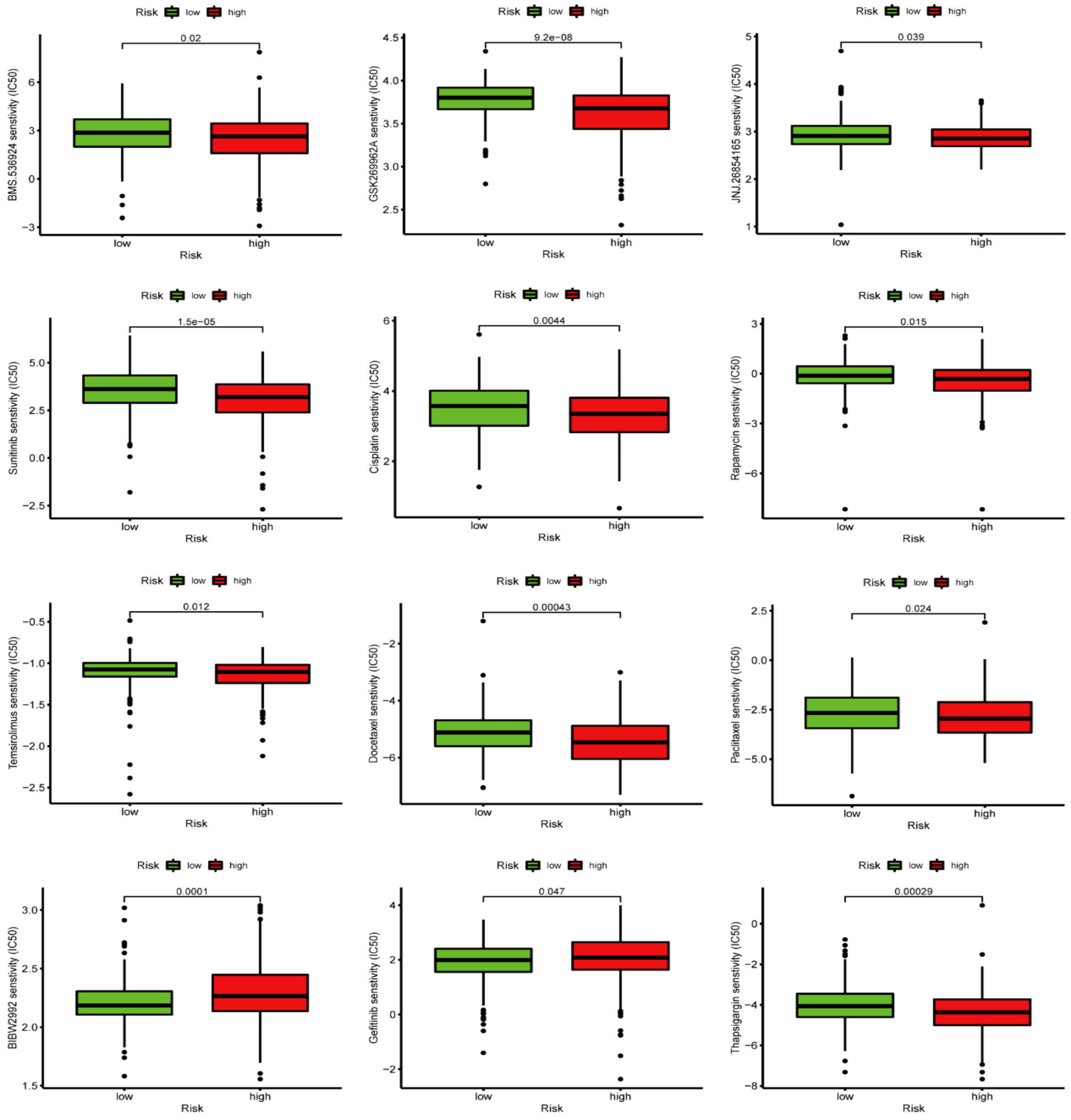


FIGURE 12
GDSC database-based chemotherapy response prediction.

The findings demonstrated that the HRG exhibited enrichment in cancer-associated pathways, including calcium, focal adhesion, ECM receptor interaction, MAPK, Wnt, Hedgehog, cancer, and TGF-β pathways, implying the existence of an immunosuppressive microenvironment. Meanwhile, the LRG genes exhibited main involvement in the regulation of autophagy, peroxisomes, and oxidative phosphorylation. Altogether, OS was inferior in the HRG patients than those in the LRG. The ICI is related to the malignant biological phenotypes and prognosis of cancer patients, which indicates that immunotherapy, particularly immune checkpoint inhibitor treatment, has become crucial for treating advanced tumors (67). CD8+ T cells are strongly correlated with the

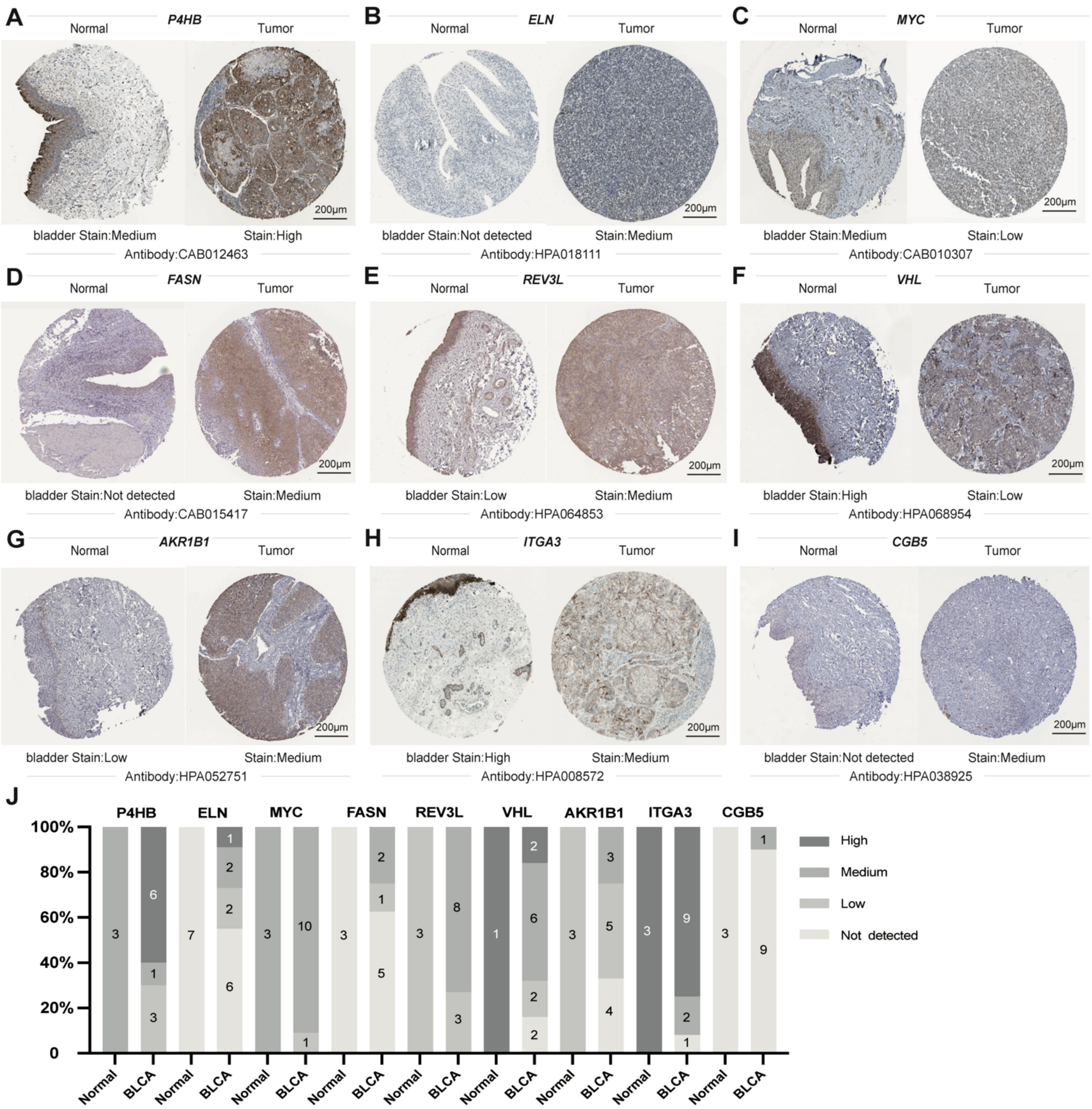


FIGURE 13

Sectional images of the differential expression of the above genes from the Human Protein Atlas. **(A–I)** representative images of P4HB **(A)**, ELN **(B)**, MYC **(C)**, FASN **(D)**, REV3L **(E)**, VHL **(F)**, AKR1B1 **(G)**, ITGA3 **(H)**, and CGB5 **(I)** protein expression from HPA databases. **(J)** Genes from HPA databases Statistical Column Stacked Plots of Characterized Protein Expression. Scale bar: 200µm.

effectiveness of cancer immunotherapy (68). An elevated CD8+ T cell infiltration level indicated a superior prognosis in BLCA patients (69), aligning with our finding that the LRG patients possessed a greater CD8+ T cell proportion and favorable outcomes. Previous research has shown that patients having a high TMB appear to possess a prolonged survival time and an improved immunotherapy response (70). However, the TMB and immunotherapy response correlation remains controversial (71). Herein, the HRG patients exhibited a lower TMB and an inferior prognosis and might benefit from cisplatin, docetaxel, temsirolimus, thapsigargin, BMS.536924, GSK269962A, JNJ.26854165, sunitinib, rapamycin, and paclitaxel. Meanwhile, LRG patients might benefit from BIBW2992 and gefitinib.

Our study had various constraints. The ROS-based signature we developed and validated was generated via retrospective research and requires confirmation through a prospective trial. However, it is

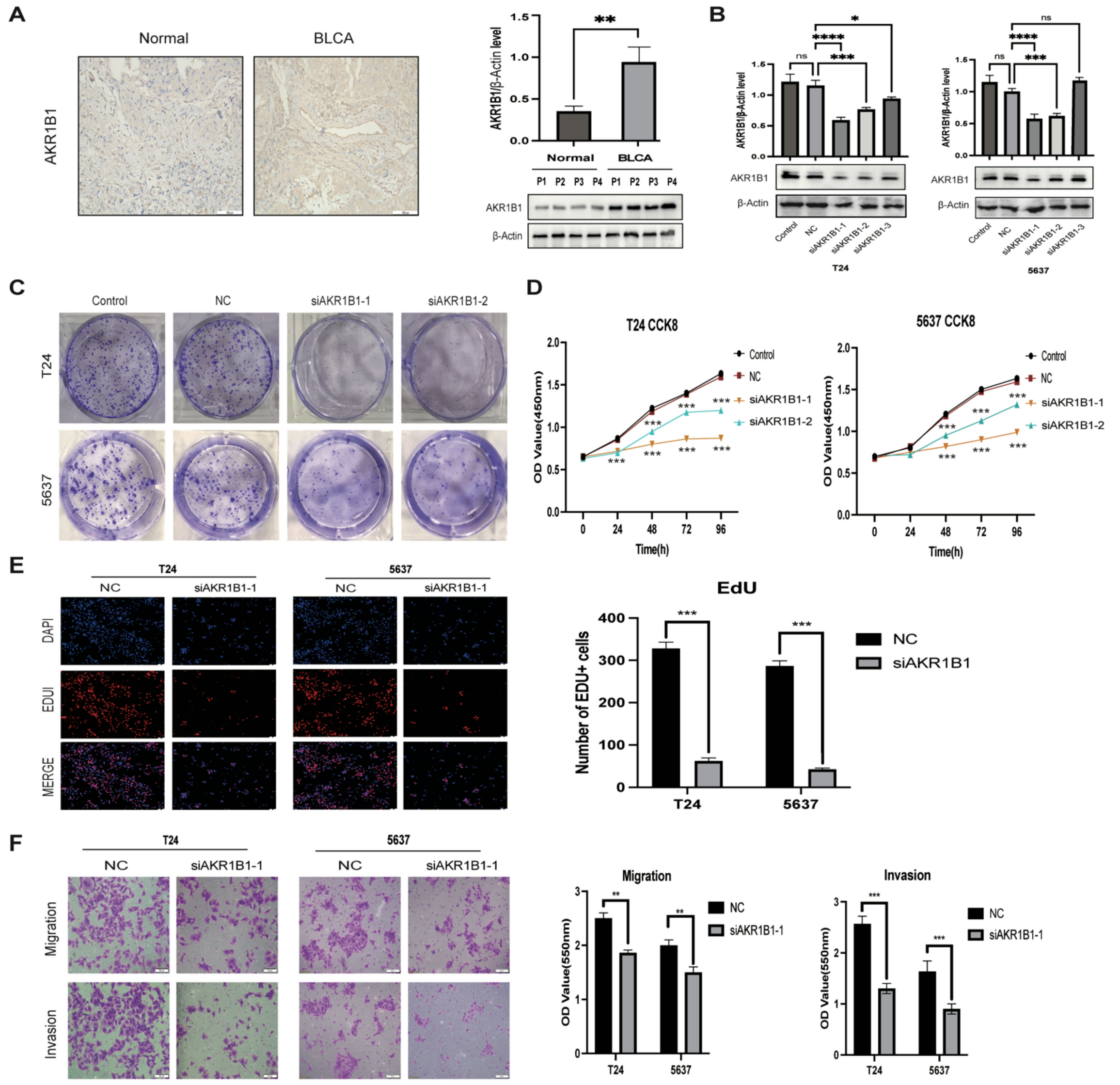


FIGURE 14
(A) IHC representation chart and western blot (WB) showed AKR1B1 expression in normal bladder tissue and BLCA tissue. Scale bar: 100µm. **(B)** WB detection of AKR1B1 relative expression in control, NC, and siAKR1B1 groups. **(C)** Colony formation experiment results with AKR1B1 expression. **(D)** Results of silencing AKR1B1 expression at different time points of CCK-8:24, 48, 72, 96h. **(E)** Edu assay showing proliferating cells (T24 and 5637); Edu (red) and DAPI (blue) staining. Scale bar: 50µm. **(F)** Transwell assay results in control, NC, and siAKR1B1 groups. Scale bar: 100µm. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$, ns $p > 0.05$.

necessary to conduct more experimental validation to confirm the probable molecular mechanisms behind the PS in BLCA.

5 Conclusions

Conclusively, we conducted a thorough investigation of the possible functions and prognostic value of ROS-associated genes in BLCA through integrated bioinformatics analyses. In addition, a ROS-dependent PS we constructed and validated with the ability to predict the outcome and chemotherapy response of BLCA patients. Moreover, we constructed a nomogram including a ROS-based PS with clinical characteristics for 3- and 5-year OS, which could aid clinicians in clinical decision-making. To verify the authenticity of the data, we detected the signature protein expression levels through HPA. *In vitro*, siRNA-mediated AKR1B1 silencing impeded BLCA cell viability, migration, and proliferation, consistent with our projections and demonstrating the constructed ROS-related gene reliability.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving humans were approved by Academic Ethics Committee of Shaoxing People's Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

YL: Conceptualization, Formal analysis, Funding acquisition, Writing – original draft, Writing – review & editing. LZ: Data curation, Formal analysis, Writing – original draft. GX(3rd author): Investigation, Methodology, Writing – original draft. GX(4th author): Methodology, Project administration, Writing – original draft. JC: Software, Supervision, Writing – review & editing. KZ(6rd author): Software, Visualization, Writing – review & editing. ML: Validation, Visualization, Writing – review & editing. JJ: Methodology, Visualization, Writing – review & editing. CP: Funding acquisition, Validation, Writing – review & editing. KW: Resources, Visualization, Writing – review & editing. SP: Funding acquisition, Investigation, Methodology, Writing – original draft, Writing – review & editing. KZ: Conceptualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by the medical and health research project of Zhejiang province to SP (Grant#2022KY1297), the university-level youth research fund project of Shaoxing People's Hospital (Grant# 2022YA07) and Shaoxing City science and technology plan project (Grant#2024A14008) and shaoxing health science and technology project (Grant#2023SKY004) to YL.

Acknowledgments

We thank the Reviewers for their comments.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1493528/full#supplementary-material.

References

1. 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 Cancer J Clin.* (2021) 71:209–49. doi: 10.3322/caac.21660

2. Jubber I, Ong S, Bukavina L, Black PC, Compérat E, Kamat AM, et al. Epidemiology of bladder cancer in 2023: A systematic review of risk factors. *Eur Urol.* (2023) 84:176–90. doi: 10.1016/j.eururo.2023.03.029

3. Ren L, Jiang M, Xue D, Wang H, Lu Z, Ding L, et al. Nitroxoline suppresses metastasis in bladder cancer via EGR1/circNDRG1/miR-520h/smad7/EMT signaling pathway. *Int J Biol Sci.* (2022) 18:5207–20. doi: 10.7150/ijbs.69373

4. Patel VG, Oh WK, Galsky MD. Treatment of muscle-invasive and advanced bladder cancer in 2020. *CA Cancer J Clin.* (2020) 70:404–23. doi: 10.3322/caac.21631

5. Jiang N, Liao Y, Wang M, Wang Y, Wang K, Guo J, et al. BUB1 drives the occurrence and development of bladder cancer by mediating the STAT3 signaling pathway. *J Exp Clin Cancer Res.* (2021) 40:378. doi: 10.1186/s13046-021-02179-z

6. Zhang XG, Zhang T, Li CY, Zhang MH, Chen FM. CD164 promotes tumor progression and predicts the poor prognosis of bladder cancer. *Cancer Med.* (2018) 7:3763–72. doi: 10.1002/cam4.2018.7.issue-8

7. Knowles MA, Hurst CD. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. *Nat Rev Cancer.* (2015) 15:25–41. doi: 10.1038/nrc3817

8. Sies H, Jones DP. Reactive oxygen species (ROS) as pleiotropic physiological signalling agents. *Nat Rev Mol Cell Biol.* (2020) 21:363–83. doi: 10.1038/s41580-020-0230-3

9. Arfin S, Jha NK, Jha SK, Kesari KK, Ruokolainen J, Roychoudhury S, et al. Oxidative stress in cancer cell metabolism. *Antioxid (Basel).* (2021) 10. doi: 10.3390/antiox10050642

10. Perillo B, Di Donato M, Pezone A, Di Zazzo E, Giovannelli P, Galasso G, et al. ROS in cancer therapy: the bright side of the moon. *Exp Mol Med.* (2020) 52:192–203. doi: 10.1038/s12276-020-0384-2

11. Chang CH, Pauklin S. ROS and TGFβ: from pancreatic tumour growth to metastasis. *J Exp Clin Cancer Res.* (2021) 40:152. doi: 10.1186/s13046-021-01960-4

12. Kirtonia A, Sethi G, Garg M. The multifaceted role of reactive oxygen species in tumorigenesis. *Cell Mol Life Sci.* (2020) 77:4459–83. doi: 10.1007/s00018-020-03536-5

13. Casas AI, Nogales C, Mucke HAM, Petraina A, Cuadrado A, Rojo AI, et al. On the clinical pharmacology of reactive oxygen species. *Pharmacol Rev.* (2020) 72:801–28. doi: 10.1124/pr.120.019422

14. Chio IIC, Tuveson DA. ROS in cancer: the burning question. *Trends Mol Med*. (2017) 23:411–29. doi: 10.1016/j.molmed.2017.03.004

15. Wang Y, Qi H, Liu Y, Duan C, Liu X, Xia T, et al. The double-edged roles of ROS in cancer prevention and therapy. *Theranostics*. (2021) 11:4839–57. doi: 10.7150/thno.56747

16. Liu D, Qiu X, Xiong X, Chen X, Pan F. Current updates on the role of reactive oxygen species in bladder cancer pathogenesis and therapeutics. *Clin Transl Oncol*. (2020) 22:1687–97. doi: 10.1007/s12094-020-02330-w

17. Wang HC, Choudhary S. Reactive oxygen species-mediated therapeutic control of bladder cancer. *Nat Rev Urol*. (2011) 8:608–16. doi: 10.1038/nrurol.2011.135

18. Iqbal MJ, Kabeer A, Abbas Z, Siddiqui HA, Calina D, Sharifi-Rad J, et al. Interplay of oxidative stress, cellular communication and signaling pathways in cancer. *Cell Commun Signal*. (2024) 22:7. doi: 10.1186/s12964-023-01398-5

19. Shi T, Dansen TB. Reactive oxygen species induced p53 activation: DNA damage, redox signaling, or both? *Antioxid Redox Signal*. (2020) 33:839–59. doi: 10.1089/ars.2020.8074

20. Xing F, Hu Q, Qin Y, Xu J, Zhang B, Yu X, et al. The relationship of redox with hallmarks of cancer: the importance of homeostasis and context. *Front Oncol*. (2022) 12:862743. doi: 10.3389/fonc.2022.862743

21. Baird L, Yamamoto M. The molecular mechanisms regulating the KEAP1-NRF2 pathway. *Mol Cell Biol*. (2020) 40. doi: 10.1128/MCB.00099-20

22. Ngo V, Duennwald ML. Nrf2 and oxidative stress: A general overview of mechanisms and implications in human disease. *Antioxid (Basel)*. (2022) 11. doi: 10.3390/antiox11122345

23. Shaw P, Kumar N, Sahun M, Smits E, Bogaerts A, Privat-Maldonado A. Modulating the antioxidant response for better oxidative stress-inducing therapies: how to take advantage of two sides of the same medal? *Biomedicines*. (2022) 10. doi: 10.3390/biomedicines10040823

24. Pisoschi AM, Pop A, Iordache F, Stanca L, Predoi G, Serban AI. Oxidative stress mitigation by antioxidants - An overview on their chemistry and influences on health status. *Eur J Med Chem*. (2021) 209:112891. doi: 10.1016/j.ejmech.2020.112891

25. Rizvi A, Farhan M, Nabi F, Khan RH, Adil M, Ahmad A. Transcriptional control of the oxidative stress response and implications of using plant derived molecules for therapeutic interventions in cancer. *Curr Med Chem*. (2021) 28:8480–95. doi: 10.2174/0929867328666210218110550

26. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. (2000) 28:27–30. doi: 10.1093/nar/28.1.27

27. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci*. (2019) 28:1947–51. doi: 10.1002/pro.v28.11

28. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res*. (2023) 51: D587–d92. doi: 10.1093/nar/gkac963

29. Zhao Q, Han B, Wang L, Wu J, Wang S, Ren Z, et al. AKR1B1-dependent fructose metabolism enhances Malignancy of cancer cells. *Cell Death Differ*. (2024). doi: 10.1038/s41418-024-01393-4

30. Li X, Qian J, Xu J, Bai H, Yang J, Chen L. NRF2 inhibits RSL3 induced ferroptosis in gastric cancer through regulation of AKR1B1. *Exp Cell Res*. (2024) 442:114210. doi: 10.1016/j.yexcr.2024.114210

31. Syamprasad NP, Jain S, Rajdev B, Panda SR, Kumar GJ, Shaik KM, et al. AKR1B1 drives hyperglycemia-induced metabolic reprogramming in MASLD-associated hepatocellular carcinoma. *JHEP Rep*. (2024) 6:100974. doi: 10.1016/j.jhepr.2023.100974

32. Zhang R, Jiang W, Wang G, Zhang Y, Liu W, Li M, et al. Parkin inhibits proliferation and migration of bladder cancer via ubiquitinating Catalase. *Commun Biol*. (2024) 7:245. doi: 10.1038/s42003-024-05935-x

33. Tuli HS, Kaur J, Vashishth K, Sak K, Sharma U, Choudhary R, et al. Molecular mechanisms behind ROS regulation in cancer: A balancing act between augmented tumorigenesis and cell apoptosis. *Arch Toxicol*. (2023) 97:103–20. doi: 10.1007/s00204-022-03421-z

34. Zeng W, Long X, Liu PS, Xie X. The interplay of oncogenic signaling, oxidative stress and ferroptosis in cancer. *Int J Cancer*. (2023) 153:918–31. doi: 10.1002/ijc.v153.5

35. Fucikova J, Spisek R, Kroemer G, Galluzzi L. Calreticulin and cancer. *Cell Res*. (2021) 31:5–16. doi: 10.1038/s41422-020-0383-9

36. Kasikova L, Hensler M, Truxova I, Skapa P, Laco J, Belicova L, et al. Calreticulin exposure correlates with robust adaptive antitumor immunity and favorable prognosis in ovarian carcinoma patients. *J Immunother Cancer*. (2019) 7:312. doi: 10.1186/s40425-019-0781-z

37. Fucikova J, Becht E, Iribarren K, Goc J, Remark R, Damotte D, et al. Calreticulin expression in human non-small cell lung cancers correlates with increased accumulation of antitumor immune cells and favorable prognosis. *Cancer Res*. (2016) 76:1746–56. doi: 10.1158/0008-5472.CAN-15-1142

38. Stoll G, Iribarren K, Michels J, Leary A, Zitvogel L, Cremer I, et al. Calreticulin expression: Interaction with the immune infiltrate and impact on survival in patients with ovarian and non-small cell lung cancer. *Oncoimmunology*. (2016) 5:e1177692. doi: 10.1080/2162402X.2016.1177692

39. Peng RQ, Chen YB, Ding Y, Zhang R, Zhang X, Yu XJ, et al. Expression of calreticulin is associated with infiltration of T-cells in stage IIIB colon cancer. *World J Gastroenterol*. (2010) 16:2428–34. doi: 10.3748/wjg.v16.i19.2428

40. Zheng Y, Li C, Xin P, Peng Q, Zhang W, Liu S, et al. Calreticulin increases growth and progression of natural killer/T-cell lymphoma. *Aging (Albany NY)*. (2020) 12:23822–35. doi: 10.18632/aging.104030

41. Lu YC, Chen CN, Wang B, Hsu WM, Chen ST, Chang KJ, et al. Changes in tumor growth and metastatic capacities of J82 human bladder cancer cells suppressed by down-regulation of calreticulin expression. *Am J Pathol*. (2011) 179:1425–33. doi: 10.1016/j.ajpath.2011.05.015

42. Zheng SS, Gao JG, Liu ZJ, Zhang XH, Wu S, Weng BW, et al. Downregulation of fatty acid synthase complex suppresses cell migration by targeting phosphor-AKT in bladder cancer. *Mol Med Rep*. (2016) 13:1845–50. doi: 10.3892/mmr.2015.4746

43. Jiang B, Li EH, Lu YY, Jiang Q, Cui D, Jing YF, et al. Inhibition of fatty-acid synthase suppresses P-AKT and induces apoptosis in bladder cancer. *Urology*. (2012) 80:484.e9–15. doi: 10.1016/j.urology.2012.02.046

44. Wang X, Bai Y, Zhang F, Yang Y, Feng D, Li A, et al. Targeted inhibition of P4HB promotes cell sensitivity to gemcitabine in urothelial carcinoma of the bladder. *Onco Targets Ther*. (2020) 13:9543–58. doi: 10.2147/OTT.S267734

45. Wang JR, Liu B, Zhou L, Huang YX. MicroRNA-124-3p suppresses cell migration and invasion by targeting ITGA3 signaling in bladder cancer. *Cancer biomark*. (2019) 24:159–72. doi: 10.3233/CBM-182000

46. Sakaguchi T, Yoshino H, Yonemori M, Miyamoto K, Sugita S, Matsushita R, et al. Regulation of ITGA3 by the dual-stranded microRNA-199 family as a potential prognostic marker in bladder cancer. *Br J Cancer*. (2017) 116:1077–87. doi: 10.1038/bjc.2017.43

47. Song J, Liu Y, Guan X, Zhang X, Yu W, Li Q. A novel ferroptosis-related biomarker signature to predict overall survival of esophageal squamous cell carcinoma. *Front Mol Biosci*. (2021) 8:675193. doi: 10.3389/fmolb.2021.675193

48. Xie J, Chen Z, Liu L, Li P, Zhu X, Gao H, et al. shRNA-mediated Slc38a1 silencing inhibits migration, but not invasiveness of human pancreatic cancer cells. *Chin J Cancer Res*. (2013) 25:514–9. doi: 10.3978/j.issn.1000-9604.2013.09.03

49. Su H, Tao T, Yang Z, Kang X, Zhang X, Kang D, et al. Circular RNA cTFRC acts as the sponge of MicroRNA-107 to promote bladder carcinoma progression. *Mol Cancer*. (2019) 18:27. doi: 10.1186/s12943-019-0951-0

50. Ji J, Xu MX, Qian TY, Zhu SZ, Jiang F, Liu ZX, et al. The AKR1B1 inhibitor epalrestat suppresses the progression of cervical cancer. *Mol Biol Rep*. (2020) 47:6091–103. doi: 10.1007/s11033-020-05685-z

51. Wu X, Li X, Fu Q, Cao Q, Chen X, Wang M, et al. AKR1B1 promotes basal-like breast cancer progression by a positive feedback loop that activates the EMT program. *J Exp Med*. (2017) 214:1065–79. doi: 10.1084/jem.20160903

52. Zhang SQ, Yung KK, Chung SK, Chung SS. Aldo-keto reductases-mediated cytotoxicity of 2-deoxyglucose: A novel anticancer mechanism. *Cancer Sci*. (2018) 109:1970–80. doi: 10.1111/cas.2018.109.issue-6

53. Yang L, Shi T, Liu F, Ren C, Wang Z, Li Y, et al. REV3L, a promising target in regulating the chemosensitivity of cervical cancer cells. *PloS One*. (2015) 10:e0120334. doi: 10.1371/journal.pone.0120334

54. Zhu X, Zou S, Zhou J, Zhu H, Zhang S, Shang Z, et al. REV3L, the catalytic subunit of DNA polymerase ζ, is involved in the progression and chemoresistance of esophageal squamous cell carcinoma. *Oncol Rep*. (2016) 35:1664–70. doi: 10.3892/or.2016.4549

55. Li J, Xu X, Jiang Y, Hansbro NG, Hansbro PM, Xu J, et al. Elastin is a key factor of tumor development in colorectal cancer. *BMC Cancer*. (2020) 20:217. doi: 10.1186/s12885-020-6686-x

56. Salesse S, Odoul L, Chazée L, Garbar C, Duca L, Martiny L, et al. Elastin molecular aging promotes MDA-MB-231 breast cancer cell invasiveness. *FEBS Open Bio*. (2018) 8:1395–404. doi: 10.1002/feb4.2018.8.issue-9

57. Wu X, Lv D, Cai C, Zhao Z, Wang M, Chen W, et al. A TP53-associated immune prognostic signature for the prediction of overall survival and therapeutic responses in muscle-invasive bladder cancer. *Front Immunol*. (2020) 11:590618. doi: 10.3389/fimmu.2020.590618

58. Fan C, Stendahl U, Stjernberg N, Beckman L. Association between orosomucoid types and cancer. *Oncology*. (1995) 52:498–500. doi: 10.1159/000227518

59. Alečković M, Wei Y, LeRoy G, Sidoli S, Liu DD, Garcia BA, et al. Identification of Nidogen 1 as a lung metastasis protein through secretome analysis. *Genes Dev*. (2017) 31:1439–55. doi: 10.1101/gad.301937.117

60. Rokavec M, Bouznad N, Hermeking H. Paracrine Induction of Epithelial-Mesenchymal Transition Between Colorectal Cancer Cells and its Suppression by a p53/miR-192/215/NID1 Axis. *Cell Mol Gastroenterol Hepatol*. (2019) 7:783–802. doi: 10.1016/j.jcmgh.2019.02.003

61. Zhang B, Xu C, Liu J, Yang J, Gao Q, Ye F. Nidogen-1 expression is associated with overall survival and temozolomide sensitivity in low-grade glioma patients. *Aging (Albany NY)*. (2021) 13:9085–107. doi: 10.18632/aging.202789

62. Jagroop R, Martin CJ, Moorehead RA. Nidogen 1 regulates proliferation and migration/invasion in murine claudin-low mammary tumor cells. *Oncol Lett*. (2021) 21:52. doi: 10.3892/ol.2020.12313

63. Gao S, Fan C, Huang H, Zhu C, Su M, Zhang Y. Effects of HCG on human epithelial ovarian cancer vasculogenic mimicry formation *in vivo*. *Oncol Lett*. (2016) 12:459–66. doi: 10.3892/ol.2016.4630

64. Xu J, Yang R, Hua X, Huang M, Tian Z, Li J, et al. lncRNA SNHG1 promotes basal bladder cancer invasion via interaction with PP2A catalytic subunit and induction of autophagy. *Mol Ther Nucleic Acids*. (2020) 21:354–66. doi: 10.1016/j.omtn.2020.06.010

65. Liu L, Ning Y, Yi J, Yuan J, Fang W, Lin Z, et al. miR-6089/MYH9/β-catenin/c-Jun negative feedback loop inhibits ovarian cancer carcinogenesis and progression. *BioMed Pharmacother*. (2020) 125:109865. doi: 10.1016/j.biopha.2020.109865

66. Jiao F, Sun H, Yang Q, Sun H, Wang Z, Liu M, et al. Identification of FADS1 through common gene expression profiles for predicting survival in patients with bladder cancer. *Cancer Manag Res*. (2020) 12:8325–39. doi: 10.2147/CMAR.S254316

67. Funt SA, Rosenberg JE. Systemic, perioperative management of muscle-invasive bladder cancer and future horizons. *Nat Rev Clin Oncol*. (2017) 14:221–34. doi: 10.1038/nrclinonc.2016.188

68. Jackson SR, Yuan J, Teague RM. Targeting CD8+ T-cell tolerance for cancer immunotherapy. *Immunotherapy*. (2014) 6:833–52. doi: 10.2217/imt.14.51

69. Byrne A, Savas P, Sant S, Li R, Virassamy B, Luen SJ, et al. Tissue-resident memory T cells in breast cancer control and immunotherapy responses. *Nat Rev Clin Oncol*. (2020) 17:341–8. doi: 10.1038/s41571-020-0333-y

70. Hellmann MD, Ciuleanu TE, Pluzanski A, Lee JS, Otterson GA, Audigier-Valette C, et al. Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *N Engl J Med*. (2018) 378:2093–104. doi: 10.1056/NEJMoa1801946

71. Snyder A, Nathanson T, Funt SA, Ahuja A, Buros Novik J, Hellmann MD, et al. Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: An exploratory multi-omic analysis. *PloS Med*. (2017) 14:e1002309. doi: 10.1371/journal.pmed.1002309

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 19 May 2025
DOI 10.3389/fimmu.2025.1481366

OPEN ACCESS

EDITED BY
Petar Ozretić,
Rudjer Boskovic Institute, Croatia

REVIEWED BY
You Zhou,
First People's Hospital of Changzhou, China
Shiqi Ren,
Affiliated Hospital of Nantong University,
China

*CORRESPONDENCE
Henghui Cheng
hhcheng2007@hust.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 15 August 2024
ACCEPTED 28 April 2025
PUBLISHED 19 May 2025

CITATION
Sun X, Jia W, Liang H and Cheng H (2025) An immune-related signature based on molecular subtypes for predicting the prognosis and immunotherapy efficacy of hepatocellular carcinoma.
Front. Immunol. 16:1481366.
doi: 10.3389/fimmu.2025.1481366

An immune-related signature based on molecular subtypes for predicting the prognosis and immunotherapy efficacy of hepatocellular carcinoma

Xuhui Sun[1†], Wenlong Jia[2†], Huifang Liang[2]
and Henghui Cheng[1*]

[1]Department of Pathology, Tongji Hospital, Tongji Medical College Huazhong University of Science and Technology, Wuhan, Hubei, China, [2]Department of Hepatology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

Background: Immunotherapy has emerged as a pivotal therapeutic modality for a multitude of malignancies, notably hepatocellular carcinoma (HCC). This research endeavors to construct a prognostic signature based on immune-related genes between different HCC molecular subtypes, offer guidance for immunotherapy application, and promote its clinical practical application through immunohistochemistry.

Methods: Distinguishing HCC subtypes through Gene set variation analysis and Consensus clustering analysis using the Kyoto Encyclopedia of Genes and Genome (KEGG) pathway. In the TCGA-LIHC cohort, univariate, Lasso, and multivariate Cox regression analyses were applied to construct a novel immune relevant prognostic signature. The Subtype-specific and Immune-Related Prognostic Signatures (SIR-PS) were validated in three prognostic cohorts, one immunotherapy cohort, different HCC cell lines and tissue chips. Further possible mechanism on immunotherapy was explored by miRNA-mRNA interactions and signaling pathway.

Results: This prognostic model, which was based on four critical immune-related genes, *STC2*, *BIRC5*, *EPO* and *GLP1R*, was demonstrated excellent performance in both prognosis and immune response prediction of HCC. Clinical pathological signature, tumor microenvironment and mutation analysis also proved the effective prediction of this model. Spatial transcriptome analysis shows that STC2 and BIRC5 are mainly enriched in liver cancer cells and their mRNA and protein expression levels were greater in higher malignant HCC cell lines than in the lower ones. Further validation on HCC tissue chips of this model also showed good correlation with cancer prognosis. The risk score of each patient demonstrated that the SIR-PS exhibited excellent 1 and 3-year survival prediction performance.

Conclusions: Our analysis demonstrates that the SIR-PS model serves as a robust prognostic and predictive tool for both the survival outcomes and the response to immunotherapy in hepatocellular carcinoma patients, which may shed light on promoting the individualized immunotherapy against hepatocellular carcinoma.

KEYWORDS

hepatocellular carcinoma, immune-related genes, prognosis, immunotherapy, immunohistochemistry, biomarker

1 Introduction

According to the 2020 Global Cancer Statistics, liver cancer is the sixth most common human malignancy and the third leading cause of cancer related deaths worldwide, in which liver hepatocellular carcinoma (HCC) accounts for the vast majority (75%-85%) (1). Characterized by nonspecific symptoms and pronounced heterogeneity in the early phases, HCC is often diagnosed at advanced stages, precluding the possibility of curative surgery for the majority of patients (2). Even with the emergence of immunotherapeutic and targeted therapies, the 5-year survival rate for HCC patients remains below 20% (3). The prognosis of patients with HCC is highly variable, which is attributable to its inherent heterogeneity (4). Consequently, there is a pressing need for a novel signature that leverages tumor heterogeneity to predict patient prognosis and select immunotherapy candidates for precision medicine.

Cancer immunotherapy activates the immune system to induce the death of cancer cells (5). The tumor microenvironment (TME), which includes immune cells, stromal cells, the extracellular matrix, and peripheral blood vessels, significantly influences tumor proliferation, metabolic processes, and metastatic potential (6). What's more, TME plays a vital role in response to cancer immunotherapy in patients with HCC. Amidst the rapid advancements in immunotherapy, its role in HCC treatment is increasingly pivotal.

High-throughput transcriptome sequencing has been widely used in recent years for both clinical and research purposes. However, stringent requirements, intricate procedures and elevated costs impeded its widespread adoption. Immunohistochemistry (IHC) offers a practical and economical alternative for determining protein expression via antibody-mediated staining. Currently, the majority of studies rely on RNA-Seq data for prognostic assessments, whereas models utilizing IHC are limited. If gene-guided predictions can ultimately be validated and applicated through IHC, that will provide a more convenient and cost-effective option.

Therefore, in this study, we established a prognostic model based on HCC subtypes and immune related genes. This model was also proofed by the immunohistochemical score to facilitate clinical prognosis and treatment. Figure 1 illustrates the methodological steps undertaken in this study. The findings might provide insights for future IHC-based studies and contribute to advanced individualized immune therapies for HCC.

2 Materials and methods

2.1 Data resources

This investigation procured RNA -Seq, clinical, and SNP data from HCC patients through the TCGA (https://portal.gdc.cancer.gov/) and ICGC (https://dcc.icgc.org/) databases, with the exclusion of subjects lacking complete overall survival (OS) data or having survival durations of less than 30 days. The GSE54236 dataset and GSE202069 dataset, sourced from the GEO (https://www.ncbi.nlm.nih.gov/geo/) database, were incorporated into this analysis. The complete TCGA-LIHC cohort served as the training set, while the ICGC- LIRI-JP, GSE54236 and GSE202069 cohorts were utilized as validation datasets.

2.2 Gene set variation analysis and consensus clustering

The GSVA algorithm, implemented in the "GSVA" package (7), was employed to derive the relative enrichment scores for the entirety of Kyoto Encyclopedia of Genes and Genome (KEGG) pathways that referenced from the MSigDB (c2.cp.kegg.v2023.1.Hs.symbols) for the comprehensive TCGA cohort (8, 9).

Unsupervised hierarchical clustering of all HCC patients from the TCGA cohort was conducted using the "ConsensusClusterPlus" package (10) to discern distinct HCC subtypes. This procedure entailed 1000 iterations, sampling 80% of the dataset per iteration, to ascertain the stability and reliability of the resulting clusters. The optimal cluster number was determined through the application of the proportion of ambiguous clustering algorithm (11, 12).

Abbreviations: HCC, hepatocellular carcinoma; IHC, immunohistochemistry; OS, overall survival; GSVA, Gene set variation analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; SDEGs, the differentially expressed genes among the two subtypes; SIRDEGs, immune-related differentially expressed genes among HCC subtypes; IRGs, immune-related genes; LASSO, least absolute shrinkage and selection operator; SIR-PS, Subtype-specific and Immune-Related Prognostic Signatures; t-ROC, time-dependent receiver operating characteristic curves; qPCR, quantitative Polymerase Chain Reaction; 97H, MHCC-97H.

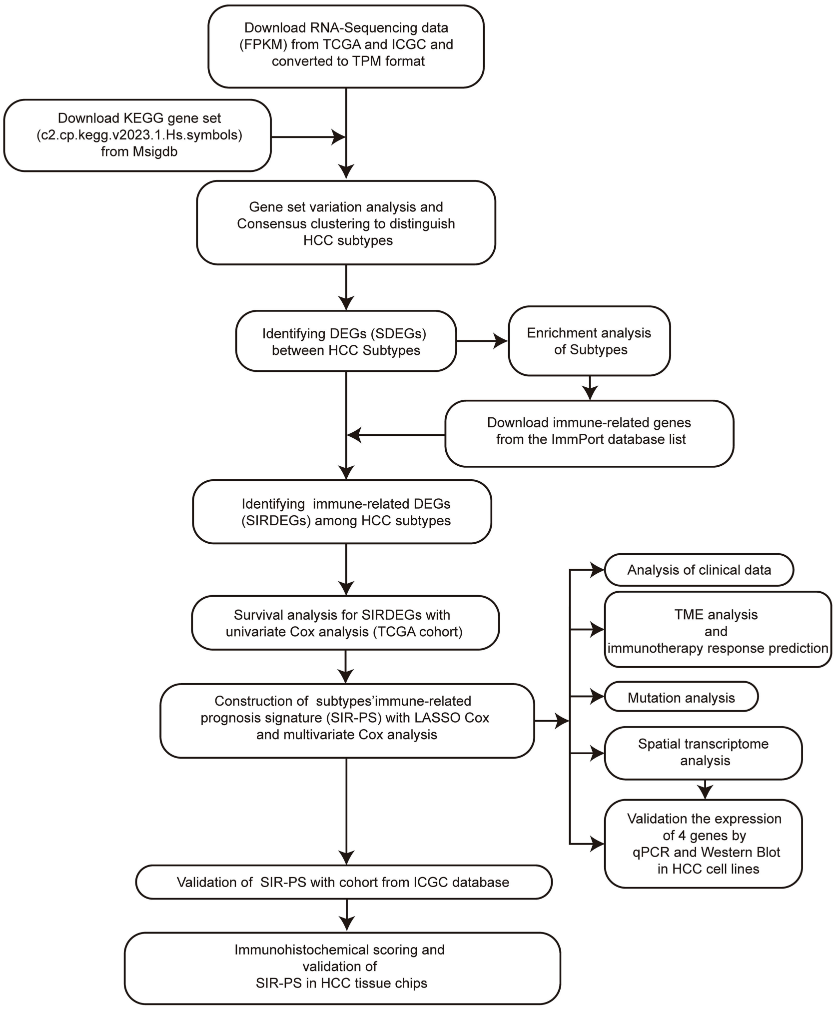


FIGURE 1
Flow diagram of the analysis procedure: data collection, preprocessing, analysis and validation.

2.3 Differential and enrichment analysis of the subtypes

Using the "DESeq2" software package for differential analysis, screening differential expressed genes between the two subtypes (SDEGs) based on adjusted P value<0.05 and absolute value of logFC>1 as criteria (13, 14). Utilizing the "clusterProfiler" R package (15), we performed enrichment analysis on above differential genes using gene sets from diverse databases, including OMIM disease gene set, OMIM expanded gene set, ClinVar 2019 gene set, and Rare Diseases GeneRIF Gene Lists gene sets (16–18).

2.4 Identification of immune-related differentially expressed genes among HCC subtypes (SIRDEGs)

Immune-related genes (IRGs) were identified from the Immunology Database and Analysis Portal (ImmPort, https://immport.niaid.nih.gov/). Intersection of SDEGs and IRGs to obtain SIRDEGs.

2.5 Construction and validation of a prognostic signature based on the SIRDEGs

The TCGA-LIHC Cohort was utilized as the training set for model development. Validation was conducted using the ICGC-LIRI Cohort, the GSE54236 and GSE202069 datasets. Univariate and least absolute shrinkage and selection operator (LASSO) Cox regression analyses were performed using the "survival" and "glmnet" packages to identify the modeling genes. Subtype-specific and Immune-Related Prognostic Signatures, designated the SIR-PS, were identified through multivariate Cox regression. The computational formula for SIR-PS is given by SIR-PS = $\sum_i^n coef_i * mRNA_i$. The R packages "survivalROC" and "survminer" were used to generate time-dependent receiver operating characteristic curves (t-ROC) and Kaplan–Meier survival curves, respectively. The samples were stratified into high-risk and low-risk groups based on the median risk score derived from the TCGA-LIHC cohort. The associations between the SIR-PS and clinicopathological parameters were assessed using the chi-square test and graphically depicted using the "ComplexHeatmap" package (19). Significant clinical parameters were further represented through a stacked bar plot.

2.6 Exploration of the tumor immune microenvironment and immunotherapy response

This study utilized the CIBERSORT algorithms for a quantitative assessment of immune cell infiltration, thereby elucidating immunological variations across different groups. Additionally, we scrutinized the expression profiles of immune checkpoint molecules, conducting a comparative analysis to delineate the distinctions between the high-risk and low-risk groups. Furthermore, we leveraged the HCC Immunotherapy Cohort (RNA-Seq data from Li et al.'s study) to substantiate the predictive efficacy of the SIR-PS in forecasting responses to immunotherapy (20).

2.7 Mutation analysis

The mutational data of patients in the TCGA-LIHC cohort were obtained from the TCGA database. The "maftools" R package was utilized to evaluate the mutational landscape and compare the mutational spectra between high-risk and low-risk groups of HCC patients (21).

2.8 Spatial transcriptome analysis

The spatial transcriptomics data were obtained from Liu et al.'s study (22). According to the authors' provided data, we calculated the model score for each cell using SIR-PS. We then used the Seurat package to visualize cell types and their corresponding scores.

2.9 Quantitative real-time reverse transcriptase polymerase chain reaction in cell lines

Hep3B, Huh7, MHCC-97H (97H), and SNU-449 cell lines (ATCC Cell Bank, United States) were cultured to verify the expression levels of these signature genes. Total RNA was isolated from the aforementioned cell lines utilizing FreeZol Reagent (Vazyme, China) followed by the synthesis of cDNA using a reverse transcription kit (Vazyme, China). qPCR was conducted with SYBR Green Mix (Q711, Vazyme) and a C1000 thermal cycler from Bio-Rad (Hercules, CA). The sequences of the primers used for the signature genes are detailed in Table 1. The relative expression levels were normalized to those of the housekeeping gene *GAPDH*.

2.10 Western blotting

Cellular lysates were prepared using RIPA lysis buffer. Equal amounts of proteins were subjected to SDS–PAGE and then transferred to polyvinylidene fluoride membranes. The membranes were blocked with a protein-free rapid blocking solution (PS108P, Epizyme) for 20 minutes to prevent nonspecific antibody binding. Primary antibodies (10314, 10508, 26196 from Proteintech, A5663 from ABclonal) were diluted according to the manufacturer's instructions and incubated at 4°C for 12 hours to allow for antibody-antigen binding. After washing with Tris-Buffered Saline with Tween, secondary antibodies (SA00001 from Proteintech) were applied and incubated for 1 h at room temperature to facilitate signal detection. After washing, the immunoreactive bands on the membranes were visualized using an enhanced chemiluminescence chromogenic substrate.

2.11 Validation of SIR-PS in HCC tissue chips

Two HCC tissue chips were obtained from the Department of Liver Surgery at Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology. Patients with incomplete clinical data or tissue loss excluded from the analysis. The immunohistochemistry staining was performed as described previously (23). The slides were incubated with primary antibodies (anti-*STC2* ab255610, Abcam, anti-*GLP1R* 26196, Proteintech, anti-*EPO* ZRB1366, Sigma, and anti-*BIRC5* ZA0530, ZSGB-BIO). Semiquantitative scores were assigned according to the staining intensity and the proportion of positively stained cells, with the following categories and corresponding scores: no staining (0), light yellow (1), medium yellow (2), dark yellow (3), and heavy yellow (4); multiple with the corresponding positive percentage of stained cells relative to the total number of cells > The composite score for each specimen was calculated as the sum of the products of the staining intensity and the percentage of positively stained cells. Immunohistochemical staining was independently evaluated by two pathologists in a double-blinded manner via microscopy. The HCC

TABLE 1 The sequences of the primers used in qPCR.

Gene name	Forward primer sequence	Reverse primer sequence
GAPDH	TCCAAAATCAAGTGGGGCGA	TGATGACCCTTTTGGCTCCC
STC2	TGAAATGTAAGGCCCACGCT	ACTGTTCGTCTTCCCACTCG
BIRC5	TCAAGGACCACCGCATCTCT	CCAAGTCTGGCTCGTTCTCA
EPO	AGGCCGAGAATATCACGACG	CAGACTTCTACGGCCTGCTG
GLP1R	AGTCCAAGCGAGGGGAAAGA	GAGGCGATAACCAGAGCAGAG

tissue chip cohort was stratified into High-risk and Low-risk groups utilizing "surv_cutpoint" from the "survminer" package. The prognostic predictive efficacy of the SIR-PS was confirmed using Kaplan–Meier analysis and t-ROC curves. In the HCC tissue chip cohort, use the "compactGroups" package to generate a three line table to statistically analyze the distribution of clinical pathologica l parameters between different groups for each indicator (24).

2.12 Exploring the potential mechanisms of SIR-PS regulating immunotherapy efficacy

Analyze the relationship between modeling genes and PDL1 expression in the TCGA-LIHC cohort. Based on the above results, miRNAs targeting PDL1 and modeling genes correlated with PDL1 expression that have been experimentally validated in the TarBase database were screened using the MultiMiR package. Take the intersection of the miRNA results of the modeling genes mentioned above with the miRNAs targeting PDL1.

2.13 Statistical analysis

All the statistical tests and bioinformatics analyses were performed with R software, version 4.0.1. The Wilcoxon rank sum test, Pearson chi-square test, t test and log-sum test were included. P <0.05 was considered to indicate statistical significance.

3 Results

3.1 Identification and enrichment analysis of subtypes based on KEGG pathway in HCC

Utilizing the enrichment scores of KEGG gene sets based on the GSVA algorithm, we conducted unsupervised hierarchical clustering to classify the samples into two distinct subtypes, which were validated by the examination of the cluster heatmap, the consensus CDF plot, the average silhouette width, and the Proportion of Ambiguous Clustering algorithm (Figures 2A-C). Consequently, the patients of TCGA-LIHC cohort was stratified into two distinct molecular subtypes (Supplementary Table 1). Subsequently, a comparative analysis of the clinical factors across different subtypes was conducted, employing heatmap for visualization (Figure 2D). Additionally, stacked bar charts were utilized to highlight factors exhibiting significant inter-subtype disparities (Figure 2E). Compared to Sub2, Sub1 is characterized by elevated levels of AFP, a higher GRADE, advanced path stage and T stage, a greater proportion of female patients, and a lower median age. As indicated by the Kaplan–Meier analysis, patients classified into Sub2 exhibited a more favorable prognosis than those classified into Sub1 (Figure 2F). In light of the observed disparities in survival outcomes, we employed the "DESeq2" package to perform a differential analysis between the two identified subtypes. Employing LogFC>1 as the criterion, Sub1 and Sub2 were found to harbor 2284 and 751 differentially expressed genes, respectively. The enrichment analysis conducted on Sub1 disclosed that in the gene sets of the four databases, OMIM disease, OMIM Expanded, ClinVar2019, and Rare Disease GeneRIF GeneLists, the genes enriched by Sub1 are unanimously associated with immunodeficiency diseases (Figure 2G). Additionally, an analysis of immune checkpoint expression levels among the subtypes was performed. This analysis indicated that the expression levels of immune checkpoint genes in Sub1 were, on the whole, markedly elevated compared to those in Sub2 (Figure 2H).

3.2 Development and validation of the SIR-PS

In the TCGA-LIHC cohort we obtained 3035 SDEGs (Supplementary Figure 1). Then, SDEGs were intersected with 1,509 immune-related genes obtained from the ImmPort database, yielding a total of 239 immune-related SDEGs (SIRDEGs) (Figure 3A). Univariate Cox regression analysis revealed 67 SIRDEGs with significant prognostic potential (Supplementary Table 2). Then, LASSO regression analysis was performed, and five SIRDEGs were further identified for modeling (Figures 3B, C). Four genes, *STC2*, *BIRC5*, *EPO*, and *GLP1R*, were identified for their substantial influence on the prognostic model. The group with high expression levels of these genes exhibited a markedly poorer prognosis than the group with low expression (Supplementary Figures 2A-D). These genes were subsequently utilized to construct a prognostic model (called SIR-PS) through multivariate Cox regression analysis, resulting in the following risk score formula: risk score = ($STC2 \times 0.22344$) + ($BIRC5 \times 0.19238$) + ($EPO \times 0.11058$) + ($GLP1R \times 0.24472$). The four-gene model demonstrated a prediction performance closely comparable to that of the five-gene model (Supplementary Figures 2E, F). Subsequently, 343 TCGA-LIHC patients were stratified into low-risk and high-risk groups based on the median risk score. In addition, we plotted ensemble plots of survival status and four signature gene expression profiles as the risk score increased (Figures 3D, E). There was a progressive increase in both mortality rates and the expression levels of the four signature genes concomitant with increasing risk scores. Kaplan–Meier analysis revealed that patients in the high-risk group experienced a more adverse clinical prognosis than did those in the low-risk group (Figure 3F). The AUC value at 1 and 3 years were 0.771 and 0.727 respectively, which is indicative of the model's robust predictive capability (Figure 3I). Leveraging the SIR-PS, we computed individual risk scores for all HCC patients within the ICGC cohort. These scores were then stratified to distinguish between high-risk and low-risk groups based on the median value of the risk scores. Consistent with the findings in the TCGA-LIHC cohort, the Kaplan–Meier analysis demonstrated that the OS of patients in the high-risk group was significantly inferior to that of patients in the low-risk group in ICGC-LIRI-JP cohort (Figure 3G). The AUCs for the ICGC-LIRI-JP cohort at 1 and 3 years were 0.791

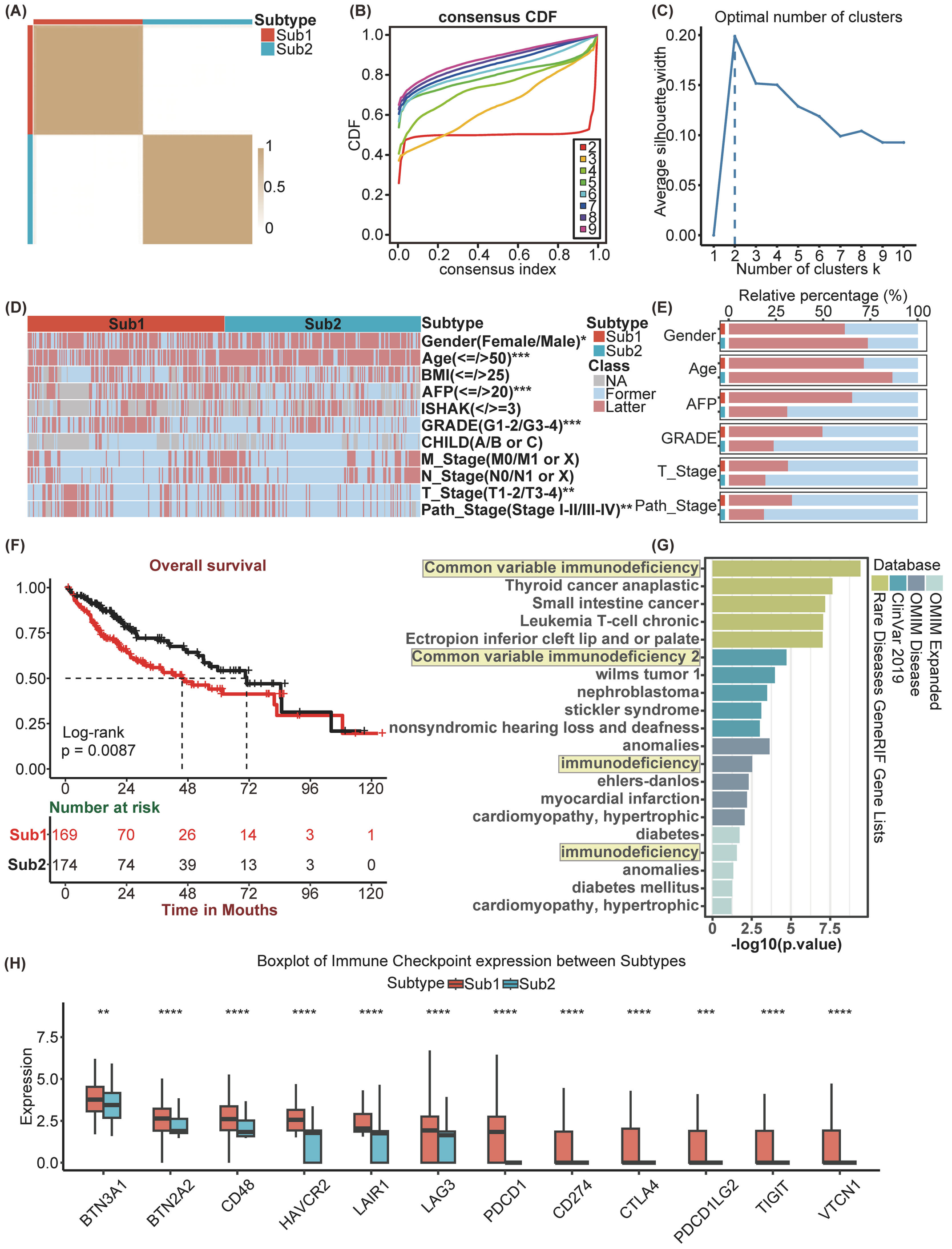


FIGURE 2
Identification and differential analysis of HCC Subtypes based on KEGG pathways. **(A)** Heatmap of sample clustering at consensus k=2. **(B)** Consensus clustering CDF for k= 2 to 9. **(C)** The Average Silhouette width Plot. **(D)** Heatmap of and **(E)** Stacked bar chart of multiple clinicopathological features between Subtypes. **(F)** Kaplan-Meier survival plots between Subtypes for Overall Survival (OS). **(G)** Enrichment analysis of diseases associated with Sub1 enrichment genes. **(H)** Immune Checkpoint genes' expression between Subtypes. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

and 0.751, respectively (Figure 3J). In the GSE54236 cohort and the GSE202069 cohort, the Kaplan–Meier analysis revealed that patients in the high-risk group experienced significantly shorter OS than those in the low-risk group (p<0.0001 and p=0.08, respectively) (Figure 3H, Supplementary Figure 3A). The AUCs for the GSE54236 cohort and the GSE202069 cohort at the 1-year were 0.838 and 0.818, respectively, while at the 3-year they were 0.67 and 0.866, respectively (Figure 3K, Supplementary Figure 3B).

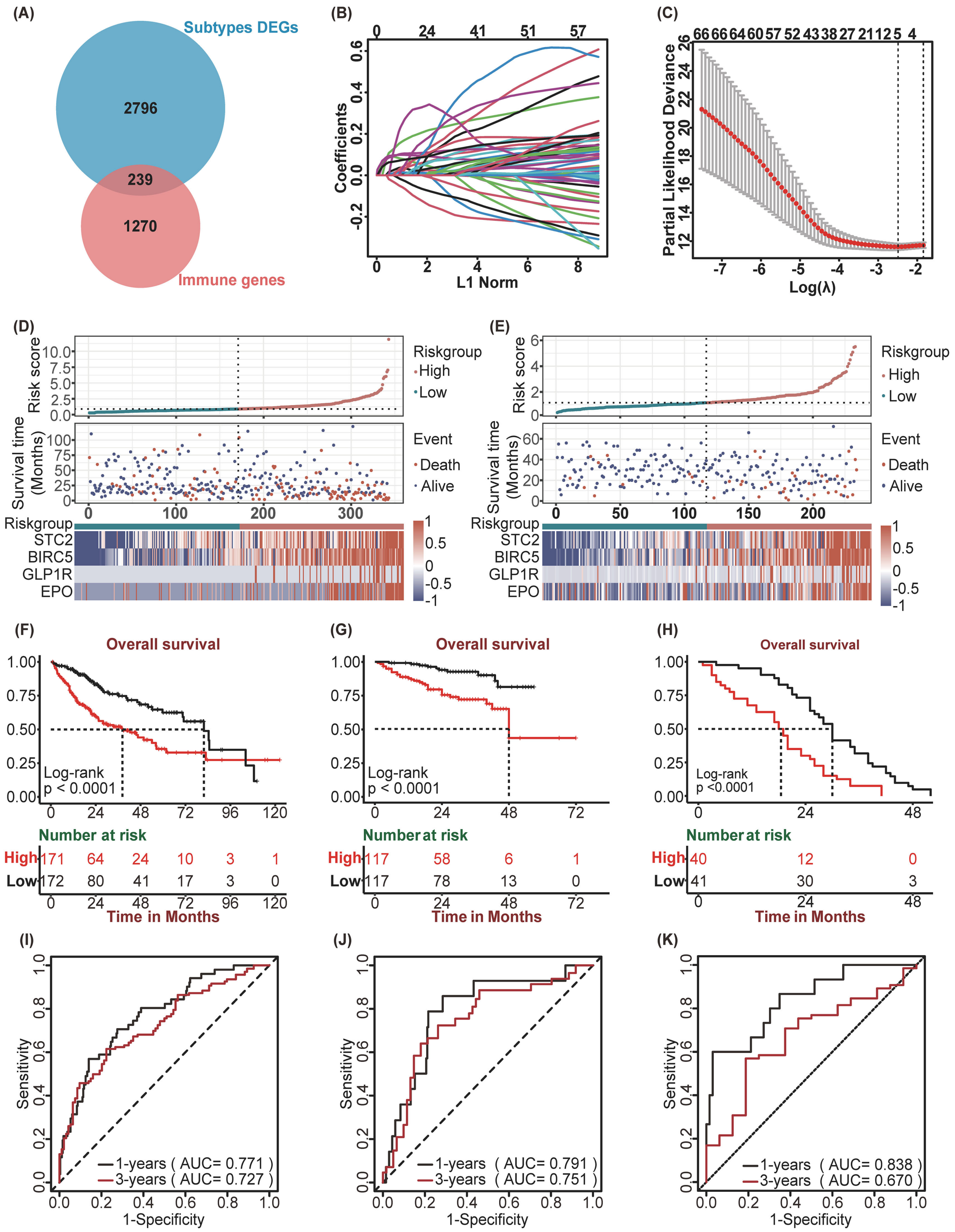


FIGURE 3

Construction and Validation of SIR-PS. **(A)** Venn plot showed 239 immune-related DEGs among subtypes. **(B)** LASSO coefficient profiles of 67 prognostic genes of HCC. **(C)** 10-fold cross validated lasso regression identified five prognostic genes with minimal λ. **(D, E)** Riskscore distribution, survival status, and expression of four SIR-PS signature genes of patients in the Low-risk and High-risk group of TCGA Cohort and ICGC Cohort, respectively. **(F–H)** Kaplan-Meier survival plots of High-risk and Low-risk group for Overall Survival in the TCGA Cohort, the ICGC Cohort and the GSE54236 Cohort. **(I–K)** Time-dependent ROC curves of SIR-PS for Overall Survival in the TCGA Cohort, the ICGC Cohort and the GSE54236 Cohort.

In addition, we compared the three-year survival prediction performance of SIR-PS with nine other prognostic models in four datasets (25–33). The results showed that the AUC value of SIR-PS had the best predicted performance in these datasets (Supplementary Figure 3C). Taking the average AUC value at 1 and 3 year of four datasets, the AUC value of SIR-PS ranked the second and first respectively, which also proofed its comprehensive prediction value.

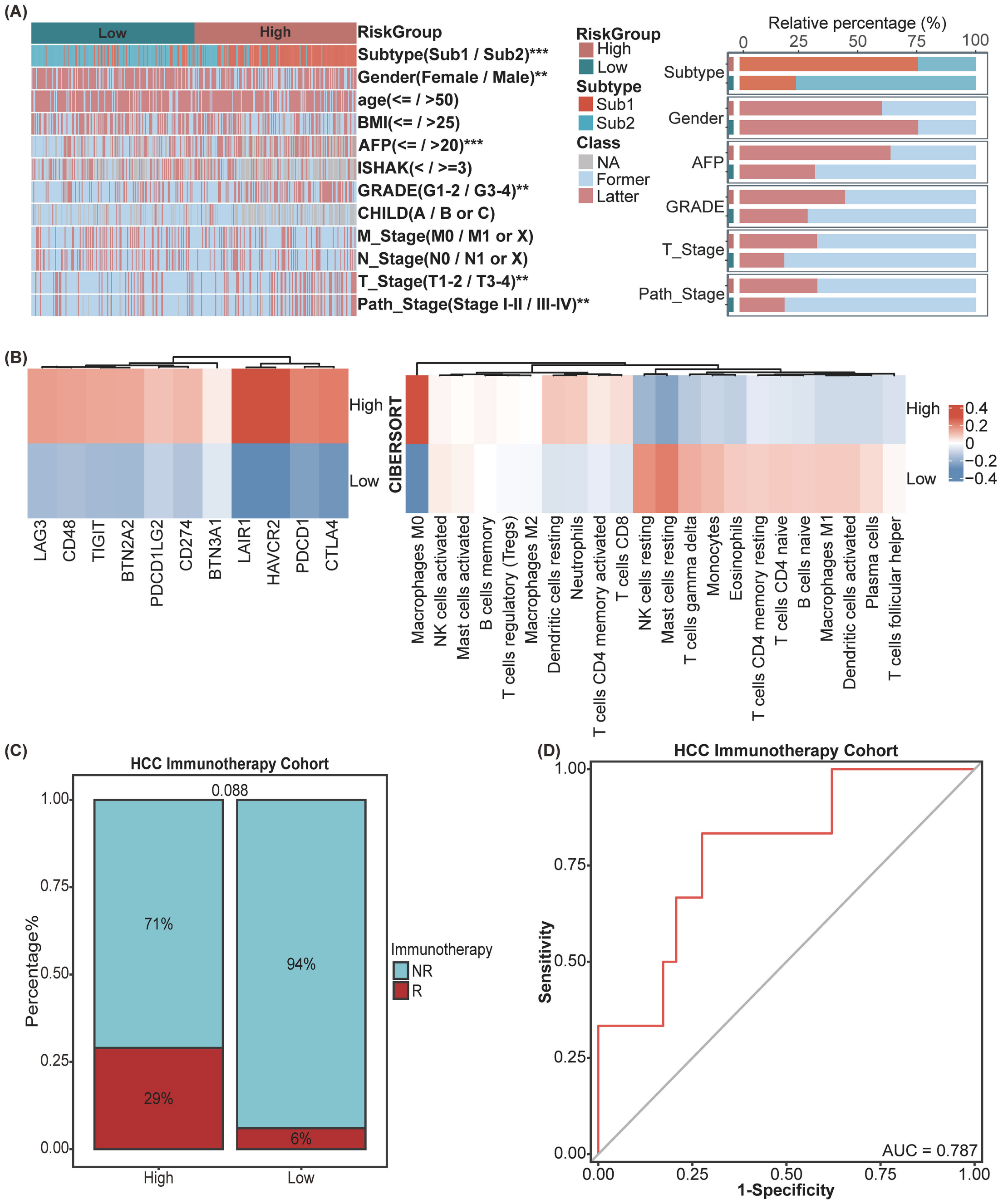


FIGURE 4

Exploration of clinical significance and tumor microenvironment of SIR-PS in the TCGA Cohort. **(A)** Heatmap and Stacked bar chart of multiple clinicopathological features between High-risk and Low-risk group of SIR-PS. **(B)** Heatmap of Immune Checkpoint expression and CIBERSORT result between High-risk and Low-risk group of SIR-PS. **(C)** Stacked bar chart of immunotherapy response between High-risk and Low-risk group of SIR-PS in our HCC Immunotherapy Cohort. **(D)** Diagnostic ROC plot of SIR-PS predicting response to immunotherapy. **p<0.01, ***p<0.001.

3.3 Exploration of the clinical significance and tumor microenvironment of the SIR-PS

To investigate the association between the SIR-PS and a range of clinicopathological characteristics, the correlation analysis was conducted and revealed significant associations between the risk groups and various HCC features (Figure 4A). The high-risk group exhibited increased levels of AFP, a greater percentage of patients within Sub1 and female patients, more advanced GRADE, and higher pathological stage and T stage than did the low-risk group (Figure 4A). Subsequently, leveraging the CIBERSORT algorithm, we quantified the infiltration levels of various immune cells across samples and delineated the comparative immune landscape between the high-risk and low-risk groups within the TCGA cohort. The analysis delineated that the high-risk group was distinguished by an enhanced infiltration of B cells memory, T cells regulatory, Dendritic cells resting, Neutrophils, T cells CD4 memory activated and T cells CD8 and a diminished presence of NK cells activated, Mast cells activated and resting, Macrophages M1, Dendritic cells activated (Figure 4B). Further analysis of immune checkpoint gene expression between risk groups within the TCGA database revealed that the high-risk group displayed elevated expression levels for the majority of these genes, in contrast to the low-risk group (Figure 4B). Concurrently, we assessed the predictive efficacy of SIR-PS concerning the response to immunotherapy within the HCC Immunotherapy Cohort, with results indicating a higher response rate among patients in the high-risk group as per the median risk score (Figure 4C). The t-ROC curve analysis revealed that the AUC value for predicting treatment responsiveness based on the risk score was 0.787 (Figure 4D).

3.4 Mutation landscape analysis of SIR-PS

Initially, we scrutinized the 10 genes exhibiting the highest mutation frequencies within the low-risk and high-risk group. Oncoplots revealed that within the TCGA database, the genes exhibiting the highest mutation frequencies in the high-risk and low-risk groups were TP53, with a 40% mutation frequency, and CTNNB1, with a 33% mutation frequency, respectively (Figures 5A, B).

3.5 Spatial transcriptome analysis of SIR-PS

To determine the cell types in which our model is active, we analyzed spatial transcriptomics data from HCC patients. Our analysis revealed that the riskscores highest in HCC cells, indicating that the SIR-PS's riskscore in patients is predominantly determined by its riskscore in these cancer cells (Figure 5C). Concurrently, *STC2* and *BIRC5* exhibit predominant expression within HCC cells.

3.6 qPCR and Western blotting in HCC cell lines

In light of the spatial transcriptome analysis findings, we chose HCC cell lines, including SNU-449, 97H, Hep3B and Huh7, to conduct cellular-level validation studies. The SNU-449 and 97H cell lines exhibited a greater degree of malignancy or transfer ability than the Hep3B and Huh7 cell lines, which commonly means a worse prognosis (34, 35). No matter in the qPCR or the western blotting detection, the expression levels of *STC2* and *BIRC5* were higher in the SNU-449 and 97H cell lines than in the Huh7 and Hep3B cell lines (Figures 5D, E), which were in accordance with their malignancies. However *EPO* and *GLP1R* showed not obvious trends in the mRNA and protein levels.

3.7 Validation of the SIR-PS based on iHC staining of the HCC tissue chips

Owing to the remarkable prognostic potential of the four signature genes, we conducted IHC staining on tissue chips sourced from HCC patients and subsequently scored the expression of these genes. Post-IHC staining revealed that *STC2*, *BIRC5*, *EPO*, and *GLP1R* exhibited increased expression in HCC tissues relative to normal controls (Figures 6A–D). Utilizing the "surv_cutpoint" function from the "survminer" package, the IHC scores for each gene were stratified into high-IHC and low-IHC groups. Kaplan–Meier analysis demonstrated that Patients in the high-IHC group for *STC2*, *BIRC5*, *EPO* and *GLP1R* exhibited a markedly poorer prognosis than did those in the low-IHC group (Figures 6E-H). Subsequently, we calculated the riskscore of each patient of HCC tissue chips using SIR-PS based on the IHC score of four genes. The riskscores of patients were subsequently categorized into high-risk and low-risk group using "surv_cutpoint" function of "survminer" package. Based on the calculated risk scores, patient stratification into high-risk and low-risk groups was determined using a cutoff value of 0.6115285. Kaplan–Meier analysis indicated that across the entire HCC tissue chip cohort, the high-risk group had a significantly worse prognosis than did the low-risk group (Figure 7A). ROC curve analysis revealed that the AUC value for the entire HCC tissue chip cohort at the 1 and 3-year was 0.711 and 0.795, respectively (Figure 7B). Furthermore, given that *GPC3* and *CK19* are commonly used prognostic markers in clinical liver cancer diagnostics, we also conducted IHC staining for these markers on HCC tissue chips and scored them accordingly. Subsequent to their score, these two prognostic indicators were evaluated independently to predict patient outcomes. Kaplan-Meier analysis revealed no significant survival disparity between the high-IHC and low-IHC groups for *CK19* and *GPC3* across the entire HCC tissue chips cohort (Figures 7C, E). Correspondingly, the 1-year AUC values of their respective t-ROC curves were 0.664 and 0.504, while the 3-year AUC values were 0.571 and 0.585, respectively (Figures 7D, F).

3.8 Exploration of clinical information between high-risk and low-risk group of patients with HCC tissue chips data

A comparative analysis of the clinical characteristics between different groups was conducted. Summary descriptives table of general clinical factors of all patients and riskgroup is shown in Table 2, while the different indicators groups are shown in Supplementary Table 3. Based on the clinical data and varying classifications of staining and risk groups, we conducted both univariate and multivariate Cox regression analysis (Table 3). The results of univariate Cox regression analysis showed that there were significant differences in survival between AST, childpugh, tumor size, vascular invasion, *BIRC5*, *EPO* and risk groups. STC2 and GLP1R cannot be subjected to Cox regression analysis due to the

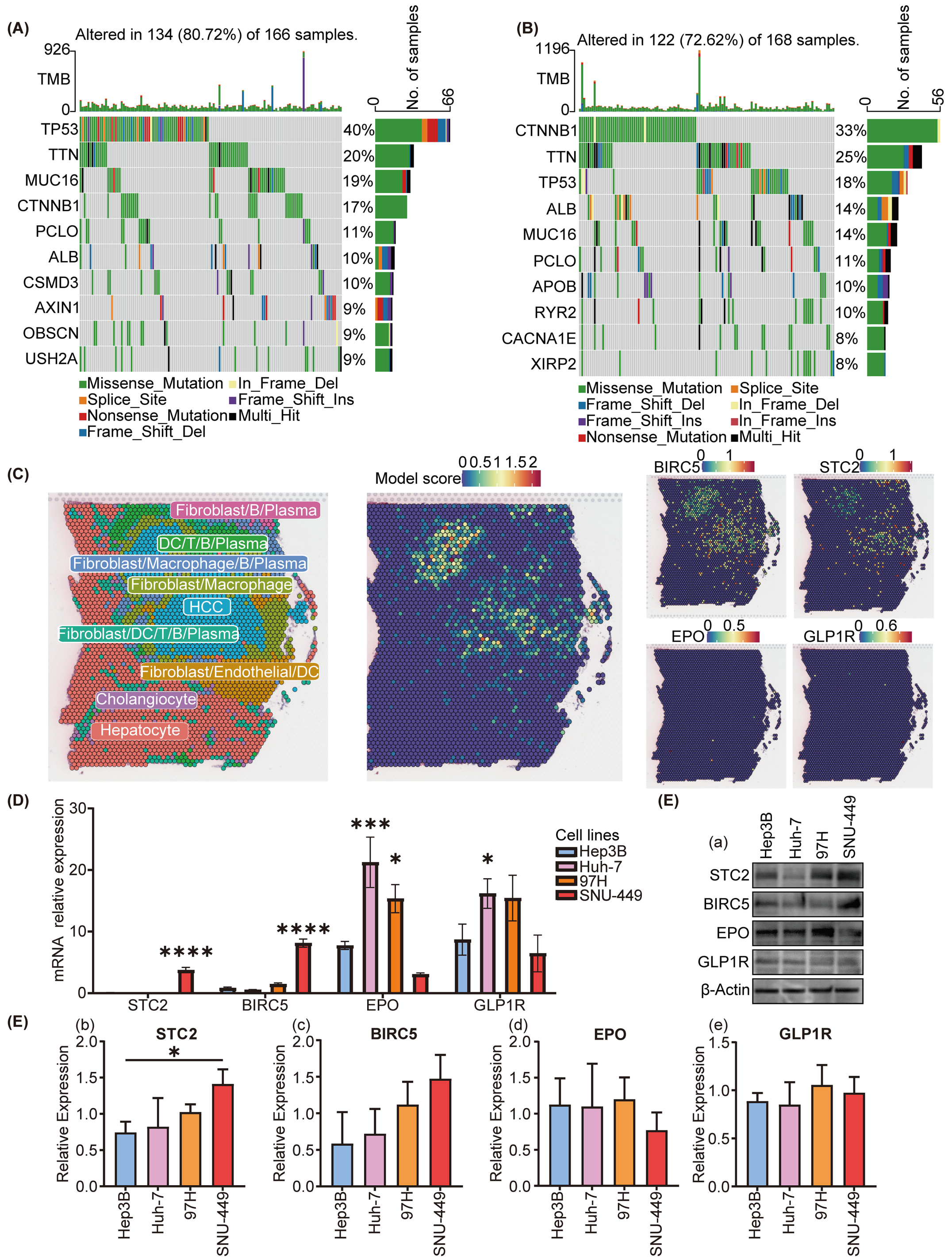


FIGURE 5

Mutational and spatial transcriptome analysis of SIR-PS risk groups and cell experiment of different cell lines. **(A, B)** Oncoplot analysis of the high-risk and low-risk group, respectively. **(C)** Spatial expression pattern of SIR-PS (including BIRC5, STC2, EPO and GLP1R). **(D)** qPCR and **(E)** Western Blotting result of Hep3B, Huh7, 97H and SNU-449 (compare with Hep3B cell lines). *p<0.05, ***p<0.001, ****p<0.0001.

fact that the number of deceased patients in one of the high and low IHC groups is less than 3. Given that risk group are determined by the expression levels of *STC2*, *BIRC5*, *GLP1R*, and *EPO*, we prioritized risk group, and other factors exhibiting significant intergroup survival differences for inclusion in the multivariate Cox regression analysis. In the multivariate Cox regression analysis, the risk group remained the sole significant predictor, with a p-value<0.05.

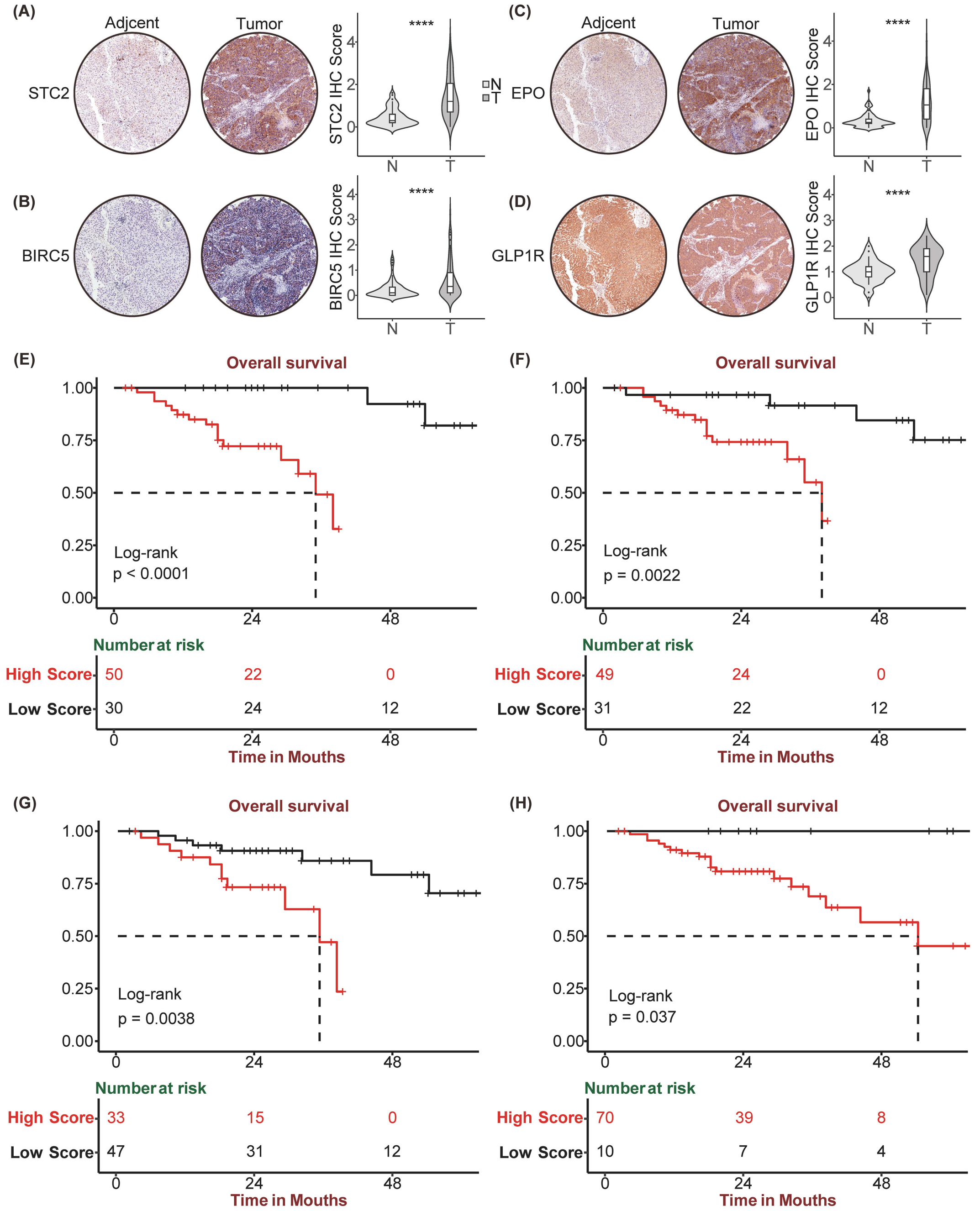


FIGURE 6
Immunohistochemistry staining and corresponding Kaplan Meier analysis of STC2, BIRC5, EPO, GLP1R. Tumor and paired Normal tissue IHC staining of HCC tissue chip by STC2 **(A)**, BIRC5 **(B)**, EPO **(C)**, GLP1R **(D)**. Kaplan-Meier curve between high and low expression of STC2 **(E)**, BIRC5 **(F)**, EPO **(G)**, GLP1R **(H)** in HCC tissue chip Cohort, respectively. ****p<0.0001.

3.9 Exploration of the mechanism by which prognostic models affect immunotherapy

Due to the good predictive effect of risk scores in the liver cancer immunotherapy queue treated with anti-PD1/PDL1, SIR-PS may affect the efficacy of immunotherapy by affecting PDL1 expression. In the TCGA-LIHC cohort, the expression levels and risk scores of STC2 and BIRC5 were positively correlated with the expression level of PDL1 (Figures 8A, B, E, F), indicated that both STC2 and BIRC5 can promote the expression of PDL1 on cancer cells, thereby promoting tumor immune escape. However, there is no obvious correlation between EPO or GLP1R and PDL1 (Figures 8C, D). Further exploration on the potential mechanism of STC2 and BIRC5 regulating PDL1 was conducted. Since the potential mutual influence of gene expression through miRNAs, the multiMIiR package was used to screen miRNAs in the TarBase database that have been experimentally validated to bind to STC2, BIRC5, and PDL1. There are a total of 87 miRNAs targeting PDL1, with 61 shared miRNAs between PDL1 and STC2, and 48 shared miRNAs between PDL1 and BIRC5 (Figures 8G, H, Table 4).

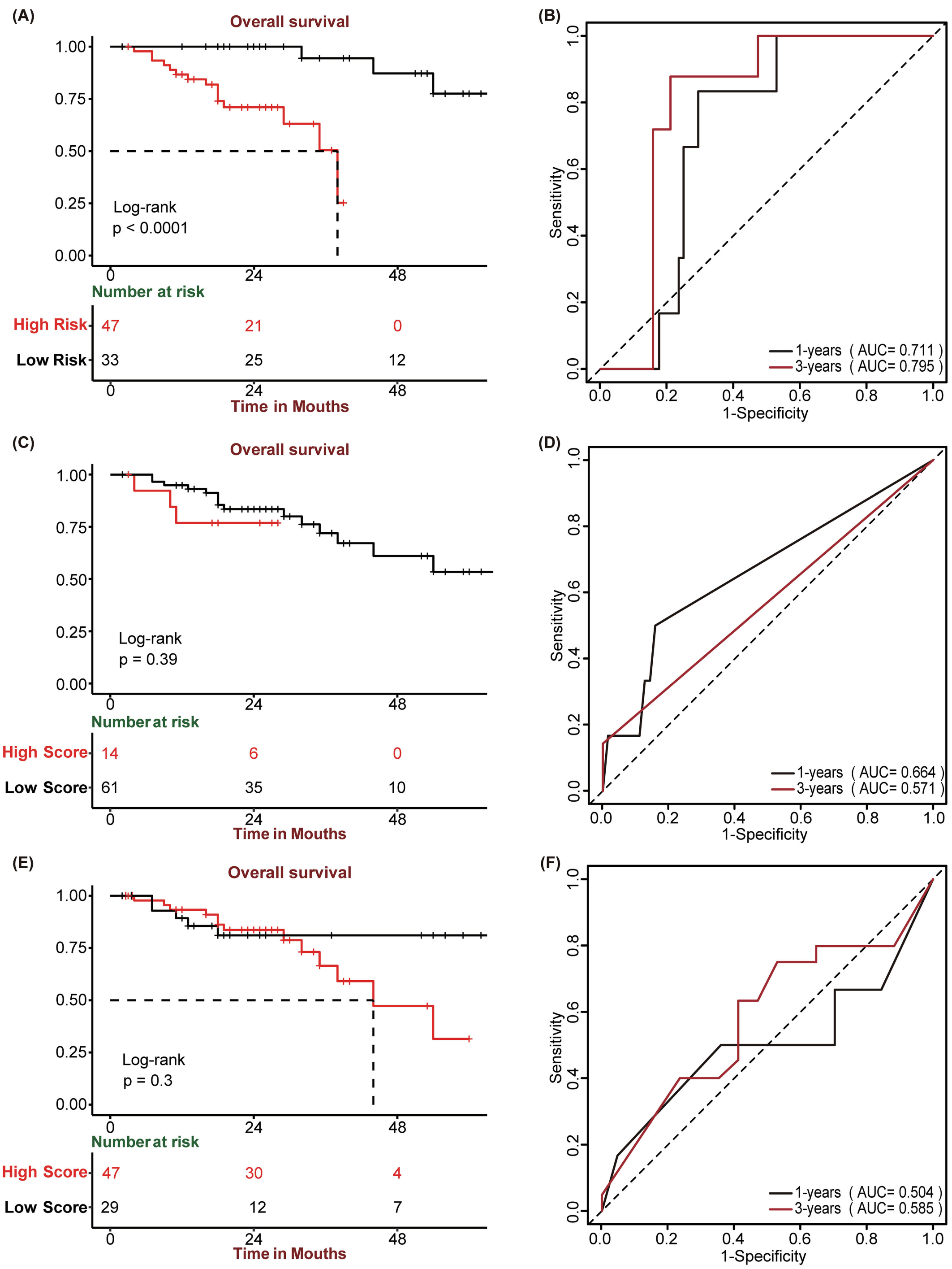


FIGURE 7
Prognostic performance of SIR-PS, GPC3, and CK19. **(A, C, E)** Kaplan-Meier survival plots of SIR-PS risk group, CK19 and GPC3 expression group for Overall Survival in the HCC tissue chip Cohort. **(B, D, F)** Time-dependent ROC curves of SIR-PS, CK19 and GPC3 for Overall Survival in the HCC tissue chip Cohort.

Discussion

Currently, nonsurgical therapeutic interventions are instrumental in the management of HCC, as the majority of patients present with advanced disease stages that preclude surgical intervention (2). As immunotherapy continues to evolve, the role of immunotherapies in the management of HCC has become increasingly pivotal, exerting a profound influence on patient prognosis. In this study, we constructed a prognostic and immunotherapy efficacy prediction model SIR-PS based on two distinct HCC molecular subtypes. This model consists of four genes: *STC2*, *BIRC5*, *EPO*, and *GLP1R*. Using a group of genes to build the

TABLE 2 Summary descriptives table of all patients and riskgroup in the tissue chips cohort.

	All	IHC Riskgroup		
		Low	High	p.overall
N:	80 (100%)	33 (41.2%)	47 (58.8%)	
Gender:				0.233
Female	16 (20.0%)	4 (12.1%)	12 (25.5%)	
Male	64 (80.0%)	29 (87.9%)	35 (74.5%)	
Age:				0.79
<=50	39 (48.8%)	15 (45.5%)	24 (51.1%)	
>50	41 (51.2%)	18 (54.5%)	23 (48.9%)	
ALT:				1
<=41	61 (76.2%)	25 (75.8%)	36 (76.6%)	
>41	19 (23.8%)	8 (24.2%)	11 (23.4%)	
AST:				0.234
<=40	56 (70.0%)	26 (78.8%)	30 (63.8%)	
>40	24 (30.0%)	7 (21.2%)	17 (36.2%)	
AFP:				1
<=20	20 (25.3%)	8 (24.2%)	12 (26.1%)	
>20	59 (74.7%)	25 (75.8%)	34 (73.9%)	
Child-Pugh:				0.139
A	76 (95.0%)	33 (100%)	43 (91.5%)	
B	4 (5.00%)	0 (0.00%)	4 (8.51%)	
Cirrhosis:				0.707
No	26 (32.5%)	12 (36.4%)	14 (29.8%)	
Yes	54 (67.5%)	21 (63.6%)	33 (70.2%)	
Tumor number:				0.933
1	59 (73.8%)	25 (75.8%)	34 (72.3%)	
>1	21 (26.2%)	8 (24.2%)	13 (27.7%)	
Tumor size:				1
<=5cm	31 (38.8%)	13 (39.4%)	18 (38.3%)	
>5cm	49 (61.3%)	20 (60.6%)	29 (61.7%)	
Vascular invasion:				1
No	66 (82.5%)	27 (81.8%)	39 (83.0%)	
Yes	14 (17.5%)	6 (18.2%)	8 (17.0%)	
Differentiation:				0.389
Moderate or High	54 (67.5%)	20 (60.6%)	34 (72.3%)	
Low or Moderately low	26 (32.5%)	13 (39.4%)	13 (27.7%)	
BCLC.stage:				0.981
A	52 (65.0%)	22 (66.7%)	30 (63.8%)	
B or C	28 (35.0%)	11 (33.3%)	17 (36.2%)	
TNM.stage:				0.475
1 or 2	61 (76.2%)	27 (81.8%)	34 (72.3%)	
3 or 4	19 (23.8%)	6 (18.2%)	13 (27.7%)	

prognostic model was successfully used in some solid tumors, such as breast cancer (36, 37). But to our knowledge, this is the first-time using SIR-PS to predict the prognosis and give suggestion of immune therapy in HCC.

Further validation on HCC cell lines revealed distinct RNA or protein expression levels of *STC2* and *BIRC5* in different malignant HCC cell lines, which were correspondence with these cells' malignances. STC2 has been revealed a marked increased expression in HCC tissues compared to normal tissues (38). Additionally, *STC2* has also been implicated in promoting tumor cell invasion and metastasis while concurrently inhibiting apoptosis in numerous tumor types (39). This heightened expression was positively correlated with an adverse patient prognosis, et al. which was consistent with our results. There was also been reported a significant overexpression of *BIRC5* in HCC tissues, contrast to its near undetectability in tissues affected by cirrhosis (40). The expression of *BIRC5* appears to be correlated with the metastatic potential of HCC, which is aligns with the findings of this study.

However, same trends didn't been observed on *EPO* and *GLP1R* in different HCC cells. In this study, we found that *EPO and GLP1R* could

TABLE 3 Cox Univariate and Multivariable regression analysis between cumulative overall survival rate and clinicopathological variables of all patients in the HCC tissue chip.

Variables	Univariate analysis		Multivariable analysis	
	HR (95% CI)	P-value	HR (95% CI)	P-value
Gender (Male/Female)	1.69 (0.385-7.41)	0.487		
Age (>50/<=50)	1.42 (0.55-3.68)	0.467		
ALT (>41/<=41)	1.33 (0.473-3.76)	0.586		
AST (>40/<=40)	4.4 (1.7-11.4)	0.00229	2.32 (0.777-6.9)	0.132
ALP (>130/<=130)	1.8 (0.589-5.48)	0.303		
AFP (>20/<=20)	0.656 (0.245-1.75)	0.401		
ChildPugh (B/A)	8.95 (2.36-34)	0.00129	2.81 (0.679-11.6)	0.154
Cirrhosis (Yes/No)	1.08 (0.403-2.87)	0.884		
Tumornumber (>1/1)	1.6 (0.596-4.3)	0.351		
Tumorsize (>5cm/<=5cm)	3.75 (1.08-13)	0.0371	2.5 (0.663-9.4)	0.176
Vascularinvasion (Yes/No)	3.66 (1.25-10.7)	0.0181	1.95 (0.584-6.52)	0.278
Differentiation (Low or Moderately low/Moderate or High)	1.64 (0.63-4.25)	0.312		
BCLCstage (B or C/A)	2.54 (0.994-6.49)	0.0514		
TNMstage (3 or 4/1 or 2)	2.39 (0.917-6.24)	0.0747		
STC2group (High/Low)	7.45e+08 (0-Inf)	0.997		
BIRC5group (High/Low)	7.99 (1.74-36.8)	0.00763		
GLP1Rgroup (High/Low)	2.71e+08 (0-Inf)	0.998		
EPOgroup (High/Low)	4.41 (1.48-13.1)	0.0076		
RiskGroup (High/Low)	22.8 (2.85-182)	0.0032	23.8 (2.74-207)	0.00405

promote HCC development and coincident with worse prognosis by bioinformation data and the validated results on HCC cohorts. However, the validation in different HCC cell lines didn't show obvious relationship with their corresponding malignances. The protein levels of *EPO* and *GLP1R* were even no statistical differences. Further exploration on the expression of these genes in HCC cell cohorts was taken out by spatial transcriptome analysis. Differ from STC2 and BIRC5 which were mainly expressed in liver cancer cells, *EPO* and *GLP1R* did not exhibit specific expression in a certain cell type, which could potentially be attributed to the fact that *EPO* and *GLP1R* may not predominantly expressed in HCC cell lines.

Analysis of the mutational landscape of genes between low-risk and high-risk groups of HCC revealed significant differences in TP53 and CTNNB1. *TP53* mutations are correlated with an unfavorable prognosis in HCC patients, and are predictive of potential responsiveness to immunotherapy (41). In various cell lines, TP53 mutations or knockdown lead to increased PDL1 expression (42, 43). Conversely, *CTNNB1* mutations, while indicative of a favorable prognosis, are linked to reduced efficacy of immunotherapy in HCC patients (44, 45). And patients with CTNNB1 mutations exhibit lower PDL1 expression (46, 47). Therefore, TP53 and CTNNB1 may influence the efficacy of immunotherapy by affecting PDL1 expression. These findings supported SIR-PS as the predictive model for HCC prognosis and immunotherapy efficacy. Meanwhile, validation on external HCC cohorts and gathering the corresponding clinical characteristics proofed SIR-PS as an apt prognostic model for HCC patients, demonstrating robust predictive accuracy in forecasting clinical outcomes. Patients categorized in the high-risk group by SIR-PS exhibited significantly adverse prognosis.

Moreover, microenvironment analysis showed this model could serves as an excellent and dependable tool for the prediction of treatment responses to immunotherapy. CD8+ T cells were the primary immune cells that exert anti-tumor effects (48). The expression of PDL1 on tumor cells often led to the exhaustion or reduced function of CD8+ T cells (49, 50). The mechanism of anti-PD-1 therapy is to restore the function of exhausted CD8 T cells and promote their proliferation (51, 52). In this study, a higher infiltration level of CD8+ T cell was observed within the high risk group. High CD8 T cells pave the way for anti-PD-1 therapy to restore those exhausted T cell function and finally killed the tumor cells. Meanwhile, we also detected the immune checkpoint gene expressions between risk groups which revealed that in contrast to the low-risk group, the high-risk group displayed elevated expression levels for most of these genes. This should be a direct clue for anti-PD-1/PDL1 efficiency.

In order to explain the potential mechanisms of prognostic models on the efficacy of immunotherapy, especially on the

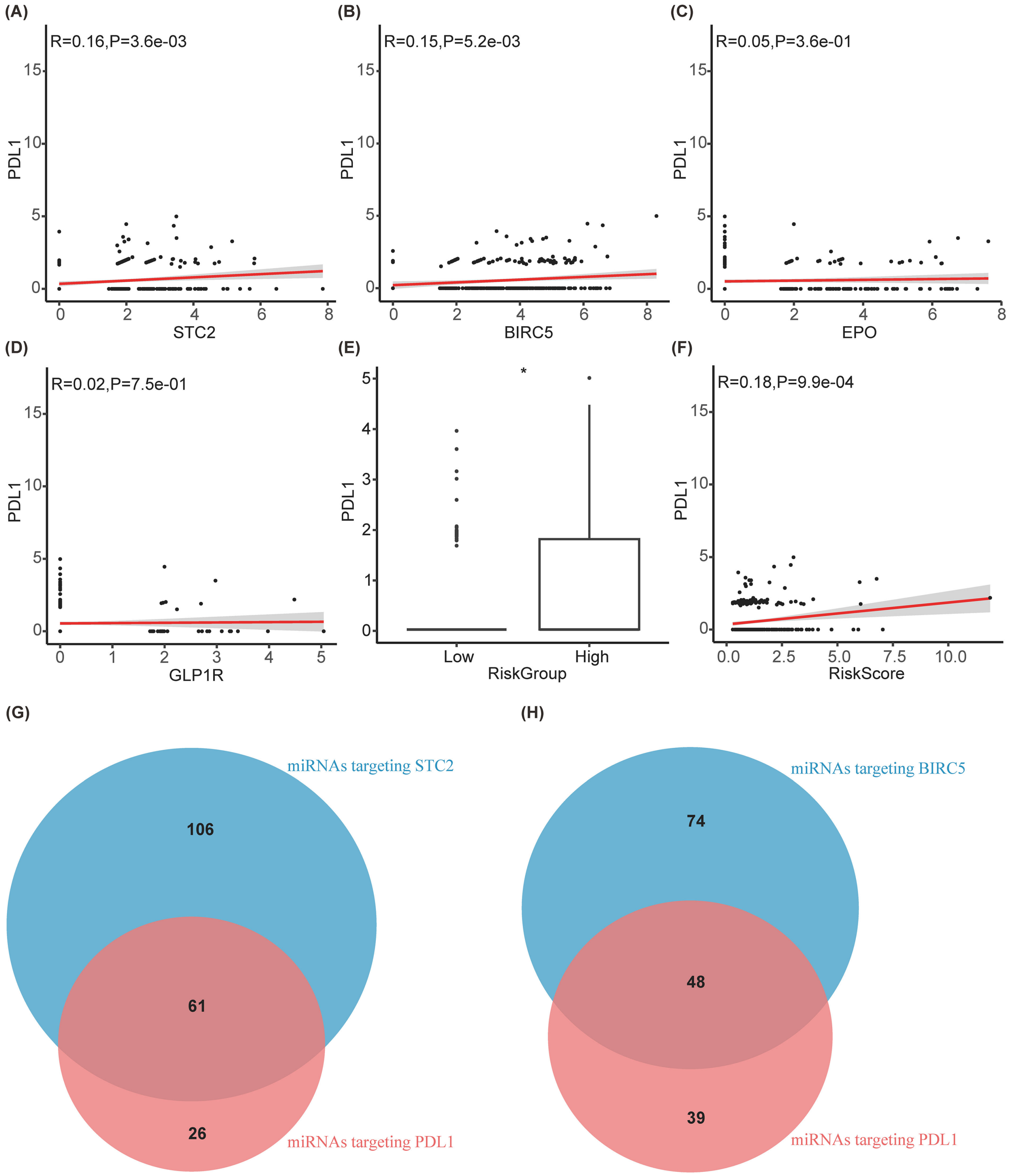


FIGURE 8
Exploration of the mechanism by which prognostic models affect immunotherapy. **(A–D)** Correlation diagram between PDL1 and STC2, BIRC5, EPO, GLP1R, respectively. **(E)** Boxplot between PDL1 and riskgroup. **(F)** Correlation diagram between PDL1 and riskscore. **(G, H)** Venn diagram of miRNAs targeting PDL1 with targeting STC2 and BIRC5, respectively.

expression of PDL1, we further explored STC2 and BIRC5 form endogenous competitive RNAs with PDL1 through multiple miRNAs, which affect the expression of PDL1. STC2 and PDL1 mRNAs can compete with each other for binding to miR-17-5p, miR-33a, miR-34a, miR-138-5p, miR-140, miR-152, miR-155, miR-197, miR-200, and miR-424 (53–68). Additionally, BIRC5 and PDL1 mRNAs also compete with each other for binding to miR-17-5p, miR-34a, miR-140, miR-142-5p, miR-152, miR-200, and miR-424 (53, 54, 56, 58, 59, 63–66, 69, 70). Consequently, an increase in the expression level of one mRNA enhances its competitive binding with miRNAs, which in turn can lead to an increase in the expression level of another mRNA to a certain

TABLE 4 The miRNAs targeting STC2, BIRC5, and PDL1.

STC2	BIRC5	PDL1
hsa-miR-106a-5p	hsa-miR-15a-5p	hsa-let-7a-5p
hsa-miR-335-5p	hsa-miR-424-5p	hsa-let-7f-5p
hsa-miR-15b-5p	hsa-miR-218-5p	hsa-miR-15a-5p
hsa-miR-20b-5p	hsa-miR-550a-3p	hsa-miR-424-5p
hsa-miR-424-5p	hsa-miR-219a-5p	hsa-miR-182-5p
hsa-miR-15a-5p	hsa-miR-21-5p	hsa-miR-15b-5p
hsa-miR-708-5p	hsa-miR-33b-5p	hsa-miR-374a-5p
hsa-miR-125b-5p	hsa-miR-320d	hsa-miR-16-5p
hsa-miR-103a-3p	hsa-miR-223-3p	hsa-miR-195-5p
hsa-miR-130a-3p	hsa-miR-219a-1-3p	hsa-miR-17-5p
hsa-miR-16-5p	hsa-miR-16-5p	hsa-miR-155-5p
hsa-miR-374b-5p	hsa-miR-195-5p	hsa-miR-302c-3p
hsa-miR-106b-5p	hsa-miR-34a-5p	hsa-miR-106a-5p
hsa-miR-576-5p	hsa-miR-20b-5p	hsa-miR-15b-3p
hsa-miR-20a-5p	hsa-let-7g-5p	hsa-miR-106b-5p
hsa-miR-17-5p	hsa-miR-452-5p	hsa-let-7b-5p
hsa-miR-124-3p	hsa-miR-181b-5p	hsa-miR-20a-5p
hsa-miR-302a-3p	hsa-miR-129-2-3p	hsa-miR-107
hsa-miR-302d-3p	hsa-miR-1225-5p	hsa-miR-1246
hsa-miR-30e-5p	hsa-miR-671-5p	hsa-miR-1292-5p
hsa-miR-876-3p	hsa-miR-30c-2-3p	hsa-miR-24-3p
hsa-miR-887-3p	hsa-miR-106a-5p	hsa-miR-34a-5p
hsa-miR-545-5p	hsa-miR-30a-5p	hsa-miR-142-5p
hsa-miR-30a-5p	hsa-miR-17-5p	hsa-miR-9-3p
hsa-miR-34a-5p	hsa-miR-182-5p	hsa-miR-130a-3p
hsa-miR-301b-3p	hsa-miR-106b-5p	hsa-miR-150-3p
hsa-miR-454-3p	hsa-miR-194-5p	hsa-miR-3928-3p
hsa-miR-155-5p	hsa-miR-576-3p	hsa-miR-93-5p
hsa-miR-130b-3p	hsa-miR-203a-3p	hsa-miR-103a-3p
hsa-miR-181b-5p	hsa-miR-7-5p	hsa-miR-301b-3p
hsa-miR-132-3p	hsa-miR-30a-3p	hsa-miR-33a-5p
hsa-miR-181a-5p	hsa-miR-20a-5p	hsa-miR-30c-1-3p
hsa-miR-4491	hsa-miR-130b-3p	hsa-miR-23a-3p
hsa-miR-181d-5p	hsa-miR-124-3p	hsa-miR-320a-3p
hsa-miR-101-3p	hsa-miR-135a-5p	hsa-miR-20b-5p
hsa-miR-139-5p	hsa-miR-130a-3p	hsa-miR-320c
hsa-miR-24-3p	hsa-miR-148a-3p	hsa-miR-18a-5p
hsa-miR-27a-3p	hsa-miR-301a-3p	hsa-miR-363-3p

(Continued)

TABLE 4 Continued

STC2	BIRC5	PDL1
hsa-miR-27b-3p	hsa-miR-301b-3p	hsa-miR-2278
hsa-miR-449b-5p	hsa-miR-454-3p	hsa-miR-183-5p
hsa-let-7a-5p	hsa-miR-10a-5p	hsa-miR-25-3p
hsa-let-7c-5p	hsa-miR-10b-5p	hsa-miR-138-5p
hsa-let-7d-5p	hsa-miR-497-5p	hsa-miR-185-5p
hsa-let-7e-5p	hsa-miR-181a-5p	hsa-miR-301a-3p
hsa-let-7f-5p	hsa-miR-142-5p	hsa-miR-374b-5p
hsa-let-7g-5p	hsa-let-7b-5p	hsa-miR-30e-3p
hsa-let-7i-5p	hsa-miR-140-3p	hsa-miR-23c
hsa-miR-196a-5p	hsa-miR-148b-3p	hsa-miR-877-5p
hsa-miR-425-5p	hsa-miR-205-5p	hsa-miR-320b
hsa-miR-7-5p	hsa-miR-1180-3p	hsa-miR-23b-3p
hsa-miR-3140-3p	hsa-miR-181d-5p	hsa-miR-32-5p
hsa-miR-625-5p	hsa-miR-200a-3p	hsa-miR-7-5p
hsa-miR-18a-5p	hsa-miR-30d-3p	hsa-miR-3934-5p
hsa-miR-18b-5p	hsa-miR-30e-3p	hsa-miR-92a-3p
hsa-miR-671-5p	hsa-miR-320b	hsa-miR-18b-5p
hsa-miR-4306	hsa-miR-542-3p	hsa-miR-590-5p
hsa-miR-3177-3p	hsa-miR-93-5p	hsa-miR-92b-3p
hsa-miR-1827	hsa-let-7d-5p	hsa-miR-320d
hsa-miR-135b-3p	hsa-miR-15b-3p	hsa-miR-19a-3p
hsa-miR-378a-3p	hsa-miR-139-5p	hsa-miR-19b-3p
hsa-miR-28-5p	hsa-miR-141-3p	hsa-miR-5000-3p
hsa-miR-19b-3p	hsa-miR-27a-3p	hsa-miR-29c-3p
hsa-miR-182-5p	hsa-miR-877-5p	hsa-miR-30a-5p
hsa-miR-423-5p	hsa-miR-25-5p	hsa-miR-30d-5p
hsa-miR-147b-3p	hsa-let-7c-5p	hsa-miR-26a-5p
hsa-miR-193b-5p	hsa-miR-671-3p	hsa-miR-26b-5p
hsa-miR-191-5p	hsa-miR-4677-3p	hsa-miR-29b-3p
hsa-miR-92a-3p	hsa-miR-1307-5p	hsa-miR-194-5p
hsa-miR-15b-3p	hsa-miR-196a-5p	hsa-miR-29c-5p
hsa-miR-218-5p	hsa-miR-423-5p	hsa-miR-584-5p
hsa-miR-98-5p	hsa-miR-22-3p	hsa-miR-4677-3p
hsa-miR-19a-3p	hsa-miR-26b-5p	hsa-let-7d-5p
hsa-miR-449c-5p	hsa-miR-375-3p	hsa-let-7c-5p
hsa-miR-30a-3p	hsa-miR-149-5p	hsa-miR-148b-3p
hsa-miR-576-3p	hsa-miR-96-5p	hsa-let-7e-5p
hsa-miR-10a-5p	hsa-miR-151b	hsa-let-7i-5p

(Continued)

TABLE 4 Continued

STC2	BIRC5	PDL1
hsa-miR-152-3p	hsa-miR-101-3p	hsa-miR-221-3p
hsa-miR-183-5p	hsa-let-7i-5p	hsa-miR-302a-3p
hsa-miR-135b-5p	hsa-miR-484	hsa-miR-196a-5p
hsa-miR-96-5p	hsa-miR-152-3p	hsa-miR-148a-3p
hsa-miR-877-5p	hsa-miR-182-3p	hsa-miR-222-3p
hsa-miR-628-5p	hsa-miR-450a-5p	hsa-miR-335-3p
hsa-let-7b-5p	hsa-miR-99b-5p	hsa-miR-191-5p
hsa-miR-107	hsa-miR-1234-3p	hsa-miR-1271-5p
hsa-miR-195-5p	hsa-miR-3184-3p	hsa-miR-340-5p
hsa-miR-503-5p	hsa-miR-328-3p	hsa-miR-34b-5p
hsa-miR-411-3p	hsa-miR-320a-3p	hsa-miR-1-3p
hsa-miR-193a-3p	hsa-miR-203b-5p	
hsa-miR-193b-3p	hsa-miR-27b-3p	
hsa-miR-205-5p	hsa-miR-19a-3p	
hsa-miR-21-5p	hsa-miR-183-5p	
hsa-miR-497-5p	hsa-miR-103a-3p	
hsa-miR-125b-2-3p	hsa-miR-15b-5p	
hsa-miR-186-5p	hsa-miR-107	
hsa-miR-320a-3p	hsa-miR-148b-5p	
hsa-miR-4677-3p	hsa-miR-29a-3p	
hsa-miR-93-5p	hsa-miR-19b-3p	
hsa-miR-29c-3p	hsa-miR-423-3p	
hsa-miR-196b-5p	hsa-miR-486-3p	
hsa-miR-29a-3p	hsa-miR-29c-3p	
hsa-miR-641	hsa-miR-30d-5p	
hsa-miR-589-3p	hsa-miR-132-3p	
hsa-miR-429	hsa-miR-103b	
hsa-miR-1301-3p	hsa-miR-17-3p	
hsa-miR-320b	hsa-miR-760	
hsa-miR-577	hsa-miR-199a-3p	
hsa-miR-532-5p	hsa-miR-199b-3p	
hsa-miR-140-3p	hsa-let-7f-5p	
hsa-miR-148a-3p	hsa-miR-185-5p	
hsa-miR-30b-3p	hsa-let-7a-5p	
hsa-miR-194-5p	hsa-miR-210-3p	
hsa-miR-3909	hsa-miR-340-5p	
hsa-miR-4446-3p	hsa-miR-708-5p	
hsa-miR-200a-5p	hsa-miR-1-3p	

(Continued)

TABLE 4 Continued

STC2	BIRC5	PDL1
hsa-miR-30d-5p	hsa-miR-1343-3p	
hsa-miR-324-5p	hsa-miR-218-1-3p	
hsa-miR-489-3p	hsa-miR-26a-5p	
hsa-miR-203a-3p	hsa-miR-147a	
hsa-miR-26b-5p	hsa-miR-345-5p	
hsa-miR-33a-5p	hsa-miR-1296-5p	
hsa-miR-33b-5p	hsa-miR-335-5p	
hsa-miR-1266-5p	hsa-miR-128-3p	
hsa-miR-181c-5p		
hsa-miR-23a-3p		
hsa-miR-25-3p		
hsa-miR-326		
hsa-miR-92b-3p		
hsa-miR-30d-3p		
hsa-miR-197-3p		
hsa-miR-3620-3p		
hsa-miR-340-3p		
hsa-miR-4728-3p		
hsa-miR-769-5p		
hsa-let-7f-2-3p		
hsa-miR-516b-5p		
hsa-miR-185-5p		
hsa-miR-182-3p		
hsa-miR-340-5p		
hsa-miR-23b-3p		
hsa-miR-4709-5p		
hsa-miR-148a-5p		
hsa-miR-548e-3p		
hsa-miR-454-5p		
hsa-miR-4429		
hsa-miR-143-3p		
hsa-miR-30c-1-3p		
hsa-miR-1225-5p		
hsa-miR-3652		
hsa-miR-1910-5p		
hsa-miR-26a-5p		
hsa-miR-3184-5p		
hsa-miR-197-5p		

(Continued)

TABLE 4 Continued

STC2	BIRC5	PDL1
hsa-miR-378i		
hsa-let-7d-3p		
hsa-miR-103b		
hsa-miR-320d		
hsa-miR-455-5p		
hsa-miR-30e-3p		
hsa-miR-423-3p		
hsa-miR-574-5p		
hsa-miR-1271-5p		
hsa-miR-21-3p		
hsa-miR-27a-5p		
hsa-miR-147a		
hsa-miR-494-3p		
hsa-miR-941		
hsa-miR-138-5p		

extent. The elevated expression of STC2 and BIRC5, can promote the binding with those competed miRNA of PDL1, which in turn upregulated the PDL1 expression. Simultaneously, the activation of the PI3K/AKT pathway is known to promote PD-L1 expression (71–73). Li and Zhu et al.'s research demonstrates that STC2 can facilitate the activation of the PI3K/AKT pathway (74, 75). Additionally, Shang et al.'s research have shown that BIRC5 expression is regulated by the PI3K/AKT pathway (76). Thus, elevated BIRC5 expression may serve as an indicator of PI3K/AKT pathway activation.

Further validation of the prognostic predictive ability of SIR-PS on HCC tissues of our own center were taken out, and the consistent results were collected. In addition, we creatively combined SIR-PS with IHC, which is more extensively utilized and offers greater convenience in clinical application in comparison to RNA-Seq technology. HCC tissue chips were performed for IHC staining and the results were scored. Utilizing these scores, we employed the SIR-PS to calculate individual patient risk scores, thereby evaluating the clinical utility of it. The SIR-PS exhibited a high degree of accuracy in prognostically assessing the 1 and 3 year survival for the HCC tissue chips' patients, with the low-risk group exhibiting a markedly more favorable prognosis than the high-risk group. In comparison to other immunohistochemical indicators, such as *GPC3* and *CK19*, the SIR-PS demonstrates superior predictive capabilities. This study has to some extent filled the gap in clinical pathological work that lacks specific IHC prognostic indicators for HCC. However, there are still limitation and deficiency in our study. Firstly, due to the lack of immune therapy results in tissue chips, we were unable to validate the predictive ability of this model for immune therapy efficacy in tissue chips through IHC. Secondly, although the datasets we included cover a wide range of ethnicities, they are still not comprehensive. Finally, as the datasets only include samples from patients who can undergo surgery, the applicability to samples from patients who cannot undergo surgery is uncertain, especially in clinical pathology work, where liver biopsy samples from non-resectable patients may not be applicable.

Taken together, the present investigation identified a novel prognostic model (SIR-PS) based on the KEGG pathway and focused on immune related genes. This model demonstrates potential as an effective tool for predicting prognosis of HCC and for assessing the efficacy of immunotherapeutic interventions. Utilizing the SIR-PS to calculate the risk score of each patient with HCC has showed a favorable efficacy in the 1 and 3 year survival rate prognostication. Given the absence of specific biomarkers for the prognostic evaluation of HCC in clinical, combination of SIR-PS with IHC promoted the clinical application of prognostic models and broadening the approach of prognostic models from databases to clinical practice.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: TCGA LIHC (https://portal.gdc.cancer.gov/); LIRI-JP of ICGC (https://dcc.icgc.org/);ImmPort (https://immport.niaid.nih.gov/);GSE54236 and GSE202069(https://www.ncbi.nlm.nih.gov/geo/).

Ethics statement

The studies involving humans were approved by the Ethics Committee of Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because This study was exempted from obtaining informed consent from research participants in accordance with the retrospective nature.

Author contributions

XS: Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. WJ: Data curation, Formal Analysis, Methodology, Visualization, Writing – original draft. HL: Data curation, Supervision, Writing – review & editing. HC: Data curation, Funding acquisition, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by National Natural Science Foundation of China (No. 81600485, 81270176, 81202063, 81302114 and 81300099).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2025.1481366/full#supplementary-material

References

1. 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 J Clinicians*. (2021) 71:209–49. doi: 10.3322/caac.21660

2. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet*. (2018) 391:1301–14. doi: 10.1016/S0140-6736(18)30010-2

3. Llovet JM, De Baere T, Kulik L, Haber PK, Greten TF, Meyer T, et al. Locoregional therapies in the era of molecular and immune treatments for hepatocellular carcinoma. *Nat Rev Gastroenterology Hepatology*. (2021) 18:293–313. doi: 10.1038/s41575-020-00395-0

4. Piñero F, Dirchwolf M, Pessôa MG. Biomarkers in hepatocellular carcinoma: diagnosis, prognosis and treatment response assessment. *Cells*. (2020) 9:1370–0. doi: 10.3390/cells9061370

5. Miura S, Mitsuhashi N, Shimizu H, Kimura F, Yoshidome H, Otsuka M, et al. Fibroblast growth factor 19 expression correlates with tumor progression and poorer prognosis of hepatocellular carcinoma. *BMC Cancer*. (2012) 12:0–0. doi: 10.1186/1471-2407-12-56

6. Casey SC, Amedei A, Aquilano K, Azmi AS, Benencia F, Bhakta D, et al. Cancer prevention and therapy through the modulation of the tumor microenvironment. *Semin Cancer Biol*. (2015) 35:S199–223. doi: 10.1016/j.semcancer.2015.02.007

7. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinf*. (2013) 14:0–0. doi: 10.1186/1471-2105-14-7

8. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. (2011) 27:1739–40. doi: 10.1093/bioinformatics/btr260

9. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci United States America*. (2005) 102:15545–50. doi: 10.1073/pnas.0506580102

10. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. (2010) 26:1572–3. doi: 10.1093/bioinformatics/btq170

11. Chen D, Huang H, Zang L, Gao W, Zhu H, Yu X. Development and verification of the hypoxia- and immune-associated prognostic signature for pancreatic ductal adenocarcinoma. *Front Immunol*. (2021) 12:728062. doi: 10.3389/fimmu.2021.728062

12. Wang Z, Wang Y, Yang T, Xing H, Wang Y, Gao L, et al. Machine learning revealed stemness features and a novel stemness-based classification with appealing implications in discriminating the prognosis, immunotherapy and temozolomide responses of 906 glioblastoma patients. *Briefings Bioinf*. (2021) 22:0–0. doi: 10.1093/bib/bbab032

13. Guo C, Tang Y, Zhang Y, Li G. Mining TCGA data for key biomarkers related to immune microenvironment in endometrial cancer by immune score and weighted correlation network analysis. *Front Mol Biosciences*. (2021) 8:645388. doi: 10.3389/fmolb.2021.645388

14. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. (2015) 43:e47–7. doi: 10.1093/nar/gkv007

15. Yu G, Wang LG, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics A J Integr Biol*. (2012) 16:284–7. doi: 10.1089/omi.2011.0118

16. OMIM® database. Available online at: https://www.omim.org (Accessed 1 May, 2024).

17. ClinVar database. Available online at: https://www.ncbi.nlm.nih.gov/clinvar (Accessed 1 May, 2024).

18. Gene Reference into Function. Available online at: https://www.ncbi.nlm.nih.gov/gene/about-generif (Accessed 1 May, 2024).

19. Gu Z. Complex heatmap visualization. *iMeta*. (2022) 1(3):e43. doi: 10.1002/imt2.43

20. Li B, Li Y, Zhou H, Xu Y, Cao Y, Cheng C, et al. Multiomics identifies metabolic subtypes based on fatty acid degradation allocating personalized treatment in hepatocellular carcinoma. *Hepatology*. (2024) 79(2):289–306. doi: 10.1097/hep.0000000000000553

21. Mayakonda A, Lin D, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. (2018) 28:1747–56. doi: 10.1101/gr.239244.118

22. Liu Y, Xun Z, Ma K-K, Liang S, Li X, Zhou S, et al. Identification of a tumor immune barrier in the HCC microenvironment that determines the efficacy of immunotherapy. *J Hepatology*. (2023) 78:770–82. doi: 10.1016/j.jhep.2023.01.011

23. Xu T, Dong M, Li H, Zhang R, Li X. Elevated mRNA expression levels of DLGAP5 are associated with poor prognosis in breast cancer. *Oncol Letters*. (2020) 19 (6):4053–65. doi: 10.3892/ol.2020.11533

24. Subirana I, Sanz H, Vila J. Building bivariate tables: the compare Groups package for R. *J Stat Software*. (2014) 57:1–16. doi: 10.18637/jss.v057.i12

25. Yang C, Lin Q-x, Xu Y, Qian F-j, Lin C-j, Zhao W-y, et al. An anoikis-related gene signature predicts prognosis and reveals immune infiltration in hepatocellular carcinoma. *Front Oncol*. (2023) 13:1158605. doi: 10.3389/fonc.2023.1158605

26. Chen Y, Xue W, Zhang Y, Gao Y, Wang Y. A novel disulfidptosis-related immune checkpoint genes signature: forecasting the prognosis of hepatocellular carcinoma. *J Cancer Res Clin Oncol*. (2023) 149:12843–54. doi: 10.1007/s00432-023-05076-4

27. Chen Y, Zhang S-m, Zhao H, Zhang J-y, Lian L-r, Liu D, et al. Identification and validation of immune and prognosis-related genes in hepatocellular carcinoma: A review. *Med (Baltimore)*. (2022) 101:e31814. doi: 10.1097/MD.0000000000031814

28. Ding X, Yao T, Liu X, Zheng F, Liu Y. A macropinocytosis-related gene signature predicts the prognosis and immune microenvironment in hepatocellular carcinoma. *Front Oncol*. (2023) 13:1143013. doi: 10.3389/fonc.2023.1143013

29. Ren M, Fan B, Cao G, Zong R, Feng L, Sun H. Exploration and validation of a combined Hypoxia and m6A/m5C/m1A regulated gene signature for prognosis prediction of liver cancer. *BMC Genomics*. (2023) 24:776. doi: 10.1186/s12864-023-09876-3

30. Su D, Zhang Z, Xia F, Liang Q, Liu Y, Liu W, et al. ICD-related risk model predicts the prognosis and immunotherapy response of patients with liver cancer. *Front Pharmacol*. (2023) 14:1202823. doi: 10.3389/fphar.2023.1202823

31. Yan Q, Zheng W, Wang B, Ye B, Luo H, Yang X, et al. A prognostic model based on seven immune-related genes predicts the overall survival of patients with hepatocellular carcinoma. *BioData Min*. (2021) 14:29. doi: 10.1186/s13040-021-00261-y

32. Yu X, Zhang H, Li J, Gu L, Cao L, Gong J, et al. Construction of a prognostic prediction model in liver cancer based on genes involved in integrin cell surface interactions pathway by multi-omics screening. *Front Cell Dev Biol*. (2024) 12:1237445. doi: 10.3389/fcell.2024.1237445

33. Zhang G, Xiao Y, Zhang X, Fan W, Zhao Y, Wu Y, et al. Dissecting a hypoxia-related angiogenic gene signature for predicting prognosis and immune status in hepatocellular carcinoma. *Front Oncol*. (2022) 12:978050. doi: 10.3389/fonc.2022.978050

34. Ding Z, Jin G-N, Wang W, Chen W, Wu Y, Ai X, et al. Reduced expression of transcriptional intermediary factor 1 gamma promotes metastasis and indicates poor prognosis of hepatocellular carcinoma. *Hepatology*. (2014) 60:1620–36. doi: 10.1002/hep.27273

35. Park SJ, Jeong SY, Kim HJ. Y chromosome loss and other genomic alterations in hepatocellular carcinoma cell lines analyzed by CGH and CGH array. *Cancer Genet cytogenetics*. (2006) 166:56–64. doi: 10.1016/j.cancergencyto.2005.08.022

36. Afghahi A, Mitani A, Desai M, Yu PP, de Bruin M, Seto T, et al. Use of the 21-gene recurrence score assay (RS) and chemotherapy (CT) across health care (HC) systems. *J Clin Oncol*. (2014) 32:6580–0. doi: 10.1200/jco.2014.32.15_suppl.6580

37. Kalinsky K, Barlow WE, Gralow JR, Meric-Bernstam F, Albain KS, Hayes DF, et al. 21-gene assay to inform chemotherapy benefit in node-positive breast cancer. *New Engl J Med*. (2021) 385:2336–47. doi: 10.1056/nejmoa2108873

38. Wang Y, Wu J, Xu J, Lin SS. Clinical significance of high expression of stanniocalcin-2 in hepatocellular carcinoma. *Bioscience Rep*. (2019) 39(4):BSR20182057. doi: 10.1042/bsr20182057

39. Kim P-H, Na S-S, Lee B, Kim J, Cho J-Y. Stanniocalcin 2 enhances mesenchymal stem cell survival by suppressing oxidative stress. *J Biochem Mol Biol*. (2015) 48:702–7. doi: 10.5483/bmbrep.2015.48.12.158

40. Zhu H, Chen XP, Zhang WG, Luo SF, Zhang BX. Expression and significance of new inhibitor of apoptosis protein surviving in hepatocellular carcinoma. *World J Gastroenterol*. (2005) 11:3855–9. doi: 10.3748/wjg.v11.i25.3855

41. Wang L, Yan K, Zhou J, Zhang N, Wang M, Song J, et al. Relationship of liver cancer with LRP1B or TP53 mutation and tumor mutation burden and survival. *J Clin Oncol*. (2019) 37:1573–3. doi: 10.1200/jco.2019.37.15_suppl.1573

42. Thiem A, Hesbacher S, Kneitz H, di Primio T, Heppt MV, Hermanns H, et al. IFN-gamma-induced PD-L1 expression in melanoma depends on p53 expression. *J Exp Clin Cancer Res*. (2019) 38:397. doi: 10.1186/s13046-019-1403-9

43. Yoon KW, Byun S-S, Kwon E, Hwang SY, Chu KK, Hiraki M, et al. Control of signaling-mediated clearance of apoptotic cells by the tumor suppressor p53. *Science*. (2015) 349:1261669. doi: 10.1126/science.1261669

44. Lu G, Lin J, Song G, Chen M. Prognostic significance of CTNNB1 mutation in hepatocellular carcinoma: a systematic review and meta-analysis. *Aging*. (2023) 15:9759–78. doi: 10.18632/aging.205047

45. Wang Z, Sheng Y, Gao X, Wang C-Q, Wang X, Xu L, et al. β-catenin mutation is correlated with a favorable prognosis in patients with hepatocellular carcinoma. *Mol Clin Oncol*. (2015) 3:936–40. doi: 10.3892/mco.2015.569

46. Mou H, Yang Q, Yu L, Wang T, Liu K, Shen R, et al. Programmed cell death-ligand 1 expression in hepatocellular carcinoma and its correlation with clinicopathological characteristics. *J Gastroenterology Hepatology*. (2021) 36:2601–9. doi: 10.1111/jgh.15475

47. Lamberti G, Spurr LF, Li Y, Ricciuti B, Recondo G, Umeton R, et al. Clinicopathological and genomic correlates of programmed cell death ligand 1 (PD-L1) expression in nonsquamous non-small-cell lung cancer. *Ann Oncol*. (2020) 31:807–14. doi: 10.1016/j.annonc.2020.02.017

48. Raskov H, Orhan A, Christensen JP, Gögenür I. Cytotoxic CD8+ T cells in cancer and cancer immunotherapy. *Br J Cancer*. (2020) 124:359–67. doi: 10.1038/s41416-020-01048-4

49. Juneja VR, McGuire KA, Manguso RT, LaFleur MW, Collins NB, Haining WN, et al. PD-L1 on tumor cells is sufficient for immune evasion in immunogenic tumors and inhibits CD8 T cell cytotoxicity. *J Exp Med*. (2017) 214:895–904. doi: 10.1084/jem.20160801

50. Wang X, Lu L, Hong X, Wu L, Yang C, Wang Y, et al. Cell-intrinsic PD-L1 ablation sustains effector CD8+ T cell responses and promotes antitumor T cell therapy. *Cell Rep*. (2024) 43:113712–2. doi: 10.1016/j.celrep.2024.113712

51. Im SJ, Hashimoto M, Gerner MY, Lee J, Kissick H, Bürger MC, et al. Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. *Nature*. (2016) 537:417–21. doi: 10.1038/nature19330

52. Barber DL, Wherry EJ, Masopust D, Zhu B, Allison J, Sharpe A, et al. Restoring function in exhausted CD8 T cells during chronic viral infection. *Nature*. (2006) 439:682–7. doi: 10.1038/nature04444

53. Audrito V, Serra S, Stingi A, Orso F, Gaudino F, Bologna C, et al. PD-L1 up-regulation in melanoma increases disease aggressiveness and is mediated through miR-17-5p. *Oncotarget*. (2017) 8:15894–911. doi: 10.18632/oncotarget.15213

54. Wang X, Li J, Dong K, Fang L, Long M, Ouyang Y, et al. Tumor suppressor miR-34a targets PD-L1 and functions as a potential immunotherapeutic target in acute myeloid leukemia. *Cell Signal*. (2015) 27:443–52. doi: 10.1016/j.cellsig.2014.12.003

55. Boldrini L, Giordano M, Niccoli C, Melfi F, Lucchi M, Mussi A, et al. Role of microRNA-33a in regulating the expression of PD-1 in lung adenocarcinoma. *Cancer Cell Int*. (2017) 17:105. doi: 10.1186/s12935-017-0474-y

56. Cortez MA, Ivan C, Valdecanas DR, Wang X, Peltier HJ, Ye Y, et al. PDL1 regulation by p53 via miR-34. *J Natl Cancer Inst*. (2015) 108:djv303. doi: 10.1093/jnci/djv303

57. Zhao L, Yu H, Yi S, Peng X, Su P, Xiao Z, et al. The tumor suppressor miR-138-5p targets PD-L1 in colorectal cancer. *Oncotarget*. (2016) 7:45370–45384. doi: 10.18632/oncotarget.v7i29

58. Xie WB, Liang L-H, Wu K-G, Wang L-X, He X, Song C, et al. MiR-140 expression regulates cell proliferation and targets PD-L1 in NSCLC. *Cell Physiol Biochem*. (2018) 46:654–63. doi: 10.1159/000488634

59. Wang Y, Wang D, Xie G, Yin Y, Zhao E, Tao K, et al. MicroRNA-152 regulates immune response via targeting B7-H1 in gastric carcinoma. *Oncotarget*. (2017) 8:28125–34. doi: 10.18632/oncotarget.15924

60. Yee D, Shah KM, Coles MC, Sharp TV, Lagos D. MicroRNA-155 induction via TNF-alpha and IFN-gamma suppresses expression of programmed death ligand-1 (PD-L1) in human primary cells. *J Biol Chem*. (2017) 292:20683–20693. doi: 10.1074/jbc.M117.809053

61. Fujita Y, Yagishita S, Hagiwara K, Yoshioka Y, Kosaka N, Takeshita F, et al. The clinical relevance of the miR-197/CKS1B/STAT3-mediated PD-L1 network in chemoresistant non-small-cell lung cancer. *Mol Ther*. (2015) 23:717–27. doi: 10.1038/mt.2015.10

62. Ahn H, Yang JM, Kim H, Chung JH, Ahn S-H, Jeong WJ, et al. Clinicopathologic implications of the miR-197/PD-L1 axis in oral squamous cell carcinoma. *Oncotarget*. (2017) 8:66178–94. doi: 10.18632/oncotarget.19842

63. Chen L, Gibbons D, Goswami S, Cortez MA, Ahn YH, Byers LA, et al. Metastasis is regulated via microRNA-200/ZEB1 axis control of tumor cell PD-L1 expression and intratumoral immunosuppression. *Nat Commun*. (2014) 5:5241. doi: 10.1038/ncomms6241

64. Xu S, Tao Z, Hai B, Huagen L, Shi Y, Wang T, et al. miR-424(322) reverses chemoresistance via T-cell immune response activation by blocking the PD-L1 immune checkpoint. *Nat Commun*. (2016) 7:11406. doi: 10.1038/ncomms11406

65. Hamilton MP, Rajapakshe KI, Bader DA, Černe J-Ž, Smith E-A, Coarfa C, et al. The landscape of microRNA targeting in prostate cancer defined by AGO-PAR-CLIP. *Neoplasia*. (2016) 18:356–70. doi: 10.1016/j.neo.2016.04.008

66. Gillen AE, Yamamoto TM, Kline E, Hesselberth JR, Kabos P. Improvements to the HITS-CLIP protocol eliminate widespread mispriming artifacts. *BMC Genomics*. (2016) 17:338. doi: 10.1186/s12864-016-2675-5

67. Jiang L, Dai Y, Liu X, Wang C, Wang A, Chen Z, et al. Identification and experimental validation of G protein alpha inhibiting activity polypeptide 2 (GNAI2) as a microRNA-138 target in tongue squamous cell carcinoma. *Hum Genet*. (2011) 129:189–97. doi: 10.1007/s00439-010-0915-3

68. Whisnant AW, Bogerd HP, Flores OGM-a, Ho P, Powers J, et al. In-depth analysis of the interaction of HIV-1 with cellular microRNA biogenesis and effector mechanisms. *mBio*. (2013) 4:e000193. doi: 10.1128/mBio.00193-13

69. Erhard F, Haas J, Lieber D, Malterer G, Jaśkiewicz Ł, Mihaela Zavolanp53 represses the oncogenic sno-miR-28, et al. Widespread context dependency of microRNA-mediated regulation. *Genome Res*. (2014) 24:906–19. doi: 10.1101/gr.166702.113

70. Yu F, Bracken CP, Pillman KA, Lawrence D, Goodall GJ, Callen DF, et al. p53 represses the oncogenic sno-miR-28 derived from a snoRNA. *PLoS One*. (2015) 10:e0129190. doi: 10.1371/journal.pone.0129190

71. Muthumani K, Shedlock DJ, Choo DK, Fagone P, Kawalekar OU, Goodman JS, et al. HIV-mediated phosphatidylinositol 3-kinase/serine–threonine kinase activation in APCs leads to programmed death-1 ligand upregulation and suppression of HIV-specific CD8 T cells. *J Immunol*. (2011) 187:2932–43. doi: 10.4049/jimmunol.1100594

72. Lastwika KJ, Wilson W 3rd, Li QK, Norris JW, Xu H, Ghazarian SR, et al. Control of PD-L1 expression by oncogenic activation of the AKT–mTOR pathway in non–small cell lung cancer. *Cancer Res*. (2016) 76:227–38. doi: 10.1158/0008-5472.can-14-3362

73. Quan Z, Yang Y, Zheng H, Zhan Y, Luo J, Ning Y, et al. Clinical implications of the interaction between PD-1/PD-L1 and PI3K/AKT/mTOR pathway in progression and treatment of non-small cell lung cancer. *J Cancer*. (2022) 13:3434–43. doi: 10.7150/jca.77619

74. Zhu C, Jiang Y, Zhu J, He Y, Yin H, Duan Q, et al. CircRNA8220 sponges miR-8516 to regulate cell viability and milk synthesis via ras/MEK/ERK and PI3K/AKT/mTOR pathways in goat mammary epithelial cells. *Anim (Basel)*. (2020) 10(8):1347. doi: 10.3390/ani10081347

75. Li D, Xiong Y, Li M, Long L, Zhang Y, Yan H, et al. STC2 knockdown inhibits cell proliferation and glycolysis in hepatocellular carcinoma through promoting autophagy by PI3K/Akt/mTOR pathway. *Arch Biochem Biophys*. (2024) 761:110149. doi: 10.1016/j.abb.2024.110149

76. Shang X, Liu G, Zhang Y-F, Tang P, Zhang H, Jiang H, et al. Downregulation of BIRC5 inhibits the migration and invasion of esophageal cancer cells by interacting with the PI3K/Akt signaling pathway. *Oncol Lett*. (2018) 16(3):3373–9. doi: 10.3892/ol.2018.8986

Frontiers in Aging

Explores the biological processes of human aging

Advances our understanding of human aging and the fundamental link with age-related diseases, ultimately leading to improved healthspans.

Discover the latest Research Topics

See more →

frontiers

Frontiers in
Aging

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

frontiers | Research Topics