frontiers | Research Topics

Multi-omics approaches for decoding heterogeneity in cancer immunotherapy

Edited by
Lin Qi, Zhigang Liu, Linhui Wang, Hongzhou Cai and Ouyang Chen

Published in
Frontiers in Pharmacology
Frontiers in Genetics

ISSN 1664-8714
ISBN 978-2-8325-3965-1
DOI 10.3389/978-2-8325-3965-1

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

Multi-omics approaches for decoding heterogeneity in cancer immunotherapy

Topic editors

Lin Qi — Central South University, China
Zhigang Liu — The Fifth Affiliated Hospital of Sun Yat-sen University, China
Linhui Wang — Second Military Medical University, China
Hongzhou Cai — Nanjing Medical University, China
Ouyang Chen — Duke University, United States

Citation

Qi, L., Liu, Z., Wang, L., Cai, H., Chen, O., eds. (2023). *Multi-omics approaches for decoding heterogeneity in cancer immunotherapy*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3965-1

Table of contents

frontiers | Frontiers in Pharmacology

TYPE Editorial
PUBLISHED 03 November 2023
DOI 10.3389/fphar.2023.1324212

OPEN ACCESS

EDITED AND REVIEWED BY
José A. G. Agúndez,
University of Extremadura, Spain

*CORRESPONDENCE
Linhui Wang,
wanglinhui@smmu.edu.cn
Lin Qi,
qi.lin@csu.edu.cn

†These authors have contributed equally to this work

RECEIVED 19 October 2023
ACCEPTED 24 October 2023
PUBLISHED 03 November 2023

CITATION
Jiang A, Liu Y, Chen O, Liu Z, Cai H, Wang L and Qi L (2023), Editorial: Multi-omics approaches for decoding heterogeneity in cancer immunotherapy. *Front. Pharmacol.* 14:1324212. doi: 10.3389/fphar.2023.1324212

Editorial: Multi-omics approaches for decoding heterogeneity in cancer immunotherapy

Aimin Jiang[1†], Ying Liu[1†], Ouyang Chen[2], Zhigang Liu[3], Hongzhou Cai[4], Linhui Wang[1*] and Lin Qi[5,6*]

[1]Department of Urology, Changhai Hospital, Naval Medical University, Shanghai, China, [2]Department of Cell Biology, Duke University Medical Center, Durham, NC, United States, [3]Dongguan Key Laboratory of Precision Diagnosis and Treatment for Tumors, Dongguan, Guangdong, China, [4]Department of Urology, Jiangsu Cancer Hospital, Jiangsu Institute of Cancer Research, The Affiliated Cancer Hospital of Nanjing Medical University, Nanjing, Jiangsu, China, [5]Department of Orthopedics, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China, [6]Hunan Key Laboratory of Tumor Models and Individualized Medicine, The Second Xiangya Hospital, Changsha, Hunan, China

KEYWORDS

multi-omics, tumor microenvironment, tumor heterogeneity, cancer immunotherapy, personalized medicine

Editorial on the Research Topic
Multi-omics approaches for decoding heterogeneity in cancer immunotherapy

The emergence of cancer immunotherapy has brought about a significant revolution in the field of oncology (Hegde and Chen, 2020). This innovative approach offers new possibilities for treating a wide range of malignancies by leveraging the immune system's power. Nevertheless, the effectiveness of such therapeutic strategies can be restricted by the inherent heterogeneity present within and between tumors. Consequently, deciphering this heterogeneity represents a critical challenge in optimizing the efficiency of immunotherapies. Tumor heterogeneity refers to the observed biological diversity among cancer cells residing in a single tumor or across different tumors. It arises from a variety of factors, including genetic variations, epigenetic modifications, transcriptional changes, and alterations in protein expression or metabolic profiles (Jia et al., 2022). This intricate interplay complicates efforts to elicit consistent and sustained responses to immunotherapy, as various cell populations may display different susceptibility to immune attack or exert diverse influences on the tumor microenvironment.

This is where the importance of multi-omics approaches comes into play. By incorporating genomics, transcriptomics, proteomics, metabolomics, and other "-omics" technologies, multi-omics provides a comprehensive view of the various biological layers present within tumors (Lee et al., 2022). It enables the characterization of genomic alterations, mRNA expression levels, protein abundance, and metabolic profiles across diverse cellular populations, allowing for a comprehensive understanding of the molecular landscape of carcinomas. Such high-resolution mapping facilitates the identification of distinctive molecular signatures linked to either response or resistance to immunotherapy. It also helps in elucidating the underlying mechanisms governing tumor-immune interactions and the discovery of novel targets for immunotherapy. Furthermore, temporal multi-omics analyses can track the dynamic changes occurring within tumors over time or in response to treatment, providing valuable insights into the evolution of tumor heterogeneity and its

impact on treatment outcomes. In summary, multi-omics represents a powerful tool in unravelling the complexity of tumor heterogeneity, a critical factor in fully unlocking the potential of cancer immunotherapy. The knowledge gained from these investigations is expected to significantly contribute to personalized medicine and ultimately improve patient outcomes in the era of immunotherapy.

Given the vast amount of multi-omics data available and the advanced approaches for high pathway data analysis, this Research Topic has been thoroughly explored through 16 original research articles authored by a total of 137 individuals. These articles provide a comprehensive overview of the role played by multi-omics approaches in decoding tumor heterogeneity and enhancing the efficacy of immunotherapy.

Uveal melanoma (UVM) represents a primary intraocular malignancy that not only significantly impairs patients' visual function and overall health but also poses a significant therapeutic challenge. The basement membrane (BM), critical in instigating and preserving diverse biological processes such as cell polarity, organ morphogenesis, and adult function, houses certain genes responsible for synthesizing basement membrane proteins which serve as valuable prognostic biomarkers across various cancer types. In their pioneering study, Li et al. adeptly employed multi-omics UVM datasets to develop an innovative risk assessment framework predicated on BM-associated genes. This system lays a robust theoretical groundwork for crafting precise, personalized treatment approaches. Immunotherapy appears particularly beneficial for high-risk group UVM patients, while those within the low-risk group enjoy enhanced survival benefits. Interestingly, *in vitro* assays established that inhibiting ITGA5 expression, a BM-related gene, effectively stymies the proliferation, migration, and invasive capabilities of UVM cells. In essence, the BM-related model proposed by Li et al. study demonstrates exceptional predictive prowess, markedly influencing patient prognosis and guiding individualized treatment strategies. Furthermore, this novel model paves the way for evaluating the effectiveness of pre-immune interventions.

Abundant empirical research has underscored the pivotal roles that cancer stem cells (CSCs) assume in driving and disseminating cancer. A study spearheaded by Chen et al. centered specifically on CSCs, identifying a selection of CSC marker genes predicated on scRNA-seq data procured from colorectal cancer samples. Through analyzing these markers' expression profiles, the team singled out 29 CSC marker genes. Two distinctive phenotypes were identified and termed as CSC1 and CSC2. CSC2 owned a shorter disease specific survival or DSS and heightened oxidative stress response. Notably, drug sensitivity analyses suggested that CSC2 was more receptive to 44 chemotherapy drugs relative to CSC1. To bolster prognostication, the researchers devised a seven-gene prognostic model, adept at distinguishing high-risk from low-risk patients. Specifically, 14 chemotherapy drugs were found to be more beneficial for the high-risk group, while 13 were more effective for the low-risk group. Collectively, these insights enhance our comprehension of CSCs' role in shaping the evolution and progression of colorectal cancer. Importantly, the seven-gene prognostic model shows promise as an indicator for predicting responsiveness to immunotherapy and chemotherapy, concurrently supplying valuable prognostic information for colorectal cancer patients. Given the global prevalence of colorectal cancer and the inherent heterogeneity-induced survival paradox in stage II/III CRC tumor biology, tumor progression is intimately linked with ferroptosis. As such, ferroptosis-related genes (FRG) may constitute novel biomarkers for predicting cancer prognosis. In their study, Wang et al. crafted a machine learning framework, which was consisted of 83 combinations of 10 algorithms to pinpoint the most robust and stable colorectal cancer model derived from FRG signatures. Strikingly, the FRG signature outperformed the clinicopathological features, and significantly correlated with BRAF mutation and microsatellite instability. Furthermore, the FRG signature was segregated into a high-risk group and a low-risk subgroup. Notably, high FRG signature corresponded to poorer prognosis across all datasets. Crucially, sensitivity analysis affirmed the FRG signature as a significant prognostic indicator, underscoring its potential in guiding clinical decision-making and facilitating personalized therapy for stage II/III colorectal cancer patients.

Recent studies have unveiled the pivotal role that microbes assume in the onset, evolution, metastasis, and treatment response of diverse tumor types. This influence is especially pronounced in the context of the tumor microenvironment and immune response. A multitude of research has demonstrated that the intratumor microbiome can shape the progression, metastases, prognosis, and immunotherapy outcomes in cancer patients. This modulation occurs via control of oxidative stress, Toll-like receptor-mediated immune response, and tumor cell metabolism, which engages several signaling pathways including mTOR, STAT3, Wnt, and MAPK. However, current understanding falls short of comprehensively elucidating the relationship between the intratumor microbiome and clinicopathological features, prognosis, heterogeneity of the tumor microenvironment, and therapeutic response in pancreatic cancer (PC). In a study spearheaded by Zhang et al., 26 prognostic genera associated with PC were identified. The PC samples were categorized into two microbiome-related subtypes: Mcluster A and B. Patients in Mcluster B exhibited poorer prognosis, higher TNM stage, and pathological grade compared to those in Mcluster A. Immune analysis exposed significantly elevated levels of infiltrated $CD8^+$ T cells, M1 and M2 macrophages, cancer-associated fibroblasts, myeloid dendritic cells, neutrophils, regulatory T cells, and activated mast cells in Mcluster B. Conversely, patients in Mcluster A were more likely to benefit from CTLA-4 blockers and exhibited heightened sensitivity to several agents including oxaliplatin, and epirubicin. Moreover, a microbe-derived model was devised to predict the outcome in PC, with ROC curves illustrating this model's high predictive performance. Single-cell analysis identified the presence of monocytes/macrophages, endothelial cells, and fibroblasts along with cancer cells within the PC tumor microenvironment. Notably, LIPH and LAMA3 displayed relatively higher expression in cancer cells and neutrophils. Collectively, these insights offer new perspectives on outcome assessment and therapeutic strategies for PC.

The 5-year survival rate for patients with advanced-stage gastric cancer (GC) remains disappointingly low. Studies have proposed a role for tryptophan metabolism in cancer progression, implicating it in eliciting immunosuppressive responses and bolstering the malignancy of cancer cells. Thus, unraveling the roles of

tryptophan and its metabolism is key to comprehending the molecular mechanisms implicated in GC evolution. In a recent study spearheaded by Luo et al., public datasets were employed to identify genes associated with tryptophan metabolism. Single sample gene set enrichment analysis, and correlation analysis were utilized to this end. Consequently, two molecular subtypes linked with tryptophan metabolism were pinpointed. The first subtype, coined C1, exhibited a more favorable prognosis than the second subtype, C2. This observation was accompanied by an uptick in CD4 positive memory T cells and activated dendritic cells (DCs) within the tumor microenvironment, along with a suppression of M2-phenotype macrophages. Furthermore, immune checkpoint activity was detected to be downregulated in the C1 subgroup. To construct a prognostic risk model, eight pivotal genes were identified as critical to predict the prognosis of GC patients. In sum, this study underscores the potential utility of tryptophan metabolism-associated genes in forecasting the prognosis of GC patients. The risk model devised in this study demonstrated remarkable precision in predicting survival outcomes in GC patients.

Hepatocellular carcinoma (HCC) represents an assertive form of liver cancer typically diagnosed at advanced stages, thereby contributing to high mortality rates. Hence, the discovery of new biomarkers for early detection and patient outcome improvement is imperative. Efferocytosis, a sophisticated process wherein one cell engulfs another, involves diverse immune cells like macrophages, dendritic cells, and natural killer (NK) cells. This process, intriguingly, plays a dual role in tumor development—sometime fostering, other times inhibiting tumorigenesis. Nevertheless, our understanding of efferocytosis-related genes' (ERGs) role in HCC progression remains incomplete, particularly their influence on immunotherapy and targeted drug treatments. In a recent study helmed by Xu et al., a risk model predicated on six ERGs was designed, leading to the identification of two distinct HCC subtypes corresponding to these genes. The identified subtypes exhibited significant disparities in the tumor immune landscape and prognosis, accentuating efferocytosis status's importance in HCC. Collectively, this investigation illuminates the critical role of efferocytosis in HCC and lays a solid groundwork for further exploration into this disease's progression and treatment response. It holds potential in shaping clinical decision-making in HCC management. Additionally, 5-FU, a traditional chemotherapeutic drug utilized in various cancers, has been demonstrated to elicit an effective response not only by targeting cancer cells but also by stimulating an anti-tumor immune response mediated by the STING pathway within the cancer cells themselves. The cGAS-STING pathway activation by 5-FU treatment instigates local production of type I interferons (IFNs), implying that 5-FU could trigger IFN production specifically in the tumor microenvironment, even in cancers traditionally unaddressed with 5-FU. In a study conducted by Gu et al., HCC subtypes were classified based on their 5-FU sensitivity. The findings corroborated the observed prognostic disparities in HCC and underscored the heterogeneity of genomic variations, the tumor immune microenvironment, and pathological pathways. Additionally, the study devised an independent prognostic risk regression model incorporating five 5-FU-related genes, promoting individualized HCC monitoring advancement. In conclusion, these studies affirm 5-FU's crucial role in bolstering anti-tumor immunity and offer valuable insights into this drug's potential application in HCC treatment.

We have gathered three research papers on breast cancer for our Research Topic analysis. One crucial mechanism that has been widely recognized as a natural defense against tumors is cellular senescence. Oncogene-induced senescence (OIS), which involves the activation of proto-oncogenes or the deactivation of tumor suppressor genes, leads to cell growth arrest. Notably, Huang et al. conducted a comprehensive study using both in-silicon and experimental approaches and categorized triple-negative breast cancer (TNBC) into two subtypes, namely, TNBCSASP1 and TNBCSASP2. They based this classification on the set of genes associated with the senescence-associated secretory phenotype. Unfortunately, the TNBCSASP1 subtype displayed a poor prognosis. It was found to exhibit immunosuppression, with reduced activity in immune-related signaling pathways and limited infiltration of immune cells. The likelihood of TP53 and TGF-β pathways mutations contributing to the unfavourable prognosis of the TNBCSASP1 subtype was suggested. Additionally, the authors uncovered FAM3B as a key biomarker influencing the prognosis of patients with TNBC. In TNBC, FAM3B expression was lower than in normal breast tissue. Survival analysis demonstrated that patients with high FAM3B expression had significantly shorter overall survival rates. Unfortunately, due to the lack of effective treatments, triple-negative breast cancer has an extremely poor prognosis. Metabolic reprogramming is a fundamental aspect of tumorigenesis, cancer diagnosis, prognosis, and treatment. Li et al. conducted a fascinating study in which they identified two metabolically distinct subtypes of TNBC. The C1 subtype exhibited high expression of immune checkpoint genes and immune and stromal scores, suggesting sensitivity to PD-1 inhibitors. Conversely, the C2 subtype exhibited significant variation in carbohydrate, lipid, and amino acid metabolism pathways. Importantly, C2 lacked immune signatures, displayed late pathological stage, low immune infiltration, and correspondingly had a poor prognosis. C2 also had a high mutation frequency in PIK3CA, KMT2D, and KMT2C, and exhibited significant activation of the PI3K and angiogenesis pathways. In conclusion, the authors successfully identified two TNBC subtypes with distinct metabolic characteristics, offering valuable insights into TNBC heterogeneity and providing a theoretical foundation for potential therapeutic strategies. Among the various types of breast cancer, estrogen receptor (ER) breast cancer is the most prevalent, characterized by the expression of estrogen receptors. Globally, it affects approximately 2.26 million women. Khan et al. conducted a study that identified differentially expressed genes and isoform switching between ER-positive and triple-negative breast cancer samples. Specifically, they pinpointed six genes predominantly associated with ER-positive breast cancer, as well as a novel set of ten genes not previously reported in ER-positive breast cancer. Additionally, alternative splicing and subsequent isoform usage in genes related to the immune system were identified. This study sheds light on the differential isoform usage occurring in cancer cells and its potential implication in immunosuppression through dysregulation of CXCR chemokine receptor binding, iron ion binding, and cytokine activity.

Copper, an indispensable mineral vital for enzyme activity and transcription factor function, can trigger proteotoxic stress and a unique form of cell death known as cuproptosis when in excess, due to the accumulation of lipoylated dihydrolipoamide S-acetyltransferase (DLAT), linked with the TCA cycle. Cuproptosis, implicated in cancer progression, presents a promising therapeutic target. In their study, Kuang et al. utilized multi-omics datasets and extensive *in vitro* experiments to probe the biological and clinical significance of cuproptosis in lung adenocarcinoma (LUAD). Employing lasso regression analysis, they developed a cuproptosis-related signature (CRS) based on 24 specific genes. According to findings, high-risk CRS patients portended poorer prognosis in both TCGA-LUAD and GSE31210 datasets. Further enrichment analysis disclosed that copper proliferation primarily transpired via chromosome-related pathways, cell cycle regulation, DNA replication, and G2M checkpoint activation. Additionally, differences in macrophage levels were observed between low and high CRS groups through immuno-infiltration analysis. Crucially, these cuproptosis-related genes might serve as potential prognostic predictors and immunotherapy effectiveness indicators for LUAD patients by influencing chromosome-related pathways and macrophage activity. Concurrently, Cong et al. study employed machine learning algorithms such as GBM, lasso, xgboost, SVM, random Forest, and Decision Trees to construct a novel risk stratification system based on fatty acid metabolism (FAM)-related signatures. Lower FAM-related pathway scores were noted in LUAD samples. Three molecular subtypes C1, C2, and C3 were defined, with differential prognostic analysis revealing the most favorable prognosis in C1 subtype, followed by C2, and the worst prognosis in C3 subtype. The C3 subtype also displayed lower levels of immune infiltration. A risk score model was developed using 12 key identified genes where high-risk score patients demonstrated significantly lower survival rates. Conversely, the low-risk score group showed higher immune scores and increased expression of immune checkpoint genes. Moreover, high-risk score patients were more likely to benefit from six anti-cancer drugs screened in this study. In essence, these studies offer valuable insights into the role of cuproptosis and fatty acid metabolism in LUAD's prognosis and treatment. They underscore the potential utility of cuproptosis-related genes and fatty acid metabolism-related signatures in predicting patient outcomes and informing therapeutic strategies, including immunotherapy and targeted drug interventions.

The amplified cardiovascular event risk experienced by cancer patients presents a substantial concern. Extensive research suggests that cancer survivors are more prone to cardiac complications, including myocardial infarction, heart failure, and arrhythmias, relative to the general populace. Yuan et al. recent study illuminates that distinct monocyte-derived biomarkers harbor significant promise in prognosticating cancer outcomes and acute myocardial infarction (AMI). The researchers applied an innovative formula to examine mRNA levels in clinical samples from AMI and cancer-diagnosed patients, leading to the construction of a novel risk score based on expression profiles. By classifying patients into high-risk and low-risk groups according to the median risk score, noticeably poorer overall survival rates were observed among high-risk patients within the cancer cohorts, as corroborated by Kaplan-Meier analysis. Key to note is the concurrent activation of the Notch signaling pathway, potentially shedding light on shared high-risk factors affiliated with both AMI and cancer. Moreover, the researchers confirmed the differential expression of these genes in cell lines and clinical samples, further reinforcing their relevance as potent biomarkers. Crucially, these findings highlight the potential utility of shared biomarkers in accurately predicting patient outcomes for both cancer and AMI.

Serine protease inhibitor clade E member 1 (SERPINE1), or PAI-1, is a pivotal modulator of the plasminogen activation system, with roles extending beyond mere plasminogen regulation. Implicated in numerous physiological processes, SERPINE1's multifaceted functions hint at its potential contribution to diverse disease processes. In an effort to demystify the role of SERPINE1, Li et al. conducted an exhaustive analysis across multiple cancer types. The study unveiled dysregulated SERPINE1 expression in cancer cells, noting enrichment in endothelial cells and fibroblasts. This irregularity can be partially attributed to copy number amplification and reduced DNA promoter methylation. Crucially, heightened SERPINE1 expression was linked with adverse prognosis in 21 disparate cancer types. Further scrutiny using gene set enrichment analysis (GSEA) presented SERPINE1 as a player in immune response regulation and tumor malignancy. Correlations were drawn between SERPINE1 expression, immunoregulator expression, and immune cell infiltration, suggesting a potential role for SERPINE1 in immunosuppression. Intriguingly, associations were also found linking SERPINE1 expression with tumor mutation burden (TMB), microsatellite instability (MSI), responses to immunotherapy, and drug sensitivity across various cancers. Collectively, this study emphasizes the anomalous expression of SERPINE1 in numerous cancer types and its consequent implications for cancer immunity and tumor behavior. Such findings furnish invaluable insights that could steer the development of personalized cancer treatments meticulously tailored to individual patients.

Metastatic castration-resistant prostate cancer (mCRPC), an intensely aggressive prostate cancer stage, exhibits progression largely driven by non-mutational epigenetic reprogramming. Super enhancers (SE), a category of epigenetic elements, are implicated in myriad tumor-promoting signaling pathways, though the precise modus operandi of SE mediation in mCRPC remains enigmatic. Zeng et al. employed a CUT&Tag assay on mCRPC C4-2B cell lines to identify SE-associated genes and transcription factors. Merging these overlapping genes, or SE-related DEGs, they developed a recurrence risk prediction model. A time-dependent receiver operating characteristic (ROC) curve analysis affirmed the robust predictive capacity of their risk score system at the 1-year (0.80), 3-year (0.85), and 5-year (0.88) stages. In sum, their study offers a comprehensive comprehension of the SE element landscape and related genes in mCRPC and furthers discourse on potential clinical implications and transitions to clinical practice.

Bladder cancer, a prevalent urologic malignancy associated with significant morbidity and mortality, presents varying patient

response rates to the promising potential of immunotherapy. The process of glycosylation, characterized by sugar molecule attachment to proteins or lipids, has been implicated in tumor development and immune regulation. Yet, the clinical implications and understanding of its role in bladder cancer remain underdeveloped. Liu et al. sought to bridge this knowledge gap by conducting a multi-cohort study of bladder cancer patients, developing a unique risk scoring system cantered on glycosylation-related genes. In particular, the authors segmented the training cohort into two clusters based on distinct patterns of glycosylation-related gene expression. Prognostic analysis unveiled worse survival outcomes for Cluster 2 than Cluster 1. Cluster 2 also showcased elevated levels of tumor microenvironment-immune cells and amplified activity in crucial phases of the cancer immune response cycle. To corroborate their findings, a standalone prognostic risk score was developed, aiding in the construction of a reliable prognostic prediction nomogram. Patients with high glycosylation risk scores demonstrated an increase in tumor immune cell infiltration, higher enrichment scores within immune related pathways. Conversely, patients with lower risk scores exhibited minimal immune cell infiltration and tended towards a luminal subtype. These findings held true even in real-world examination, as evidenced by the Xiangya cohort. In essence, the study underscores that a multi-omics glycosylation score derived from these identified genes can reliably elucidate bladder cancer heterogeneity and predict immunotherapy effectiveness and molecular subtypes. Such pivotal insight empowers refined individual treatment decisions for bladder cancer patients.

Tumorigenesis, a multifaceted process marked by the perturbation of assorted genetic and non-genetic mechanisms that amass over a duration, is characterized by the inherent genomic instability of neoplastic cells. Such unstable features often yield serendipitous tumorigenic events throughout disease progression. These stochastic occurrences play a crucial role in molding an eclectic immune microenvironment—spatially or temporally—which introduces a degree of heterogeneity. The introduction of multi-omics profiling technologies has heralded a transformation in our comprehension of the labyrinthine and heterogeneous character of the Tumor Microenvironment (TME) at unparalleled resolution. These advanced multi-omics modalities offer unmatched granularity in deciphering the orchestrated modifications in composition and status of immune and stromal components within the TME. This results from cancer treatment responses, thus enabling a comprehensive and nuanced insight into cancer progression and immunotherapeutic outcomes.

Author contributions

AJ: Writing–original draft, Writing–review and editing, Data curation, Methodology, Resources. YL: Writing–original draft, Writing–review and editing, Data curation, Methodology, Software. OC: Data curation, Formal Analysis, Writing–review and editing. ZL: Investigation, Methodology, Writing–review and editing. HC: Methodology, Software, Writing–review and editing. LW: Conceptualization, Supervision, Writing–review and editing. LQ: Conceptualization, Supervision, Writing–review and editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the National Natural Science Foundation of China (81902560, 81730073, and 81872074), the Fundamental Research Funds for the Central Universities of Central South University (2023ZZTS0024) and the Postgraduate Scientific Research Innovation Project of Hunan Province (CX20230114).

Acknowledgments

We deeply thank all the authors and reviewers who have participated in this Research Topic.

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.

The authors declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Hegde, P. S., and Chen, D. S. (2020). Top 10 challenges in cancer immunotherapy. *Immunity* 52 (1), 17–35. doi:10.1016/j.immuni.2019.12.011

Jia, Q., Wang, A., Yuan, Y., Zhu, B., and Long, H. (2022). Heterogeneity of the tumor immune microenvironment and its clinical relevance. *Exp. Hematol. Oncol.* 11 (1), 24. doi:10.1186/s40164-022-00277-y

Lee, J. Y., Kannan, B., Lim, B. Y., Li, Z., Lim, A. H., Loh, J. W., et al. (2022). The multi-dimensional biomarker landscape in cancer immunotherapy. *Int. J. Mol. Sci.* 23 (14), 7839. doi:10.3390/ijms23147839

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 11 May 2023
DOI 10.3389/fphar.2023.1191910

OPEN ACCESS

EDITED BY
Linhui Wang,
Second Military Medical University, China

REVIEWED BY
Hangcheng Fu,
University of Louisville, United States
Guang Lei,
University of Texas MD Anderson Cancer Center, United States

*CORRESPONDENCE
Zheng Wang,
wilsonwangzheng@163.com
Xiaosong Chen,
chenxiaosong0156@hotmail.com
Kunwei Shen,
kwshen@medmail.com.cn

†These authors have contributed equally to this work

RECEIVED 22 March 2023
ACCEPTED 17 April 2023
PUBLISHED 11 May 2023

CITATION
Huang R, Wang H, Hong J, Wang Z, Wu J, Huang O, He J, Chen W, Li Y, Chen X and Shen K (2023), A senescence-associated signature refines the classification of different modification patterns and characterization of tumor immune microenvironment infiltration in triple-negative breast cancer.
Front. Pharmacol. 14:1191910.
doi: 10.3389/fphar.2023.1191910

A senescence-associated signature refines the classification of different modification patterns and characterization of tumor immune microenvironment infiltration in triple-negative breast cancer

Renhong Huang†, Han Wang†, Jin Hong†, Zheng Wang*, Jiayi Wu, Ou Huang, Jianrong He, Weiguo Chen, Yafen Li, Xiaosong Chen* and Kunwei Shen*

Department of General Surgery, Comprehensive Breast Health Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Background: Recent studies have found that senescence-associated genes play a significant role in cancer biological processes. We aimed to analyze the characteristics and role of senescence-associated genes in triple-negative breast cancer (TNBC).

Methods: We systematically screened senescence-associated secretory phenotype (SASP) genes based on the gene expression information in the TCGA database. According to the expression levels of senescence-associated genes, TNBC was classified into two subtypes, namely, TNBCSASP1 and TNBCSASP2, using an unsupervised cluster algorithm. We then performed gene expression, enrichment pathway, immune infiltration, mutational profile characterization, drug sensitivity and prognostic value analyses for the two subtypes. The reliability and prognostic predictive utility of this classification model were validated. The most prognostically relevant gene, FAM3B, was comprehensively identified and validated by tissue microarray in TNBC.

Results: TNBC was classified into two senescence-associated subtypes, TNBCSASP1 and TNBCSASP2, based on the set of senescence-associated secretory phenotype genes, among which the TNBCSASP1 subtype had a poor prognosis. The TNBCSASP1 subtype was immunosuppressed, with suppressed immune-related signaling pathways and low immune cell infiltration. The effect of the mutation on the TP53 and TGF-β pathways could be related to the poor prognosis of the TNBCSASP1 subtype. Drug sensitivity analysis showed that AMG.706, CCT007093, and CHIR.99021 were potential targeted drugs for the TNBCSASP1 subtype. Finally, FAM3B was a key biomarker affecting the prognosis of patients with triple-negative breast cancer. Compared to normal breast tissue, the expression of FAM3B was reduced in triple-negative breast cancer. Survival analysis showed that overall survival was significantly shorter in triple-negative breast cancer patients with high FAM3B expression.

Conclusion: A senescence-associated signature with different modification patterns has critical potential for providing a better understanding of TNBC biological processes, and FAM3B might serve as an applicable target for TNBC therapy.

KEYWORDS

senescence-associated genes, triple-negative breast cancer, modification patterns, tumor immune microenvironment, FAM3B

Introduction

Triple-negative breast cancer is a special subtype of breast cancer, accounting for approximately 15%–20% of all breast cancers (Garrido-Castro et al., 2019). Compared with other subtypes of breast cancer, triple-negative breast cancer is more aggressive, with an earlier onset age, larger tumor volume, higher histological grade, early recurrence and distant metastasis, and poor overall survival (Garrido-Castro et al., 2019). With advances in research, it has been gradually revealed that triple-negative breast cancer is highly heterogeneous in biological aspects. In recent years, with the rapid development of multiomics technology, the heterogeneity of triple-negative breast cancer has been further confirmed at the genomic level, transcriptome level, and proteome level. Therefore, it is necessary to study different classification methods and identify the specific subtypes of triple-negative breast cancer to provide direction for the development of new targeted therapy strategies.

Cellular senescence is a permanent state of cell cycle arrest accompanied by changes in cell secretory characteristics (Campisi and d'Adda di Fagagna, 2007). Cellular senescence is a stress response that can be induced by various internal or external damage signals, including telomere dysfunction, oncogene activation, oxidative stress, and persistent DNA damage. Cellular senescence has long been considered a natural antitumor mechanism. Among them, one of the key mechanisms is oncogene-induced senescence (OIS) (Serrano et al., 1997), in which the activation of proto-oncogenes or the inactivation of tumor suppressor genes triggers cell growth arrest. Increasing evidence shows that cell senescence is also closely related to the occurrence and development of tumors. Through the mechanism of the senescence-associated secretory phenotype, senescent cells can act on surrounding tumor cells in a paracrine manner and change the tumor microenvironment to promote the occurrence and development of tumors. In the tumor microenvironment, cellular senescence is immunogenic, enhances MHC-1 antigen delivery, and can activate anti-tumor immune responses mediated by dendritic cells and $CD8^+$ T cells (Marin et al., 2023). Based on the above effects, senescent cells have the potential to become emerging markers of cancer (Hanahan, 2022). The senescence-related secretory phenotype is one of the important characteristics of senescence (Rodier and Campisi, 2011). Senescent cells undergo significant changes at the secretion level and secrete many substances dominated by proinflammatory factors, including cytokines, chemokines, growth factors and extracellular matrix proteases. By activating the above transcriptome, senescent cells can enhance their own senescence process in an autocrine manner and send signals to neighboring cells in a paracrine manner, thereby affecting the tissue microenvironment. Senescence cells can enhance macrophage display of senescence characteristics by secreting SASP factors. Senescence macrophages can affect other parts of the immune system to evade immune surveillance and clearance of senescence cells (Prieto and Baker, 2019). In addition, the accumulation of senescent cells in the tumor microenvironment will also promote the release of SASP factors and promote the growth of tumor cells. The relationship between senescence-related secretory phenotypes and tumor development is complex and is manifested by the pleiotropic effects of senescence-related secretory phenotypes. With the development of multiomics technology, the genes that play a key role in cellular senescence have been gradually identified by researchers, and senescence-related genomes have been formed. The proposal of senescence-related genomes provide data support for further research on the relationship between senescence and tumors. Using senescence-related datasets from the literature, a comprehensive pancancer analysis was conducted (Zhao et al., 2022). The results showed that senescence-related genes are widely different in different cancers and that cell senescence has an important impact on the tumor immune microenvironment. Wang et al. (Wang et al., 2022) used the senescence-related gene set to define a cellular senescence score in pancancer and demonstrated that this score can represent the degree of immune activation in the tumor microenvironment and can identify groups with better prognosis. In specific cancers, researchers have also constructed classification models and prognostic prediction models based on senescence gene sets, including clear cell renal cell carcinoma (Lu et al., 2022), hepatocellular carcinoma (Luo et al., 2022), glioblastoma (Tan et al., 2022), etc., and explored the relationship between senescence-related genes and clinical characteristics, tumor microenvironment, immunotherapy, etc. These studies demonstrate the importance of senescence genes in cancer therapy.

The aim of this study was to construct a classification model for triple-negative breast cancer based on the expression profile of senescence-associated secretory phenotype genes and to explore the prognostic value of this classification model. The differences in different senescence-related subtypes were compared at the multiomics level, including gene enrichment pathways, immune infiltration, genomic mutations, drug sensitivity, etc. The classification model was validated on external datasets. Furthermore, based on the senescence-related subtypes of triple-negative breast cancer, the key genes affecting the prognosis of different subtypes were explored, and the prognostic role of these genes in triple-negative breast cancer was analyzed.

Method and material

Data collection and processing

The pan cancer expression profiles and survival information, which included 110 triple-negative breast cancer patients from TCGA-BRCA, were retrieved from the XENA datasets

(http://xena.ucsc.edu/) (Tomczak et al., 2015). Genomics information, containing copy number variation and SNV of BRCA, was downloaded from The Cancer Genome Atlas (TCGA) (Blum et al., 2018). Three independent TNBC cohorts were used: GSE21653 (252 samples), GSE25066 (205 samples) and GSE103091 (111 samples) (Hatzis et al., 2011; Sabatier et al., 2011; Jézéquel et al., 2015). For cohorts downloaded from public datasets, instructional review board approval or informed consent was not needed. A novel senescence-associated signature was acquired from the Supplementary Material from the Saul et al. search (Supplementary Table S1).

Unsupervised cluster analysis

First, we performed univariable Cox analysis to filter prognosis-related genes from the senescence-associated signature. Finally, a three-signature-based expression matrix of TCGA-TNBC, including CXCL1, CCL13 and ACVR1B, was maintained to perform unsupervised clustering analysis to identify novel senescence subtypes of TNBC with the use of the ConsensusClusterPlus package (Wilkerson and Hayes, 2010). The optimal cluster number k of TNBC was evaluated by the proportion of ambiguous clustering (PAC score) and consensus cumulative distribution function (CDF) curve.

Differential expression and enrichment analysis

The count mRNA expression matrix of TNBC was used to conduct differential expression analysis through the DEseq2 package and visualized by the EnhancedVolcano package (Love et al., 2014). The threshold was set as abstract log-fold change = 1.5 and p-adjusted value <0.01. In addition, common enrichment, including Gene Ontology (GO), gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA), was also performed with the ClusterProfiler package (Yu et al., 2012). For differential expression gene annotation, we downloaded cancer-related hallmarks from the MSigDB dataset and IOBR package (Liberzon et al., 2011).

Immune infiltration analysis

Five immune-related signature gene sets, containing chemokines, chemokine receptors, MHCs, immunoinhibitors and immune stimulators, were compared between subtypes. Several immune-related deconvolution algorithms, including TIMER, CIBERSORT, QUANTISEQ, MCPCOUNTER, XCELL and EPIC, were adopted to compare the different immune components. The tumor immune dysfunction and exclusion (TIDE) algorithm was further utilized to estimate the immunotherapy response score for TNBC. In addition, the Tracking Tumor Immunotype (TIP) algorithm was also applied to compare antitumor immunity differences between subtypes, which included 7 steps as follows: release of cancer cell antigens (step 1), cancer antigen presentation (step 2), priming and activation (step 3), trafficking of immune cells to tumors (step 4), infiltration of immune cells into tumors (step 5), recognition of cancer cells by T cells (step 6), and killing of cancer cells (step 7) (Xu et al., 2018). The Estimate package was further adopted to verify the immune infiltration degree difference between subtypes.

Mutation spectrum characteristics

We downloaded genomic mutation, except for germline mutation, profiles of TNBC from TCGA GDC database, then we compared and visualized the difference between subtype through Maftools package (Mayakonda et al., 2018). Multilevel differences in the genomic profile, including onco-pathway, somatic interaction, mutation prognostic impact and drug categories, were also compared by Maftools and ggpubr packages.

Validation of remodeling results in validation cohorts

After identifying subtype biomarkers of each subtype, we next applied the nearest template prediction (NTP) algorithm to perform remodeling analysis in three independent TNBC cohorts. After estimating each sample's cluster tendency, we compared the prognostic difference to verify the reproductivity of unsupervised cluster results.

Potential implications of preclinical treatment agent analysis

For therapeutic sensitivity analysis, we first collected the subtype and mRNA expression of TNBC; then, the cell line's expression profile and therapeutic information from genomic of drug sensitivity in cancer (GDSC) datasets were also downloaded (Cokelaer et al., 2018). We next applied Ridge's regression and 10-fold cross validation to identify subtype molecular agents by the pRRophetic package (Geeleher et al., 2014). Half maximal inhibitory concentration (IC50) values were applied to compare different sensitivities between subtypes.

Clinical data collection and follow-up

The clinicopathological data of the patients in this study were obtained from the Shanghai Jiao Tong University-Breast Cancer Database (SJTU-BCDB). The included clinicopathological features were the following: age, menstrual status, tumor location, tumor size, number of lymph node metastases, TNM stage, histological grade, lymph vascular invasion, Ki-67 index, and adjuvant therapy information. A Ki-67 index >30% was defined as high Ki-67 expression, and a Ki-67 index ≤30% was defined as low Ki-67 expression. All breast cancer patients received regular follow-up in the clinic or by telephone. Patients were followed up every 3 months for 2 years after surgery. Follow-up was performed every 6 months for 3–5 years, and follow-up was performed annually 5 years after surgery until event and death. The

follow-up data were recorded and summarized by the breast specialist nurses and clerks in our department. Overall survival (OS) and disease-free survival (DFS) were defined as reported in our previous study (Huang et al., 2022). The last follow-up was in January 2023. Patients' inclusion and exclusion criteria were described in the Supplementary Figure S2. The detail information of these patients was shown in Supplementary Table S5.

Immunohistochemistry (IHC) of tissue microarray

IHC staining of FAM3B protein expression in the tissue microarray was performed by incubation with rabbit polyclonal antibodies against human FAM3B antibody (27131-1-AP, Proteintech, 1:200) overnight, followed by incubation with goat monoclonal antibody against rabbit antibody (111-035-003, JACKSON, 1:1,000) for 1 h at room temperature. The immunohistochemical evaluation of FAM3B was employed and analyzed by two individual pathologists, Anqi Li and Miao Ruan, who were blinded to the clinical information of the patients. The immunohistochemical staining results were analyzed by ImageJ processing software. Protein expression was evaluated based on the optical density (OD) value of the images. Immunohistochemical staining was evaluated at the same magnification (×40) in five randomly selected fields of tumor tissue from each patient. After setting a specific threshold for each image, the integrated optical density (IOD) value was calculated in the positive area. The average optical density (AOD) was calculated by the ratio of IOD to the area of the positive area. The average AOD value of the five regions was taken as the result of the AOD value of the patient, which represented the expression level of the markers.

Statistical analysis

All multiomics dataset processing, plotting and statistical tests were performed using R software (Version 4.1.0). Student's t tests and Mann–Whitney U tests were applied for continuous variables with normal or skewed distributions. The chi-square test or Fisher's exact test was utilized to compare categorical variables. Spearman correlation was used to calculate the correlation index between quantitative variables by the corrplot package. Kaplan–Meier and time ROC curves were depicted by the survival package. All statistical tests were two-sided with a level of significance set as $p < 0.05$.

Results

Expression of senescence-associated genes in pancancer

To initially explore the expression of senescence-related genes in tumors, the expression levels of senescence-related genes in tumor tissues and normal tissues were compared in 20 cancer types (Figure 1A). The results showed that in several cancer types, including breast cancer, the expression of senescence-related genes was lower in tumor tissues than in normal tissues, indicating that the expression of senescence-related genes was suppressed. Activating senescence-related pathways may be a potential way to treat breast cancer.

Next, the enrichment of senescence-related genes in classical tumor pathways was investigated. Senescence-related genes were associated with multiple immune pathways, including the KRAS upregulation pathway, IFNγ response pathway, IFNα response pathway, immune response pathway, IL-6/JAK/STAT3 pathway, and IL-2/STAT5 pathway (Figure 1B). The activation status was consistent in many cancer types. Therefore, targeting senescence-related pathways may activate immune regulation and other pathways, thereby promoting the occurrence and development of tumors.

Construction and functional enrichment analysis of senescence-associated subtypes in TNBC

By analyzing the senescence-related genes, three genes significantly associated with prognosis were screened out among the 125 genes in the senescence-related gene set using univariate Cox regression analysis ($p < 0.05$) (Figure 2A). CXCL1 ($p = 0.026$, HR = 0.723) and CCL13 ($p = 0.035$, HR = 0.761) were protective factors related to prognosis, while ACVR1B ($p = 0.049$, HR = 1.945) was a risk factor related to prognosis. Unsupervised cluster analysis was performed using the Consensus Cluster Plus package in the TNBC population of the TCGA database. According to the cumulative distribution function (CDF) and the proportion of ambiguous clustering (PAC) (Figures 2B–D), by analyzing the distribution of the CDF curve and the change in the area under the curve, the classification reliability was evaluated, and the optimal number of Clusters $k = 2$ was obtained. According to the clustering results, the population in TCGA-TNBC was divided into two senescence-related subtypes, named TNBCSASP1 and TNBCSASP2. The results of principal component analysis (PCA) based on senescence-related genotyping were analyzed (Figure 2E). Further analysis of the gene expression profiles of CXCL1, CCL13, and ACVR1B in the two senescence-related subtypes and normal breast tissues (Figure 2F) showed that CXCL1 and CCL13 were significantly downregulated in the TNBCSASP1 subtype.

Next, the clinical significance of senescence-related subtyping in terms of prognosis was evaluated by comparing the survival outcomes of the two subgroups. Kaplan–Meier survival analysis showed that patients in the TNBCSASP1 group had poor survival outcomes (Figure 3), and the OS, DSS, and PFI of the TNBCSASP1 group were significantly lower than those of the TNBCSASP2 group ($p < 0.05$). TNBCSASP classification based on senescence-related genes can predict the prognosis of TNBC patients to a certain extent. Based on the different prognoses of the TNBCSASP1 and TNBCSASP2 groups, it was necessary to further explore the specific biological differences between the two subtypes. The differentially expressed genes of the two subtypes were analyzed using the DEseq2 package (Figure 4). Next, functional enrichment analysis of differentially expressed genes was performed using the ClusterProfiler package. GO (Gene Ontology) enrichment analysis

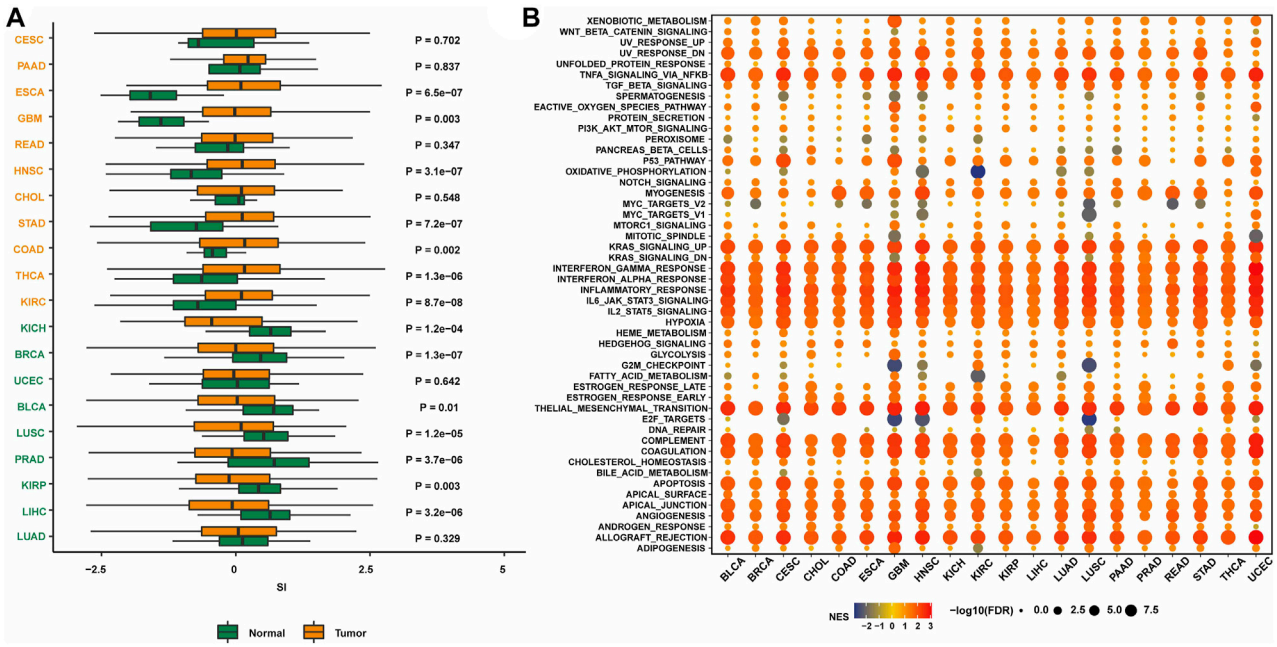


FIGURE 1
The expression of SASP genes across cancers. **(A)** The landscape of SASP gene set scores (based on NES and *p* values) in 20 cancer types. **(B)** The association between SASP gene set score and expression with HALLMARKS enrichment scores in 20 cancer types.

showed that in terms of biological process (BP), differentially expressed genes were mainly involved in antibacterial humoral response, antimicrobial peptide-mediated humoral immune response, negative regulation of peptidase activity and other processes. In terms of cell component (CC), the differentially expressed genes were mainly located in the tertiary granule cavity, specific granule cavity and platelet α granule cavity. In terms of molecular function (MF), the differentially expressed genes were mainly involved in receptor ligand activity, signal receptor activator activity, peptidase inhibitor activity and other functions. GSEA of senescence-related differentially expressed genes showed that oxidative phosphorylation, cAMP signaling pathway, and estrogen signaling pathway were activated, while cytokine receptor interaction pathway, primary immunodeficiency pathway, and antigen processing and presentation pathway were inhibited in TNBCSASP1 subtypes (Figure 5). These results suggest that the TNBCSASP1 subtype tumors may be in a state of poor immune response due to the inhibition of immune-related pathways, leading to a poor prognosis of patients with this subtype.

Analysis of immune infiltration in senescence-associated subtypes of TNBC

Based on the above functional enrichment analysis results, the TNBCSASP1 and TNBCSASP2 isoforms showed significant differences in immune-related pathways. To further characterize the immune profile of the senescence-related subtypes of TNBC, five immune-related signature gene sets were analyzed (Figure 6A), including chemokines, chemokine receptors, MHC, immunosuppressive factors, and immune stimulating factors. The results showed that CXCL1, CXCL13, CXCL10, CCL13, CCL18, CCL8, IDO1 and other immune-related characteristic genes were significantly downregulated in TNBCSASP1 subtypes. Subsequently, deconvolution algorithms such as TIMER, CIBERSORT, CIBERSORT-ABS, MCPCOUNTER, QUANTISEQ, EPIC and XCELL were used to compare the composition of immune-infiltrating cells in the tumor microenvironment of the TNBCSASP1 and TNBCSASP2 subtypes. The results showed that the TNBCSASP1 subtype had less infiltration of immune cells, including monocytes, macrophages, myeloid dendritic cells, $CD8^+$ T cells, and others, than the TNBCSASP2 subtype (Figure 6B). Therefore, based on the results of the above analysis, the TNBCSASP1 isoform can be defined as an senescence-related isoform in an immunosuppressed state.

Next, the ESTIMATE algorithm was used to explore components of the tumor microenvironment in the two senescence-related subtypes to assess tumor purity (Figures 7A–C). These include the stromal score and immune score, which are used to evaluate the composition of stromal cells and immune cells in tumor samples, respectively. The results showed that the immunoscore and ESTIMATE score of the TNBCSASP1 subtype were significantly lower than those of the TNBCSASP2 subtype, indicating that the TNBCSASP1 subtype had a lower degree of immune cell infiltration and a higher proportion of tumor cells in the tumor microenvironment. Comparing the antitumor activities of the TNBCSASP1 and TNBCSASP2 isoforms (Figure 7D), the TNBCSASP1 isoform was found to be less active during most antitumor steps. Cancer cell antigen release (step 1), CD4 T cells, CDB T cells, macrophages, and other cell recruitment (step 4), T-cell-to-tumor cell recognition (step 6), tumor cell clearance (step 7), etc. Further analysis of the

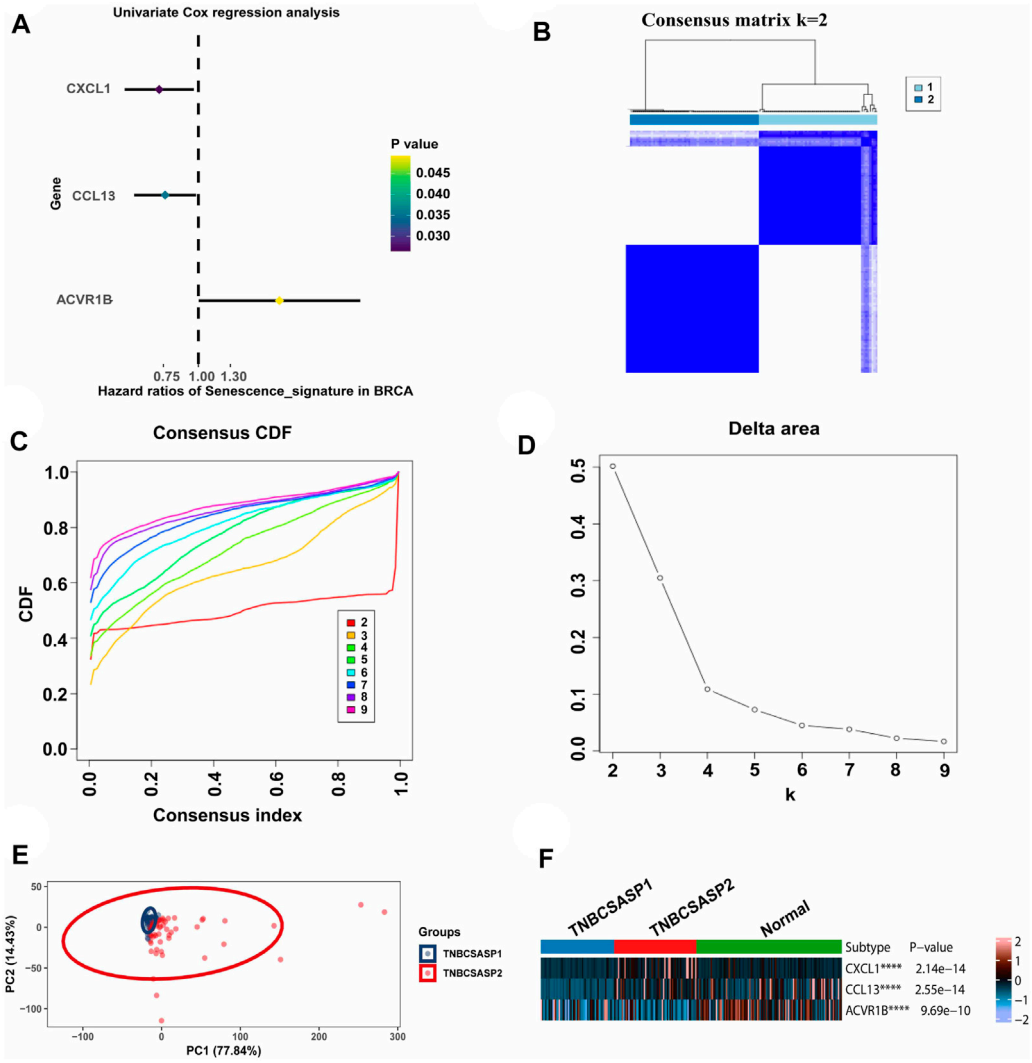


FIGURE 2
Clustering of SASP genes for the identification of different cancer modification patterns in TNBC. **(A)** Univariate Cox regression analysis of SASP genes in TNBC. **(B)** Consensus matrix heatmap displaying the scale for two cleanly separated clusters. **(C)** CDF plot shows flattening of the consensus index curve for every consensus matrix from two to nine. **(D)** Delta area curve of consensus clustering, which indicates the relative change in area under the CDF curve for each category number k compared to that of k-1. The horizontal axis represents the category number k, and the vertical axis represents the relative change in area under the CDF curve. **(E)** Principal component analysis was carried out based on the results of consensus clustering. **(F)** The expression profiles of SASP regulator genes among the two subtypes and normal tissues.

composition of immune-infiltrating cells in the tumor microenvironment of senescence subtypes of TNBC showed that the TNBCSASP1 subtype had a lower infiltration of DCs than the TNBCSASP2 subtype (Figure 7E).

Somatic mutations in senescence-related subtypes of TNBC

Based on the genome mutation spectrum of TNBC in the TCGA GDC database, Maftools and ggpubr packages were used to analyze the differences in somatic mutations between the TNBCSASP1 and TNBCSASP2 subtypes and to explore the potential carcinogenic factors in the two senescence subtypes. We also identified the top 20 most frequently mutated genes in the two senescence subtypes, with similar mutation rates in the two groups (96% vs. 95%) and high-frequency mutations in TP53 and TTN. MUC17, PIK3CA, ABCA13, ZKSCAN7 and other genes had higher mutation frequencies in the TNBCSASP1 subtype than in the TNBCSASP2 subtype (Figures 8A, B). The somatic comutations of the two senescence subtypes were shown (Figures 8C, D). The TNBCSASP1 subtype had comutations in MUC16-SI, ABCA1-MUC16, CACNA1F-PCDH15, F5-SI, and MAP1A-NF1 ($p < 0.01$). There were comutations of CSMD3-HMCN1, FAT3-PKHD1L1, FAT3-LRP2, RELN-CASR, APOB-DNAH2, and PTEN-AHNAK ($p < 0.01$) in the TNBCSASP2 subtype. TP53-TTN mutations were mutually exclusive. Subsequently, somatic

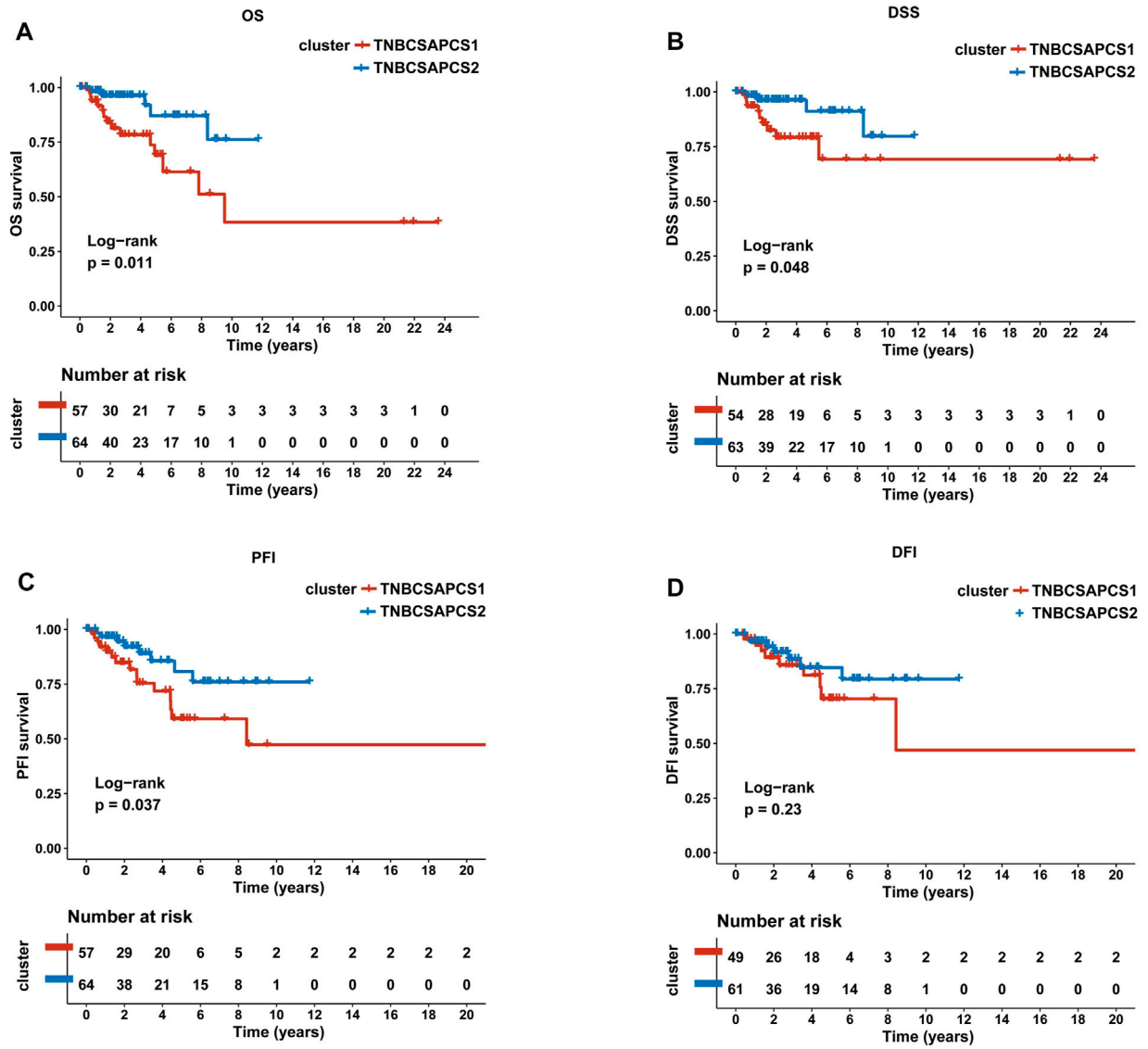


FIGURE 3
Prognostic value of TNBCSASP1 and TNBCSASP2 in TNBC. Survival analysis for OS **(A)**, DSS **(B)**, PFI **(C)** and DFI **(D)** among the TNBCSASP1 and TNBCSASP2 subtypes of TCGA data.

mutations in common tumor-related pathways in both subtypes were evaluated, including the RTK-RAS, NOTCH, WNT, Hippo, PI3K, TP53, TGF-Beta, NRF2, MYC, and Cell Cycle pathways. The results showed that TP53 and TGF-beta pathway receptor cell mutations had the greatest impact on the TNBCSASP1 subtype, whereas TP53, PI3K, and RTK-RAS pathway receptor cell mutations had the greatest impact on the TNBCSASP2 subtype (Figures 8E, F). Next, based on data from the somatic mutant gene set, the DGIdb database and the Maftools package were used to analyze drug-gene interactions in the two senescence-related subtypes to identify potential drug-target genes (Figures 8G, H). The drug target genes of the two senescence-related subtypes could be divided into 18 and 23 types, including druggable genomes, clinically operable genomes, kinase genomes, transport-related genomes, etc.

Drug sensitivity analysis of senescence-related subtypes in TNBC

The analysis of drug sensitivity to chemotherapy agents (Figures 9A, B) showed that the semi-inhibitory concentration of the TNBCSASP1 subtype was higher when treated with either paclitaxel or cisplatin, indicating that the TNBCSASP1 subtype was less sensitive to chemotherapy agents, which is consistent with the previous results of worse prognosis of the TNBCSASP1 subtype. We further analyzed the drug responsiveness of the TNBCSASP1 and TNBCSASP2 isoforms to molecular inhibitors and presented the top 10 potential drugs with the most significant therapeutic differences between the two senescence-related isoforms (Figures 9C, D). TNBCSASP1 was more sensitive to AMG.706, CCT007093, and CHR.99021, while

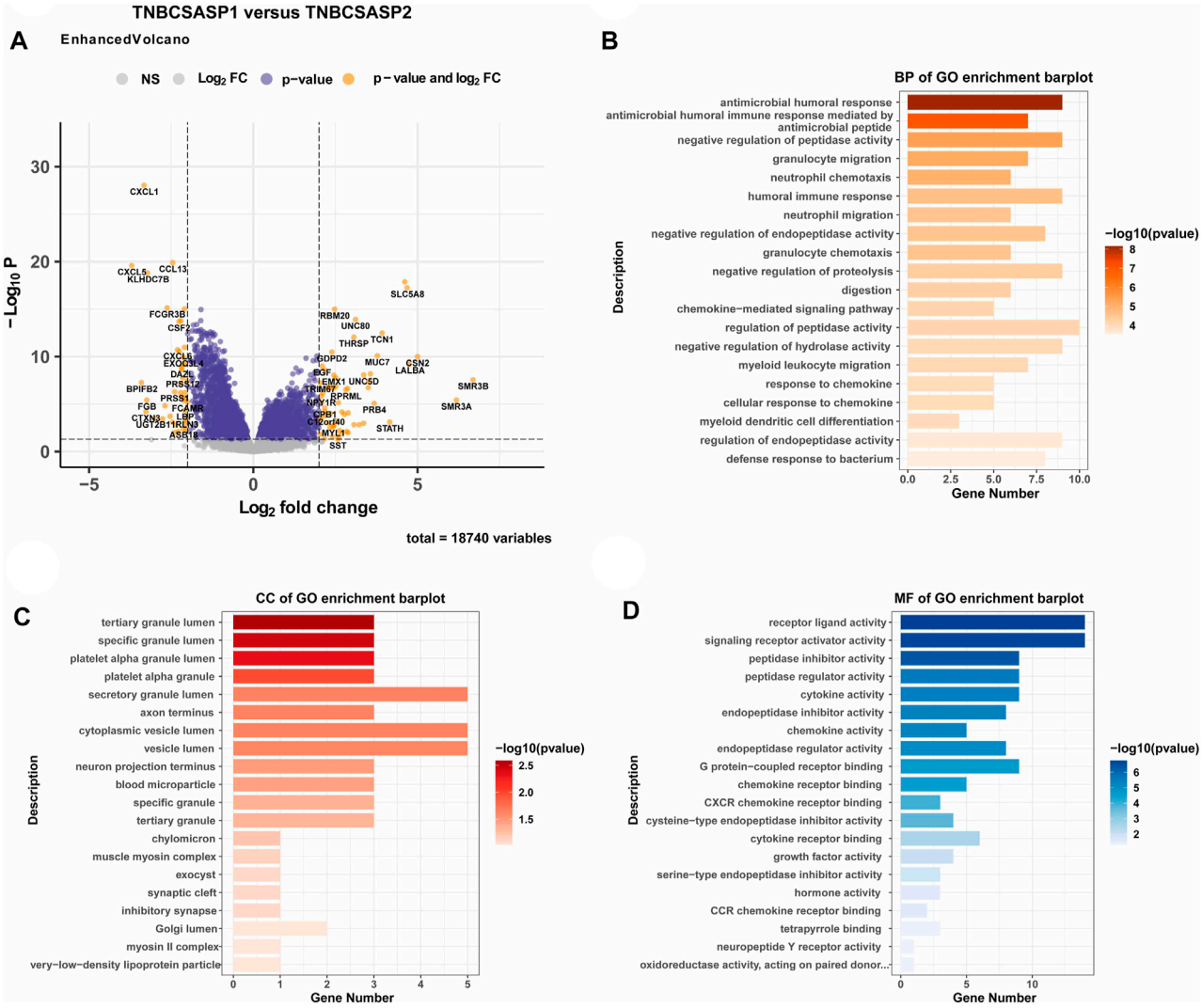


FIGURE 4
Functional enrichment analysis of the TNBCSASP1 and TNBCSASP2 subtypes. **(A)** Volcano plot of the landscape of gene expression changes between the 2 subtypes. **(B–D)** The GO terms of the BP, CC, and MF categories enriched in the senescence-associated differentially expressed genes. genes.

TNBCSASP2 was more sensitive to sunitinib, mitomycin C, and camptothecin. It can be predicted that drugs such as AMG.706, CCT007093, and CHIR.99021 are expected to become new development targets for TNBCSASP1, a subtype with poor prognosis, and provide more precise treatment strategies for TNBC patients.

The prognostic value of senescence-related subtypes and the hub gene FAM3B in TNBC

To confirm the reliability of the senescence-related classification, external datasets were used to validate the classification model. The independent TNBC cohort GSE103091 was divided into two distinct subtypes using the nearest template prediction (NTP) algorithm using the senescence subtype-specific genes identified in the TCGA-TNBC dataset and the classification model reconstruction (Supplementary Figure S1A). A subsequent survival analysis showed that the TNBCSASP1 subtype had a significantly lower survival rate than the TNBCSASP2 subtype (Supplementary Figure S1B, $p = 0.007$), which was consistent with the results of the previous analysis. Therefore, the reliability and stability of senescence-related typing were verified. Based on the above two subtypes, we identified the key differentially expressed gene FAM3B. We used random forest algorithm to identify the gene FAM3B which is most relevant to the prognosis of TNBC (Figure 10A). The expression of FAM3B in triple-negative breast cancer was further analyzed. The relevant data in the GSE21653 cohort were obtained, and the expression level of FAM3B in different subtypes was analyzed based on PAM50 typing (Figure 10B). Compared with normal tissues, the expression level of FAM3B in basal-like subtypes was lower. In addition, proteomic information was obtained from the CPTAC database, and the results similarly showed that the expression level of FAM3B was lower in triple-negative breast cancer tissues

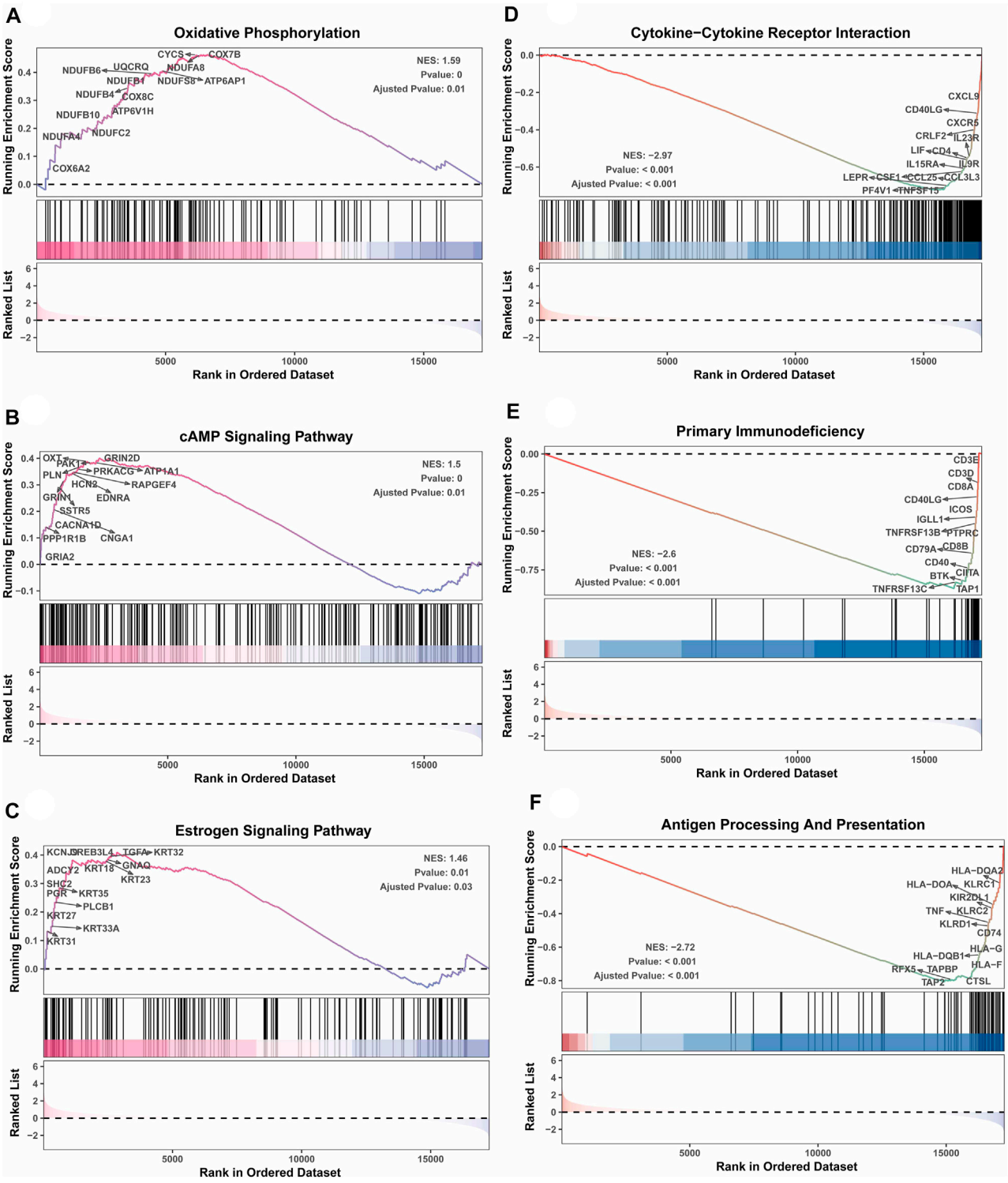


FIGURE 5
GSEA depicting the enriched pathways of senescence-associated differentially expressed genes in TNBC. These pathways include oxidative phosphorylation **(A)**, cAMP signaling pathway **(B)**, estrogen signaling pathway **(C)**, cytokine–cytokine receptor interaction **(D)**, primary immunodeficiency **(E)**, and antigen processing presentation **(F)**.

(Figure 10C). Single-cell analysis of triple-negative breast cancer was performed in different independent datasets, and FAM3B expression in different cells was investigated (Figure 10D). The results of the analysis of different datasets all showed that FAM3B was specifically highly expressed in malignant epithelial cells in the triple-negative breast cancer population.

From January 2009 to March 2021, 1469 patients with triple-negative invasive breast cancer underwent surgery at our center,

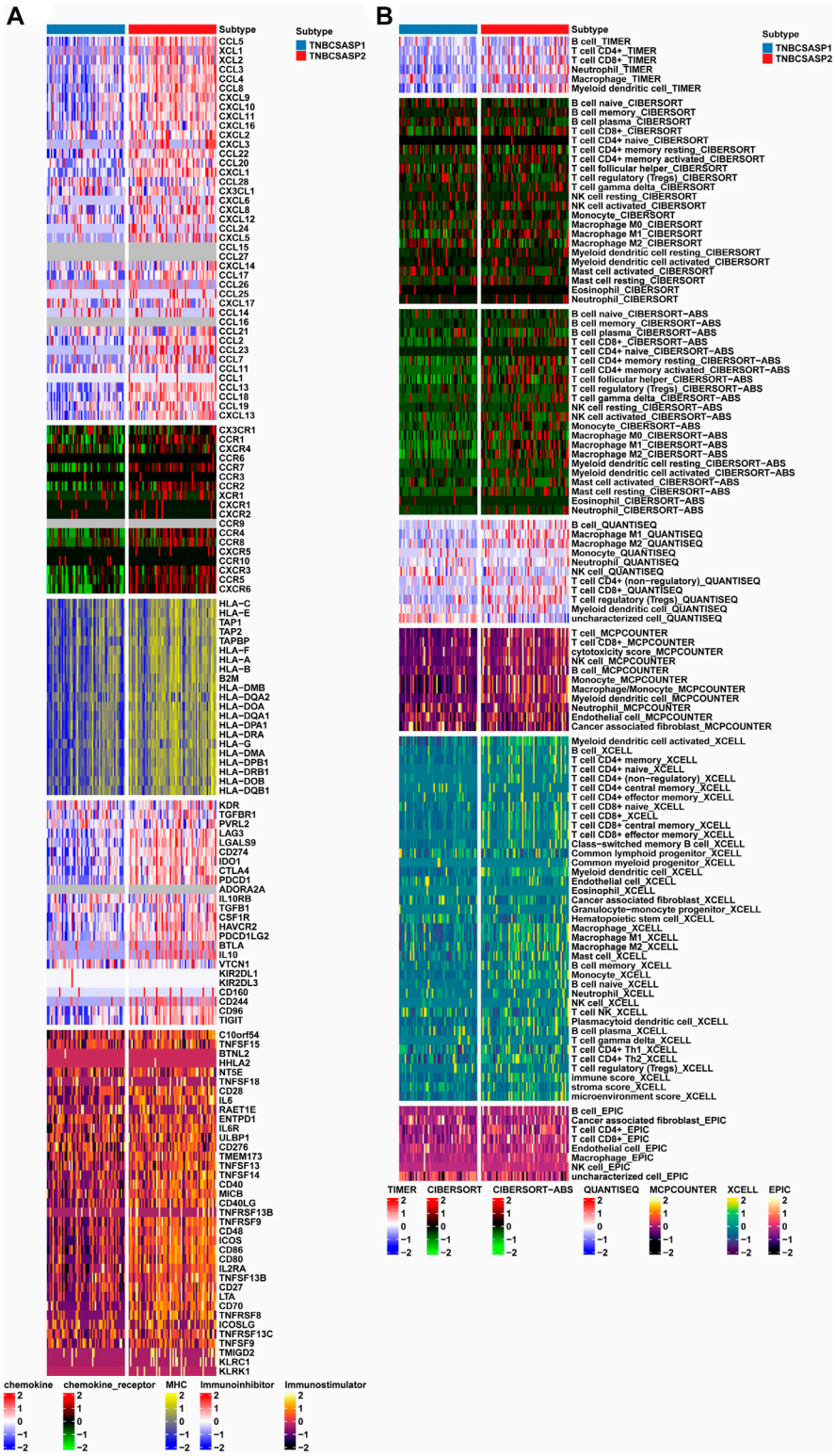


FIGURE 6
Comprehensive analysis of immune landscapes in TNBCSASP1 and 2 subtypes. **(A)** Heatmap showing differential immune-related genes among the TNBCSASP1 and TNBCSASP2 subtypes. **(B)** Heatmap plot of tumor-related infiltrating immune cells based on TIMER, CIBERSORT, CIBERSORT-ABS, MCPCOUNTER, QUANTISEQ, EPIC, and XCELL algorithms among the TNBCSASP1 and TNBCSASP2 subtypes.

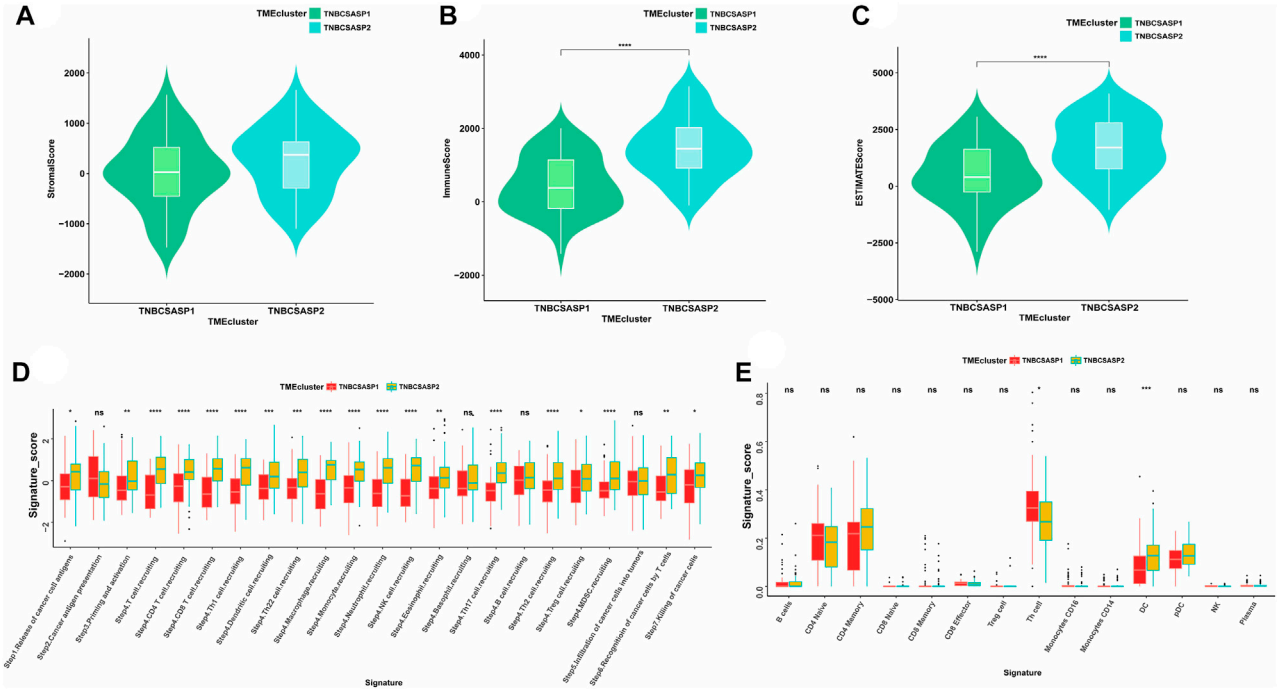


FIGURE 7
Immune landscapes in TNBCSASP1 and 2 subtypes. Comparison of stromal score **(A)**, immune score **(B)**, and ESTIMATE score **(C)** among the TNBCSASP1 and TNBCSASP2 subtypes. **(D–E)** The immune pathways and anticancer steps among the TNBCSASP1 and TNBCSASP2 subtypes.

and 122 patients were finally included for survival analysis in the study. Univariate analysis of clinicopathological features of patients with different FAM3B expression levels was further performed, and the results showed that there were no significant differences between the FAM3B high expression group and the FAM3B low expression group in each clinicopathological feature (Supplementary Table S2). Univariate analysis of clinicopathological features for predicting prognosis showed that lower clinicopathological stage was associated with longer OS. Higher clinicopathological stage and lymph node metastasis were associated with DFS events. Low FAM3B expression was associated with longer OS and DFS. Other pathological factors did not predict the prognosis of patients (Supplementary Tables S3, S4). Further Kaplan–Meier survival analysis of the FAM3B high and low expression groups showed that OS ($p = 0.036$) and DFS ($p = 0.031$) in the FAM3B high expression group were significantly lower than those in the FAM3B low expression group (Figures 10E, F). Survival analysis showed that high FAM3B expression was associated with poor prognosis in triple-negative breast cancer.

Discussion

Triple-negative breast cancer (TNBC) is a subtype of breast cancer with poor prognosis due to its high invasiveness and lack of specific therapeutic targets. As triple-negative breast cancer is highly heterogeneous, it is necessary to characterize different molecular subtypes from a variety of biological perspectives to identify patient risk stratification, predict disease prognosis, and provide potential therapeutic targets.

Based on the existing senescence-associated secretory phenotype gene sets, we constructed a senescence-related classification model for TNBC and defined the two groups as the TNBCSASP1 and TNBCSASP2 subtypes. Survival analysis and multiomics analysis showed that the TNBCSASP1 subtype tumors were in an immunosuppressed state. Immune-related pathways were inhibited, the expression of immune-related factors was lower, the number of immune infiltrating cells was lower, and the TNBCSASP1 subtype showed lower antitumor activity and worse prognosis. Based on the above two subtypes, we identified the key differentially expressed gene FAM3B. Compared with normal breast tissue, FAM3B expression was different in different subtypes of breast cancer. The results showed that high expression of FAM3B in TNBC tumor tissues was related to poor prognosis in patients, suggesting that FAM3B may play a carcinogenic role in TNBC.

Cellular senescence is an important mechanism to maintain tissue homeostasis and aims to eliminate the stress response of damaged cells. Cellular senescence can be induced by damage factors such as DNA damage, reactive oxygen species, activation of oncogenes and inactivation of tumor suppressor genes (Herranz and Gil, 2018). However, recent studies have found that cellular senescence plays an important role in the tumor-promoting process and has been included as one of the emerging markers of cancer (Hanahan, 2022). SASP is characterized by the secretion of many inflammatory factors involved in the immune response, and its dynamic components affect the tumor microenvironment and participate in the regulation of the immune response. SASP can

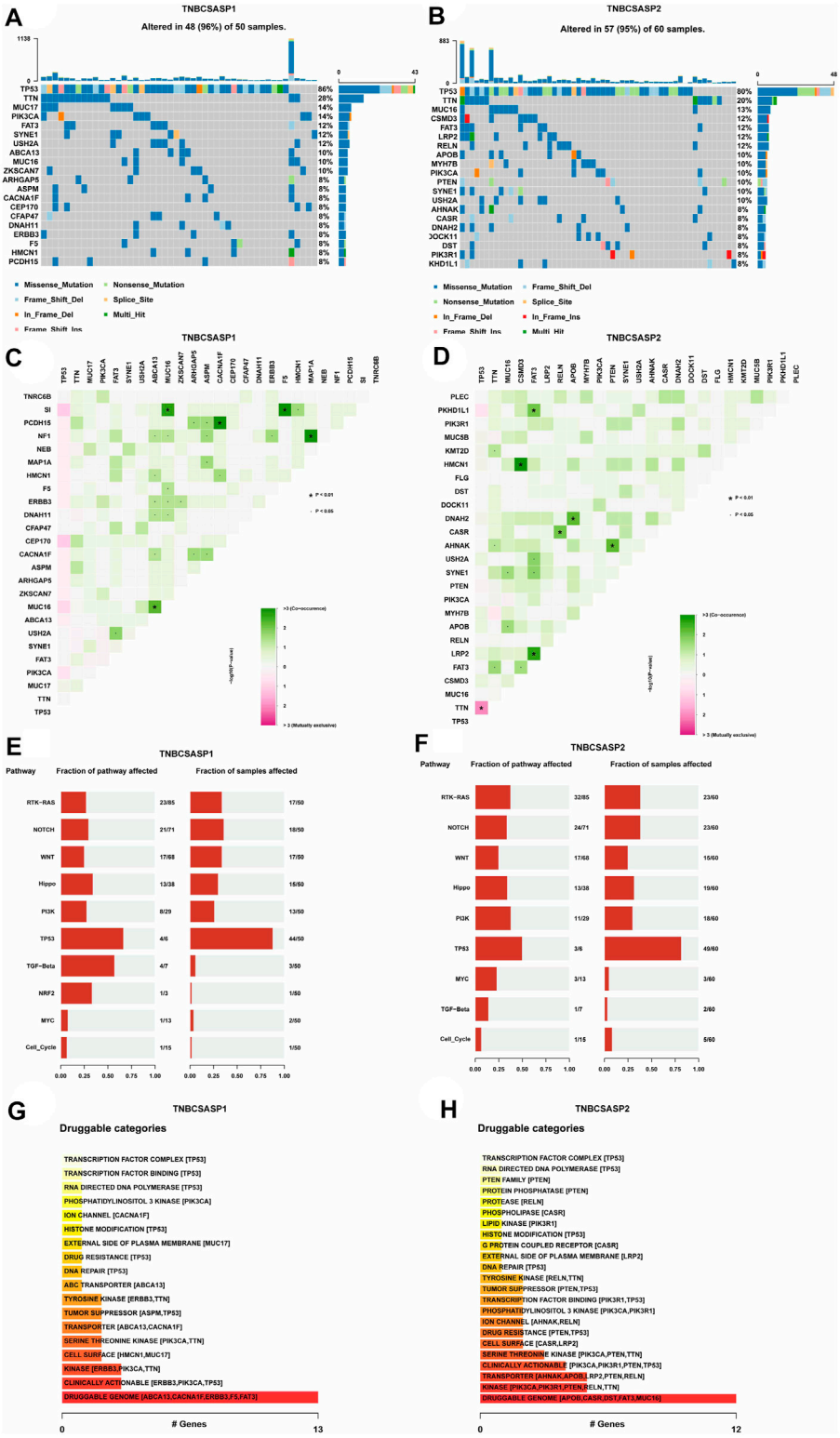


FIGURE 8
Profiles of somatic mutations and potential targets among the TNBCSASP subtypes. **(A,B)** Waterfall plot showing the mutation patterns of the top 20 most frequently mutated genes. **(C,D)** Cooccurring mutations in TNBCSASP1 and TNBCSASP2. **(E,F)** The fraction of pathways or samples of oncogenic signaling pathways among TNBCSASP1 and TNBCSASP2. **(G,H)** Potential druggable gene categories from the mutation dataset among TNBCSASP1 and TNBCSASP2.

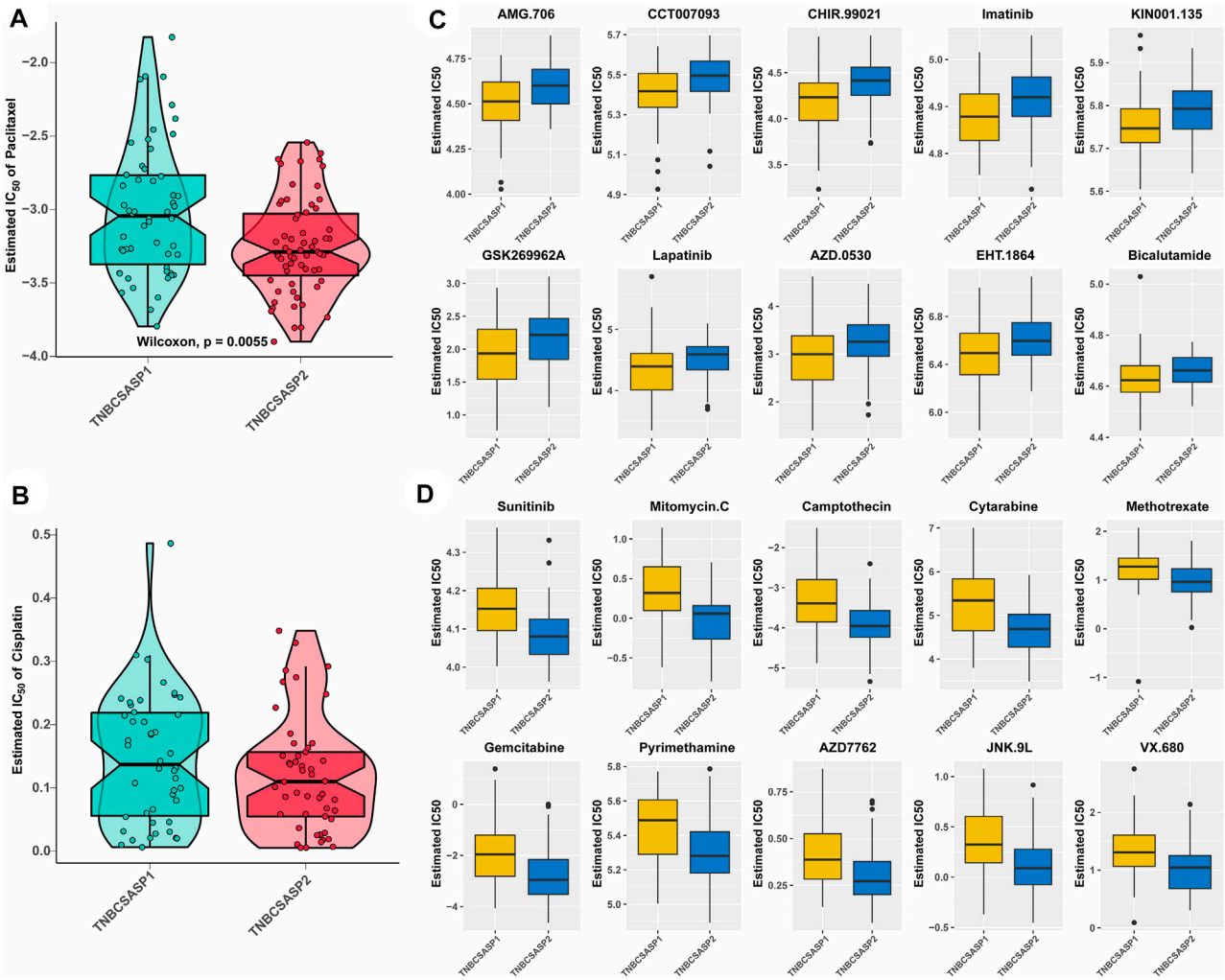


FIGURE 9
Drug sensitivity analysis of the TNBCSASP subtypes. **(A,B)** Estimated IC50 of the indicated molecular-targeted drugs among TNBCSASP1 and TNBCSASP2. **(C,D)** Estimated IC50 of the potential molecular inhibitors in TNBCSASP1 and TNBCSASP2.

continuously secrete a variety of inflammatory factors, maintain a low adaptive immune response in the tissue microenvironment, and form a "chronic inflammatory" environment (López-Otín et al., 2013). Important related factors include IL-1, IL-6, GM-CSF, IFNγ, TNF and CRP (Zinger et al., 2017). *In vitro* experiments showed that the senescence-related proinflammatory cytokines IL-6 and IL-8 were important for maintaining the invasive properties of the triple-negative breast cancer cell line MDA-MB-231. In a study of the breast cancer population (Knüpfer and Preiss, 2007), IL-6, as a key inflammatory factor, plays an important role in promoting tumor progression, and high circulating levels of IL-6 and age are associated with poor prognosis in breast cancer patients. *In vivo* experiments on breast cancer (Wang et al., 2019), the number of MDSs in the tumor microenvironment of senescence mice was significantly increased, and depletion of MDSCs significantly reduced tumor growth in senescence mice. *In vitro* experiments confirmed that these tumor-specific MDSCs had a highly immunosuppressive effect. In addition, factors in the tumor microenvironment can also induce T-cell senescence and inhibit the immune response. Preclinical studies have shown that the senescence of $CD8^+$ T cells plays a key role in the development of breast cancer (Onyema et al., 2015). According to our study, the degree of infiltration of $CD8^+$ T cells with immune response function in the TNBCSASP1 subtype was significantly reduced, reflecting the immunosuppressive state of this subtype. On the other hand, the expression of proinflammatory factors and immunosuppressive MDSCs was also generally low in this subtype. TNBCSASP1 subtype has an immune desert phenotype and is more likely to be defined as a "cold tumor" in which effector T cells are unable to infiltrate into the tumor microenvironment, making it difficult for them to exert anti-tumor immune effects. Activating immune cells infiltration may be an effective therapeutic strategy to address the immune escape mechanism in TNBCSASP1 subtype. Since senescent cell types, senescence-inducing factors, tumor progression and other factors affect the pleiotropic effects of SASP, the specific mechanism of SASP on the TNBCSASP1 subtype tumor microenvironment needs to be further explored. Based on the results in this study, we identified

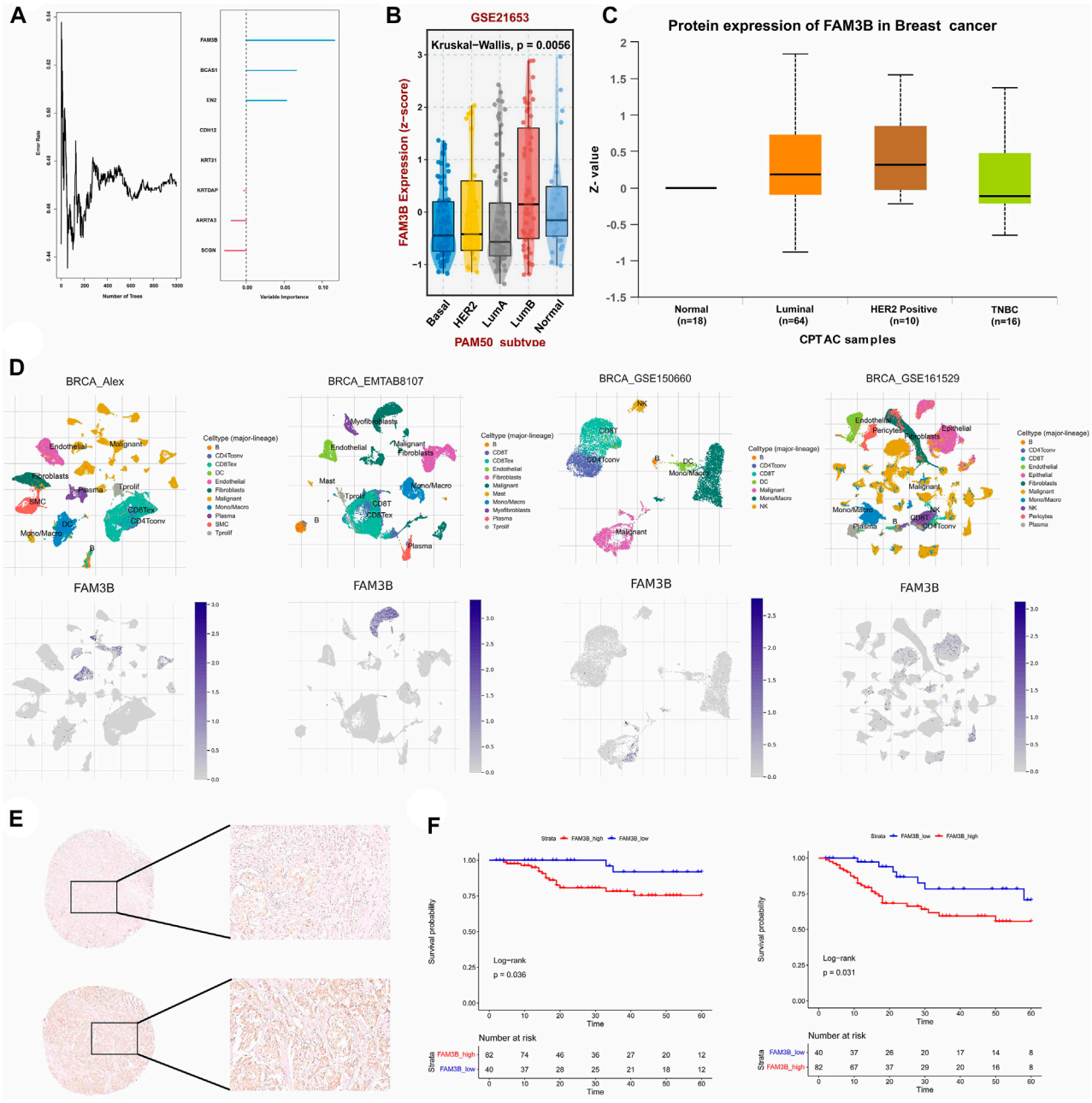


FIGURE 10
The hub gene FAM3b and its prognostic value in TNBC. **(A)** Number of trees showing the importance proportion of SASP regulator genes. **(B)** FAM3B expression in paired tumor and normal tissues in BRCA from the TCGA database. **(C)** FAM3B expression according to the molecular subtypes of breast cancer. **(D)** UMAP plot of intratumoral immune cells and FAM3B showing the correlation between infiltration of different immune cells and FAM3B expression in the Alex, EMTAB8107, GSE150660 and GSE161529 datasets. **(E,F)** The impact of FAM3B on OS and DFS in breast cancer tissue microarray using Kaplan-Meier analysis.

TNBCSASP1, a subtype with a worse prognosis based on the difference in expression of SASP-related genes, whose expression of immune-related factors and immune cell infiltration were both at a lower level, indicating that this subtype is in an immunosuppressive state. It can be hypothesized that the difference in expression of SASP factors to some extent makes the two subtypes have different immune microenvironment, which involves the regulation of inflammatory factors, crosstalk of related pathways and gene mutations. The specific mechanism needs to be further investigated.

In addition to the stage of tumor development, senescence also affects the therapeutic response of diseases through its complex mechanism in the treatment stage. Antitumor therapies, including chemotherapy and radiotherapy, can induce cellular senescence in tumor tissues and normal tissues and cause the accumulation of senescent cells (Roberson et al., 2005), namely, therapy-induced

senescence (TIS). Senescence-related markers such as p16INK4a, p21, p53 and SA-β-gal can be detected in the tumor tissues of breast cancer patients receiving chemotherapy (te Poele et al., 2002). Another study (Sanoff et al., 2014) showed that cytotoxic chemotherapy induced cellular senescence in hematopoietic tissues of breast cancer patients and prolonged the increase in SASP factors VEGFA and CCL2 levels. In addition, targeted therapies can induce tumor cell senescence. CDK4/6 inhibitors (Goel et al., 2016) can reverse the resistance of HER2-positive breast cancer to anti-HER2-targeted therapy and induce tumor cells to enter cell cycle arrest and have a senescent cell phenotype by inhibiting Rb and S6RP activity. The above evidence suggests that TIS is one of the mechanisms by which antineoplastic therapy works and that cellular senescence is an outcome produced by antineoplastic therapy. However, related studies have shown that TIS cells can produce SASP with cancer-promoting and immunosuppressive functions, which affects the clinical outcome of patients. In non-small cell lung cancer (Wang et al., 2013), TIS is associated with lower overall survival. In the p53 wild-type breast cancer model (Jackson et al., 2012), chemotherapy-induced cell senescence causes tumor regression, and the accumulated senescent cells can secrete cancer-promoting SASP, leading to early recurrence of breast cancer. After neoadjuvant chemotherapy (Morales-Valencia et al., 2022), the exposure of residual breast cancer cells to SASP can lead to the upregulation of intracellular LCN2, enhance tumor invasiveness and is related to tumor chemotherapy resistance. In another study (Muñoz et al., 2019), therapy-induced senescent breast cancer cells could evade immune clearance by paracrine inhibition of NKG2D receptor-mediated immune surveillance. In conclusion, TIS has both antitumor and protumor effects in tumor treatment. Therefore, more studies are needed to quantify the degree of cellular senescence and further evaluate the long-term role of cellular senescence in antitumor therapy.

Current studies have shown that senescence plays an important and complex role in tumors. Therefore, it is necessary to develop new therapeutic regimens for senescence-related targets and incorporate related therapies into future antitumor treatment strategies. In our study, triple-negative breast cancer patients with different senescence subtypes had different drug sensitivities, and small molecule drugs such as the VEGFR inhibitor AMG.706, PPM1D inhibitor CCT007093, and GSK-3α/β inhibitor CHR.99021 were more effective in the TNBCSASP1 subtype. It is expected that senescence-related genes will be used as targets for the development of new drugs in the future. At present, therapies targeting senescent cells have been gradually studied, including the elimination of senescent cells and regulation of senescent cell phenotype. Senolytic drugs are a class of drugs that eliminate senescent cells. Drugs such as ABT-737 and ABT-263 can induce senescent cells to initiate an apoptotic program by inhibiting the activity of BCL-2 family members. *In vivo* experiments have shown (Leverson et al., 2015) that the addition of the BCL-2 family inhibitor A-1331852 enhances the therapeutic effect of docetaxel in a triple-negative metastatic breast cancer model. Senomorphism plays a role by regulating the characteristics of senescent cells, including regulating the SASP and changing the state of cellular senescence. Targeting the mTOR pathway using rapamycin (Zhang et al., 2018) reduces the secretion of the protumorigenic SASP and prevents cellular senescence. Inhibitors of the p38MAPK/MK2 pathway can reduce SASP secretion, thereby inhibiting tumor metastasis in the TIS breast cancer mouse model (Murali et al., 2018). These two approaches to targeting senescence inhibit the tumor-promoting effects of senescence based on different mechanisms. In addition, some new senescence-related markers have also been identified in recent studies, which will help to develop new senescence-related therapeutic targets. Differentiated Embryonic Chondrocyte Gene 1 (DEC1) is one of the target genes of p53, which mediates cell senescence by regulating the phosphorylation of Rb2 protein and acting on the Rb pathway. *In vitro* experiments showed that overexpression of DEC1 could arrest the cell cycle and induce cell senescence. DEC1 downregulation can attenuate DNA damage-induced cellular senescence (Qian et al., 2008). During the progression from normal breast to breast cancer (Chakrabarti et al., 2004), increased expression of DEC1 was observed, suggesting that DEC1 may contribute to the progression of breast cancer to an aggressive phenotype. Human tumor necrosis factor-related apoptosis-inducing ligand receptor 4 (TRAFE-R4) is also one of the target genes of p53. p53 acts on TRAFE-R4 to inhibit its induction of apoptosis. It has been found (Ganten et al., 2009) that breast cancer patients with high expression of TRAFE-R4 have reduced overall survival and disease-free survival. Therapeutic regimens targeting senescence have potential application value, and the development of related drugs still needs to be further verified by experiments.

FAM3B plays an important role in tumor development, disease prognosis, and drug resistance. He et al. (2019) demonstrated that the expression of FAM3B in human esophageal squamous cell carcinoma (ESCC) was higher than that in adjacent tissues, its expression level was significantly related to the clinical stage of ESCC patients, and the high expression level of FAM3B was related to the poor prognosis of patients. Further studies confirmed that FAM3B overexpression could inhibit ESCC cell death by regulating the AKT-MDM2-p53 signaling pathway and affect the epithelial-mesenchymal transition process by regulating Snail and E-cadherin to promote ESCC cell migration and invasion. Cisplatin is one of the main chemotherapy drugs for patients with gastric cancer, but it easily leads to drug resistance during treatment. Song and Duan (2020) found that FAM3B expression was upregulated in cisplatin-resistant gastric cancer cell lines, and FAM3B overexpression affected the epithelial-mesenchymal transition process and induced drug resistance in gastric cancer cells by regulating Snail. Knockdown of FAM3B increased the drug sensitivity of drug-resistant cell lines. Alternatively, the potential role of FAM3B in tumor suppression has been reported. Liao et al. (Liao et al., 2022) used public databases to explore the feasibility of the FAM3 family as a prognostic factor for head and neck squamous cell carcinoma, and the analysis results showed that the low expression level of FAM3B in head and neck squamous cell carcinoma was related to poor prognosis of patients. FAM3B may be related to the increase in immune cell infiltration in the tumor microenvironment, the inhibition of the epithelial-mesenchymal transition process and cytochrome P450 and other targets. In a study on bladder cancer, FAM3B was listed as one of the immune-related genes involved in the construction of a prognosis-related index for bladder cancer, indirectly indicating the association of FAM3B with tumor immunity (Tian et al., 2020).

This study still has some limitations. First, the sample size in this report is limited, and the findings are not fully representative of the overall population of triple-negative breast cancer. Second,

the results of this study are based on multiomics bioinformatics analysis, and further experimental verification is needed to explore the specific mechanism of senescence. In addition, the classification model obtained in this study may be affected by some confounding factors, such as race and region, so more independent datasets are needed to validate the classification model. This retrospective study of FAM3B is based on the clinical data of a single center. Complete follow-up data is limited, and there is a bias in population selection. In addition, this study only preliminarily described the expression of FAM3B in triple-negative breast cancer and its effect on prognosis, and the specific mechanism needs to be explored by basic experiments to evaluate the reliability of the results of this study.

Therefore, based on the expression profile of senescence-related secretory phenotype genes, this study created a classification model for TNBC and explored the differential biomarkers of senescence subtypes, which can provide theoretical guidance for the treatment of TNBC.

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 human participants were reviewed and approved by the Ethical Committee of Ruijin Hospital, Shanghai Jiao Tong University School of Medicine. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

KS, XC, ZW, and RH mainly conceived and designed the article. RH, HW, JHo, JW, OH, JHe, WC, and YL collected the clinical samples for tissue microarray. RH, HW, JHo, and ZW analyzed the data and made manuscript draft. KS revised the final version. All authors listed in the paper have contributed to the study and approved the final submitting.

Funding

This work was supported by the National Natural Science Foundation of China (82002773, 82072897, and 82072937), Innovative Research Team of High-Level Local Universities in Shanghai (SHSMU-ZDCX20212200), and Shanghai Municipal Education Commission-Gaofeng Clinical Medicine Grant Support (20172007).

Acknowledgments

We must show sincere thanks two pathologists in our hospital, Anqi Li and Miao Ruan, for their analysis of the results of the tissue microarray.

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/fphar.2023.1191910/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Validation of SASP subtyping by different external datasets. **(A)** Heatmap of SASP subtype-specific signature genes in the external dataset GSE103091. **(B)** Kaplan–Meier survival curves of the TNBCSASP subtypes in the external dataset GSE103091.

SUPPLEMENTARY FIGURE S2
Patients' inclusion and exclusion flow chart.

References

Blum, A., Wang, P., and Zenklusen, J. C. (2018). SnapShot: TCGA-analyzed tumors. *Cell* 173, 530. doi:10.1016/j.cell.2018.03.059

Campisi, J., and D'Adda di Fagagna, F. (2007). Cellular senescence: When bad things happen to good cells. *Nat. Rev. Mol. Cell Biol.* 8, 729–740. doi:10.1038/nrm2233

Chakrabarti, J., Turley, H., Campo, L., Han, C., Harris, A. L., Gatter, K. C., et al. (2004). The transcription factor DEC1 (stra13, SHARP2) is associated with the hypoxic response and high tumour grade in human breast cancers. *Br. J. Cancer* 91, 954–958. doi:10.1038/sj.bjc.6602059

Cokelaer, T., Chen, E., Iorio, F., Menden, M. P., Lightfoot, H., Saez-Rodriguez, J., et al. (2018). GDSCTools for mining pharmacogenomic interactions in cancer. *Bioinforma. Oxf. Engl.* 34, 1226–1228. doi:10.1093/bioinformatics/btx744

Ganten, T. M., Sykora, J., Koschny, R., Batke, E., Aulmann, S., Mansmann, U., et al. (2009). Prognostic significance of tumour necrosis factor-related apoptosis-inducing ligand (TRAIL) receptor expression in patients with breast cancer. *J. Mol. Med. (Berlin, Ger.)* 87, 995–1007. doi:10.1007/s00109-009-0510-z

Garrido-Castro, A. C., Lin, N. U., and Polyak, K. (2019). Insights into molecular classifications of triple-negative breast cancer: Improving patient selection for treatment. *Cancer Discov.* 9, 176–198. doi:10.1158/2159-8290.CD-18-1177

Geeleher, P., Cox, N., and Huang, R. S. (2014). pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PloS One* 9, e107468. doi:10.1371/journal.pone.0107468

Goel, S., Wang, Q., Watt, A. C., Tolaney, S. M., Dillon, D. A., Li, W., et al. (2016). Overcoming therapeutic resistance in HER2-positive breast cancers with CDK4/6 inhibitors. *Cancer Cell* 29, 255–269. doi:10.1016/j.ccell.2016.02.006

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

Hatzis, C., Pusztai, L., Valero, V., Booser, D. J., Esserman, L., Lluch, A., et al. (2011). A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. *JAMA* 305, 1873–1881. doi:10.1001/jama.2011.593

He, S.-L., Wang, W.-P., Yang, Y.-S., Li, E.-M., Xu, L.-Y., and Chen, L.-Q. (2019). FAM3B promotes progression of oesophageal carcinoma via regulating the AKT-MDM2-p53 signalling axis and the epithelial-mesenchymal transition. *J. Cell. Mol. Med.* 23, 1375–1385. doi:10.1111/jcmm.14040

Herranz, N., and Gil, J. (2018). Mechanisms and functions of cellular senescence. *J. Clin. Invest.* 128, 1238–1246. doi:10.1172/JCI95148

Huang, R., Yang, Z., Liu, Q., Liu, B., Ding, X., and Wang, Z. (2022). CircRNA DDX21 acts as a prognostic factor and sponge of miR-1264/QKI axis to weaken the progression of triple-negative breast cancer. *Clin. Transl. Med.* 12, e768. doi:10.1002/ctm2.768

Jackson, J. G., Pant, V., Li, Q., Chang, L. L., QuintáS-Cardama, A., Garza, D., et al. (2012). p53-mediated senescence impairs the apoptotic response to chemotherapy and clinical outcome in breast cancer. *Cancer Cell* 21, 793–806. doi:10.1016/j.ccr.2012.04.027

JéZéQUEL, P., Loussouarn, D., GuéRIN-Charbonnel, C., Campion, L., Vanier, A., Gouraud, W., et al. (2015). Gene-expression molecular subtyping of triple-negative breast cancer tumours: Importance of immune response. *Breast cancer Res. BCR* 17, 43. doi:10.1186/s13058-015-0550-y

KnüPFER, H., and Preiss, R. (2007). Significance of interleukin-6 (IL-6) in breast cancer (review). *Breast Cancer Res. Treat.* 102, 129–135. doi:10.1007/s10549-006-9328-3

Leverson, J. D., Phillips, D. C., Mitten, M. J., Boghaert, E. R., Diaz, D., Tahir, S. K., et al. (2015). Exploiting selective BCL-2 family inhibitors to dissect cell survival dependencies and define improved strategies for cancer therapy. *Sci. Transl. Med.* 7, 279ra40. doi:10.1126/scitranslmed.aaa4642

Liao, C., Wang, Q., An, J., Wang, H., Xiao, L., Long, Q., et al. (2022). FAM3 family as prognostic factors for head and neck squamous cell carcinoma. *Comb. Chem. High Throughput Screen.* 26, 539–558. doi:10.2174/1386207325666220509191153

Liberzon, A., Subramanian, A., Pinchback, R., ThorvaldsdóTTIR, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinforma. Oxf. Engl.* 27, 1739–1740. doi:10.1093/bioinformatics/btr260

LóPEZ-OtíN, C., Blasco, M. A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. *Cell* 153, 1194–1217. doi:10.1016/j.cell.2013.05.039

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

Lu, C., Wang, Y., Nie, L., Chen, L., Li, M., Qing, H., et al. (2022). Comprehensive analysis of cellular senescence-related genes in the prognosis, tumor microenvironment, and immunotherapy/chemotherapy of clear cell renal cell carcinoma. *Front. Immunol.* 13, 934243. doi:10.3389/fimmu.2022.934243

Luo, Y., Liu, H., Fu, H., Ding, G.-S., and Teng, F. (2022). A cellular senescence-related classifier based on a tumorigenesis- and immune infiltration-guided strategy can predict prognosis, immunotherapy response, and candidate drugs in hepatocellular carcinoma. *Front. Immunol.* 13, 974377. doi:10.3389/fimmu.2022.974377

Marin, I., Boix, O., Garcia-Garijo, A., Sirois, I., Caballe, A., Zarzuela, E., et al. (2023). Cellular senescence is immunogenic and promotes antitumor immunity. *Cancer Discov.* 13, 410–431. doi:10.1158/2159-8290.CD-22-0523

Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28, 1747–1756. doi:10.1101/gr.239244.118

Morales-Valencia, J., Lau, L., Martí-Nin, T., Ozerdem, U., and David, G. (2022). Therapy-induced senescence promotes breast cancer cells plasticity by inducing Lipocalin-2 expression. *Oncogene* 41, 4361–4370. doi:10.1038/s41388-022-02433-4

MuñOZ, D. P., Yannone, S. M., Daemen, A., Sun, Y., Vakar-Lopez, F., Kawahara, M., et al. (2019). Targetable mechanisms driving immunoevasion of persistent senescent cells link chemotherapy-resistant cancer to aging. *JCI insight* 5, e124716. doi:10.1172/jci.insight.124716

Murali, B., Ren, Q., Luo, X., Faget, D. V., Wang, C., Johnson, R. M., et al. (2018). Inhibition of the stromal p38MAPK/MK2 pathway limits breast cancer metastases and chemotherapy-induced bone loss. *Cancer Res.* 78, 5618–5630. doi:10.1158/0008-5472.CAN-18-0234

Onyema, O. O., Decoster, L., Njemini, R., Forti, L. N., Bautmans, I., De Waele, M., et al. (2015). Chemotherapy-induced changes and immunosenescence of CD8+ T-cells in patients with breast cancer. *Anticancer Res.* 35, 1481–1489.

Prieto, L. I., and Baker, D. J. (2019). Cellular senescence and the immune system in cancer. *Gerontology* 65, 505–512. doi:10.1159/000500683

Qian, Y., Zhang, J., Yan, B., and Chen, X. (2008). DEC1, a basic helix-loop-helix transcription factor and a novel target gene of the p53 family, mediates p53-dependent premature senescence. *J. Biol. Chem.* 283, 2896–2905. doi:10.1074/jbc.M708624200

Roberson, R. S., Kussick, S. J., Vallieres, E., Chen, S.-Y. J., and Wu, D. Y. (2005). Escape from therapy-induced accelerated cellular senescence in p53-null lung cancer cells and in human lung cancers. *Cancer Res.* 65, 2795–2803. doi:10.1158/0008-5472.CAN-04-1270

Rodier, F., and Campisi, J. (2011). Four faces of cellular senescence. *J. Cell Biol.* 192, 547–556. doi:10.1083/jcb.201009094

Sabatier, R., Finetti, P., Cervera, N., Lambaudie, E., Esterni, B., Mamessier, E., et al. (2011). A gene expression signature identifies two prognostic subgroups of basal breast cancer. *Breast Cancer Res. Treat.* 126, 407–420. doi:10.1007/s10549-010-0897-9

Sanoff, H. K., Deal, A. M., Krishnamurthy, J., Torrice, C., Dillon, P., Sorrentino, J., et al. (2014). Effect of cytotoxic chemotherapy on markers of molecular age in patients with breast cancer. *J. Natl. Cancer Inst.* 106, dju057. doi:10.1093/jnci/dju057

Serrano, M., Lin, A. W., Mccurrach, M. E., Beach, D., and Lowe, S. W. (1997). Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* 88, 593–602. doi:10.1016/s0092-8674(00)81902-9

Song, C., and Duan, C. (2020). Upregulation of FAM3B promotes cisplatin resistance in gastric cancer by inducing epithelial-mesenchymal transition. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* 26, e921002. doi:10.12659/MSM.921002

Tan, C., Wei, Y., Ding, X., Han, C., Sun, Z., and Wang, C. (2022). Cell senescence-associated genes predict the malignant characteristics of glioblastoma. *Cancer Cell Int.* 22, 411. doi:10.1186/s12935-022-02834-1

Te Poele, R. H., Okorokov, A. L., Jardine, L., Cummings, J., and Joel, S. P. (2002). DNA damage is able to induce senescence in tumor cells *in vitro* and *in vivo*. *Cancer Res.* 62, 1876–1883.

Tian, Z., Meng, L., Long, X., Diao, T., Hu, M., Wang, M., et al. (2020). Identification and validation of an immune-related gene-based prognostic index for bladder cancer. *Am. J. Transl. Res.* 12, 5188–5204.

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). The cancer genome Atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol. (Poznan, Pol.)* 19, A68–A77. doi:10.5114/wo.2014.47136

Wang, Q., Wu, P. C., Dong, D. Z., Ivanova, I., Chu, E., Zeliadt, S., et al. (2013). Polyploidy road to therapy-induced cellular senescence and escape. *Int. J. Cancer* 132, 1505–1515. doi:10.1002/ijc.27810

Wang, Y., Ding, Y., Guo, N., and Wang, S. (2019). MDSCs: Key criminals of tumor pre-metastatic niche formation. *Front. Immunol.* 10, 172. doi:10.3389/fimmu.2019.00172

Wang, X., Ma, L., Pei, X., Wang, H., Tang, X., Pei, J.-F., et al. (2022). Comprehensive assessment of cellular senescence in the tumor microenvironment. *Briefings Bioinforma.* 23, bbac118. doi:10.1093/bib/bbac118

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinforma. Oxf. Engl.* 26, 1572–1573. doi:10.1093/bioinformatics/btq170

Xu, L., Deng, C., Pang, B., Zhang, X., Liu, W., Liao, G., et al. (2018). Tip: A web server for resolving tumor immunophenotype profiling. *Cancer Res.* 78, 6575–6580. doi:10.1158/0008-5472.CAN-18-0689

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

Zhang, B., Fu, D., Xu, Q., Cong, X., Wu, C., Zhong, X., et al. (2018). The senescence-associated secretory phenotype is potentiated by feedforward regulatory mechanisms involving Zscan4 and TAK1. *Nat. Commun.* 9, 1723. doi:10.1038/s41467-018-04010-4

Zhao, Q., Hu, W., Xu, J., Zeng, S., XI, X., Chen, J., et al. (2022). Comprehensive pan-cancer analysis of senescence with cancer prognosis and immunotherapy. *Front. Mol. Biosci.* 9, 919274. doi:10.3389/fmolb.2022.919274

Zinger, A., Cho, W. C., and Ben-Yehuda, A. (2017). Cancer and aging - the inflammatory connection. *Aging Dis.* 8, 611–627. doi:10.14336/AD.2016.1230

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 24 May 2023
DOI 10.3389/fphar.2023.1191129

OPEN ACCESS

EDITED BY
Linhui Wang,
Second Military Medical University, China

REVIEWED BY
Yuexiang Wang,
Chinese Academy of Sciences (CAS), China
Zheng Wang,
Shanghai Jiao Tong University, China

*CORRESPONDENCE
Weide Zhong,
zhongwd2009@live.cn
Shumin Fang,
Fangshumin20160104@163.com

RECEIVED 21 March 2023
ACCEPTED 25 April 2023
PUBLISHED 24 May 2023

CITATION
Zeng J, Chen J, Li M, Zhong C, Liu Z, Wang Y, Li Y, Jiang F, Fang S and Zhong W (2023), Integrated high-throughput analysis identifies super enhancers in metastatic castration-resistant prostate cancer.
Front. Pharmacol. 14:1191129.
doi: 10.3389/fphar.2023.1191129

Integrated high-throughput analysis identifies super enhancers in metastatic castration-resistant prostate cancer

Jie Zeng[1], Jiahong Chen[2], Maozhang Li[2], Chuanfan Zhong[3], Zezhen Liu[4], Yan Wang[1], Yuejiao Li[1], Funeng Jiang[1], Shumin Fang[2]* and Weide Zhong[1]*

[1]Department of Urology, The Second Affiliated Hospital, School of Medicine, South China University of Technology, Guangzhou, Guangdong, China, [2]Department of Urology, Huizhou Municipal Central Hospital, Huizhou, Guangdong, China, [3]Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong, China, [4]Department of Urology, Minimally Invasive Surgery Center, The First Affiliated Hospital of Guangzhou Medical University, and Guangdong Key Laboratory of Urology, Guangzhou, Guangdong, China

Background: Metastatic castration-resistant prostate cancer (mCRPC) is a highly aggressive stage of prostate cancer, and non-mutational epigenetic reprogramming plays a critical role in its progression. Super enhancers (SE), epigenetic elements, are involved in multiple tumor-promoting signaling pathways. However, the SE-mediated mechanism in mCRPC remains unclear.

Methods: SE-associated genes and transcription factors were identified from a cell line (C4-2B) of mCRPC by the CUT&Tag assay. Differentially expressed genes (DEGs) between mCRPC and primary prostate cancer (PCa) samples in the GSE35988 dataset were identified. What's more, a recurrence risk prediction model was constructed based on the overlapping genes (termed SE-associated DEGs). To confirm the key SE-associated DEGs, BET inhibitor JQ1 was applied to cells to block SE-mediated transcription. Finally, single-cell analysis was performed to visualize cell subpopulations expressing the key SE-associated DEGs.

Results: Nine human TFs, 867 SE-associated genes and 5417 DEGs were identified. 142 overlapping SE-associated DEGs showed excellent performance in recurrence prediction. Time-dependent receiver operating characteristic (ROC) curve analysis showed strong predictive power at 1 year (0.80), 3 years (0.85), and 5 years (0.88). The efficacy of his performance has also been validated in external datasets. In addition, FKBP5 activity was significantly inhibited by JQ1.

Conclusion: We present a landscape of SE and their associated genes in mCPRC, and discuss the potential clinical implications of these findings in terms of their translation to the clinic.

KEYWORDS

prostate cancer, super enhancer, CUT&Tag, mCRPC, transcription factor

Introduction

Prostate cancer (PCa) is a prevalent malignant tumor affecting the male genitourinary system (Siegel, et al., 2021). According to the World Health Organization (WHO) 2020 GLOBOCAN statistics, it ranks second in incidence among malignancies in men globally, after lung cancer, and has the fifth highest mortality rate among all malignancies in men (Sung, et al., 2021). Progression from local PCa to castration-resistant prostate cancer (CRPC) is inevitable (Davies, et al., 2019). Approximately 35% of patients with early-stage localized cancer and 50% of locally advanced prostate cancer have recurrence and metastasis (Djavan, et al., 2003). And the 5-year survival rate of metastatic prostate cancer is 28% (Nandana and Chung, 2014).

Epigenetic programming has emerged as a critical step in the activation and maintenance of aberrant transcriptional programs in CRPC pathogenesis (Cimadamore, et al., 2017; Yegnasubramanian, et al., 2019; Sugiura, et al., 2021). Recent studies have shown that hypermethylation of androgen receptor (AR) leads to loss of AR expression in CRPC patients (up to 30%) (Suzuki, et al., 2003; Chmelar, et al., 2007), driving CRPC progression. Epigenetic regulatory heterogeneity could lead to intratumoral phenotypic plasticity (Ateeq, et al., 2016). Phenotypic plasticity increases the probability of tumor cells successfully migrating to and surviving in different metastatic environments (Klein, 2013). In addition, nonmutational epigenetic reprogramming, an emerging feature, has been added to the list of hallmarks of cancer (Hanahan, 2022).

Super enhancers (SEs) are important elements of epigenetic regulation (Hah, et al., 2015; Thandapani, 2019), the concept of which was first proposed by Professor Young R.A. (Whitehead Institute for Biomedical Research) in 2013. SEs are a large cluster of active transcriptional enhancers spanning a long-range region of genomic DNA. SE binding sites are occupied by high-density transcription factors (TFs), coactivators (mediators), and histone modification marks (Ing-Simmons, et al., 2015; Sengupta and George, 2017). Compared to typical enhancers (TEs), which only recruit one TF, SEs recruit multiple TFs to one site and drive stronger transcriptional activation. Therefore, SEs participate in multiple signaling pathways and facilitate tumor-promoting gene changes (Hnisz, et al., 2015). SEs are essential for maintaining the stemness of embryonic stem cells (Whyte, et al., 2013) and maintaining tumor characteristics by facilitating special gene expression patterns (Hnisz, et al., 2013; Thandapani, 2019).

Specifically, SEs facilitate the dysregulation of transcriptional programs mediated by BRD4 (Urbanucci and Mills, 2018), CDK7, ERG, and other factors in PCa cells. BRD4, a member of the bromodomain and extraterminal domains (BETs) family (Chen et al., 2020), is a critical SE-related protein in PCa (Donati, et al., 2018; Shafran, et al., 2019). It acts as an epigenetic "reader" that binds to specific acetylated lysine residues on histone tails, facilitating the assembly of transcription complexes. BRD4 exhibits dense binding activity in SE and drives cell-identical gene expression (Lee, et al., 2017). In particular, BRD4 physically interacts with the N-terminal domain of AR, driving AR-mediated gene transcription. AR signaling remains the most common resistance mechanism in most CRPC patients (Dai, et al., 2017; Aurilio, et al., 2020). Moreover, BET inhibitors could resensitize drug-resistant CRPCs to enzalutamide (Shah, et al., 2017). The above evidence suggests that SE may be closely related to mCRPC.

To reveal the epigenetic dysregulation mechanism of mCRPC, SEs and TFs were first screened from C4-2B cells by CUT&Tag. We present a landscape of SE and their associated genes in mCPRC.

Materials and methods

Cell culture

The human PCa cell line C4-2B, a subline of human PCa LNCaP-derived C4-2 cells, was purchased from the BeNa Culture Collection (BNCC) (BNCC341733). Cells were cultured at 37°C in an atmosphere humidified with 20% O2 and 5% CO_2 and were cultivated in RPMI-1640 medium (MA0215, Meilunbio) containing 10% fetal bovine serum (A3160802, Gibco) and 1% (100 μg/mL) penicillin/streptomycin (15140-122, Gibco).

Cell line treatment conditions

To verify the hub SE-associated DEGs, C4-2B cells were treated with 500 nM and 2 μM JQ1 (CAS No.:1268524-70-4, MedChemExpree) for 24 h. Equal volumes of the carrier (DMSO) were used as control.

High-throughput CUT&Tag

The Cleavage Under Targets and Tag mentation (CUT&Tag) assay was performed as previously described (Kaya-Okur, et al., 2019). Briefly, 1×10^5 C4-2B cells were carefully washed twice with wash buffer (20 mM HEPES pH 7.5; 150 mM NaCl; 0.5 mM spermidine; 1× protease inhibitor cocktail). Ten microliters of concanavalin A-coated magnetic beads (Bangs Laboratories) were added to each sample and incubated at RT for 10 min. The unbound supernatant was removed, and bead-bound cells were resuspended in dig wash buffer (20 mM HEPES pH 7.5; 150 mM NaCl; 0.5 mM spermidine; 1× protease inhibitor cocktail; 0.05% digitonin; 2 mM EDTA) and a 1: 50 dilution of primary antibody (ab4729 for H3K27ac, Abcam; ab8895 for H3K4me1, Abcam). Then, the cells were incubated overnight at 4°C on a rotating platform. The primary antibody was removed using a magnet stand. Secondary antibody (goat monoclonal: Millipore AP132) was diluted 1:100 in digitonin wash buffer, and the cells were incubated at room temperature for 1 h. The cells were washed 3 times with a magnet stand in digitonin wash buffer. A 1:100 dilution of the pA-Tn5 adapter complex was prepared in Dig-med Buffer (0.01% digitonin; 20 mM HEPES pH 7.5; 300 mM NaCl; 0.5 mM spermidine; 1× protease inhibitor cocktail) and incubated with cells at room temperature for 1 h. The cells were washed 3× for 5 min in 1 mL Dig-med buffer, resuspended in tagmentation buffer (10 mM $MgCl_2$ in Dig-med Buffer) and incubated at 37°C for 1 h. DNA was purified using phenol–chloroform-isoamyl alcohol extraction and ethanol precipitation. For amplification of the libraries, 21 μL DNA was mixed with 2 μL of a universal i5 primer and a uniquely barcoded i7 primer. A volume of 25 μL NEBNext HiFi 2× PCR Master mix was added, and the

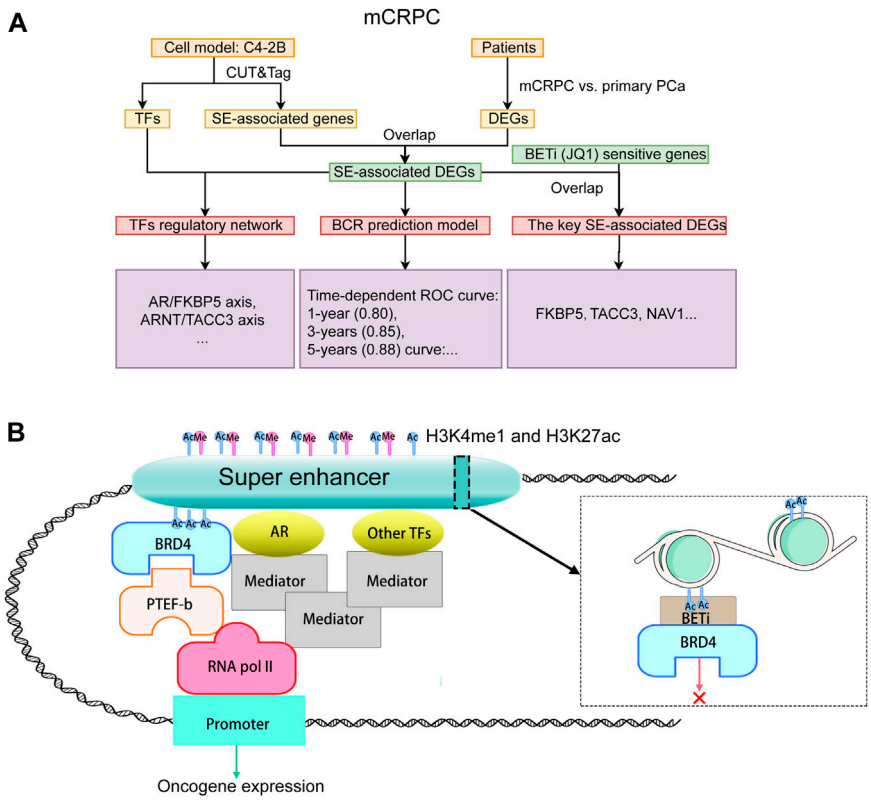


FIGURE 1
Study design and pathway diagram. **(A)** Study design. **(B)** Pathway diagram.

sample was mixed. The sample was placed in a thermocycler with a heated lid, and the following cycling conditions were applied: 72°C for 5 min (gap filling); 98°C for 30 s; 14 cycles of 98°C for 10 s and 63°C for 30 s; final extension at 72°C for 1 min and holding at 8°C. The library clean-up was performed with XP beads (Beckman Counter).

Identification of predicted SEs and TFs

H3 lysine 4 monomethylation (H3K4me1) and H3K27 acetylation (H3K27ac) (Creyghton, et al., 2010), both active histones, are associated with active enhancers and the presence of actively transcribing Pol II (McVicker, et al., 2013). Therefore, antibodies against active histones were used to bind target chromatin proteins between nucleosomes in the genome. Then, target peaks were detected in purified DNA by CUT&Tag (Kaya-Okur, et al., 2019). Active enhancers were those enriched in both H3K4me1 and H3K27ac (Li, et al., 2020). H3K4me1 can mark active or poised enhancers. H3K27ac distinguishes active enhancers from inactive/poised enhancer elements. Thus, the H3K4me1 peaks file identified by MACS2 and the H3K27ac BAM file were used as input to the algorithm firstly. Then, sort the enhancer according to the signal value of H3K27ac by SEs software (ROSE). The regions of SE or TE units and the 200 bp extending on both sides were used as input regions. Then, hypergeometric optimization of motif enrichment (HOMER) was used to predict enriched motifs. Detailed information about TFs was obtained from the Human TFs website (http://humantfs.ccbr.utoronto.ca/) (Lambert, et al., 2018). Information on over 1,600 human TFs has been deposited on this website.

Data acquisition

GSE35988 (Grasso, et al., 2012), containing mCRPC patients, was obtained from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo). Then, DEGs between mCRPC ($n = 27$) and localized PCa ($n = 49$) tissue were compared via the online tool PCaDB (http://bioinfo.jialab-ucr.org/PCaDB/). The analysis method was limma, and the thresholds were |fold change| > 2 and p-value <0.01.

The training cohort (Taylor) and the validation cohorts (TCGA-PRAD, DKZF, GSE54460, CancerMap, Cambridge, Belfast, CPC-Gene, and Stockholm) of SE-associated DEGs prognostic model were downloaded from PCaDB.

Single-cell RNA-seq of mCRPC tissue was performed by the Dana-Farber/Harvard Cancer Center Institutional Review Board with ethics approval and informed consent (He, et al., 2021).

Gene expression values of C4-2B, PC-3, LNCaP under JQ1 treatment were extracted from GSE98069 (Coleman, et al., 2019) using the edgeR package (Shen, et al., 2021).

Functional enrichment analysis

SE-associated DEGs are the intersection of SE-associated genes and DEGs of mCRPC. Enrichment analyses of SE-associated genes, DEGs and SE-associated DEGs were performed based on various online databases. Hallmark enrichment analysis based on The Molecular Signatures Database (MSigDB) was performed using gene set enrichment analysis (GSEA). In addition, genes were subjected to Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analysis. |Fold change|>1 and p-value < 0.05 were considered to indicate statistical significance.

Ratio model training

For SE-associtated DEGs prognostic model training, 94 upregulated SE-associated DEGs and 48 downregulated SE-associated DEGs were used as candidates; CoxRidge, a package for fitting Cox models with penalized ridge-type partial likelihood, was used as the method.

$$\text{SE} - \text{associated DEGs score} = \sum_{i=1}^{N} (\text{Coefficient}i \times \text{Expression level of mRNA}i)$$

Where "N" (N = 13) represents the total number of the SE-associated DEGs in the training model, "Coefficient$_i$" denotes a specific mRNA's coefficient of SE-associated genes, and "Expression level of mRNAi" refers to the relative expression level of a certain mRNA.

Given the median scores in Taylor dataset, the low risk group and high risk group were defined. The Kaplan-Meier (KM) survival analysis depicted the BCR-free survival probability curves between the low risk group and high risk group. ROC curves of 1-, 3-, and 5-year evaluate the predictive power of SE-associated DEGs score.

Subsequently, 8 independent validation cohorts (TCGA-PRAD, DKZF, GSE54460, CancerMap, Cambridge, Belfast, CPC-Gene, and Stockholm) were used for depicted the BCR-free survival probability, which are the classic PCa datasets currently retrievable. Forest plots (http://vip.sangerbox.com/home.html) were used for visualization.

RNA isolation and RT–qPCR

Total RNA was extracted from C4-2B cells using the RNA Quick Purification Kit (RN001; esunbio) according to the manufacturer's instructions. The RNA concentration was measured using a DS -11+ Spectrophotometer (DeNovix, United States). cDNA was synthesized from the extracted RNA by reverse transcription using HiScript II Q RT SuperMix for qPCR (+gDNA wiper) (R223; Vazyme Biotech) in the Veriti™ 96-Well Fast Thermal Cycler (catalog number: 4375305, Thermo Fisher Scientific). RT–qPCR was performed using ChamQ Universal SYBR qPCR Master Mix (Q711-02; Vazyme Biotech) and a QuantStudio™ 1 Real-Time PCR System (catalog number: A40427, Thermo Fisher Scientific). Primers were chemically synthesized by Tsingke Biotechnology Co., Ltd. (Guangzhou, China).

Primer sequences for amplification are listed in the supplemental information (Supplementary Table S9).

Interaction network analyses

TFs were predicted by HOMER's findMotifsGenome.pl tool based on the C4-2B CUT&Tag peak file. The interaction network between TFs and SE-associated DEGs was determined by utilizing the STRING (http://string-db.org) online database and then constructed using Cytoscape software.

Results

Flowchart of this study

The study design is illustrated in Figure 1A. To reveal the epigenetic dysregulation mechanism of mCRPC, super-enhancer (SE) associated genes and transcription factors were first screened from C4-2B cellsusing CUT&Tag assay. The differentially expressed genes related to mCRPC were identified from the GSE35988 dataset, and the SE-associated DEGs were obtained by taking the intersection of SE-associated genes and DEGs. Subsequently, a recurrence prediction model was constructed based on SE-associated DEGs, and a SE-mediated TF regulatory network was built using TFs and SE-associated DEGs. JQ1, a BET inhibit BCR or, was added to block SE-mediated TF transcription and identify key SE-associated DEGs, which were found to include FKBP5 and TACC3 genes. The hypothesis diagram depicted in Figure 1B suggests that H3K4me1 and H3K27ac histone marks can recruit active SEs, which can bind multiple TFs to drive stronger transcriptional activation.

Identification of SE-associated genes

SEs can be found according to H3K4me1 and H3K27ac histone marks. Specifically, H3K4me1 enrichment indicates regions related to poised or less active enhancers. Enrichment of H3K27ac is a marker of active regulatory elements, including enhancers and promoters. Therefore, to identify SEs of C4-2B, a CUT&Tag assay was performed with H3K4me1 and H3K27ac. H3K4me1 can mark active enhancers as well as those in a poised or predetermined state. And H3K27ac distinguishes active enhancers from inactive/poised enhancer elements containing H3K4me1 alone (Creyghton, et al., 2010).

In this study, we firstly evaluated the epigenetic landscape based on two active histone marks (H3K4me1 and H3K27ac). Annotations of the peaks showed that H3K4me1 modification was mainly found in intron (43.06%) and distal intergenic (23.06%) regions. In contrast, H3K27ac modification was mainly found in promoter (47.02%) and intron (32.2%) regions (Figure 2A). The H3K4me1 and H3K27ac marks showed the same profile surrounding the transcription start site (TSS) (Figures 2B, C). H3K4me1 and H3K27ac histone marks identify regions that likely contain enhancers. Importantly, SEs could be found in regions with H3K4me1 and H3K27ac histone marks. Moreover, the peaks of H3K4me1 and H3K27ac were highly colocalized in the promoter and distal intergenic regions (Figure 2D), suggesting that H3K4me1 and H3K27ac histone marks were also highly enriched in the promoter and distal intergenic regions.

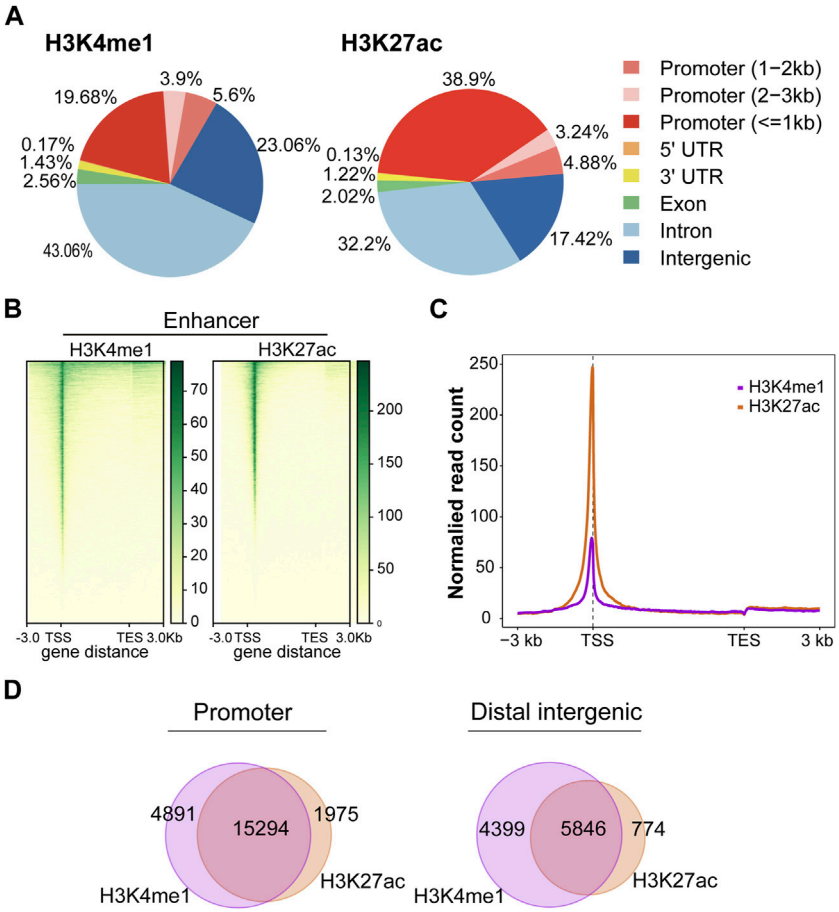


FIGURE 2
Active histone mark distributions in C4-2B cells. **(A)** The distribution of genomic regions modified by H3K4me1 and H3K27ac, which were classified into six region types (promoter, 5′UTR, 3′UTR, exon, intron, and intergenic). **(B)** Heatmap of active histone modifications detected in C4-2B cells. **(C)** Density profiles of H3K4me1 (purple) and H3K27ac (orange) at C4-2B. Profiles included 3 kb upstream of TSS and 3 kb downstream of TES. **(D)** The overlapping region between H3K4me1 and H3K27ac in promoters and distal intergenic regions.

Immediately afterward, we identified SEs and their associated genes. H3K4me1 is used to find out poised enhancer. Then, sort the enhancer according to the signal value of H3K27ac by SE software (ROSE) (Figure 3A). According to the rank ordering of ROSE, 867 activated SEs were identified, and the cutoff value was 13537.2441 (Figure 3B). These SEs could be annotated to 867 genes (Supplementary Table S1). Among them, there were 486 protein-coding RNAs, 68 ncRNAs, 50 pseudogenes and 1 snoRNA. In the present study, only protein-coding RNA was selected for further investigation. Therefore, we defined 486 SE-related protein-coding RNAs as SE-associated genes (Figure 3C). Hallmark enrichment analysis showed that C4-2B may be greatly influenced by the early estrogen response, P53 signaling pathway, and G2/M checkpoint (Figure 3D). KEGG analysis revealed that the SE-associated genes are involved in the Rap1 signaling pathway, tight junctions, the AMPK signaling pathway, and adherens junctions (Figure 3E). GO enrichment analysis confirmed the perturbation of Wnt signaling pathways (Figure 3F).

Overall, we not only screened 486 SE-associated genes related to mCRPC but also proposed a possible mechanism connecting mCRPC and the Rap1 signaling pathway, tight junctions, the AMPK signaling pathway, or adherens junctions.

Analysis of the differential expression pattern of mCRPC

The GSE35988 dataset contains information on 49 localized PCa patients and 27 mCRPC patients (Figure 4A) and was used for differential expression pattern analysis of mCRPC. First, we observed that localized PCa patients and mCRPC patients could be divided into two groups according to principal component analysis (PCA) (Figure 4B). Then, we identified 5,798 statistically significant DEGs between the two groups (Figure 4C; Supplementary Table S2). These DEGs of mCRPC could be divided into 5,417 protein-coding RNAs, 262 ncRNA, 111 pseudogenes, and 1 snoRNA. and 1 snRNA (Figure 4D). Subsequently, 5,417 protein-coding RNAs were used for enrichment analysis. In the hallmark enrichment analysis, DEGs were found to be enriched in epithelial-mesenchymal transformation, E2F targets, the G2/M checkpoint, and the

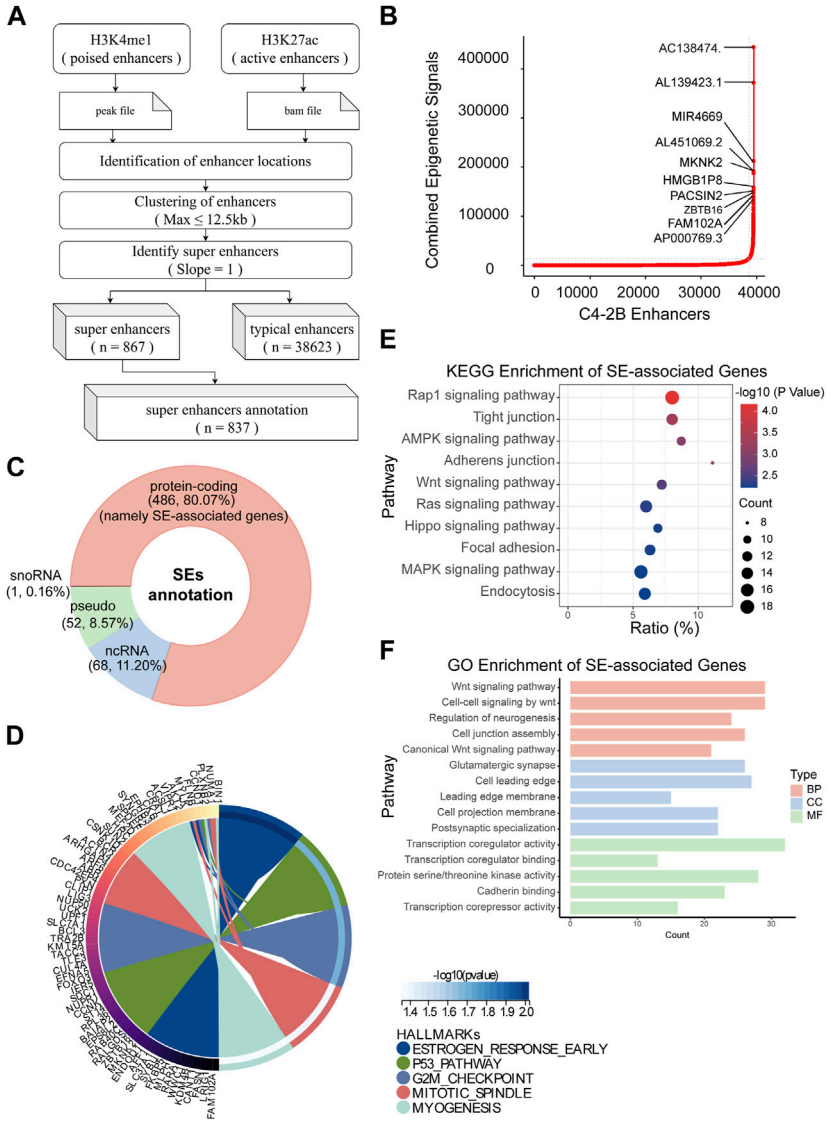


FIGURE 3
Identification of SEs in C4-2B cells. **(A)** The flow of SE identification. **(B)** Ranking of enhancers using the ROSE algorithm. **(C)** Classification of SE-related genes; protein-coding RNAs, namely, SE-associated genes, were the focus. **(D–F)** Hallmark, KEGG and GO enrichment analyses of SE-associated genes in C4-2B cells.

androgen response (Figure 4E). KEGG enrichment analysis showed that DEGs were involved in the MAPK-type pathway, focal adhesion and the Rap1 signaling pathway (Figure 4F). Moreover, GO enrichment analysis demonstrated that Neurogenesis was the main enriched term for DEGs in mCRPC (Figure 4G).

Application of SE-associated DEGs to predict biochemical recurrence (BCR) in PCa patients

SEs have been considered important epigenetic regulators. Therefore, we aimed to investigate the clinical significance of SEs in mCRPC. First, we took the intersection of 486 specific SE-associated genes, 2,418 upregulated DEGs and 3,269 downregulated DEGs. We obtained 94 upregulated SE-associated DEGs and 48 downregulated SE-associated DEGs (Figure 5A; Supplementary Table S3). These SE-associated DEGs were presumed to be involved in the progression. Thus, we performed hallmark, KEGG, and GO enrichment analyses of SE-associated DEGs. Hallmark enrichment analysis showed that estrogen response signaling and androgen response signaling were the most prominent signaling pathways (Figure 5B). KEGG enrichment analyses showed that the Hippo signaling pathway, MAPK signaling pathway, and Ras signaling pathway may be vital (Figure 5C). The GO enrichment analysis supported the Wnt signaling pathway result (Figure 5D).

BCR, which is characterized by an increase in prostate-specific antigen (PSA) after completed surgery (Moul, 2000) or radiation (Roach et al., 2006), was used to reflect the treatment effect. Here, we

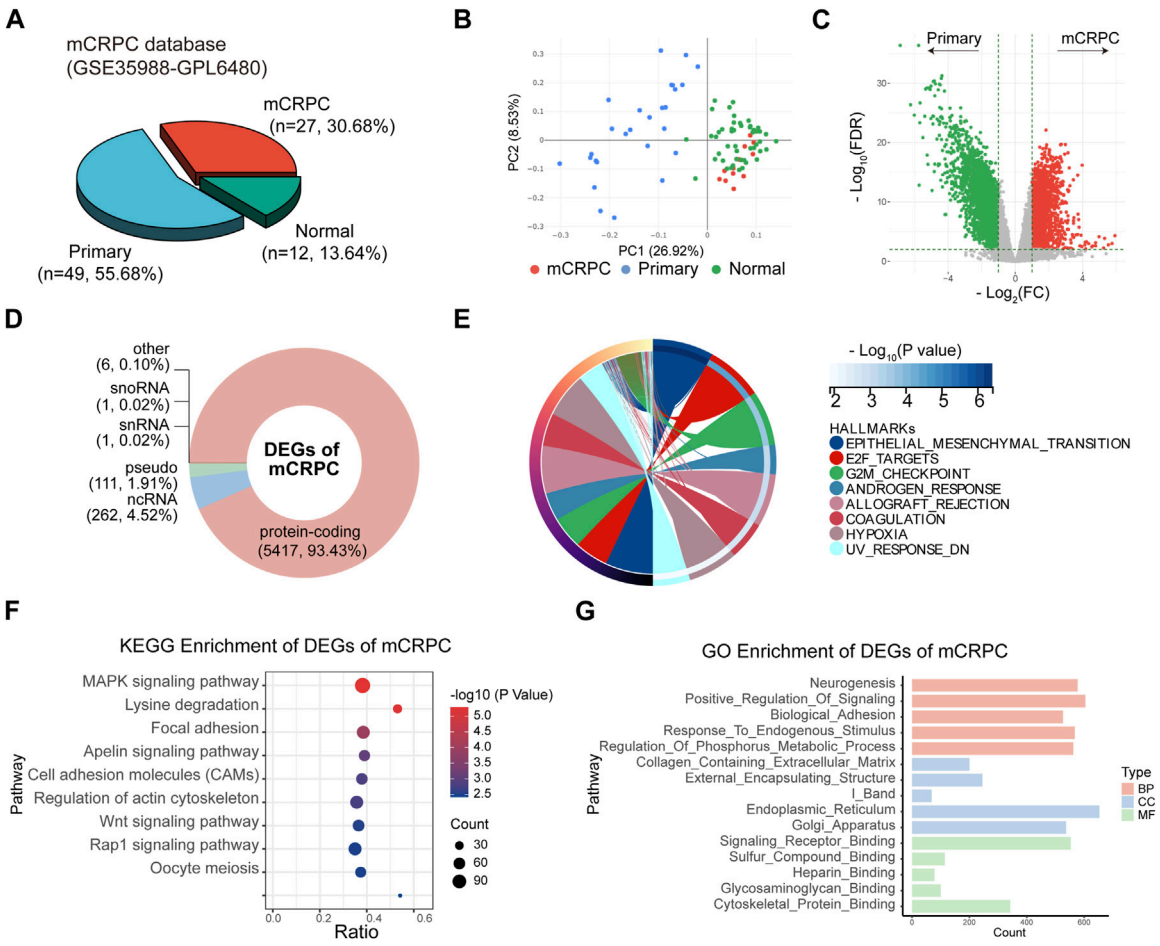


FIGURE 4
Identification of DEGs between mCRPC and primary PCa patients in GSE35988. **(A)** Composition of sample types in GSE35988 dataset. **(B)** Principal component analysis (PCA) of GSE35988 dataset. **(C)** Volcano plot of DEGs between mCRPC patients and primary PCa patients. ($p < 0.01$, |fold change| > 2, limma methods). **(D)** Classification of DEGs. **(E–G)** Hallmark, KEGG and GO enrichment analyses of DEGs of mCRPC.

established a BCR prediction model based on SE-associated DEGs. First, the above SE-associated DEGs were selected for training in the Taylor dataset using the CoxRidge. Kaplan–Meier survival analysis of biochemical recurrence-free survival (BRFS) showed a significantly worse prognosis in the high-risk group, implying that the high-risk group was likely to relapse earlier than the low-risk group in the Taylor cohort (HR = 4.98) (Figures 5E, F; Supplementary Tables S4, S5). Time-dependent ROC curve analysis revealed that the BCR model performed well in predicting outcomes at 1 year (0.80), 3 years (0.85), and 5 years (0.88) (Figure 5F). Consistently, the model of SE-associtated DEGs also predicted BRFS well in multiple established validation cohorts (Figure 5G, Supplementary Table S6), such as the TCGA-PRAD, DKZF, GSE54460, CancerMap, Cambridge, Belfast, CPC-Gene, and Stockholm cohorts.

TF regulatory network of mCRPC

Together, SEs, TFs, and multiple genes form a transcriptional regulatory loop. Notably, TFs in particular play a crucial role in SE-mediated transcriptional regulation. To explore the mechanism of epigenetic dysregulation of mCRPC, we predicted TFs to establish a TF regulatory network, which consisted of three layers of elements of regulation: SEs, TFs and SE-associated DEGs.

First, TFs of mCRPC were predicted by HOMER based on the DNA-binding motifs of C4-2B. HOMER motif discovery revealed 14 known motifs and 36 *de novo* motifs as the highest-scoring motifs ($p < 0.01$). Analysis of these known motifs revealed 3 human TFs: AR, NKRF and RFX2. Moreover, *de novo* motif analysis identified 6 probable human TFs, including NFIC, NKX2-5, SP2, AHR, ARNT and FOXL1. According to DNA-binding domain (DAB) classification, AR is a TF of nuclear receptors; RFX2 is a TF of regulatory factor binding to the X-box (RFX); NFIC is a TF of SMAD; NKX2-5 is a TF of homeodomain genes; SP2 is a TF of Cis2-His2 zinc finger (C2H2-ZF); AHR and ARNT are basic helix-loop-helix (bHLH) TFs; and FOXL1 is a forkhead box (FOX) TF (Figure 6A).

Subsequently, 3 human TFs revealed by known motif analysis and 6 human TFs identified by *de novo* motif analysis were found to interact with SE-associated DEGs using the STRING database and visualized by Cytoscape. The TF regulatory network of mCRPC was

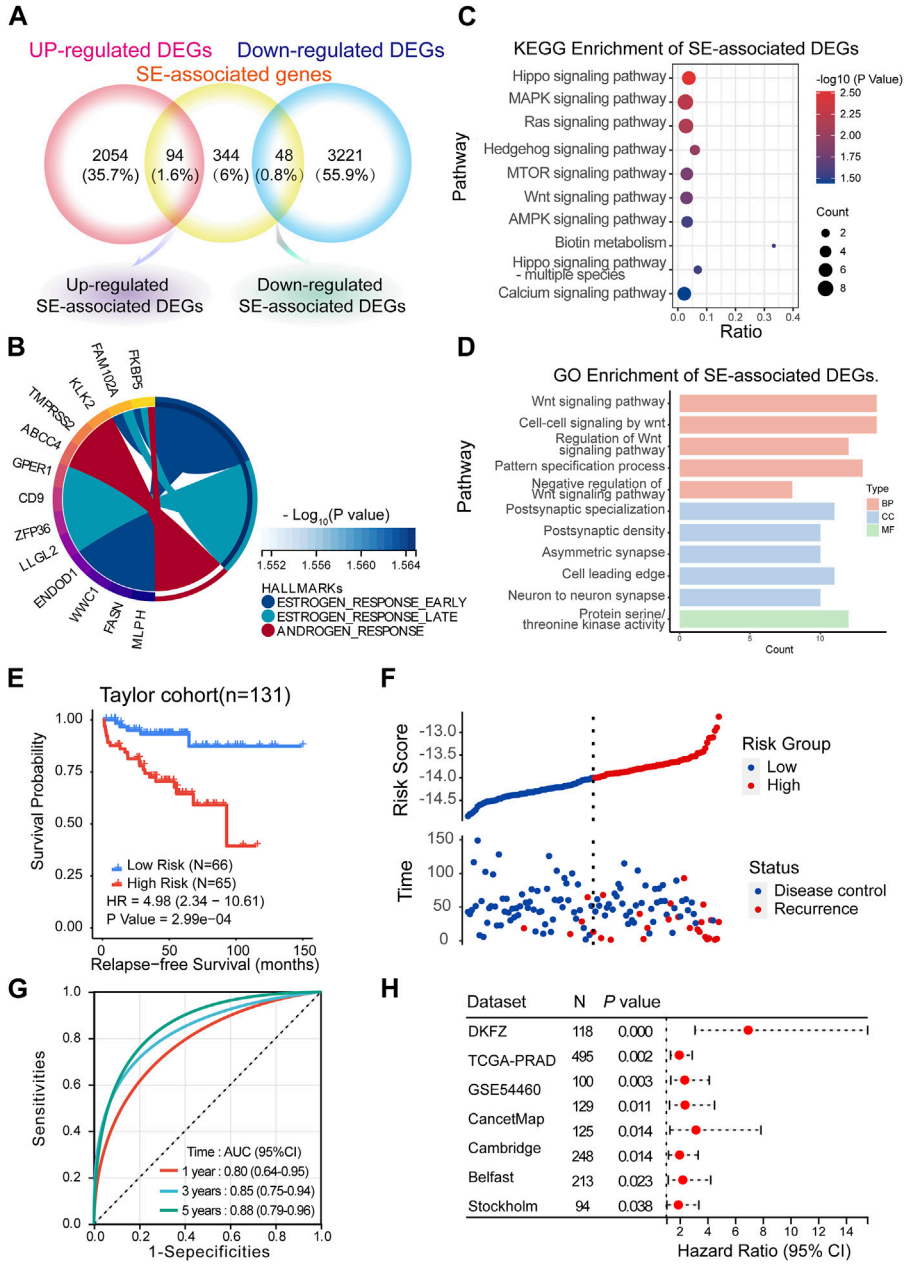


FIGURE 5
Predicting BCR based on the SE-associated DEGs model. **(A)** Screening of SE-associated DEGs. **(B–D)** Hallmark, KEGG and GO enrichment analyses of SE-associated DEGs. **(E,F)** Recurrence-free survival (RFS) in the training cohort (Taylor) (Method: CoxRidge). **(G)** ROC curves test the predictive value of the risk score in the training cohort. **(H)** Forest plot analysis of RFS in the validation cohort.

constructed from TFs and their potential regulatory SE-associated DEGs (Figure 6B).

Consistent with a previous report (Mulholland, et al., 2011), AR was associated with FKBP5 in CRPC. FKBP5 is considered to be an androgen-inducible gene that physically interacts with AR (Zheng, et al., 2015). Moreover, the ARNT/TACC3 axis was apparent in the TF regulatory network. An association study showed that TACC3 was markedly upregulated in CRPC (Qie, et al., 2020). In addition, the ARNT/TACC3 complex is active in a hypoxic environment (Guo, et al., 2013), a signature of the tumor environment. Thus, it is reasonable to assume that progression to mCRPC may be regulated by the AR/FKBP5 and ARNT/TACC3 axis.

Harnessing inhibitor sensitivity reveals hub genes

The bromodomain and extraterminal (BET) family, including BRD2, BRD3, and BRD4, which are partially enriched at termed SEs

A

Motif	TF	P-value	DBD	EntrezGene ID	PDB ID
	AR	1.0E03	Nuclear receptor	367	1R4I
	NKRF	1.0E02	Unknown	55922	None
	RFX2	1.0E02	RFX	5990	None
	NFIC	1.0E11	SMAD	4782	None
	NKX2-5	1.0E08	Homeodomain	1482	3RKQ
	SP2	1.0E06	C2HZF	6668	None
	AHR	1.0E03	bHLH	196	None
	ARNT	1.0E03	bHLH	405	4ZPR
	FOXL1	1.0E02	Forkhead	2300	None

B

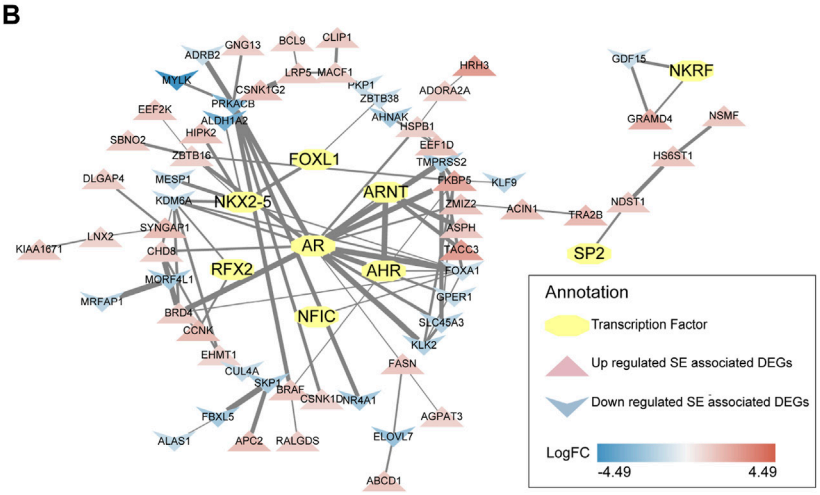


FIGURE 6
Predicted TFs and their interaction with SE-associated DEGs. **(A)** Detailed information of selected human TFs. **(B)** TFs regulatory network of mCRPC.

(Ribeiro, et al., 2012; Jin, et al., 2019), can regulate the activity of SEs and promote SE-mediated transcriptional regulatory programs (Shankar, et al., 2015). In other words, some SE-related genes could be affected by BET inhibitors. Surprisingly, JQ1, a known BET inhibitor, was reported to block the growth of PCa cells (Chen and Song, 2016; Coleman, et al., 2019). Therefore, JQ1 was used in the present study for further screening of SE-associated CRPC recurrence genes.

First, the publicly available dataset GSE98069 was searched and used to rapidly detect the JQ1-sensitive genes of C4-2B. Treatment with 500 nM JQ1 for 24 h, the appropriate dose and time period, significantly reduced proliferation (Urbanucci and Mills, 2018). A total of 1423 genes were upregulated and 1680 genes were downregulated in C4-2B (Figure 7A; Supplementary Table S7). Of SE-associated CRPC recurrence genes, a total of 10 highly expressed SE-associated CRPC recurrence genes were downregulated by JQ1 treatment (Figure 7B). Among them, TACC3, FKBP5 and NAV1 had the largest fold changes and greatest significant. Next, the intersecting genes among the top 10 SE-associated genes (in terms of fold change) and JQ1-sensitive genes, which are considered to be key SE-mediated regulatory genes, were selected for PCR validation. Strikingly, KLK3, the protein-coding RNA of AR, was selected as a positive control and has been reported to be downregulated by JQ1 in PCa cells (Chen and Song, 2016) (Figure 7C; Supplementary Table S8). Our RT–qPCR results were consistent with the RNA-Seq results of C4-2B cells treated with 500 nM JQ1. To further validate the expression changes of intersecting genes, we treated C4-2B cells with a higher dose of JQ1 (2 mM). These results indicated that even a high dose caused expression changes (Figure 7D).

To simplify, TACC3, FKBP5, and NAV1 were the most likely SE-mediated regulatory genes, and they were expressed at higher levels in mCRPC than in localized PCa. We further investigated the localization of TACC3, FKBP5, and NAV1 in single cells isolated from prostate tissue of CRPC patients. FKBP5 was obviously highly expressed in PCa cells. TACC3 was highly expressed in $CD4^+$ Treg cells and $CD14^+$ monocytes. NAV1 was mainly expressed in plasmablasts (Figures 7E, G). Surprisingly, we found that FKBP5 interacts closely with AR (Figure 7F). Strikingly, FKBP5 is considered to be an androgen-inducible gene that physically interacts with AR (Zheng, et al., 2015).

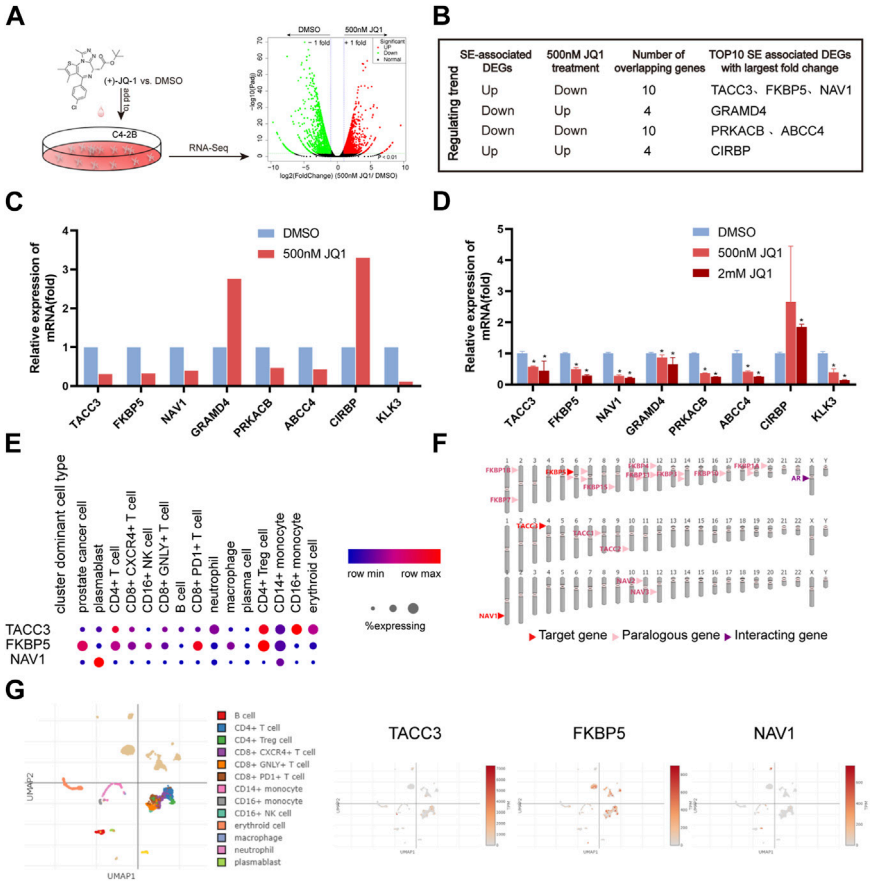


FIGURE 7
Verification of SE-associated CRPC recurrence genes. **(A)** Fast screening of BET inhibitor-sensitive SE-associated DEGs. **(B)** Selection of SE-associated CRPC recurrence genes. **(C)** Relative expression of selected genes based on GSE98069 (***$p < 0.001$, 500 nM JQ1 vs. DMSO). **(D)** Verification of selected genes by RT–qPCR (***$p < 0.001$, 500 nM or 2 mM JQ1 vs. DMSO). **(E–G)** Single-cell analysis of TACC3, FKBP5 and NAV1.

Discussion

The therapeutic options are limited in the stage of mCRPC patients. SEs, epigenetic regulators, have been implicated in tumorigenesis (Cucchiara, et al., 2017; Hankey, et al., 2020). Therefore, inhibtion of SEs might be a valid strategy in the treatment of patients with refractory mCRPC.

In this study, we characterized the landscape of active histone modifications in C4-2B cells representative of mCRPC. H3K4me1 and H3K27ac are known markers of active enhancers and important indicators of enhancer activity. In particular, H3K4me1 is located in many type-specific enhancer sites. Most enhancer regions labeled with H3K4me1 are active (Calmasini, et al., 2020). H3K27ac is a candidate for distinguishing active from inactive enhancer elements (Xu, et al., 2018) A total of 867 SEs were identified in C4-2B cells by detecting both H3K4me1-and H3K27ac-enriched regions (Xu, et al., 2018). Our data showed that SE-associated genes involved in the early estrogen response and G2/M checkpoint contribute to the pathogenesis of mCRPC. Analysis of mCRPC tissues also supported this conclusion. To further prove this hypothesis, we overlapped SE-associated genes in C4-2B cells *versus* normal control cells and DEGs in mCRPC *versus* primary PCa. The overlapping genes were involved in the estrogen response, androgen response, Hippo signaling pathway, MAPK signaling pathway, and so on. These genes are closely related to mCRPC.

Local therapies, including radical prostatectomy or primary definitive radiotherapy, are the initial treatments for primary PCa. However, many patients will eventually reach the mCRPC stage (Cornford, et al., 2021). BCR reflects the effects of treatment. Therefore, SE-associated DEGs, namely, SE-associated mCRPC recurrence genes, were used to construct the BCR prediction model. This model displayed a significantly excellent predictive power for mCRPC recurrence at the 1-year, 3-year and 5-year follow-up. This implies that mCRPC patients with high predictive scores might benefit from BET inhibitors.

SEs activate transcription by recruiting TFs. Therefore, we further sought TFs of mCRPC and then established a TF regulatory network. Reliable studies have summarized 1,639 known or probable human TFs (Lambert, et al., 2018). Fortunately, 9 human TFs were explored in C4-2B cell lines.

Undoubtedly, it has been widely assumed that AR signaling is related to CRPC biology (Watson, et al., 2015). The AR gene body and/or the enhancer are amplified in 81% of mCRPC cases (Quigley, et al., 2018; Zhao, et al., 2020). NF-κB repressing factor (NKRF), an inhibitor of NF-KB-mediated gene transcription, has been associated with tumor invasion, migration, and progression (Hsu, et al., 2012; Xu, et al., 2022). Regulatory factor X2 (RFX2) has been identified as the major master TF in regulating the angiogenesis signature in renal carcinoma (Zheng, et al., 2021). Furthermore, the RFX2/RFX3 complex could only be detected in the nuclear extract of FGF-1B-positive cells; it directly binds the 18-bp cis-element (−484 to −467) and contributes to the regulation of the fibroblast growth factor 1 (FGF1) promoter, which has been shown to regulate cell proliferation and cell division (Hsu, et al., 2012).

Acetylated chromatin, particularly in SE regions, is associated with bromodomain and extraterminal (BET) proteins to facilitate transcriptional activation (Li, et al., 2020). BET inhibitors disrupt the interaction of bromodomains with acetylated histones and result in the loss of enhancer-promoter long-range interactions. Several BET inhibitors are currently being investigated in phase I/II in mCRPC patients, including ZEN003694 (NCT02711956, NCT02705469), GS-5829(NCT02607228), and OTX015/MK-8628 (NCT02259114) (Zheng, et al., 2015). JQ1, a well-known BET inhibitor, was reported to block PCa cell growth (Chen and Song, 2016; Coleman, et al., 2019). Thus, JQ1 was used in the present study to further screen the hub SE-associated DEGs.

FKBP5 and TACC3 were still the key players deactivated by JQ1, while they were upregulated in mCRPC. Thus, we confirmed that FKBP5 and TACC3 might be the key candidates regulated by SEs in the pathogenic process of mCRPC. The TF regulatory network of mCRPC, the AR/FKBP5 axis and the ARNT/TACC3 axis were the focuses of our study. Consistent with a previous report (Mulholland, et al., 2011), AR was associated with FKBP5 in CRPC. Moreover, ARNT is a TF of basic helix-loop-helix/per-arnt-sim (PAS) family members, and predominantly heterodimerizes with the aryl hydrocarbon receptor (AHR) or hypoxia-inducible factor-1 alpha (HIF-1α), which has been linked to PCa angiogenesis (Fritz, et al., 2008). TACC3, which is involved in the pathogenesis of several cancers (Qie, et al., 2020), was markedly upregulated in CRPC. The ARNT/TACC3 complex is active in a hypoxic environment (Guo, et al., 2013). Our data suggest that CRPC progression may be regulated by the ARNT/TACC3 axis.

Single-cell analysis revealed that FKBP5 is mainly expressed in PCa cells of CRPC tissues. Therefore, FKBP5 might be a crucial SE-mediated gene in mCRPC. FKBP5, an androgen-inducible gene, physically interacts with AR (Zheng, et al., 2015). Recent studies have identified widespread activation of FKBP5 transcription in PCa cells by AR via distal intronic enhancers (Zheng, et al., 2015), which is related to chemoresistance in cancer (Li, et al., 2019). By contrast, little attention has been paid to the effects on ARNT/TACC3 axis in prostate cancer. Our previous study has demonstrated that CCR8-ARNT increased lactate production, and promoted aerobic glycolysis, which was related to poor outcomes of patients with advanced PCa (Chen G et al., 2020; Chen X et al., 2020). What's more, recent studies showed that ARNT was related to tumor heterogeneity (Watkins, et al., 2020) and resistance of enzalutamide (Zhang, et al., 2022). ARNT, also known as hypoxia-inducible factor-1β (HIF-1β), has been recognized as an oncoprotein that promotes tumor growth in response to hypoxia (Harris, 2002), including multiple myeloma (Hassen, et al., 2015; Huang, et al., 2021), lymphoid cancer (Gardella, et al., 2016) and melanoma (Leick, et al., 2019). ARNT PAS-B domain was reported to interact directly with TACC3, which is a necessary step for transcriptional responses to hypoxia (Partch and Gardner, 2011). TACC3 gene is a centrosomal protein that is, involved in mitotic spindle assembly and chromosome segregation. Recent studies have shown that TACC3 is over-expressed in prostate cancer, and is associated with tumor progression and poor prognosis. Overexpression of TACC3 in prostate cancer cells has been shown to promote cell proliferation and migration. Therefore, it is reasonable to speculate that AR/FKBP5 axis and ARNT/TACC3 axis might play important roles in the progression of mCRPC.

Conclusion

Our study provides valuable insights into the role of SEs in mCRPC development and suggests potential clinical applications for SEs.

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.

Author contributions

ML and WZ designed this work. JC, CZ and JZ integrated and analyzed the data. JZ wrote this manuscript. JC, ML, CZ, ZL, YW, YL, FJ, SF, and JZ edited and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the National Natural Science Foundation of China (No.82072813), Natural Science Foundation of Guangdong Province, China (No. 2022A1515010344), Huizhou High Level Hospital Construction Science and Technology Special Project (No. 2022CZ010004), and Guangzhou Municipal Science and Technology Project (202102080029, 202201020386).

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

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2023.1191129/full#supplementary-material

References

Ateeq, B., Bhatia, V., and Goel, S. (2016). Molecular discriminators of racial disparities in prostate cancer. *Trends cancer* 2 (3), 116–120. doi:10.1016/j.trecan.2016.01.005

Aurilio, G., Cimadamore, A., Mazzucchelli, R., Lopez-Beltran, A., Verri, E., Scarpelli, M., et al. (2020). Androgen receptor signaling pathway in prostate cancer: From genetics to clinical applications. *Cells* 9 (12), 2653. doi:10.3390/cells9122653

Calmasini, F. B., McCarthy, C. G., Wenceslau, C. F., Priviero, F., Antunes, E., and Webb, R. C. (2020). Toll-like receptor 9 regulates metabolic profile and contributes to obesity-induced benign prostatic hyperplasia in mice. *Pharmacol. Rep. P. R.* 72 (1), 179–187. doi:10.1007/s43440-019-00010-3

Chen, G., Cai, Z. D., Lin, Z. Y., Wang, C., Liang, Y. X., Han, Z. D., et al. (2020). ARNT-dependent CCR8 reprogrammed LDH isoform expression correlates with poor clinical outcomes of prostate cancer. *Mol. Carcinog.* 59 (8), 897–907. doi:10.1002/mc.23201

Chen, J., and Song, H. (2016). Protective potential of epigallocatechin-3-gallate against benign prostatic hyperplasia in metabolic syndrome rats. *Environ. Toxicol. Pharmacol.* 45, 45315–45320. doi:10.1016/j.etap.2016.06.015

Chen, X., Ma, Q., Shang, Z., and Niu, Y. (2020). Super-enhancer in prostate cancer: Transcriptional disorders and therapeutic targets. *NPJ Precis. Oncol.* 4 (1), 31. doi:10.1038/s41698-020-00137-0

Chmelar, R., Buchanan, G., Need, E. F., Tilley, W., and Greenberg, N. M. (2007). Androgen receptor coregulators and their involvement in the development and progression of prostate cancer. *Int. J. cancer* 120 (4), 719–733. doi:10.1002/ijc.22365

Cimadamore, A., Gasparrini, S., Scarpelli, M., Doria, A., Mazzucchelli, R., Massari, F., et al. (2017). Epigenetic modifications and modulators in prostate cancer. *Crit. Rev. Oncog.* 22 (5-6), 439–450. doi:10.1615/CritRevOncog.2017020964

Coleman, D. J., Gao, L., Schwartzman, J., Korkola, J. E., Sampson, D., Derrick, D. S., et al. (2019). Maintenance of MYC expression promotes de novo resistance to BET bromodomain inhibition in castration-resistant prostate cancer. *Sci. Rep.* 9 (1), 3823. doi:10.1038/s41598-019-40518-5

Cornford, P., van den Bergh, R., Briers, E., Van den Broeck, T., Cumberbatch, M. G., De Santis, M., et al. (2021). EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer. Part II-2020 update: Treatment of relapsing and metastatic prostate cancer. *Eur. Urol.* 79 (2), 263–282. doi:10.1016/j.eururo.2020.09.046

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* 107 (50), 21931–21936. doi:10.1073/pnas.1016071107

Cucchiara, V., Yang, J. C., Mirone, V., Gao, A. C., Rosenfeld, M. G., and Evans, C. P. (2017). Epigenomic regulation of androgen receptor signaling: Potential role in prostate cancer therapy. *Cancers* 9 (1), 9. doi:10.3390/cancers9010009

Dai, C., Heemers, H., and Sharifi, N. (2017). Androgen signaling in prostate cancer. *Cold Spring Harb. Perspect. Med.* 7 (9), a030452. doi:10.1101/cshperspect.a030452

Davies, A., Conteduca, V., Zoubeidi, A., and Beltran, H. (2019). Biological evolution of castration-resistant prostate cancer. *Eur. Urol. focus* 5 (2), 147–154. doi:10.1016/j.euf.2019.01.016

Djavan, B., Moul, J. W., Zlotta, A., Remzi, M., and Ravery, V. (2003). PSA progression following radical prostatectomy and radiation therapy: New standards in the new millennium. *Eur. Urol.* 43 (1), 12–27. doi:10.1016/s0302-2838(02)00505-5

Donati, B., Lorenzini, E., and Ciarrocchi, A. (2018). BRD4 and cancer: Going beyond transcriptional regulation. *Mol. cancer* 17 (1), 164. doi:10.1186/s12943-018-0915-9

Fritz, W. A., Lin, T. M., and Peterson, R. E. (2008). The aryl hydrocarbon receptor (AhR) inhibits vanadate-induced vascular endothelial growth factor (VEGF) production in TRAMP prostates. *Carcinogenesis* 29 (5), 1077–1082. doi:10.1093/carcin/bgn069

Gardella, K. A., Muro, I., Fang, G., Sarkar, K., Mendez, O., and Wright, C. W. (2016). Aryl hydrocarbon receptor nuclear translocator (ARNT) isoforms control lymphoid cancer cell proliferation through differentially regulating tumor suppressor p53 activity. *Oncotarget* 7 (10), 10710–10722. doi:10.18632/oncotarget.7539

Grasso, C. S., Wu, Y. M., Robinson, D. R., Cao, X., Dhanasekaran, S. M., Khan, A. P., et al. (2012). The mutational landscape of lethal castration-resistant prostate cancer. *Nature* 487 (7406), 239–243. doi:10.1038/nature11125

Guo, Y., Partch, C. L., Key, J., Card, P. B., Pashkov, V., Patel, A., et al. (2013). Regulating the ARNT/TACC3 axis: Multiple approaches to manipulating protein/protein interactions with small molecules. *ACS Chem. Biol.* 8 (3), 626–635. doi:10.1021/cb300604u

Hah, N., Benner, C., Chong, L. W., Yu, R. T., Downes, M., and Evans, R. M. (2015). Inflammation-sensitive super enhancers form domains of coordinately regulated enhancer RNAs. *Proc. Natl. Acad. Sci. U. S. A.* 112 (3), E297–E302. doi:10.1073/pnas.1424028112

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

Hankey, W., Chen, Z., and Wang, Q. (2020). Shaping chromatin states in prostate cancer by pioneer transcription factors. *Cancer Res.* 80 (12), 2427–2436. doi:10.1158/0008-5472.CAN-19-3447

Harris, A. L. (2002). Hypoxia--a key regulatory factor in tumour growth. *Nat. Rev. Cancer* 2 (1), 38–47. doi:10.1038/nrc704

Hassen, W., Kassambara, A., Reme, T., Sahota, S., Seckinger, A., Vincent, L., et al. (2015). Drug metabolism and clearance system in tumor cells of patients with multiple myeloma. *Oncotarget* 6 (8), 6431–6447. doi:10.18632/oncotarget.3237

He, M. X., Cuoco, M. S., Crowdis, J., Bosma-Moody, A., Zhang, Z., Bi, K., et al. (2021). Transcriptional mediators of treatment resistance in lethal prostate cancer. *Nat. Med.* 27 (3), 426–433. doi:10.1038/s41591-021-01244-6

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-Andre, V., Sigova, A. A., et al. (2013). Super-enhancers in the control of cell identity and disease. *Super-Enhancers Control Cell Identity Dis.* 155 (4), 934–947. doi:10.1016/j.cell.2013.09.053

Hnisz, D., Schuijers, J., Lin, C. Y., Weintraub, A. S., Abraham, B. J., Lee, T. I., et al. (2015). Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol. Cell* 58 (2), 362–370. doi:10.1016/j.molcel.2015.02.014

Hsu, Y. C., Kao, C. Y., Chung, Y. F., Chen, M. S., and Chiu, I. M. (2012). Ciliogenic RFX transcription factors regulate FGF1 gene promoter. *J. Cell. Biochem.* 113 (7), 2511–2522. doi:10.1002/jcb.24127

Huang, C. R., Chang, T. W., Lee, C. T., Shen, C. J., Chang, W. C., and Chen, B. K. (2021). ARNT deficiency represses pyruvate dehydrogenase kinase 1 to trigger ROS production and melanoma metastasis. *Oncogenesis* 10 (1), 11. doi:10.1038/s41389-020-00299-3

Ing-Simmons, E., Seitan, V. C., Faure, A. J., Flicek, P., Carroll, T., Dekker, J., et al. (2015). Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. *Genome Res.* 25 (4), 504–513. doi:10.1101/gr.184986.114

Jin, S., Xiang, P., Liu, J., Yang, Y., Hu, S., Sheng, J., et al. (2019). Activation of cGMP/PKG/p65 signaling associated with PDE5-Is downregulates CCL5 secretion by CD8 (+) T cells in benign prostatic hyperplasia. *Prostate* 79 (8), 909–919. doi:10.1002/pros.23801

Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., et al. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* 10 (1), 1930. doi:10.1038/s41467-019-09982-5

Klein, C. A. (2013). Selection and adaptation during metastatic cancer progression. *Nature* 501 (7467), 365–372. doi:10.1038/nature12628

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., et al. (2018). The human transcription factors. *Cell* 172 (4), 650–665. doi:10.1016/j.cell.2018.01.029

Lee, J. E., Park, Y. K., Park, S., Jang, Y., Waring, N., Dey, A., et al. (2017). Brd4 binds to active enhancers to control cell identity gene induction in adipogenesis and myogenesis. *Nat. Commun.* 8 (1), 2217. doi:10.1038/s41467-017-02403-5

Leick, K. M., Obeid, J. M., Bekiranov, S., and Slingluff, C. L., Jr. (2019). Systems analysis of barrier molecule and ARNT-related gene expression regulation in melanoma. *Oncoimmunology* 8 (12), e1665978. doi:10.1080/2162402X.2019.1665978

Li, Y., Shi, B., Dong, F., Zhu, X., Liu, B., and Liu, Y. (2019). Effects of inflammatory responses, apoptosis, and STAT3/NF-κB- and Nrf2-mediated oxidative stress on benign prostatic hyperplasia induced by a high-fat diet. *Aging* 11 (15), 5570–5578. doi:10.18632/aging.102138

Li, Y. Z., Shi, B. K., Li, J. Y., Zhu, X. W., Liu, J., and Liu, Y. L. (2020). Role of p-ERK1/2 in benign prostatic hyperplasia during hyperinsulinemia. *Urology J.* 18 (2), 225–229. doi:10.22037/uj.v16i7.5694

McVicker, G., van de Geijn, B., Degner, J. F., Cain, C. E., Banovich, N. E., Raj, A., et al. (2013). Identification of genetic variants that affect histone modifications in human cells. *Hist. Modifications Hum. Cells* 342 (6159), 747–749. doi:10.1126/science.1242429

Moul, J. W. (2000). Prostate specific antigen only progression of prostate cancer. *J. urology* 163 (6), 1632–1642. doi:10.1097/00005392-200006000-00003

Mulholland, D. J., Tran, L. M., Li, Y., Cai, H., Morim, A., Wang, S., et al. (2011). Cell autonomous role of PTEN in regulating castration-resistant prostate cancer growth. *Cancer Cell* 19 (6), 792–804. doi:10.1016/j.ccr.2011.05.006

Nandana, S., and Chung, L. W. (2014). Prostate cancer progression and metastasis: Potential regulatory pathways for therapeutic targeting. *Am. J. Clin. Exp. Urol.* 2 (2), 92–101.

Partch, C. L., and Gardner, K. H. (2011). Coactivators necessary for transcriptional output of the hypoxia inducible factor, HIF, are directly recruited by ARNT PAS-B. *Proc. Natl. Acad. Sci. U. S. A.* 108 (19), 7739–7744. doi:10.1073/pnas.1101357108

Qie, Y., Wang, L., Du, E., Chen, S., Lu, C., Ding, N., et al. (2020). TACC3 promotes prostate cancer cell proliferation and restrains primary cilium formation. *Exp. Cell Res.* 390 (2), 111952. doi:10.1016/j.yexcr.2020.111952

Quigley, D. A., Dang, H. X., Zhao, S. G., Lloyd, P., Aggarwal, R., Alumkal, J. J., et al. (2018). Genomic hallmarks and structural variation in metastatic prostate cancer. *Cell* 174 (3), 758–769. doi:10.1016/j.cell.2018.06.039

Ribeiro, D. L., Pinto, M. E., Maeda, S. Y., Taboga, S. R., and Góes, R. M. (2012). High fat-induced obesity associated with insulin-resistance increases FGF-2 content and causes stromal hyperplasia in rat ventral prostate. *Cell tissue Res.* 349 (2), 577–588. doi:10.1007/s00441-012-1420-x

Roach, M., 3rd, Hanks, G., Thames, H., Jr, Schellhammer, P., Shipley, W. U., Sokol, G. H., et al. (2006). Defining biochemical failure following radiotherapy with or without hormonal therapy in men with clinically localized prostate cancer: Recommendations of the RTOG-ASTRO phoenix consensus conference. *Int. J. Radiat. Oncol. Biol. Phys.* 65 (4), 965–974. doi:10.1016/j.ijrobp.2006.04.029

Sengupta, S., and George, R. E. (2017). Super-enhancer-driven transcriptional dependencies in cancer. *Trends cancer* 3 (4), 269–281. doi:10.1016/j.trecan.2017.03.006

Shafran, J. S., Andrieu, G. P., Györffy, B., and Denis, G. V. (2019). BRD4 regulates metastatic potential of castration-resistant prostate cancer through AHNAK. *Mol. cancer Res. MCR* 17 (8), 1627–1638. doi:10.1158/1541-7786.MCR-18-1279

Shah, N., Wang, P., Wongvipat, J., Karthaus, W. R., Abida, W., Armenia, J., et al. (2017). Regulation of the glucocorticoid receptor via a BET-dependent enhancer drives antiandrogen resistance in prostate cancer. *eLife* 6, e27861. doi:10.7554/eLife.27861

Shankar, E., Bhaskaran, N., MacLennan, G. T., Liu, G., Daneshgari, F., and Gupta, S. (2015). Inflammatory signaling involved in high-fat diet induced prostate diseases. *J. urology Res.* 2 (1), 1018.

Shen, J. Z., Qiu, Z., Wu, Q., Finlay, D., Garcia, G., Sun, D., et al. (2021). FBXO44 promotes DNA replication-coupled repetitive element silencing in cancer cells. *Cell* 184 (2), 352–369.e23. doi:10.1016/j.cell.2020.11.042

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer statistics, 2021. *CA a cancer J. Clin.* 71 (1), 7–33. doi:10.3322/caac.21654

Sugiura, M., Sato, H., Kanesaka, M., Imamura, Y., Sakamoto, S., Ichikawa, T., et al. (2021). Epigenetic modifications in prostate cancer. *Int. J. urology official J. Jpn. Urological Assoc.* 28 (2), 140–149. doi:10.1111/iju.14406

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA a cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Suzuki, H., Ueda, T., Ichikawa, T., and Ito, H. (2003). Androgen receptor involvement in the progression of prostate cancer. *Endocrine-related cancer* 10 (2), 209–216. doi:10.1677/erc.0.0100209

Thandapani, P. (2019). Super-enhancers in cancer. *Pharmacol. Ther.*, 199129–199138. doi:10.1016/j.pharmthera.2019.02.014

Urbanucci, A., and Mills, I. G. (2018). Bromodomain-containing proteins in prostate cancer. *Mol. Cell. Endocrinol.* 462, 31–40. doi:10.1016/j.mce.2017.06.007

Watkins, T., Lim, E. L., Petkovic, M., Elizalde, S., Birkbak, N. J., Wilson, G. A., et al. (2020). Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature* 587 (7832), 126–132. doi:10.1038/s41586-020-2698-6

Watson, P. A., Arora, V. K., and Sawyers, C. L. (2015). Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat. Rev. Cancer* 15 (12), 701–711. doi:10.1038/nrc4016

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., et al. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153 (2), 307–319. doi:10.1016/j.cell.2013.03.035

Xu, C., Xu, Y., Shen, Z., Zhou, H., Xiao, J., and Huang, T. (2018). Effects of metformin on prostatic tissue of rats with metabolic syndrome and benign prostatic hyperplasia. *Int. urology Nephrol.* 50 (4), 611–617. doi:10.1007/s11255-018-1826-9

Xu, Y., Ren, W., Li, Q., Duan, C., Lin, X., Bi, Z., et al. (2022). LncRNA uc003xsl.1-mediated activation of the nf?b/IL8 Axis promotes progression of triple-negative Breast cancer. *Cancer Res.* 82 (4), 556–570. doi:10.1158/0008-5472.CAN-21-1446

Yegnasubramanian, S., De Marzo, A. M., and Nelson, W. G. (2019). Prostate cancer epigenetics: From basic mechanisms to clinical implications. *Cold Spring Harb. Perspect. Med.* 9 (4), a030445. doi:10.1101/cshperspect.a030445

Zhang, M., Moreno-Rodriguez, T., and Quigley, D. A. (2022). Why ARNT prostate tumors responding to enzalutamide. *Cancer Discov.* 12 (9), 2017–2019. doi:10.1158/2159-8290.CD-22-0702

Zhao, S. G., Chen, W. S., Li, H., Foye, A., Zhang, M., Sjöström, M., et al. (2020). The DNA methylation landscape of advanced prostate cancer. *Nat. Genet.* 52 (8), 778–789. doi:10.1038/s41588-020-0648-8

Zheng, J. X., Xiao, Y. C., Hu, Y. R., Hao, M., and Kuang, H. Y. (2015). Exendin-4 shows no effects on the prostatic index in high-fat-diet-fed rat with benign prostatic hyperplasia by improving insulin resistance. *Andrologia* 47 (2), 236–242. doi:10.1111/and.12252

Zheng, W., Zhang, S., Guo, H., Chen, X., Huang, Z., Jiang, S., et al. (2021). Multi-omics analysis of tumor angiogenesis characteristics and potential epigenetic regulation mechanisms in renal clear cell carcinoma. *Cell Commun. Signal. CCS* 19 (1), 39. doi:10.1186/s12964-021-00728-9

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 12 June 2023
DOI 10.3389/fphar.2023.1200017

OPEN ACCESS

EDITED BY
Linhui Wang,
Second Military Medical University, China

REVIEWED BY
Aimin Jiang,
Fudan University, China
Yunzhao Xu,
Affiliated Hospital of Nantong University, China

*CORRESPONDENCE
Jianjun Li,
leejjun2007@163.com
Shuo Huang,
18180920941@163.com
Zhe Wang,
wangzhe@tmmu.edu.cn

†These authors have contributed equally to this work

RECEIVED 04 April 2023
ACCEPTED 30 May 2023
PUBLISHED 12 June 2023

CITATION
Chen J, Wu S, Peng Y, Zhao Y, Dong Y, Ran F, Geng H, Zhang K, Li J, Huang S and Wang Z (2023), Constructing a cancer stem cell related prognostic model for predicting immune landscape and drug sensitivity in colorectal cancer. *Front. Pharmacol.* 14:1200017. doi: 10.3389/fphar.2023.1200017

Constructing a cancer stem cell related prognostic model for predicting immune landscape and drug sensitivity in colorectal cancer

Jianfang Chen†, Shuang Wu†, Yu Peng†, Yang Zhao, Yan Dong, Fengwei Ran, Haofei Geng, Kang Zhang, Jianjun Li*, Shuo Huang* and Zhe Wang*

Department of Oncology and Southwest Cancer Center, Southwest Hospital, Third Military Medical University (Army Medical University), Chongqing, China

Background: Colorectal cancer (CRC) ranks the second malignancy with high incidence and mortality worldwide. Cancer stem cells (CSCs) function critically in cancer progression and metastasis via the interplay with immune cells in tumor microenvironment. This study aimed to identify important CSC marker genes and parsed the role of these marker genes in CRC.

Materials and methods: CRC samples' single-cell RNA sequencing data and bulk transcriptome data were utilized. Seurat R package annotated CSCs and identified CSC marker genes. Consensus clustering subtyped CRC samples based on CSC marker genes. Immune microenvironment, pathway and oxidative stress analysis was performed using ESTIMATE, MCP-counter analysis and ssGSEA analysis. A prognostic model was established by Lasso and stepAIC. Sensitivity to chemotherapeutic drugs was determined by the biochemical half maximal inhibitory concentration with pRRophetic R package.

Results: We identified a total of 29 CSC marker genes related to disease-specific survival (DSS). Two clusters (CSC1 and CSC2) were determined, and CSC2 showed shorter DSS, a larger proportion of late-stage samples, and higher oxidative stress response. Two clusters exhibited differential activation of biological pathways associated with immune response and oncogenic signaling. Drug sensitivity analysis showed that 44 chemotherapy drugs were more sensitive to CSC2 that those in CSC1. We constructed a seven-gene prognostic model (DRD4, DPP7, UCN, INHBA, SFTA2, SYNPO2, and NXPH4) that was effectively to distinguish high-risk and low-risk patients. 14 chemotherapy drugs were more

Abbreviations: CSCs, cancer stem cells; CAF, cancer-associated fibroblasts; CRC, colorectal cancer; CNV, copy number variation; CDF, cumulative distribution function; DEGs, differentially expressed genes; DFS, disease-free survival; DSS, disease-specific survival; EMT, epithelial–mesenchymal transformation; FDR, false discovery rate; GEO, Gene Expression Omnibus; Lasso, least absolute shrinkage and selection operator; MCP, Microenvironment Cell Populations; mRNAsi, mRNA stemness index; MDSC, myeloid-derived suppressor cells; OCLR, one-class logistic regression; PCA, principle component analysis; ROC, receiver operation characteristic; RFS, recurrence-free survival; SNV, Single nucleotide variation; ssGSEA, single sample gene set enrichment analysis; scRNA-seq, Single-cell RNA sequencing; stepAIC, stepwise Akaike information criterion; TCGA, The Cancer Genome Atlas; TGF, transforming growth factor; TIDE, Tumor Immune Dysfunction and Exclusion; TNF, tumor necrosis factor; TAM, tumor-associated macrophages.

sensitive to high-risk group and 13 chemotherapy drugs were more sensitive to low-risk group. Combination of higher oxidative stress and risk score indicated dismal prognosis.

Conclusion: The CSC marker genes we identified may help to further decipher the role of CSCs in CRC development and progression. The seven-gene prognostic model could serve as an indicator for predicting the response to immunotherapy and chemotherapy as well as prognosis of CRC patients.

KEYWORDS

colorectal cancer, cancer stem cells, single-cell analysis, CSC marker genes, clustering, prognostic model, oxidative stress, drug sensitivity

Introduction

According to global cancer statistics in 2020, colorectal cancer (CRC) is one of the most contributable malignancies worldwide, resulting in around 9.8% of new cancer cases and 9.2% of new cancer deaths worldwide (Sung et al., 2021). Males have both higher incidence and mortality rates than females, which may result from more frequent smoking in males. The gender disparity also varies greatly by age. For instance, incidence of aging from 55–74 is 40%–50% higher in male population than in females, while the close incidence is shown between men and women in ages beneath 45 years (Murphy et al., 2011). The incidence markedly escalates with the increasing age from 40 years, presented with almost or even over double increase per 10 years (Siegel et al., 2020). With the developing and popularization of screening methods like colonoscopy, the incidence rate of CRC drastically declined from late 20th century to 2017 (Laiyemo et al., 2010; Fedewa et al., 2017). Nevertheless, survival rate for 5 years is still extremely low, about 12% for metastatic CRC patients (Siegel et al., 2019). Further investigation on molecular mechanisms and screening or prognosis predicting methods is needed for facilitating survival of CRC patients.

In recent years, molecular stratification therapy based on tumor biological characteristics has improved the prognosis of patients with advanced colorectal cancer to some extent. For example, anti-D-1 and anti-CTLA-4 monoclonal antibodies for metastatic disease with MSI or high TMB (Hong et al., 2016; Tamura, 2018) and verofinil for colorectal cancer with BRAFV600E. Dienstmann et al. (Dienstmann et al., 2017) pointed out that precision therapy for colorectal cancer will shift from single-gene single-drug to multi-gene-multi-drug as well as multi-molecular multi-drug, i.e., from a clonal perspective to a clone-stromal-immune perspective, which represents the future direction of colorectal cancer treatment.

The substantial proliferation and invasion of cancer cells are tightly linked to cancer stem cells (CSCs). CSCs possess a solid self-renewal ability to expand cancer cell growth and promote tumorigenesis (Bjerkvig et al., 2005). In addition to the self-renew, CSCs can also differentiate into other cell types such as endothelial cells that are responsible for angiogenesis (Xiong et al., 2009; Ricci-Vitiani et al., 2010). Therefore, CSCs generate intra-tumor heterogeneity by differentiating a range of different cell types. On top of that, CSCs interact with epithelial–mesenchymal transformation (EMT) process to promote cancer cell invasion and migration (Kong et al., 2011). CSCs share some of same pathways with normal stem cells, such as Hedgehog signaling, Wnt/β-catenin, and Notch signaling pathways that maintain their self-renewal ability as well as confer the resistance to chemotherapy and radiotherapy in CRC (Baumann et al., 2008; Dylla et al., 2008; Colak et al., 2014; Yang et al., 2020). A Phase II clinical trial study confirmed the effect of metformin on CSCs in ovarian cancer, suggesting that epigenetic changes in tumor stroma may drive platinum sensitivity *in vitro* (Brown et al., 2020). Overall, the properties of CSCs endow them to complicate tumor microenvironment and enhance resistance to clinical therapy. Consequently, targeting CSCs could be a promising strategy for CRC treatment. For example, a phase I/II clinical trial employed CSC-loaded dendritic cells as vaccine using in metastatic CRC patients (https://clinicaltrials.gov/ct2/show/NCT02176746).

Given that CSCs function critically in cancer progression and therapy, we sought to emphasize and decipher the role of CSC markers in CRC development and therapy. Single-cell RNA sequencing data of CRC samples was analyzed for accurately annotating CRC marker genes. We identified two clusters through molecular subtyping based on CRC marker genes and parsing the difference of two clusters from various aspects including prognosis, immune microenvironment, biological pathways, and response to clinical therapy. Importantly, we established a CRC-based prognostic model which was reliable and effective for the survival prediction of CRC.

Materials and methods

The acquisition and preprocessing of bulk transcriptome data

The bulk RNA sequencing (RNA-seq) data of CRC samples and para-cancerous (normal) samples were obtained from The Cancer Genome Atlas (TCGA) database through Sangerbox platform in 30 September 2022 (named as TCGA dataset) (Tomczak et al., 2015; Shen et al., 2022). Microarray data of CRC samples (GSE17538 and GSE39582) were downloaded from Gene Expression Omnibus (GEO, specific links please see https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17538, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39582) (Clough and Barrett, 2016).

For RNA-seq data of TCGA dataset, we removed the samples without clinical information and survival information. Ensembl IDs were transferred to gene symbols, and averaged expression levels were selected in the condition that one gene had multiple Ensembl IDs. Finally, 438 CRC samples were included in TCGA dataset

(Supplementary Table S1). For microarray data, only samples with survival information were remained. Probes were transferred to gene symbols. We eliminated probes matching to multiple genes and selected averaged expression value when one gene had multiple probes. After preprocessing, a total of 232 and 556 CRC samples were remained in GSE17538 and GSE39582 datasets, respectively (Supplementary Tables S2, S3).

The acquisition and processing of single-cell RNA sequencing data

Single-cell RNA sequencing (scRNA-seq) dataset (GSE200997) was downloaded from GEO. We retained 16 CRC samples in the dataset. ScRNA-seq data was filtered under following conditions: 1) each gene expressed at least in three cells; 2) each cell expressed at least 250 genes; 3) the percentage of mitochondria is less than 10%; 4) UMI of each cell >500 and log10(GenesPerUMI) > 0.8. After preprocessing, we analyzed the scRNA-seq data using Seurat R package according to following procedures (Gribov et al., 2010). Firstly, the expression profiles were log-normalized. Then we removed the batch effects of 16 tumor samples using FindVariableFeatures and FindIntegrationAnchors functions, and integrated data through IntegrateData function. Next, ScaleData function was conducted to scale data and identify the anchor for principal component analysis (PCA). Single cells were clustered with dim = 40 and Resolution = 0.5 based on FindNeighbors and FindClusters functions. Subsequently, we annotated the cell clusters according to the cell markers of eight cell types (B cells, T cells, CSCs, endothelial cells, fibroblasts, mast cells, myeloid cells, NK cells, and T cells) from CellMarker 2.0 and previous studies (Supplementary Table S4) (Peng et al., 2019; Zhang et al., 2019; Lee et al., 2021; Su et al., 2021). Finally, FindAllMarkers function was performed to discriminate differentially expressed genes (DEGs) among eight cell types.

Analysis of cancer stemness

We used mRNA stemness index (mRNAsi) to measure cancer stemness at RNA expression level. Following a previous study, one-class logistic regression (OCLR) machine-learning algorithm was used to calculate the mRNAsi (Malta et al., 2018). The DEGs of CSCs were determined as CSC marker genes. Single sample gene set enrichment analysis (ssGSEA) calculated the score of CSC marker genes through GSVA R package (Hänzelmann et al., 2013). The mRNAsi and ssGSEA score of CSC marker genes were calculated for each tumor and normal sample in TCGA, GSE17538 and GSE39582 datasets. Pearson correlation analysis assessed the correlation between mRNAsi and CSC marker genes using Hmisc R package.

Mutation analysis

Copy number variation (CNV) and single nucleotide variation (SNV) data were obtained from TCGA dataset, where SNV data had been processed by mutect2 software. Genes mutated in more than three tumor samples were retained and examined by Fisher's exact test to determine significantly mutated genes ($p < 0.05$). The top 15 highly mutated genes were visualized.

Molecular subtyping based on CSC marker genes

First of all, to identify disease-specific survival (DSS)-associated CSC marker genes ($p < 0.05$), we performed univariate Cox regression analysis. Then based on the expression profiles of DSS-associated CSC genes, tumor samples were subtyped by unsupervised consensus clustering in ConsensusClusterPlus R package with parameter settings were as follows: reps = 50, pItem = 0.8, pFeature = 1, and distance = Euclidean (Wilkerson and Hayes, 2010). We determined the optimal cluster number k referring to cumulative distribution function (CDF) curves, relative area change under CDF curves, and consensus matrix.

Immune and pathway analysis

We obtained a group of gene sets of 28 immune cells, innate and adaptive immunity from previous research (Charoentong et al., 2017; He et al., 2018), and measured their enrichment scores using ssGSEA. ESTIMTAE algorithm evaluated the enrichment scores of immune cells and stromal cells, and outputted an ESTIMATE score representing the combined immune and stromal scores (Yoshihara et al., 2013). Microenvironment Cell Populations (MCP)-counter method was employed to assess the enrichment scores of nine immune cells and fibroblasts (Becht et al., 2016). We obtained a total of 47 immune checkpoint genes from a previous study (Danilova et al., 2019). For pathway analysis, hallmark pathways (h.all.v7.4. symbols.gmt) were collected from Molecular Signature Database (MSigDB) (Liberzon et al., 2015). The ssGSEA score for each pathway was calculated and compared between different groups.

Assessment of oxidative stress

Oxidative stress related genes were collected from "GOBP_RESPONSE_TO_OXIDATIVE_STRESS" in MSigDB. Distribution of this GOBP gene set was analyzed in GSE17538, GSE39582, and TCGA. Pearson's correlation analysis was performed to evaluate the relationship between risk score and oxidative stress. surv_cutpoint function embedded in survminer package was employed to determine the optimal cutoff and generate survival curves.

Predicting the response to immunotherapy and chemotherapy

Tumor Immune Dysfunction and Exclusion (TIDE) algorithm (http://tide.dfci.harvard.edu/) was implemented to estimate the potential response of tumor samples to immune checkpoint inhibitors (ICIs) (Jiang et al., 2018). A higher TIDE score is positively correlated with a higher possibility of immune escape

from ICIs. T cell exclusion and T cell dysfunction were examined by TIDE, and the enrichment scores of immunosuppressive cells including tumor-associated macrophages (TAM), myeloid-derived suppressor cells (MDSC), cancer-associated fibroblasts (CAF) were also calculated. The sensitivity to chemotherapeutic drugs was determined by the biochemical half maximal inhibitory concentration with pRRophetic R package (Geeleher et al., 2014).

Constructing and validating a prognostic model

Under the threshold of |fold change| > 1.5 and false discovery rate (FDR) < 0.05, DEGs between different clusters were identified by limma R package (Ritchie et al., 2015). WebGestaltR package was used to annotate significantly enriched KEGG pathways for DEGs (Liao et al., 2019). Then TCGA dataset was randomly assigned at a ratio of 1:1 into training and testing groups. We screened DSS-associated DEGs through univariate Cox regression analysis in the training group ($p < 0.01$). To reach an optimal prognostic model, we conducted least absolute shrinkage and selection operator (Lasso) with glmnet package and stepwise Akaike information criterion (stepAIC) with MASS package to determine the most contributable genes to the model (Friedman et al., 2010; Zhang, 2016). The prognostic model was defined as: risk score = Σβi×Expi, where β indicates Lasso coefficients and Exp indicates the expression levels of prognostic genes (i).

According to the optimal cut-off determined by survminer R package, each tumor sample obtained a risk score and was classified into high-risk and low-risk groups. Survival time between two risk groups was shown by Kaplan-Meier survival analysis. Receiver operation characteristic (ROC) curve analysis was used to predict the efficiency of the prognostic model in predicting different survival time through timeROC R package (Blanche et al., 2013). The effectiveness and reliability of the model was validated in TCGA and GSE17538 datasets.

Statistical analysis

The statistical analysis in this study was conducted and outputted by R software (version 4.1.0). Two-group statistical difference was examined by Wilcoxon test. Log-rank test was used in survival analysis and univariate Cox regression analysis. We considered $p < 0.05$ as statistically significant.

Results

Identification of CSC markers and their relation with mRNAsi

First of all, we used scRNA-seq data to identify different cell types based on their markers. Single cells were filtered to ensure the quality of data (see details in materials and methods). The gene counts, UMI counts, and mitochondrial percentage of 16 CRC samples before and after quality control were shown in Supplementary Figure S1. After quality control, we normalized the data and removed the batch effects based on highly variable genes (Supplementary Figure S2). Then single cells were further scaled and grouped into 21 clusters (Supplementary Figure S3). Using cell markers from CellMarker 2.0 and based on previous research, we annotated cells into eight cell types including B cells, T cells, endothelial cells, fibroblasts, mast cells, cancer stem cells, myeloid cells, and NK cells (Figures 1A, B; Supplementary Table S4). T cells contributed the largest proportion followed by B cells and CSCs in most tumor samples (Figure 1C). Subsequently, differentially expressed genes (DEGs) were identified from each cell type and the top five DEGs (bright yellow) were visualized (Figure 1D). For CSCs, we identified a total of 257 DEGs (Supplementary Table S5).

To evaluate the reliability of 257 DEGs as marker genes of CSCs, we introduced mRNA stemness index (mRNAsi) to assess the correlation between 257 DEGs and mRNAsi. We firstly calculated the ssGSEA score of 257 DEGs and mRNAsi score in three independent datasets (TCGA, GSE17538, and GSE39582). By comparing the ssGSEA score and mRNAsi score in normal and cancer samples, we observed that cancer samples had higher scores of both ssGSEA and mRNAsi than normal samples in TCGA and GSE39582 datasets ($p < 0.001$, Figures 2A, B). In addition, the ssGSEA score of CSC marker genes was significantly positively related to mRNAsi score, with coefficients of 0.443, 0.380, and 0.477 in TCGA, GSE39582, and GSE17538, respectively ($p < 0.0001$, Figures 2A–C). Therefore, it is reasonable to determine the 257 DEGs as CSC marker genes.

Identification of molecular subtypes based on CSC marker genes

To identify which CSC marker genes were associated with CRC progression, we performed univariate Cox regression based on DSS time. Of 257 CSC marker genes, we identified a total of 29 genes (20 risk genes and 9 protective genes) significantly associating with DSS (Supplementary Figure S4A; Supplementary Table S6). Within these 29 genes, 22 of them were differently expressed in cancer and para-cancerous samples (Supplementary Figure S4B). We also analyzed the gene mutations and genomic variations of 29 genes in cancer samples. PLEC, PLCG2, and LENG8 were the top three frequently mutated genes, with mutation frequencies of 10%, 7%, and 5%, respectively (Supplementary Figure S4C). CNV results showed that the frequency of gain of CNVs was larger than that of loss of CNVs (Supplementary Figure S4D). Especially, BRI3, CEBPB, HSPB1, and PLEC had frequencies of gain of CNVs over than 25%.

Given that 29 CSC marker genes were closely related to patients' prognosis, we then studied the role of these marker genes in CRC. Therefore, the expression profiles of 29 CSC marker genes in TCGA dataset were used in consensus clustering on CRC samples. According to CDF curve and consensus matrix, cluster number k = 2 was determined as the optimal and samples were classified into two clusters (CSC1 and CSC2) (Figures 3A–C). In GSE17538 and GSE39582 datasets, we used the same method to cluster samples and consensus matrix results showed that samples were evidently divided into two clusters (Figures 3D, E). Then we compared the prognosis of two clusters in three datasets. In TCGA dataset,

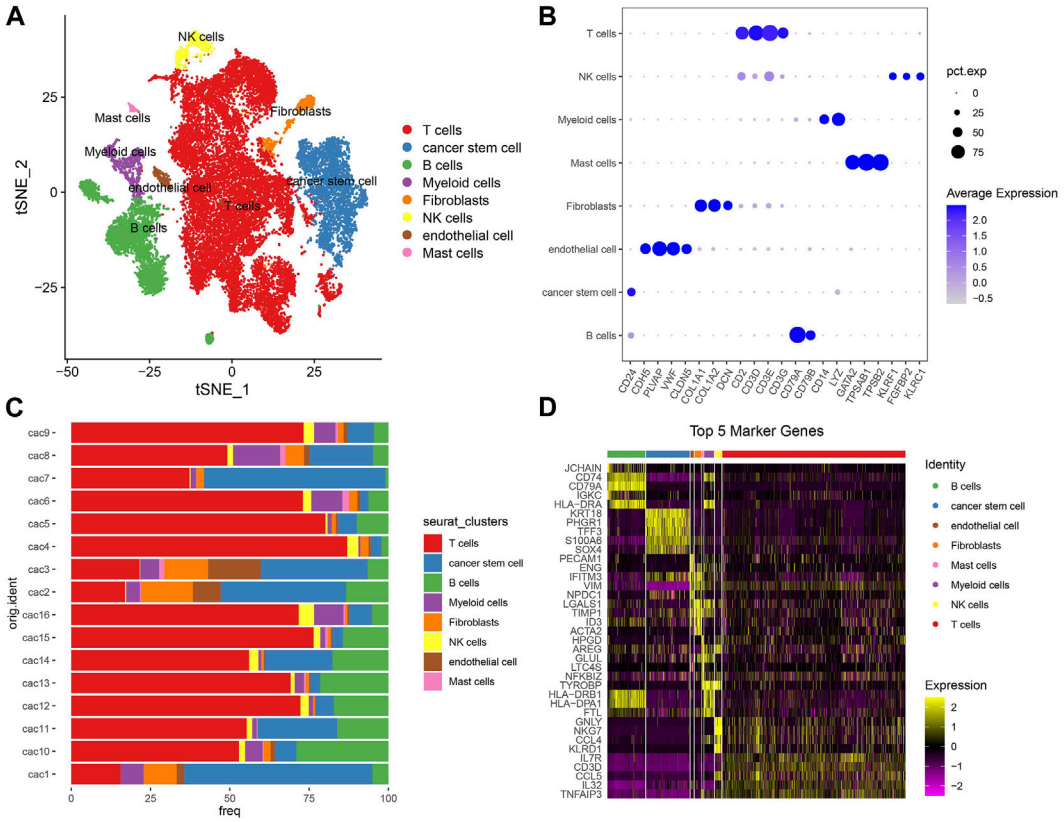


FIGURE 1
Analysis of scRNA-seq data. **(A)** T-SNE plot showed the distribution of eight cell types with different colors. **(B)** The expression of cell markers in different cell types. Pct. exp (dot) indicates the percentage of cells expressing marker genes. Blue color from light to dark indicates the expression from low to high. **(C)** The percentage of eight cell types in 16 tumor samples. 16 samples were indicated in the left and percentage was indicated in the bottom. **(D)** The top five DEGs of eight cell types. Yellow and purple represents high and low normalized expression respectively.

CSC1 and CSC2 showed significantly different disease-specific survival (DSS) ($p < 0.0001$), progression-free interval (PFI) ($p = 0.0011$), and overall survival (OS) ($p = 0.00023$) (Figure 3F). In GSE39582 dataset, CSC1 and CSC2 had different prognosis on recurrence-free survival (RFS) ($p = 0.026$) and OS ($p = 0.018$) (Figure 3G). In GSE17538 dataset, two clusters had different DSS ($p = 0.048$) and disease-free survival (DFS) ($p = 0.005$), but no significant difference on OS (Figure 3H). Overall, CSC1 had better prognosis than CSC2. PCA plot presented that two clusters were evidently separated (Figure 3I). Therefore, we considered that the clustering of CRC samples based on 29 CSC marker genes was effective and reliable.

Mutation and clinical characteristics of two clusters

We assessed the mutation data of TCGA dataset, and identified a total of 380 genes that had significantly higher mutation frequencies in CRC samples that in normal samples. The top 15 mutated genes were visualized, where XIRP2 and SCN1A had frequencies of over than 10% (Supplementary Figure S5A). However, there was no significant difference on tumor mutation burden between cancer and normal samples (Supplementary Figure S5B).

In addition, we compared the clinical characteristics including gender, age, stage I to IV, TNM stage in two clusters. The distribution of different ages and genders did not show significant differences between two clusters (Supplementary Figure S5C, D). Noteworthy, CSC2 had markedly larger proportions of the samples with late stages than CSC1, with ratios of 0.14 and 0.09 in T4 stage, 0.24 and 0.12 in N2 stage, 0.22 and 0.10 in M1 stage, 0.21 and 0.09 in stage IV for CSC2 and CSC1 respectively (Supplementary Figure S5E–H). The findings suggested that CSC marker genes may have an influence on the progression of CRC.

Immune microenvironment and oxidative stress differences of CSC1 and CSC2 clusters

We applied different methods to evaluate the immune microenvironment in CSC1 and CSC2. SsGSEA on the gene sets of 28 immune cells showed that 14 immune cells were differently enriched in two clusters, and CSC1 had higher enrichment scores in most of them such as natural killer cells, activated CD4 T cells, memory B cells (Figure 4A). In the response of adaptive and innate immunity, CSC1 also performed higher enrichment score than CSC2 but the difference was not significant in the innate immune response (Figure 4B). ESTIMATE analysis revealed higher

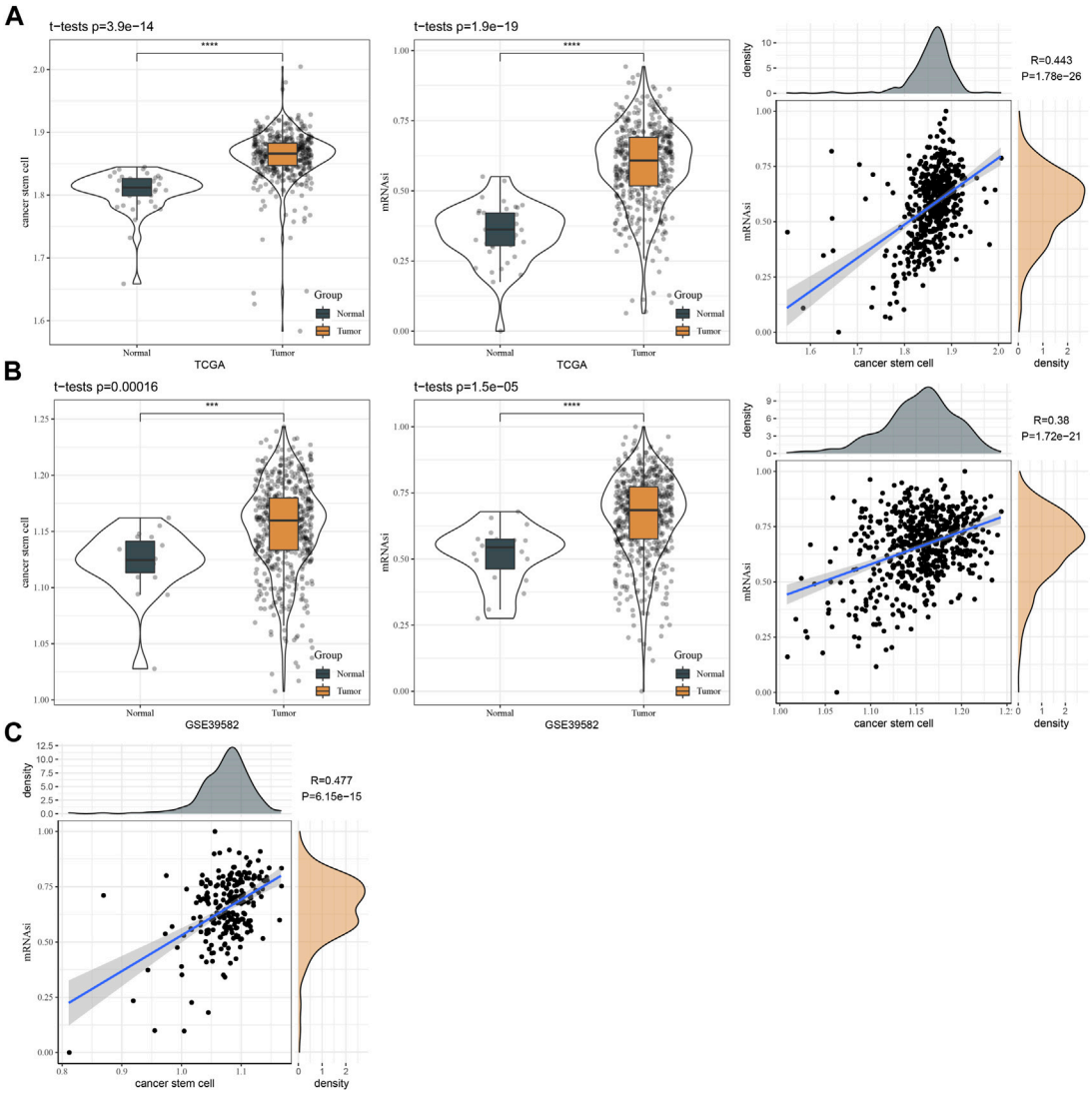


FIGURE 2
The relation between CSC markers and mRNAsi. **(A)** The ssGSEA score of CSC markers and mRNAsi in normal and tumor samples in TCGA dataset. Pearson correlation analysis between CSC markers and mRNAsi. **(B)** The ssGSEA score of CSC markers and mRNAsi in normal and tumor samples in GSE39582 dataset. Pearson correlation analysis between CSC markers and mRNAsi. **(C)** Pearson correlation analysis between CSC markers and mRNAsi in GSE17538 dataset.

infiltration of immune cells and stromal cells in CSC1 than that in CSC2 (Figure 4C). Moreover, MCP-counter manifested that of 10 immune-related cells, three cell types including monocytic lineage, endothelial cells, myeloid dendritic cells had noticeably higher enrichment scores in CSC1 compared with CSC2 (Figure 4D). The above results assessed by different methods were consistent with each other, suggesting a difference of immune cell infiltration and immune microenvironment between two clusters. Immune checkpoints are essential linkage of different immune cells for enhancing or inhibiting the cytotoxicity of immune cells. The expression levels of a total of 47 immune checkpoints were compared in two clusters. As a result, 26 of 47 immune checkpoints showed a significant difference between two clusters, with most of them were higher expressed in CSC1 (Figure 4E). Different expression levels of these immune checkpoints may contribute to the difference immune response between two clusters.

Furthermore, we reckoned the scores of hallmark pathways using ssGSEA to unveil the potential molecular mechanisms resulting in different prognosis in two clusters. As a result, 25 pathways were differently enriched between two clusters (Figure 4F). CSC1 displayed relatively enhanced activation of immune-correlated pathways, for example, complement and inflammatory response, IL2-STAT5 signaling, in accordant with the result of immune analysis. In addition, reactive oxygen species pathway, p53 signaling pathway and Wnt signaling pathway that were associated with oncogenesis were more activated in CSC2 compared with CSC1. Moreover, Figure 5 revealed that the score of "GOBP_RESPONSE_TO_OXIDATIVE_STRESS" was significantly increased in CSC2 in GSE17538 and GSE39582.

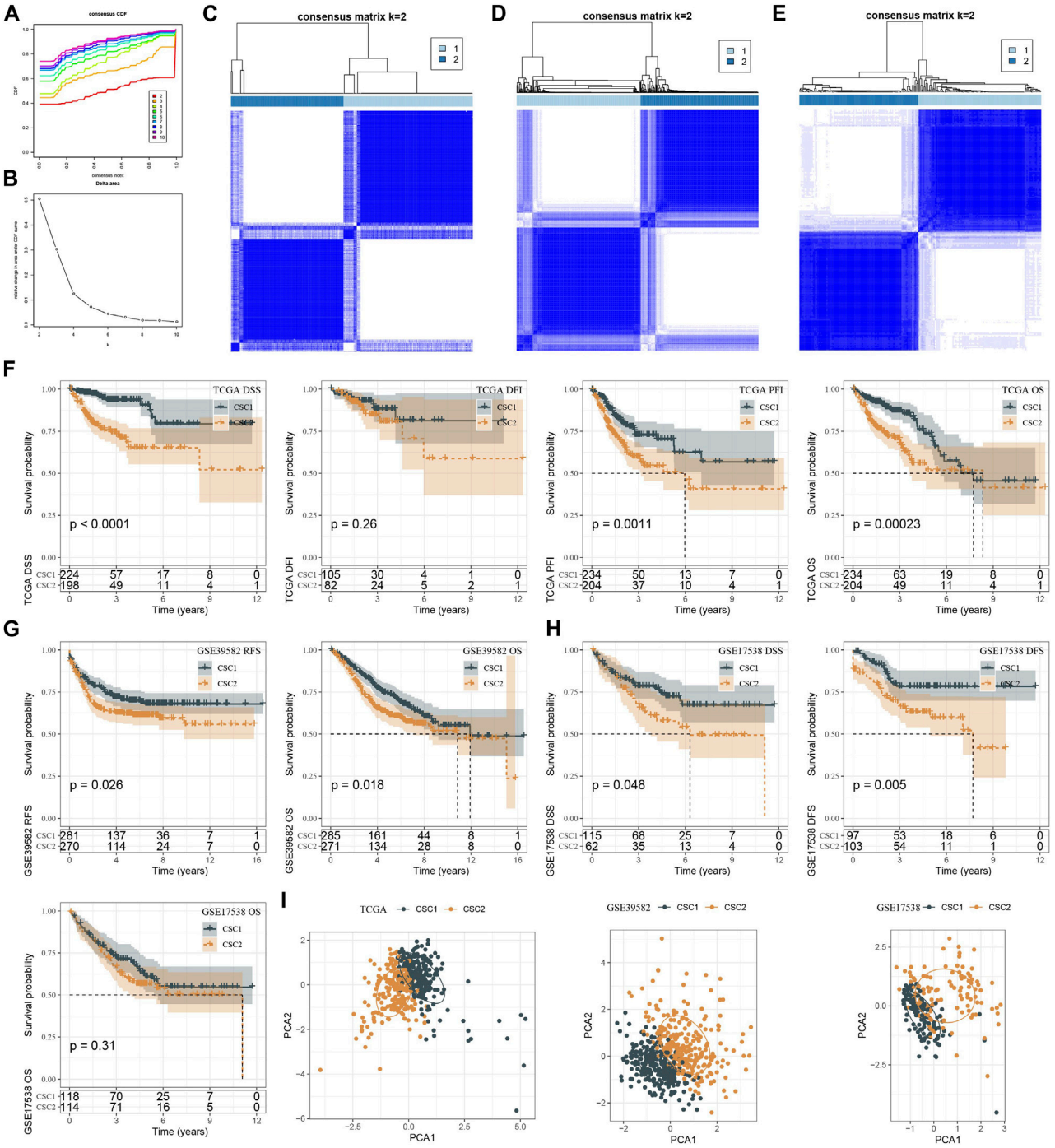


FIGURE 3
Molecular subtyping based on CSC markers. **(A, B)** CDF curves and relative change under CDF curves when cluster number k was 2–10 in TCGA dataset. **(C–E)** Consensus matrix when k = 2 in TCGA (C), GSE39582 **(D)** and GSE17538 **(E)** datasets. **(F–H)** Kaplan-Meier survival curves of CSC1 and CSC2 for different survival time in TCGA **(F)** GSE39582 **(G)** and GSE17538 **(H)** datasets. **(I)** PCA plots of CSC1 and CSC2 in three datasets. OS, overall survival; DSS, disease-specific survival; DFI, disease-free interval; PFI, progression-free interval; RFS, recurrence-free survival; DFS, disease-free survival.

The predicted response of two clusters to immunotherapy and chemotherapy

We employed TIDE analysis to estimate the response to immunotherapy for two clusters. No significant difference was detected in TIDE score between two clusters. Higher TIDE score suggested lower sensitivity to immunotherapy. Although two clusters showed similar response to immunotherapy, CSC2 had higher score of T cell dysfunction and higher enrichment of MDSC, but lower score of CAF than CSC1

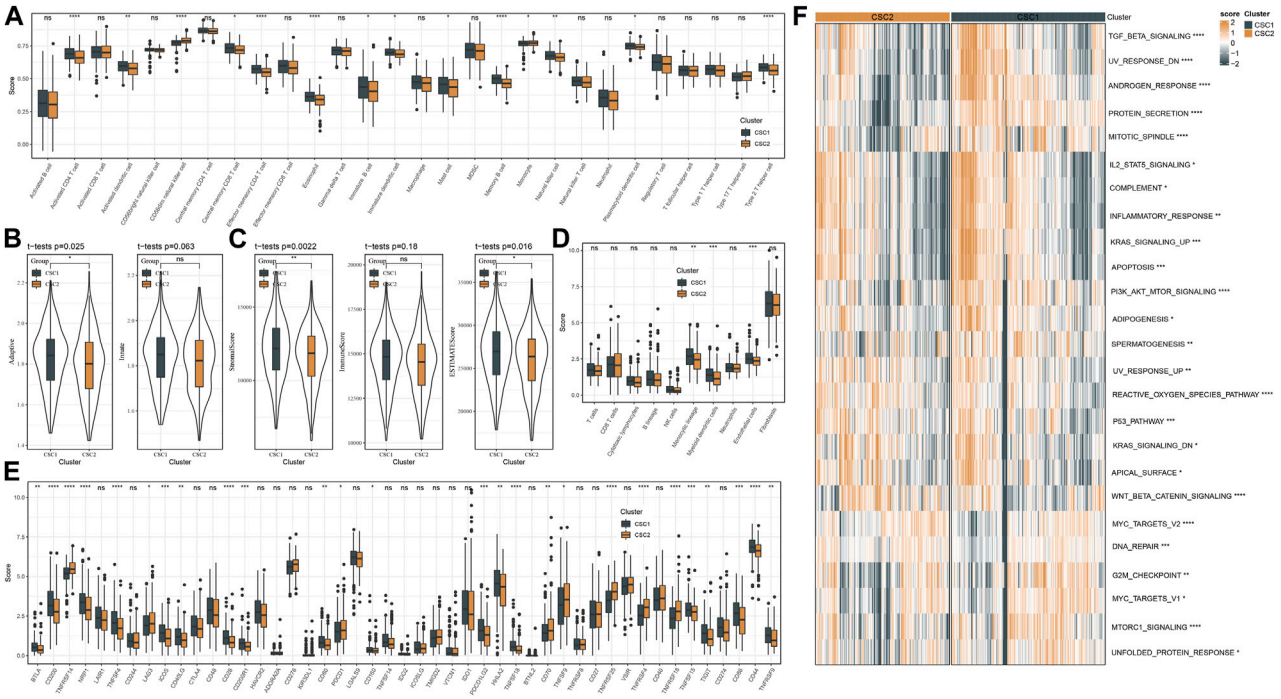


FIGURE 4
Immune microenvironment and pathway analysis of two clusters in TCGA dataset. **(A)** The estimated proportion of 28 immune-related cells by ssGSEA. **(B)** The ssGSEA score of adaptive and innate immune cells. **(C)** The stromal and immune scores measured by ESTIMATE. **(D)** MCP-counter assessed the enrichment score of 10 immune-related cells. **(E)** The expression levels of immune checkpoints. **(F)** A heatmap showed the z-score expression levels of differentially enriched pathways between two clusters. ns, not significant. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

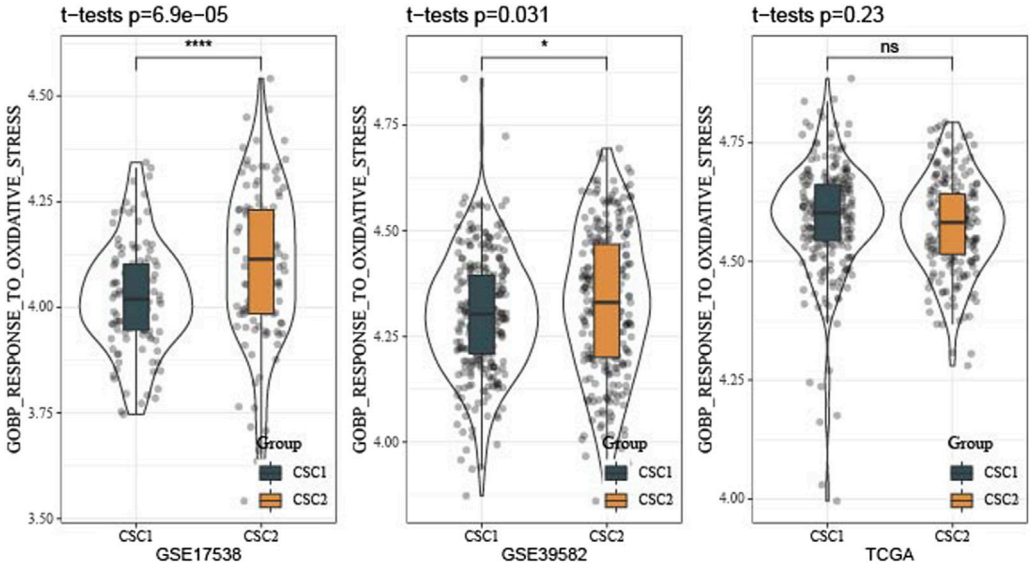


FIGURE 5
Difference of oxidative stress response of two clusters in TCGA, GSE17538 and GSE39582 datasets.

(Figure 6A). The function of T cells and infiltration levels of immunosuppressive cells (MDSC and CAF) can affect the response to immunotherapy.

In the predicted response of two clusters to chemotherapy, we evaluated a number of chemotherapeutic drugs using pRRophetic package. We identified a total of 51 chemotherapeutic drugs with

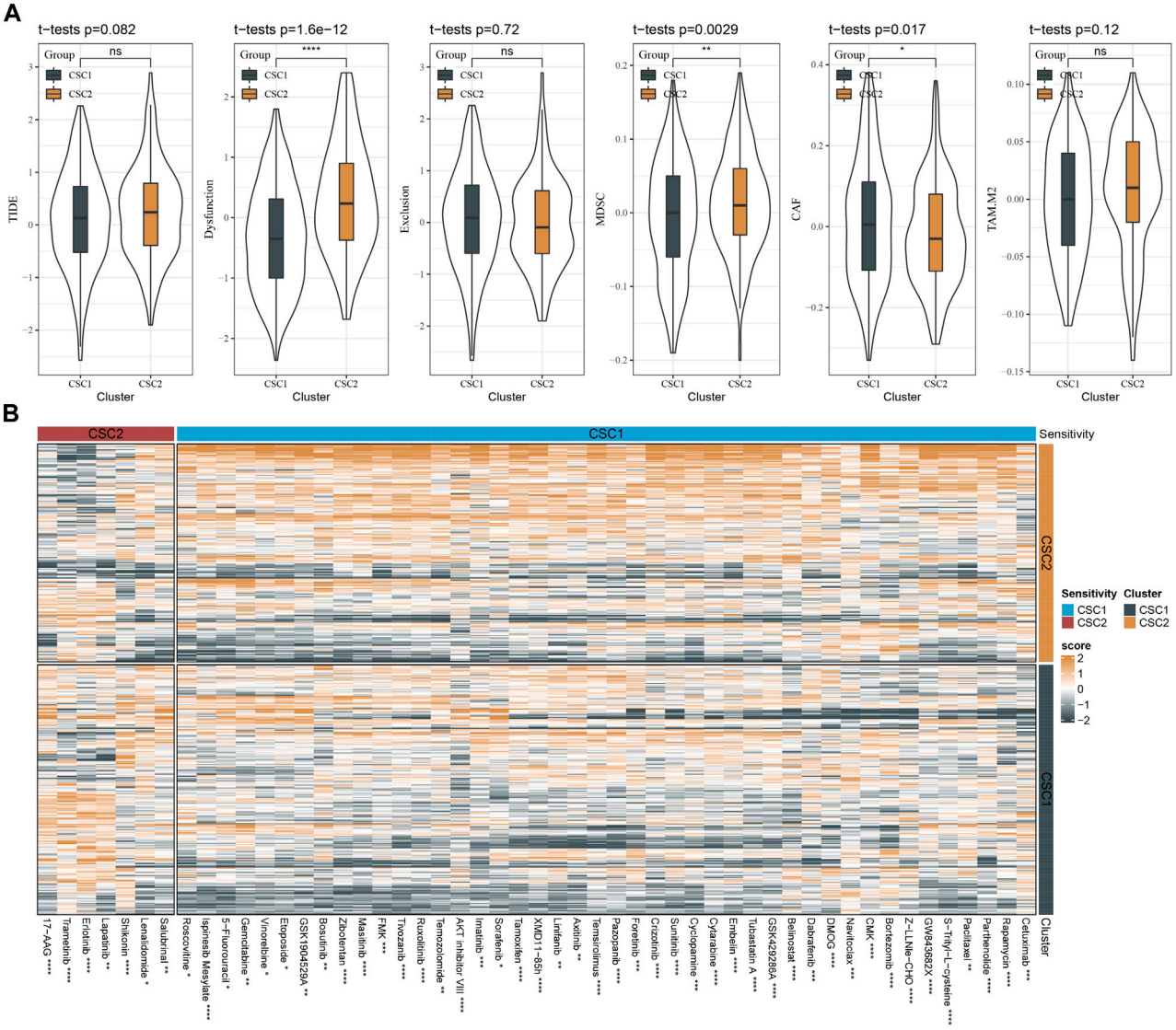


FIGURE 6
The sensitivity of two clusters to immunotherapy and chemotherapy. **(A)** TIDE analysis showed the scores of TIDE, T cell function, and infiltration of immunosuppressive cells. **(B)** A heatmap showed the estimated half maximal inhibitory concentration (IC50) of two clusters to different chemotherapeutic drugs. The drugs with significantly different IC50 in two clusters were visualized. MDSC, myeloid-derived suppressor cells; CAF, cancer-associated fibroblasts; TAM. M2, M2 tumor-associated macrophages. ns, not significant. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

different sensitivities to two clusters, where 44 drugs were more sensitive to CSC2 and 7 drugs were more sensitive to CSC1 (Figure 6B). Therefore, we inferred that CSC marker genes for molecular subtyping may be involved in the response to these chemotherapeutic drugs.

Constructing a prognostic model based on DEGs between CSC1 and CSC2

In the above sections, we illustrated that CSC1 and CSC2 exhibited different prognosis, immune microenvironment and activated pathways. To identify which genes had a difference to the outcome of clusters, we performed differential analysis on the expression profiles between CSC1 and CSC2and screened DEGs under |log Foldchange (FC)| > 1.5 and FDR <0.05. Consequently, 598 DEGs including 214 downregulated genes and 384 upregulated genes were identified in CSC1 (Supplementary Figure S6A). The DEGs were significantly enriched in pathways like drug metabolism, TGF-β signaling pathway, and gap junction, as shown by KEGG pathway analysis (Supplementary Figure S6B).

TCGA dataset was randomly divided into two groups, training and testing groups at a ratio of 1:1. To determine prognostic genes, we performed univariate Cox regression on 598 DEGs in the training group and screened 26 genes significantly related to DSS (Figure 7A). Furthermore, we used Lasso and stepAIC to decrease the number of prognostic genes for constructing a prognostic model efficiently applied in clinics. Lasso regression

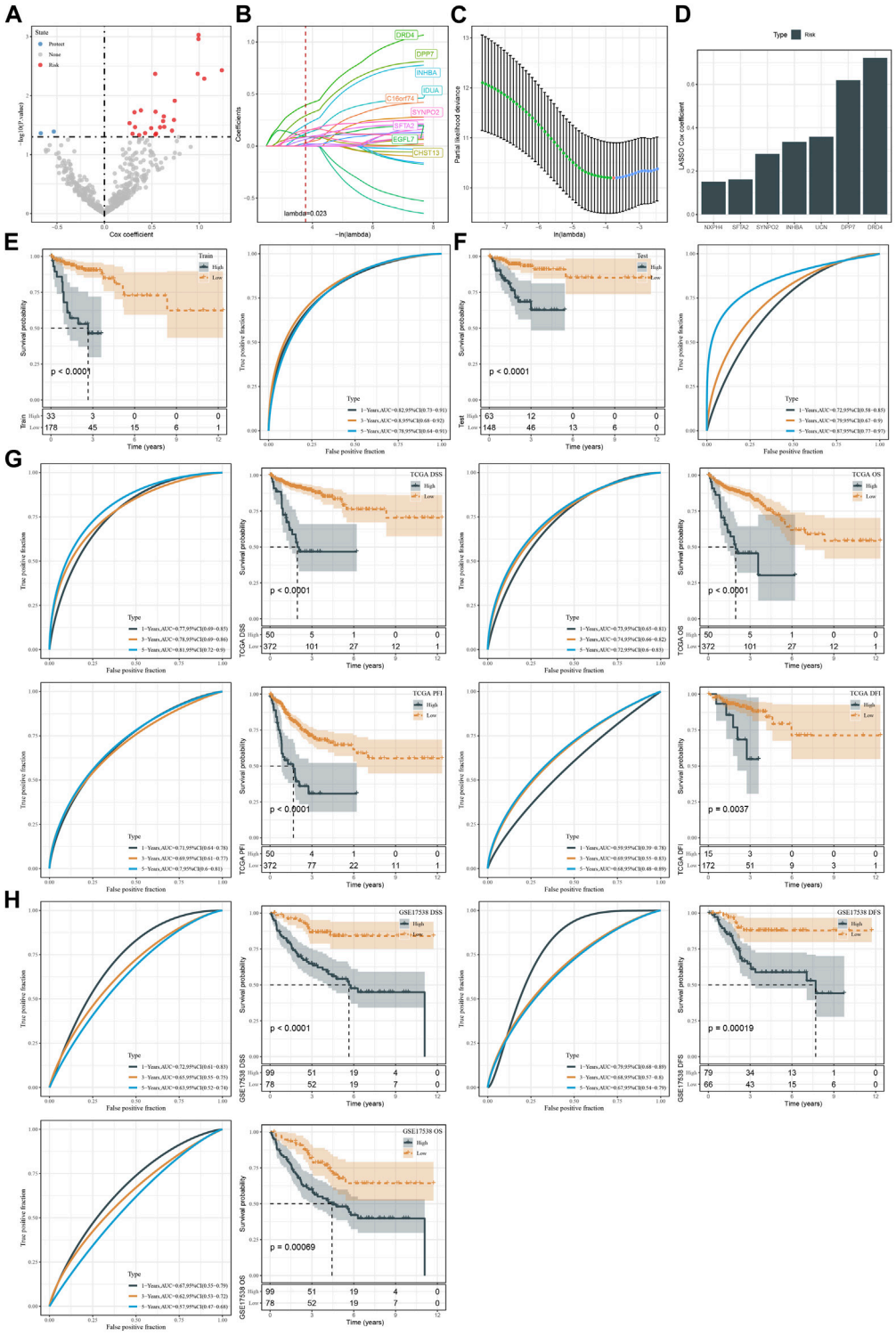


FIGURE 7
Construction and validation of the prognostic model. **(A)** Volcano plot of 26 CSC marker genes significantly associated with DSS in the training group. **(B, C)** Lasso regression analysis on 26 CSC marker genes. The coefficients of marker genes close to zero with the increasing value of lambda. Red dotted line and red dot represents the optimal lambda value of the model. **(D)** The Lasso coefficients of seven prognostic genes in the prognostic model. **(E, F)** Kaplan-Meier survival plots based on DSS of high-risk and low-risk groups in the TCGA training and testing groups. **(G)** Kaplan-Meier survival plots for DSS, OS, PFI, and DFI of high-risk and low-risk groups in TCGA dataset. **(H)** Kaplan-Meier survival plots for DSS, DFS and OS of high-risk and low-risk groups in GSE17538 dataset.

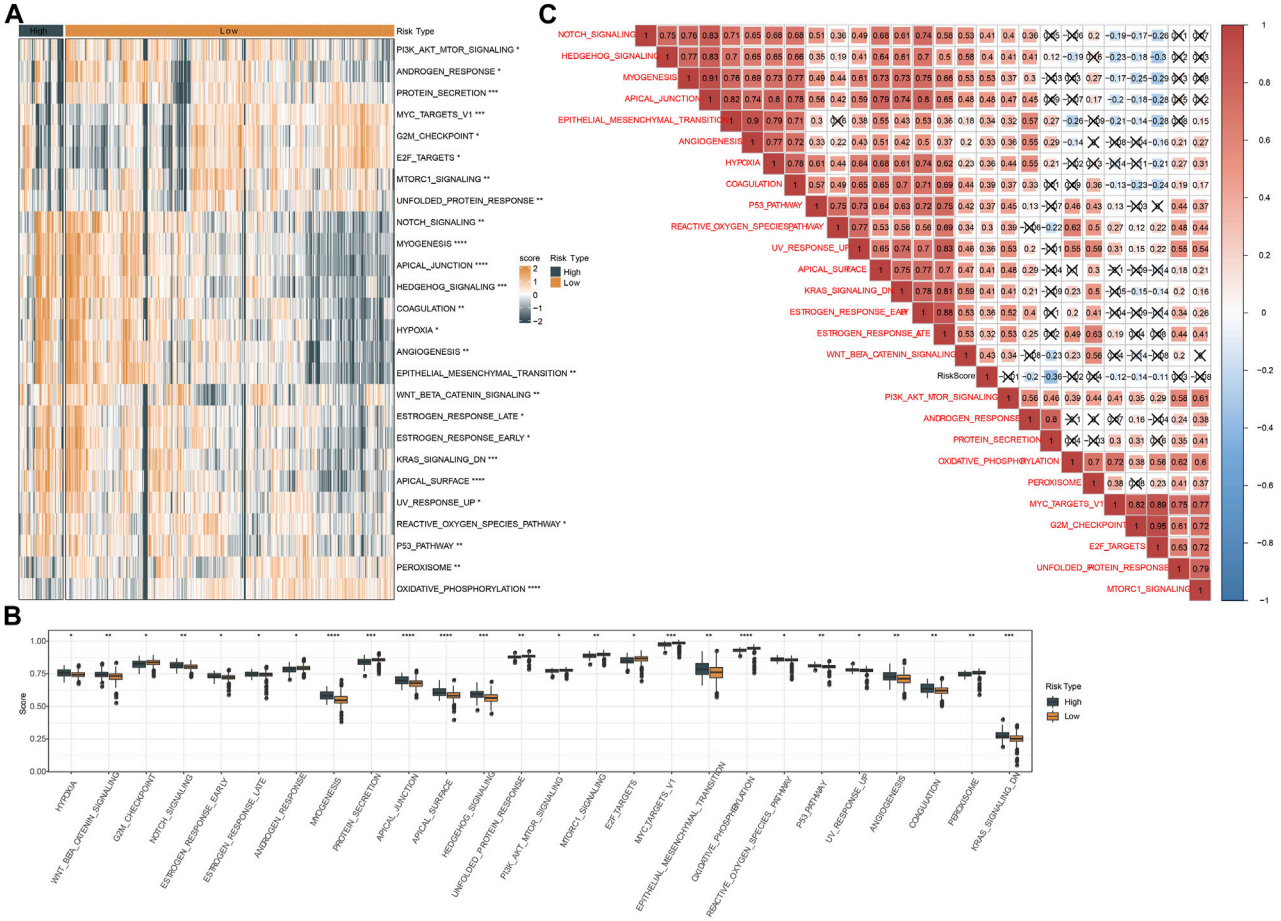


FIGURE 8
Analysis of hallmark pathways in two risk groups in TCGA dataset. **(A)** A heatmap displayed the normalized ssGSEA score of 26 pathways in two risk groups. **(B)** Box plots of ssGSEA score of 26 pathways in two risk groups. **(C)** Pearson correlation analysis between risk score and 26 pathways. Red and blue represents positive and negative correlation respectively. Fork indicates not significant.

analysis determined 14 prognostic genes when the lambda value reached the optimal (lambda = 0.023, Figures 7B, C). Then stepAIC compressed 14 genes to 7 for the final prognostic genes in the model. Finally, the prognostic model was defined as: risk score = 0.722*DRD4 + 0.619*DPP7 +0.358*UCN +0.335*INHBA +0.162*SFTA2 + 0.279*SYNPO2 + 0.151*NXPH4.

We calculated risk score for each sample in TCGA dataset and classified samples into two groups (high risk and low risk) by the optimal cut-off determined by survminer package. In both training and testing groups, high-risk group showed evidently inferior DSS than low-risk group ($p < 0.0001$, Figures 7D, E). ROC curve analysis presented favorable AUC values of the model in predicting 1-year, 3-year, and 5-year DSS with over than 0.70 (Figures 7E, F). We verified the effectiveness of the prognostic model in the whole TCGA dataset. The model showed a good performance in predicting patient survival with different status (Figure 7G). Moreover, we used an independent dataset (GSE17538) to validate the reliability of the prognostic model (Figure 7H). In the DSS, DFS, and OS prediction and classification, the model showed a good efficiency (Figure 7H). In addition, samples with advanced T stage, N stage, M stage and Stage had a higher risk score in TCGA dataset, and similarly situation was observed in GSE17538 dataset samples along with Stage and Grade (Supplementary Figure S7).

Pathway analysis of two risk groups

Next, we assessed the enrichment of biological pathways in two risk groups to identify key pathways in tumor progression. Using ssGSEA we distinguished a total of 26 pathways that were differentially enriched in high-risk and low-risk groups ($p < 0.05$, Figures 8A, B). High-risk group exhibited relatively more activated oncogenic pathways than low-risk group, such as P53 signaling, angiogenesis, EMT, hypoxia, and Notch signaling pathways. Also, we examined the correlation between risk score and these pathways delineated by a heatmap. The result showed that risk score was positively correlated with most of these pathways, such as Notch signaling (R = 0.41), Hedgehog signaling (R = 0.40), apical junction (R = 0.48), EMT (R = 0.34), angiogenesis (R = 0.33), hypoxia (R = 0.36), P53 signaling (R = 0.37), reactive oxygen species pathway (R = 0.30), KRAS signaling down (R = 0.41) and Wnt-beta catnin signaling (R = 0.43) (Figure 8C).

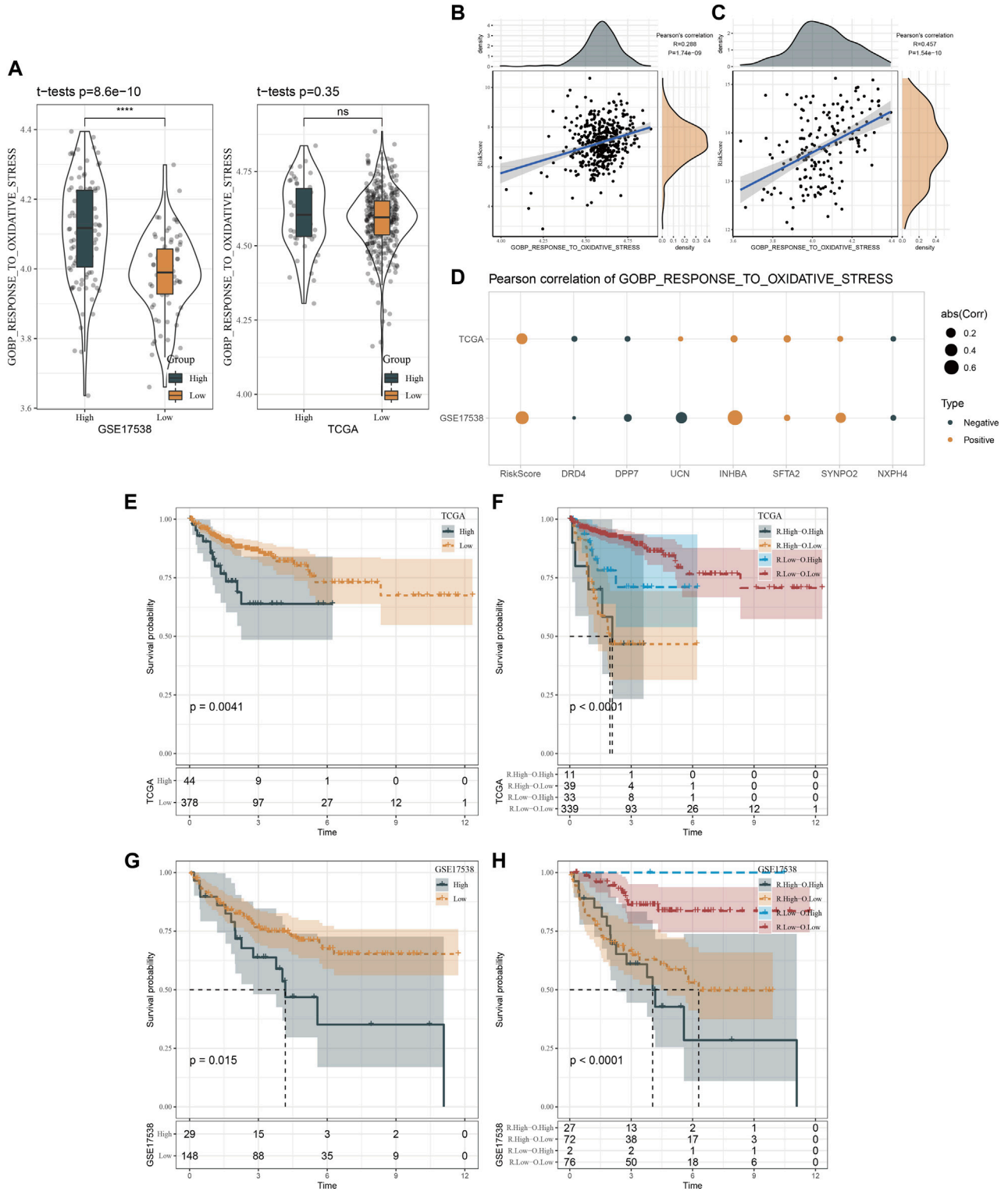


FIGURE 9
Oxidative stress analysis of two risk groups. **(A)**, Distribution of GOBP gene set in TCGA and GSE17538 datasets. **(B, C)**, Pearson correlation analysis of risk score with the response to oxidative stress in TCGA and GSE17538 datasets. **(D)**, Pearson correlation analysis of risk score and risk genes with the response to oxidative stress in TCGA and GSE17538 datasets. **(E, F)**, Kaplan-Meier survival curves of two risk groups or combination of risk score and oxidative stress in TCGA. **(G, H)**, Kaplan-Meier survival curves of two risk groups or combination of risk score and oxidative stress in GSE17538.

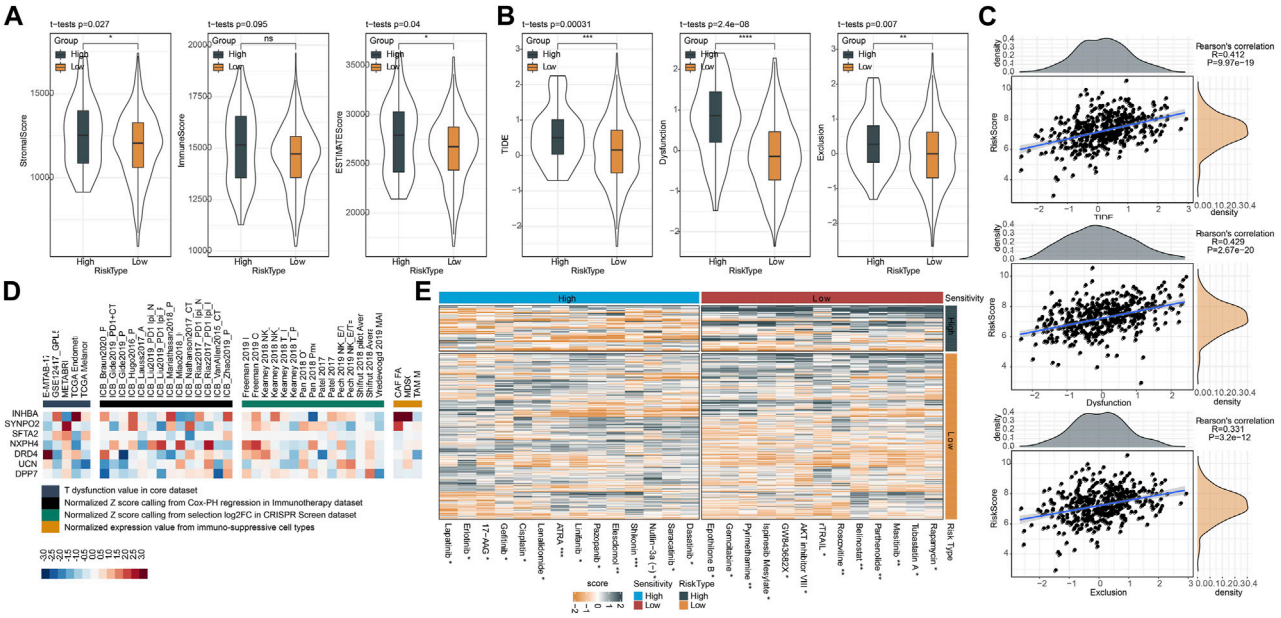


FIGURE 10
Prediction of response of two risk groups to immunotherapy and chemotherapy in TCGA dataset. **(A)** ESTIMATE analysis calculated the stromal score and immune score of two risk groups. **(B)** TIDE analysis predicted the response to immune checkpoint inhibitors. **(C)** Pearson correlation analysis of risk score with TIDE score, T cell dysfunction and T cell exclusion. **(D)** Enrichment of seven prognostic genes in T cell dysfunction score (how a gene interacts with cytotoxic T cells to influence patient survival outcome), ICB outcome (genes whose activities are correlated with ICB benefit), log-fold change (logFC) in CRISPR screens (the efficacy of lymphocyte-mediated tumor killing in cancer models) and T cell exclusion score (the gene expression levels in immunosuppressive cell types). Colors from red to blue indicates expression levels from high to low. **(E)** The estimated IC50 of two risk groups shown as heatmap. ns, not significant. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

Oxidative stress analysis of two risk groups

In addition, we emphatically analyzed the relation between risk score and response to oxidative stress. Figure 9A showed that high risk patients in GSE17538 possessed higher oxidative stress score that of low risk patients (p = 8.6e-10). Not surprisingly, risk score was positively correlated with GOBP response to oxidative stress in both TCGA (R = 0.288, p = 1.74e-09) and GSE17538 datasets (R = 0.457, p = 1.54e-10) (Figures 9B, C). Besides, GOBP response to oxidative stress was positively correlated with INHBA, SFTA2, and SYNPO2 both in TCGA and GSE17538 datasets (Figure 9D). Furthermore, we found that high patients exhibited dismal prognosis in TCGA (p = 0.0041) and GSE17538 (p = 0.015). Meanwhile, patients with high risk combined with high oxidative stress had the poorest prognosis (Figures 9E–H).

The responses of two risk groups to immunotherapy and chemotherapy

Similarly, we applied ESTIMATE algorithm to evaluate immune cell infiltration and stromal cell infiltration in two risk groups. High-risk group manifested both higher stromal score and immune score than low-risk group, but immune score was not significantly different (Figure 10A). Generally, high immune infiltration is beneficial to immune response and prognosis. To address this puzzle, we further used TIDE analysis to predict the T cell function. As a result, high-risk group displayed more significant impairment of T cell function, where higher scores of T cell dysfunction and exclusion were shown in high-risk group in comparison to low-risk group (Figure 10B). Accordingly, high-risk group had higher TIDE score, indicating a higher immune escape possibility in the high-risk group. Supportively, risk score had a highly positive correlation with TIDE, T cell exclusion, T cell dysfunction, in accordant with the above findings (Figure 10C). The result also demonstrated that risk score was a potential indicator to predict the response to immunotherapy and T cell function. Furthermore, TIDE analysis generated the association of seven prognostic genes with T cell dysfunction, T cell exclusion, immune checkpoint blockade (ICB) outcome, and the efficiency of tumor killing in CRISPR-based models (Figure 10D). but, no significance of TMB was observed in high-risk group and low-risk group (Supplementary Figure S8A). 17 of 47 immune checkpoint genes expressions were enhanced in high-risk group (Supplementary Figure S8B). Drug sensitivity analysis revealed that two risk groups had different sensitivity to 27 chemotherapeutic drugs in which 14 drugs were more sensitive to high-risk group and 13 drugs were more sensitive to low-risk group (Figure 10E). Based on the above findings, we could speculate that risk score was predictive to indicate the response of CRC patients to different chemotherapeutic drugs.

Discussion

The important roles of CSCs in cancer development and metastasis have been substantially demonstrated in the previous

studies (Ayob and Ramasamy, 2018; Prager et al., 2019). In this study, we focused on CSCs and screened a group of CSC marker genes based on scRNA-seq data of CRC samples. Using the expression profiles of CSC marker genes, we subtyped CRC samples into two clusters (CSC1 and CSC2). We compared clinical characteristics, immune microenvironment and biological pathways in two clusters. Based on DEGs between CSC1 and CSC2, we established a prognostic model for predicting the prognosis and therapeutic response of CRC patients.

From scRNA-seq data, we detected 257 CSC marker genes that showed a consistent performance with mRNAsi score. The mRNAsi score denotes the stemness degree using expression profiles (Malta et al., 2018). Compared with the normal samples, the ssGSEA score of CSC marker genes and mRNAsi score were both significantly higher in CRC samples. Moreover, CSC marker genes was noticeably positively related to mRNAsi, which proved the reliability of identification method for CSC marker genes. Of 257 CSC marker genes, 29 of them were found to be significantly related to DSS. Some CSC marker genes showed extremely disparate expression levels between normal and tumor samples, such as PLCG2, DDX11, IER5L, LENG8, HAGHL and CPNE7 (Supplementary Figure S4). They were reported to contribute cancer progression and metastasis. For example, small cell lung cancer cells with PLCG2-high phenotype had stem-like and pro-metastatic features (Chan et al., 2021). DDX11 is essential for DNA replication and genomic stability, and is considered to have an oncogenic role (Mahtab et al., 2021). Some marker genes had a large percentage of gain of CNVs, particularly BRI3, CEBPB, HSPB1, and PLEC. CEBPB was identified as a prognostic biomarker in CRC and was found to participate CRC metastasis (Rahman et al., 2019; Shao et al., 2021). HSPB1 was highly expressed in tumor tissues correlating with poor prognosis in CRC (Nagaraja et al., 2012). However, a few studies reported the roles of these CSC marker genes in cancer stemness. We considered these CSC marker genes as important candidates for exploiting the mechanisms of CSCs in CRC.

To figure out the effects of 29 CSC marker genes in CRC prognosis and tumor microenvironment, we used consensus clustering to subtype tumor samples into two clusters (CSC1 and CSC2) based on the expression profiles of 29 genes. CSC1 had evidently longer disease-specific survival than CSC2 in both TCGA and GSE17538 datasets, indicating these CSC marker genes were involved in CRC progression. The speculation was further demonstrated by the distribution of clinical characteristics in two clusters. Tumor samples with late stages like T4, N2, M1, and stage IV had substantially higher proportion in CSC2 than that in CSC1. Therefore, the 29 CSC marker genes played important roles in CRC progression and metastasis.

Previous studies have outlined the intense linkage between CSCs and tumor microenvironment (Zhang et al., 2018; Khosravi et al., 2020), which enables targeting CSCs as a possible strategy to eradicate CRC (Jahanafrooz et al., 2020). The inflammatory cytokines, for instance, interferons (IFN), transforming growth factor (TGF)-β, tumor necrosis factor (TNF)-α secreted from immune cells especially TAMs of M2 phenotype exert profound effects on maintaining the stemness of CSCs and promoting immunosuppression through pathways such as NF-κB, STAT3, and Notch (Zhang et al., 2018; Bayik and Lathia, 2021). Reciprocally, CSCs can recruit TAMs through expressing immunomodulatory factors thereby intertwining with CSC stemness programming and transcriptional activity. In comparison on immune microenvironment between CSC1 and CSC2, we observed discrepant immune infiltration and stromal infiltration. CSC1 showed higher infiltration of immune cells such as monocytic lineage, dendritic cells, activated CD4 T cells, and natural killer cells than CSC2. Although two clusters had similar proportions of CD8 T cells and cytotoxic lymphocytes, CSC2 presented more severely impaired T cell function, which resulted in its poor prognosis. Notably, CSC2 also showed a higher proportion of MDSCs and M2 TAMs than CSC1. CSC-TAM and CSC-MDSC crosstalk promoting stemness and immunosuppression have been underlined by previous studies. TAMs can facilitate CSC phenotypes by mediators such as IL-6, TGF-β, and WNT ligands (Jinushi et al., 2011; Fan et al., 2014; Wan et al., 2014). Mechanistic analysis suggested that Nos2 and nitric oxide (NO) produced by MDSCs fostered CSC phenotypes via activating Notch and STAT3 pathways in cancer cells (Peng et al., 2016; Ouzounova et al., 2017). In addition, pathway analysis revealed that tumor-associated pathways such as TGF-β and Wnt-β catenin signaling, cell cycle-related pathways such as MYC, and immune-related pathways such as inflammatory response and IL2-STAT5 signaling were distinctly enriched in CSC1 and CSC2, which was responsible for their different anti-cancer response and prognosis. It's worth noting that difference in TIDE score between two clusters was not been observed. Although two clusters showed similar response to immunotherapy, CSC2 had higher score of T cell dysfunction and higher enrichment of MDSC, but lower score of CAF than CSC1. The function of T cells and infiltration levels of immunosuppressive cells (MDSC and CAF) can affect the response to immunotherapy (Tamura, 2018).

Given the discrepant clinical characteristics and molecular features between CSC1 and CSC2, we established a prognostic model based on DEGs between two clusters. Finally, we confirmed seven prognostic genes in the model, including DRD4, DPP7, UCN, INHBA, SFTA2, SYNPO2, and NXPH4. DRD4 belongs to dopamine receptor (DR) family that is associated with the progressive phenotypes of cancer (Wang et al., 2019). A machine learning study identified DRD4 as a survival-related candidate gene for CRC patients (Lee et al., 2022). DPP7 is a member of dipeptidyl peptidase (DPP) family, a high expression of which was related to a favorable prognosis in breast cancer (Choy et al., 2021). Ahluwalia et al. developed a four-gene signature where DPP7 was included for predicting survival of CRC patients (Ahluwalia et al., 2019). INHBA is a member of TGF-β superfamily and can accelerate migration and invasion of gastric cancer cells via TGF-β signaling pathway (Chen et al., 2019). INHBA was identified as an independent risk factor for both OS and DFS in colon cancer (Li et al., 2020). SFTA2 was also identified as a prognostic gene for colon cancer (Li et al., 2018; Gong et al., 2020). Other three genes were few reported in the relation with CRC.

The seven-gene prognostic model effectively classified CRC patients into two risk groups. Specifically, high-risk group showed evidently inferior OS and DSS than low-risk group. In addition to different prognosis, two risk groups also manifested

different activation of biological pathways and different response to immunotherapy and chemotherapy. Oncogenic pathways such as Wnt-β catenin, hypoxia, EMT, angiogenesis, Hedgehog signaling, and Notch signaling were more activated in high-risk group than in low-risk group. Simultaneously, risk score showed a positive correlation with the above pathways. High-risk group was less responsive to ICB therapy, resulting from T cell exclusion and dysfunction. Moreover, two risk groups performed different sensitivity to different chemotherapeutic drugs.

Conclusion

In conclusion, this study harnessed scRNA-seq data to identify CSC marker genes in CRC and demonstrated the important roles of CSC marker genes in CRC progression by delineating CSC-based subtyping (CSC1 and CSC2). The 29 CSC marker genes were considered as candidate genes for further exploring the mechanism of CSC in CRC. Importantly, we developed a seven-gene prognostic model for not only predicting OS and DSS of CRC patients, but also guiding immunotherapy and chemotherapy in clinics for CRC treatment.

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.

Author contributions

All authors contributed to this present work: JC, SW, YP, and YZ designed the study, YD, FR, and HG revised the manuscript, KZ, SH, ZW, and JL acquired the data and drafted the manuscript. All authors contributed to the article and approved the submitted version.

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

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2023.1200017/full#supplementary-material

References

Ahluwalia, P., Mondal, A. K., Bloomer, C., Fulzele, S., Jones, K., Ananth, S., et al. (2019). Identification and clinical validation of a novel 4 gene-signature with prognostic utility in colorectal cancer. *Int. J. Mol. Sci.* 20 (15), 3818. doi:10.3390/ijms20153818

Ayob, A. Z., and Ramasamy, T. S. (2018). Cancer stem cells as key drivers of tumour progression. *J. Biomed. Sci.* 25 (1), 20. doi:10.1186/s12929-018-0426-4

Baumann, M., Krause, M., and Hill, R. (2008). Exploring the role of cancer stem cells in radioresistance. *Nat. Rev. Cancer* 8 (7), 545–554. doi:10.1038/nrc2419

Bayik, D., and Lathia, J. D. (2021). Cancer stem cell-immune cell crosstalk in tumour progression. *Nat. Rev. Cancer* 21 (8), 526–536. doi:10.1038/s41568-021-00366-w

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17 (1), 218. doi:10.1186/s13059-016-1070-5

Bjerkvig, R., Tysnes, B. B., Aboody, K. S., Najbauer, J., and Terzis, A. J. (2005). Opinion: The origin of the cancer stem cell: Current controversies and new insights. *Nat. Rev. Cancer* 5 (11), 899–904. doi:10.1038/nrc1740

Blanche, P., Dartigues, J. F., and Jacqmin-Gadda, H. (2013). Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics Med.* 32 (30), 5381–5397. doi:10.1002/sim.5958

Brown, J. R., Chan, D. K., Shank, J. J., Griffith, K. A., Fan, H., Szulawski, R., et al. (2020). Phase II clinical trial of metformin as a cancer stem cell-targeting agent in ovarian cancer. *JCI insight* 5 (11), e133247. doi:10.1172/jci.insight.133247

Chan, J. M., Quintanal-Villalonga, Á., Gao, V. R., Xie, Y., Allaj, V., Chaudhary, O., et al. (2021). Signatures of plasticity, metastasis, and immunosuppression in an atlas of human small cell lung cancer. *Cancer Cell* 39 (11), 1479–1496.e18. doi:10.1016/j.ccell.2021.09.008

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Chen, Z. L., Qin, L., Peng, X. B., Hu, Y., and Liu, B. (2019). INHBA gene silencing inhibits gastric cancer cell migration and invasion by impeding activation of the TGF-β signaling pathway. *J. Cell. physiology* 234 (10), 18065–18074. doi:10.1002/jcp.28439

Choy, T. K., Wang, C. Y., Phan, N. N., Khoa, Ta H. D., Anuraga, G., Liu, Y. H., et al. (2021). Identification of dipeptidyl peptidase (DPP) family genes in clinical breast cancer patients via an integrated bioinformatics approach. *Diagn. (Basel, Switz.* 11 (7), 1204. doi:10.3390/diagnostics11071204

Clough, E., and Barrett, T. (2016). The gene expression Omnibus database. *Methods Mol. Biol. Clift. NJ)* 1418, 93–110. doi:10.1007/978-1-4939-3578-9_5

Colak, S., Zimberlin, C. D., Fessler, E., Hogdal, L., Prasetyanti, P. R., Grandela, C. M., et al. (2014). Decreased mitochondrial priming determines chemoresistance of colon cancer stem cells. *Cell death Differ.* 21 (7), 1170–1177. doi:10.1038/cdd.2014.37

Danilova, L., Ho, W. J., Zhu, Q., Vithayathil, T., De Jesus-Acosta, A., Azad, N. S., et al. (2019). Programmed cell death ligand-1 (PD-L1) and CD8 expression profiling identify an immunologic subtype of pancreatic ductal adenocarcinomas with favorable survival. *Cancer Immunol. Res.* 7 (6), 886–895. doi:10.1158/2326-6066.CIR-18-0822

Dienstmann, R., Vermeulen, L., Guinney, J., Kopetz, S., Tejpar, S., and Tabernero, J. (2017). Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat. Rev. Cancer* 17 (2), 268–292. doi:10.1038/nrc.2017.24

Dylla, S. J., Beviglia, L., Park, I. K., Chartier, C., Raval, J., Ngan, L., et al. (2008). Colorectal cancer stem cells are enriched in xenogeneic tumors following chemotherapy. *PloS one* 3 (6), e2428. doi:10.1371/journal.pone.0002428

Fan, Q. M., Jing, Y. Y., Yu, G. F., Kou, X. R., Ye, F., Gao, L., et al. (2014). Tumor-associated macrophages promote cancer stem cell-like properties via transforming growth factor-beta1-induced epithelial-mesenchymal transition in hepatocellular carcinoma. *Cancer Lett.* 352 (2), 160–168. doi:10.1016/j.canlet.2014.05.008

Fedewa, S. A., Flanders, W. D., Ward, K. C., Lin, C. C., Jemal, A., Goding Sauer, A., et al. (2017). Racial and ethnic disparities in interval colorectal cancer incidence: A population-based cohort study. *Ann. Intern. Med.* 166 (12), 857–866. doi:10.7326/M16-1154

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

Geeleher, P., Cox, N., and Huang, R. S. (2014). pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PloS one* 9 (9), e107468. doi:10.1371/journal.pone.0107468

Gong, B., Kao, Y., Zhang, C., Sun, F., Gong, Z., and Chen, J. (2020). Identification of hub genes related to carcinogenesis and prognosis in colorectal cancer based on integrated bioinformatics. *Mediat. Inflamm.* 2020, 5934821. doi:10.1155/2020/5934821

Gribov, A., Sill, M., Lück, S., Rücker, F., Döhner, K., Bullinger, L., et al. (2010). Seurat: Visual analytics for the integrated analysis of microarray data. *BMC Med. genomics* 3, 21. doi:10.1186/1755-8794-3-21

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

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

Hong, D. S., Morris, V. K., El Osta, B., Sorokin, A. V., Janku, F., Fu, S., et al. (2016). Phase IB study of vemurafenib in combination with irinotecan and cetuximab in patients with metastatic colorectal cancer with BRAFV600E mutation. *Cancer Discov.* 6 (12), 1352–1365. doi:10.1158/2159-8290.CD-16-0050

Jahanafrooz, Z., Mosafer, J., Akbari, M., Hashemzaei, M., Mokhtarzadeh, A., and Baradaran, B. (2020). Colon cancer therapy by focusing on colon cancer stem cells and their tumor microenvironment. *J. Cell. physiology* 235 (5), 4153–4166. doi:10.1002/jcp.29337

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

Jinushi, M., Chiba, S., Yoshiyama, H., Masutomi, K., Kinoshita, I., Dosaka-Akita, H., et al. (2011). Tumor-associated macrophages regulate tumorigenicity and anticancer drug responses of cancer stem/initiating cells. *Proc. Natl. Acad. Sci. U. S. A.* 108 (30), 12425–12430. doi:10.1073/pnas.1106645108

Khosravi, N., Mokhtarzadeh, A., Baghbanzadeh, A., Hajiasgharzadeh, K., Shahgoli, V. K., Hemmat, N., et al. (2020). Immune checkpoints in tumor microenvironment and their relevance to the development of cancer stem cells. *Life Sci.* 256, 118005. doi:10.1016/j.lfs.2020.118005

Kong, D., Li, Y., Wang, Z., and Sarkar, F. H. (2011). Cancer stem cells and epithelial-to-mesenchymal transition (EMT)-Phenotypic cells: Are they cousins or twins? *Cancers* 3 (1), 716–729. doi:10.3390/cancers30100716

Laiyemo, A. O., Doubeni, C., Pinsky, P. F., Doria-Rose, V. P., Bresalier, R., Lamerato, L. E., et al. (2010). Race and colorectal cancer disparities: Health-care utilization vs different cancer susceptibilities. *J. Natl. Cancer Inst.* 102 (8), 538–546. doi:10.1093/jnci/djq068

Lee, C. J., Baek, B., Cho, S. H., Jang, T. Y., Jeon, S. E., Lee, S., et al. (2022). Machine learning with *in silico* analysis markedly improves survival prediction modeling in colon cancer patients. *Cancer Med.* 12, 7603. doi:10.1002/cam4.5420

Lee, J. J., Bernard, V., Semaan, A., Monberg, M. E., Huang, J., Stephens, B. M., et al. (2021). Elucidation of tumor-stromal heterogeneity and the ligand-receptor interactome by single-cell transcriptomics in real-world pancreatic cancer biopsies. *Clin. cancer Res. official J. Am. Assoc. Cancer Res.* 27 (21), 5912–5921. doi:10.1158/1078-0432.CCR-20-3925

Li, C., Shen, Z., Zhou, Y., and Yu, W. (2018). Independent prognostic genes and mechanism investigation for colon cancer. *Biol. Res.* 51 (1), 10. doi:10.1186/s40659-018-0158-7

Li, X., Yu, W., Liang, C., Xu, Y., Zhang, M., Ding, X., et al. (2020). INHBA is a prognostic predictor for patients with colon adenocarcinoma. *BMC cancer* 20 (1), 305. doi:10.1186/s12885-020-06743-2

Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. *Nucleic acids Res.* 47 (W1), W199–w205. doi:10.1093/nar/gkz401

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Mahtab, M., Boavida, A., Santos, D., and Pisani, F. M. (2021). The Genome stability maintenance DNA helicase DDX11 and its role in cancer. *Genes* 12 (3), 395. doi:10.3390/genes12030395

Malta, T. M., Sokolov, A., Gentles, A. J., Burzykowski, T., Poisson, L., Weinstein, J. N., et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* 173 (2), 338–354.e15. doi:10.1016/j.cell.2018.03.034

Murphy, G., Devesa, S. S., Cross, A. J., Inskip, P. D., McGlynn, K. A., and Cook, M. B. (2011). Sex disparities in colorectal cancer incidence by anatomic subsite, race and age. *Int. J. cancer* 128 (7), 1668–1675. doi:10.1002/ijc.25481

Nagaraja, G. M., Kaur, P., and Asea, A. (2012). Role of human and mouse HspB1 in metastasis. *Curr. Mol. Med.* 12 (9), 1142–1150. doi:10.2174/156652412803306701

Ouzounova, M., Lee, E., Piranlioglu, R., El Andaloussi, A., Kolhe, R., Demirci, M. F., et al. (2017). Monocytic and granulocytic myeloid derived suppressor cells differentially regulate spatiotemporal tumour plasticity during metastatic cascade. *Nat. Commun.* 8, 14979. doi:10.1038/ncomms14979

Peng, D., Tanikawa, T., Li, W., Zhao, L., Vatan, L., Szeliga, W., et al. (2016). Myeloid-derived suppressor cells endow stem-like qualities to breast cancer cells through IL6/STAT3 and NO/NOTCH cross-talk signaling. *Cancer Res.* 76 (11), 3156–3165. doi:10.1158/0008-5472.CAN-15-2528

Peng, J., Sun, B. F., Chen, C. Y., Zhou, J. Y., Chen, Y. S., Chen, H., et al. (2019). Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* 29 (9), 725–738. doi:10.1038/s41422-019-0195-y

Prager, B. C., Xie, Q., Bao, S., and Rich, J. N. (2019). Cancer stem cells: The architects of the tumor ecosystem. *Cell stem Cell* 24 (1), 41–53. doi:10.1016/j.stem.2018.12.009

Rahman, M. R., Islam, T., Gov, E., Turanli, B., Gulfidan, G., Shahjaman, M., et al. (2019). Identification of prognostic biomarker signatures and candidate drugs in colorectal cancer: Insights from systems biology analysis. *Med. Kaunas. Lith.* 55 (1), 20. doi:10.3390/medicina55010020

Ricci-Vitiani, L., Pallini, R., Biffoni, M., Todaro, M., Invernici, G., Cenci, T., et al. (2010). Tumour vascularization via endothelial differentiation of glioblastoma stem-like cells. *Nature* 468 (7325), 824–828. doi:10.1038/nature09557

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Shao, K., Pu, W., Zhang, J., Guo, S., Qian, F., Glurich, I., et al. (2021). DNA hypermethylation contributes to colorectal cancer metastasis by regulating the binding of CEBPB and TFCP2 to the CPEB1 promoter. *Clin. epigenetics* 13 (1), 89. doi:10.1186/s13148-021-01071-z

Shen, W., Song, Z., Xiao, Z., Huang, M., Shen, D., Gao, P., et al. (2022). Sangerbox: A comprehensive, interaction-friendly clinical bioinformatics analysis platform. *iMeta* 1 (3), e36.

Siegel, R. L., Miller, K. D., Goding Sauer, A., Fedewa, S. A., Butterly, L. F., Anderson, J. C., et al. (2020). Colorectal cancer statistics, 2020. *CA a cancer J. Clin.* 70 (3), 145–164. doi:10.3322/caac.21601

Siegel, R. L., Miller, K. D., and Jemal, A., 2019. Cancer statistics, *CA a cancer J. Clin.*, 69(1):7–34. doi:10.3322/caac.21551

Su, C., Lv, Y., Lu, W., Yu, Z., Ye, Y., Guo, B., et al. (2021). Single-cell RNA sequencing in multiple pathologic types of renal cell carcinoma revealed novel potential tumor-specific markers. *Front. Oncol.* 11, 719564. doi:10.3389/fonc.2021.719564

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA a cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Tamura, H. (2018). Immunopathogenesis and immunotherapy of multiple myeloma. *Int. J. Hematol.* 107 (3), 278–285. doi:10.1007/s12185-018-2405-7

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). The cancer Genome atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol. (Poznan, Pol.* 19 (1a), A68–A77. doi:10.5114/wo.2014.47136

Wan, S., Zhao, E., Kryczek, I., Vatan, L., Sadovskaya, A., Ludema, G., et al. (2014). Tumor-associated macrophages produce interleukin 6 and signal via STAT3 to promote expansion of human hepatocellular carcinoma stem cells. *Gastroenterology* 147 (6), 1393–1404. doi:10.1053/j.gastro.2014.08.039

Wang, X., Wang, Z. B., Luo, C., Mao, X. Y., Li, X., Yin, J. Y., et al. (2019). The prospective value of dopamine receptors on bio-behavior of tumor. *J. Cancer* 10 (7), 1622–1632. doi:10.7150/jca.27780

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinforma. Oxf. Engl.* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Xiong, Y. Q., Sun, H. C., Zhang, W., Zhu, X. D., Zhuang, P. Y., Zhang, J. B., et al. (2009). Human hepatocellular carcinoma tumor-derived endothelial cells manifest increased angiogenesis capability and drug resistance compared with normal endothelial cells. *Clin. cancer Res. official J. Am. Assoc. Cancer Res.* 15 (15), 4838–4846. doi:10.1158/1078-0432.CCR-08-2780

Yang, L., Shi, P., Zhao, G., Xu, J., Peng, W., Zhang, J., et al. (2020). Targeting cancer stem cell pathways for cancer therapy. *Signal Transduct. Target. Ther.* 5 (1), 8. doi:10.1038/s41392-020-0110-5

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

Zhang, S., Yang, X., Wang, L., and Zhang, C. (2018). Interplay between inflammatory tumor microenvironment and cancer stem cells. *Oncol. Lett.* 16 (1), 679–686. doi:10.3892/ol.2018.8716

Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., Deng, C., et al. (2019). CellMarker: A manually curated resource of cell markers in human and mouse. *Nucleic acids Res.* 47 (D1), D721-D728–d8. doi:10.1093/nar/gky900

Zhang, Z. (2016). Variable selection with stepwise and best subset approaches. *Ann. Transl. Med.* 4 (7), 136. doi:10.21037/atm.2016.03.35

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 13 June 2023
DOI 10.3389/fphar.2023.1218244

OPEN ACCESS

EDITED BY
Ouyang Chen,
Duke University, United States

REVIEWED BY
Chen Li,
Free University of Berlin, Germany
Zhengrui Li,
Shanghai Jiao Tong University, China
Biao Zhang,
Dalian Medical University, China

*CORRESPONDENCE
Tengfei Wang,
ti_3601@163.com

RECEIVED 06 May 2023
ACCEPTED 02 June 2023
PUBLISHED 13 June 2023

CITATION
Xu K, Liu Y, Luo H and Wang T (2023), Efferocytosis signatures as prognostic markers for revealing immune landscape and predicting immunotherapy response in hepatocellular carcinoma. *Front. Pharmacol.* 14:1218244. doi: 10.3389/fphar.2023.1218244

Efferocytosis signatures as prognostic markers for revealing immune landscape and predicting immunotherapy response in hepatocellular carcinoma

Ke Xu[1], Yu Liu[1], Huiyan Luo[1] and Tengfei Wang[2]*

[1]Department of Oncology, Chongqing General Hospital, Chongqing, China, [2]Department of Equipment, Bishan Hospital of Chongqing, Chongqing, China

Background: Hepatocellular carcinoma (HCC) is a highly lethal liver cancer with late diagnosis; therefore, the identification of new early biomarkers could help reduce mortality. Efferocytosis, a process in which one cell engulfs another cell, including macrophages, dendritic cells, NK cells, etc., plays a complex role in tumorigenesis, sometimes promoting and sometimes inhibiting tumor development. However, the role of efferocytosis-related genes (ERGs) in HCC progression has been poorly studied, and their regulatory effects in HCC immunotherapy and drug targeting have not been reported.

Methods: We downloaded efferocytosis-related genes from the Genecards database and screened for ERGs that showed significant expression changes between HCC and normal tissues and were associated with HCC prognosis. Machine learning algorithms were used to study prognostic gene features. CIBERSORT and pRRophetic R packages were used to evaluate the immune environment of HCC subtypes and predict treatment response. CCK-8 experiments conducted on HCC cells were used to assess the reliability of drug sensitivity prediction.

Results: We constructed a prognostic prediction model composed of six genes, and the ROC curve showed good predictive accuracy of the risk model. In addition, two ERG-related subgroups in HCC showed significant differences in tumor immune landscape, immune response, and prognostic stratification. The CCK-8 experiment conducted on HCC cells confirmed the reliability of drug sensitivity prediction.

Conclusion: Our study emphasizes the importance of efferocytosis in HCC progression. The risk model based on efferocytosis-related genes developed in our study provides a novel precision medicine approach for HCC patients, allowing clinicians to customize treatment plans based on unique patient characteristics. The results of our investigation carry noteworthy implications for the development of individualized treatment approaches involving immunotherapy and chemotherapy, thereby potentially facilitating the realization of personalized and more efficacious therapeutic interventions for HCC.

KEYWORDS

HCC, efferocytosis, immune, therapy, biomarker

1 Introduction

HCC is the most common type of liver cancer and accounts for a significant proportion of cancer-related deaths worldwide (Yang et al., 2019a). Despite advances in medical treatment, the overall survival (OS) of HCC patients remains suboptimal, and the precise molecular mechanisms underlying HCC prognosis are poorly understood. Currently, HCC prognostic models rely on clinical indicators such as grading and TNM staging, which may have limited accuracy (Icard et al., 2021; Zhai et al., 2022; Wang et al., 2023a; Conche et al., 2023). Therefore, it is imperative to identify novel and effective prognostic biomarkers for HCC, which may help to determine specific therapeutic targets. Molecular immune-targeted therapy represents a promising avenue for future HCC treatment.

"Efferocytosis" refers to the process by which one cell engulfs another cell, usually referring to macrophages engulfing apoptotic cells (Zhou et al., 2020a; Wang et al., 2023b). In addition, there are other cells such as neutrophils, which release DNA fiber networks during the inflammatory process and can engulf these DNA fiber networks and cell debris on DNA (Bukong et al., 2018; Lee et al., 2022); natural killer cells (NK cells), which are usually part of the immune system and can kill infected cells or cancer cells, and can also clear dead cells through efferocytosis (Jensen et al., 2020); malignant tumor cells, some of which can express efferocytosis-related receptors and ligands, enabling them to engulf surrounding apoptotic cells and evade the immune system's attack (Brightwell et al., 2016; Zhang et al., 2022); dendritic cells and some other immune cells also play a role in efferocytosis (Maschalidi et al., 2022; Nino-Castano et al., 2022). The function of efferocytosis in tumors is complex and can sometimes promote tumor development and other times inhibit it (Banerjee et al., 2021; Tajbakhsh et al., 2021; Lin et al., 2022). In the early stages of tumor development, efferocytosis can promote tumor growth by reducing immune system attacks and promoting the growth of tumor cells by clearing apoptotic cells around them. In addition, tumor cells can further promote tumor development by expressing efferocytosis-related receptors and ligands to evade the immune system's attack. However, in the late stages of tumor development, efferocytosis can inhibit tumor growth by promoting the immune system's attack on the tumor.

After engulfing apoptotic cells, antigens in the cells can be presented to T cells by macrophages, thereby activating T cells to attack the tumor (Lu et al., 2022; Zhou et al., 2023). Macrophages play a significant role in the progression of HCC. Recent investigations have revealed significant findings regarding the impact of tumor-derived alpha-fetoprotein (tAFP) on macrophage polarization and its influence on HCC cells. Specifically, tAFP has been shown to promote the differentiation of M0 macrophages into M2 macrophages, while concurrently suppressing the efferocytosis of M1 macrophages towards HCC cells (Zhang et al., 2023a). Polarization of M1 macrophages contributes to the protection against HCC, while M2 macrophages emerges as a prominent factor driving HCC development (Liu et al., 2022a). In view of this, it is necessary to study the role of efferocytosis in the progression of HCC. In addition, efferocytosis can also promote anti-inflammatory responses, thereby inhibiting tumor development (Zhou et al., 2020b). Therefore, the role of efferocytosis in tumors is different from its role in normal physiological conditions and needs to be analyzed according to specific circumstances. In the treatment of tumors, efferocytosis can be used as an important target for intervention to achieve treatment goals (Lahey et al., 2022; Mehrotra and Ravichandran, 2022). However, the role of efferocytosis-related genes in the progression and prognosis of HCC remains poorly understood.

We have developed a risk model based on six efferocytosis-related genes and identified two ERG-associated subtypes that exhibit significant differences in tumor immune landscape and prognostic stratification, highlighting the importance of efferocytosis status in HCC. Importantly, our study reveals patterns of immune therapy and chemotherapy response, and *in vitro* validation confirms the predictive ability of the prognostic model for drug response. These findings underscore the significance of efferocytosis in HCC and suggest potential therapeutic strategies for patients with different efferocytosis statuses. This study may provide a basis for future research on the mechanisms underlying HCC progression and treatment response, as well as inform clinical decision-making in HCC management.

2 Materials and methods

2.1 Acquisition of TCGA-LIHC data

The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/) has aggregated and scrutinized genomic, transcriptomic, epigenomic, and proteomic data obtained from thousands of individuals afflicted with various forms of cancer, culminating in an extensive data repository exceeding 2.5 petabytes. This compendium of knowledge has unveiled potential hereditary drivers of cancer, identified plausible pharmacological targets, and catalyzed the development of customized cancer therapeutics (Wang et al., 2016). We obtained the TCGA-LIHC cohort, comprising transcriptome data of 374 HCC tumor patients and 50 normal liver tissue samples, from TCGA. Additionally, clinical data of 374 HCC tumor patients were downloaded. After rigorous selection, we retained clinical data for a total of 370 HCC patients with comprehensive clinical information.

2.2 ERGs from genecards portal

GeneCards is a portal website and database that furnishes a wealth of information on more than 155,000 human genes, encompassing details on gene expression, function, protein domains, and interactions (Safran et al., 2021). Given its comprehensiveness and timeliness, GeneCards represents a valuable resource for investigating the intricacies of human genes and their implications for disease (Sun et al., 2023; Zhong et al., 2023). We employed the following approach to obtain the efferocytosis-related genes. Firstly, we utilized highly relevant keywords and gene descriptions provided in Genecards, such as "efferocytosis," "phagocytosis of apoptotic cells," and "clearance of dying cells." Subsequently, we reviewed the literature to carefully screen and manually confirm these keywords and descriptions to

ensure that the final selected genes are indeed closely related to the efferocytosis process. Finally, we obtained a total of 111 genes related to efferocytosis (ERGs) from the GeneCards database.

2.3 Prognostic ERGs signature identification

Through the use of univariate Cox regression analysis, we identified a set of 13 genes that displayed a significant correlation with the survival rates of patients with HCC. Then, optimal lambda (λ) was determined to be the ideal value by 10-fold cross validation when performing the LASSO Cox regression analysis to screen the core ERGs that were strongly linked with HCC patients' prognosis (Chi et al., 2022a; Wang et al., 2022). Using the "glmnet" R package, 6 core genes were subsequently utilized to create a risk signature (Engebretsen, 2019). The risk score was calculated by integrating the expression profile of ERGs with the paired multivariate Cox regression values (β) (Ni et al., 2022; Xu et al., 2022; Zhao et al., 2023a). Based on their respective gene expression profiles, we computed a risk score for each patient in the cohort as follows: Risk score = e(Exp.GAPDH*0.1481 + Exp.ADAM9*0.1581 + Exp.SIRT6*0.1247 + Exp.LGALS3*0.0666 - Exp.CD5L*0.0144 - Exp.IL33*0.0985).

2.4 Evaluating infiltration of immune cells

We employed the CIBERSORT and ssGSEA R scripts to assess the levels of infiltrating immune cells (Newman et al., 2015; Chi et al., 2023a). The CIBERSORT algorithm was used to calculate the immune cell type scores for individual samples, and then the corresponding scores for each sample were calculated based on the estimated immune cell type scores (Chi et al., 2022b). In addition, spearman correlation analysis was used to investigate the relationship between immune cell and risk scores. Using the immune cell profiles of HCC patients, we used the ssGSEA method to distinguish individuals classified as different risks (Zhao et al., 2023b).

2.5 Evaluating the accuracy of chemotherapy response predictions

We employed the "pRRophetic" R software package for evaluating the therapeutic response in patient subgroups classified as high-risk and low-risk, based on the half-maximal inhibitory concentration (IC50) values obtained from each individual with HCC from the Genomics of Drug Sensitivity in Cancer (GDSC) dataset (Geeleher et al., 2014; Chi et al., 2023b). Further, the transcriptional profiles of HCC cell lines were obtained from the CCLE website, and risk scores for different HCC cell lines were calculated using the ERGs risk scoring formula. Based on the computed results, Huh7 was identified as having a high risk score, while HepG2 exhibited a comparatively lower risk score. Then, the sensitivity of HCC cells to the drug was evaluated through implementation of the CCK-8 assay (Zhang et al., 2023b).

2.6 KEGG and GO analysis

Two frequently utilized bioinformatics resources for investigating the functional and metabolic pathways of genes and proteins, as well as other biological features, are the KEGG and GO databases. Annotations provided by these tools can facilitate a more comprehensive comprehension of gene and protein function, ultimately leading to enhanced insights into gene expression and metabolic regulation. In this study, we performed enrichment analysis using Gene Set Variation Analysis (GSVA) and utilized the "c2.cp.kegg.v7.4.symbols.gmt" data set derived from the MSigDB database (Hanzelmann et al., 2013; Liu et al., 2023a).

2.7 Statistical analysis

All data analyses were conducted using R version 4.1.3. For variables that exhibited a normal distribution, the Student's t-test was employed, whereas Pearson's correlation coefficient was used to evaluate the association between variables. The levels of statistical significance were set at $p < 0.05$*, $p < 0.01$**, and $p < 0.001$***, respectively.

3 Results

3.1 Efferocytosis-based gene signature construction

We retrieved 111 genes associated with efferocytosis from the Genecards website. The HCC dataset comprising 370 tumor samples and 50 adjacent normal tissue samples was sourced from the TCGA database. We employed the "limma" R package to identify ERGs that were differentially expressed between HCC tumor and adjacent normal samples. This analysis identified 20 ERGs with significant differences (Figure 1A). Next, we utilized the "survival" and "survminer" R packages to investigate the association between ERGs and survival in HCC patients. Thirteen out of the 20 ERGs were significantly linked to survival in HCC patients based on a *p*-value cutoff of less than 0.05 and a km score less than 0.05 (Figure 1B). All ERGs except for CD5L, PLG, and IL33 were found to be poor prognostic factors. To develop an HCC prognostic model, we conducted Lasso analysis using these 13 ERGs (Figure 1C; Supplementary Table S1). The time-dependent ROC curve illustrated the favorable predictive accuracy of the model at 1, 3, and 5 years (Figure 1F). Based on the median riskscore, we divided the 370 HCC patients into high-risk and low-risk subgroups, and the high-risk subgroup displayed a shorter overall survival time than the low-risk subgroup (Figure 1E), with median survival times of 2.7 and 6.7 years, respectively. Furthermore, we generated a heatmap to depict the expression levels of the top 10 ERGs in various riskscore groups (Figure 1D).

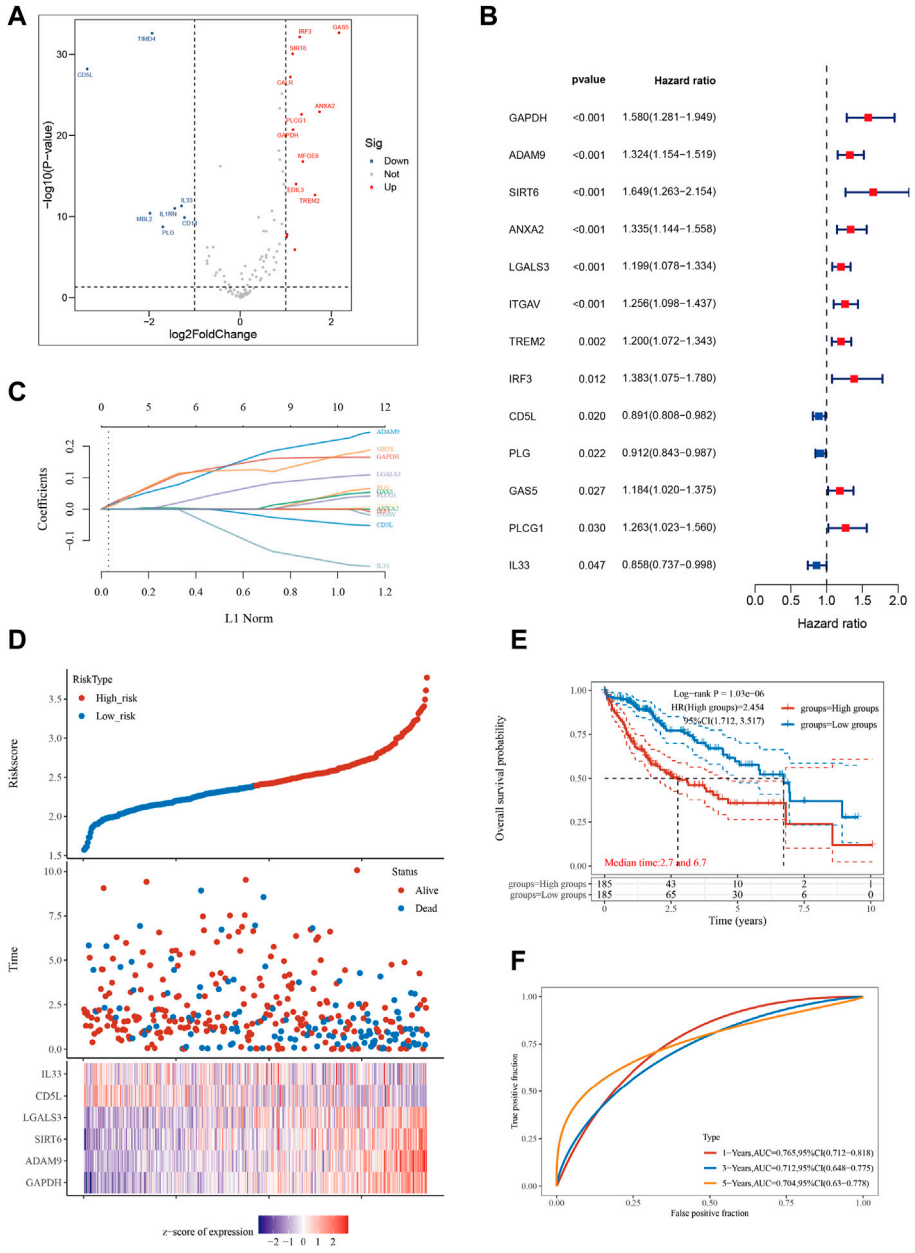


FIGURE 1
Constructing a prognostic model based on efferocytosis-related genes in HCC. **(A)** Differential gene screening was conducted to identify ERGs associated with hepatocellular carcinoma (HCC). **(B)** 13 genes of prognostic significance, which we refer to as ERGs, were identified from the differential gene screening analysis. These ERGs demonstrated an association with survival in HCC patients. **(C)** Utilizing the Lasso method, a prognostic model was constructed based on the identified ERGs. **(D)** The risk scores, survival status, and expression levels of the top 6-ERGs were plotted to visualize the distribution of prognostic risk. **(E)** Kaplan-Meier (KM) analysis was performed to further investigate the prognostic significance of the 6-ERGs in different HCC subtypes. **(F)** The predictive efficiency of the prognostic model was evaluated using ROC analysis.

3.2 ERGs expression variations among subtypes

Using mRNA expression levels as a metric, we proceeded to assess the expression levels of the six ERGs in both normal and tumor tissues (Figure 2A). Notably, we observed a significantly higher expression of all six ERGs in tumor tissues compared to their adjacent non-tumor counterparts ($p < 0.001$), with GAPDH exhibiting the highest level of expression. To further elucidate the biological significance of these findings, we also examined the expression levels of the six ERGs in high-risk versus low-risk subgroups. Interestingly, we found that the expression trend of the six ERGs in this subgroup mirrored that of Figure 1A (Figure 2B). Moreover, we employed Kaplan-Meier curves to establish the correlation between each key ERG gene and the prognosis of HCC patients, and our analysis revealed that all six ERGs were significantly linked to a poor prognosis ($p < 0.05$).

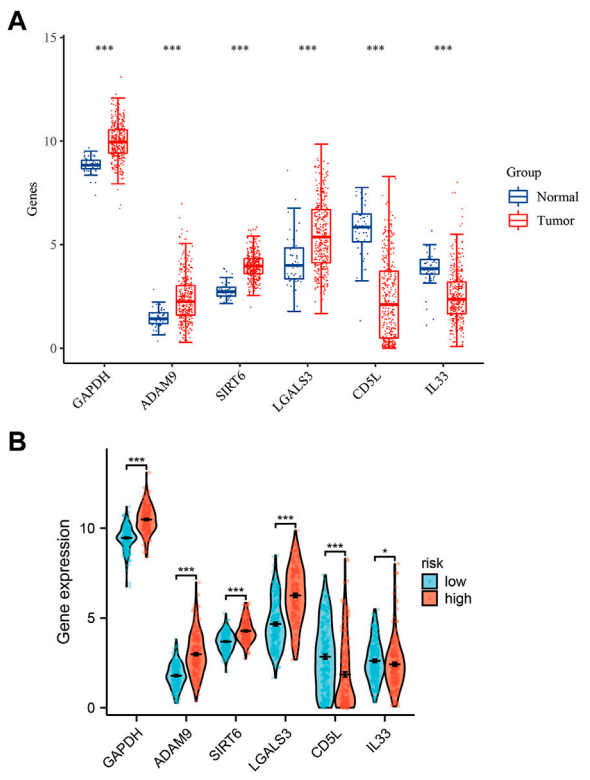


FIGURE 2
Expression levels of 6-ERGs. **(A)** Expression levels of 6-ERGs in HCC tumor tissues and adjacent tissues. **(B)** Expression levels of 6-ERGs in HCC risk subgroups. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

3.3 GO and KEGG enrichment analysis

We examined the impact of various signaling pathway activations on the growth and progression of tumor cells, as well as their effect on the tumor microenvironment. To identify genes that were differentially expressed between high-risk and low-risk groups, we conducted a comparative analysis of gene expression levels (Figure 3A). In high-risk patients with HCC, several pathways including Cytoplasmic, Ribosomal Protein, Eukaryotic Translation Elongation, Developmental Biology, and Infectious Disease were significantly enriched (Figure 3B). Additionally, our Gene Ontology enrichment analysis revealed that the humoral immune response process was notably upregulated in the high-risk subgroup (Figure 3C). Furthermore, we investigated the GO pathways that corresponded to the most differentially expressed genes between the high-risk and low-risk subgroups (Figures 3D, E).

3.4 Differential immune infiltration levels in HCC patients with diverse risk profiles

Using the Lasso method, we employed dimensionality reduction and clustering on a set of 6-ERGs selected for HCC patients, and our results indicate that these 6-ERGs effectively differentiate between HCC patients of varying risk levels (Figure 4A). Subsequently, we investigated the immune infiltration patterns in HCC patients with distinct prognostic risks (Figures 4B, C). The riskscore values were sorted in ascending order to represent the proportion of each immune cell type (Figure 4B). Remarkably, our analyses revealed significant infiltration of Macrophage M2, activated CD4 memory T cells, and Tregs in HCC patients classified as high-risk, whereas Macrophage M1 was notably decreased in this group (Figure 4C). Additionally, Neutrophils were increased in the high-risk group, suggesting that the HCC patients with high-risk scores may be experiencing an immune-suppressed state, which may be associated with immune checkpoint expression. During our analysis of HCC, we discovered noteworthy distinctions in the expression of both macrophage M1 and macrophage M2 between the high- and low-risk subgroups.

In order to explore this finding further, we sought to investigate the association between macrophage M1 and macrophage M2, and ERGs which have been linked to poor prognosis (Figure 4D). According to the data presented in Figure 4D, the expression levels of ADAM9, GAPDH, and LGALS3 were observed to be positively associated with the abundance of Macrophage M2, while conversely associated with the levels of Macrophage M1. Furthermore, we conducted a more detailed examination of the relationship between the 6-ERGs and immune cells (Figure 5A). Notably, we observed that ADAM9 and GAPDH expression levels exhibited associations with the concentrations of several distinct immune cell types (Figures 5B, C).

3.5 Investigating the correlation between ERG expression and immunotherapy efficacy

The results of the prior analysis indicate that high-risk and low-risk groups display dissimilar immune microenvironments, characterized by increased infiltration of Tregs, activated CD4 memory T cells, and Macrophage M2 in the high-risk group. These changes create an immunosuppressive microenvironment, which influences the efficacy of immunotherapy differently between the groups. Notably, patients with elevated expression levels of 6-ERGs are more likely to respond to Anti-PD-L1 and Anti-PD-1 therapy (Figure 6). Moreover, 6-ERGs can serve as a predictive tool for the accuracy of immune checkpoint blockade (ICB) in HCC patients (Figures 6A, B). ADAM9 expression gradually increases in cancer tissue and is recognized as a negative prognostic biomarker for prostate cancer patients. Elevated ADAM9 expression is shown to regulate the inflammatory state of the tissue by modulating the efferocytosis of macrophages *in vitro* and *in vivo*. In HCC patient tissues, ADAM9 expression is significantly upregulated (Figure 2A), indicating a higher immune response compared to lower ADAM9 expression subgroups (Figure 7A). To investigate the response of high-risk and low-risk HCC patients to ICB, we used the TIDE algorithm to combine ADAM9 expression levels with HBV infection factors (Figure 7B). We found that high ADAM9 expression predicts a higher immune response score, independent of HBV infection status. Given the effect of ADAM9 on the immune response score, we further explored the expression levels of immune checkpoints in HCC patients with different ERGs riskscores. Surprisingly, we observed that most

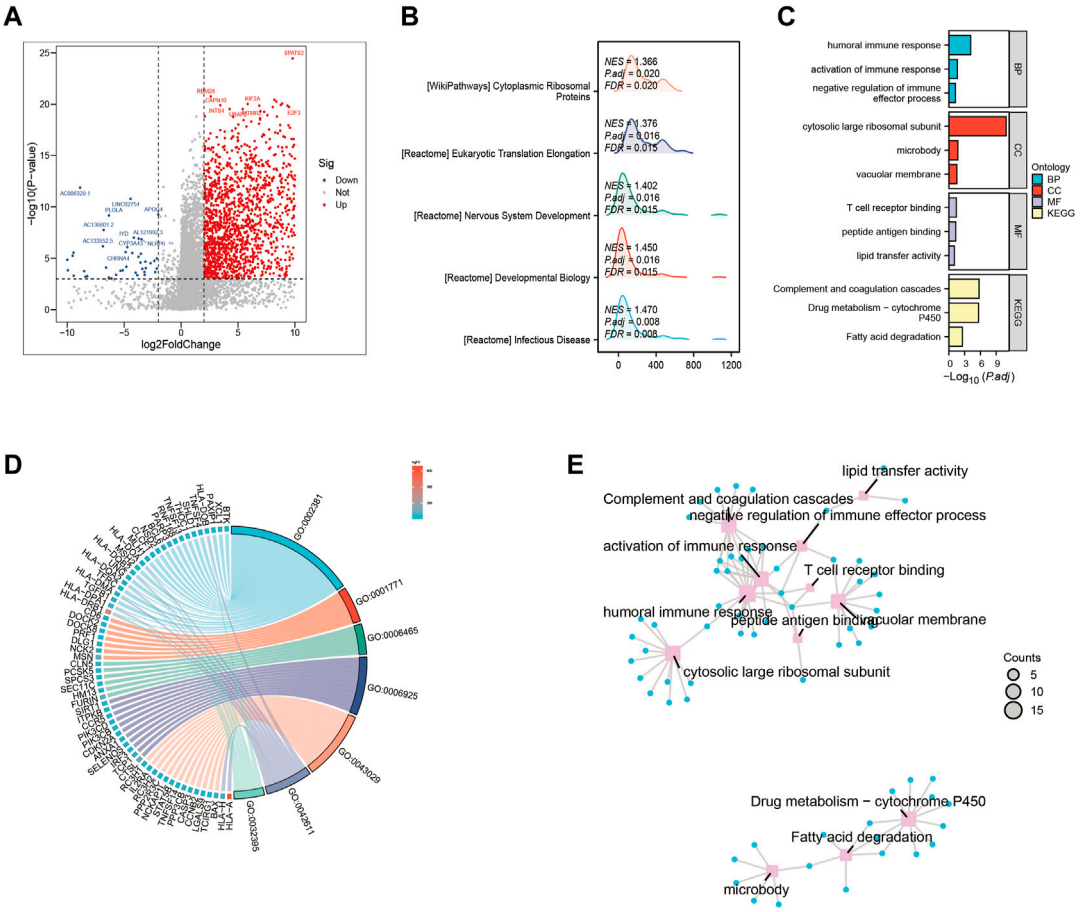


FIGURE 3
Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis **(A)** Volcano map screening for differential genes. **(B)** Mountain map showing the enriched KEGG pathway. **(C–E)** GO enrichment analysis.

immune checkpoints, including PDCD1, CTLA4, and PDCD1LG2, are significantly upregulated in the high ADAM9 expression group (Figures 7C, D). Finally, we employed Cibersort to analyze the level of immune cell infiltration in tissue samples categorized into normal, low-risk, and high-risk groups (Figures 7E, F), which revealed significant differences in infiltration levels among the groups.

3.6 Prediction and authentication of drug sensitivity

To develop targeted therapies for patients with HCC, we investigated variations in chemotherapy drug sensitivity between subgroups with high- and low-risk scores. Our analysis compared the IC50 levels of sixteen chemotherapy drugs in the high-risk score and low-risk score subgroups (Figure 8). The results revealed significant differences in IC50 values for some drugs, such as Etoposide, suggesting that patients with high-risk scores may be more responsive to this type of chemotherapy (Supplementary Table S1). To further validate our findings, we assessed riskscores in various HCC cell lines based on gene expression profiles (Figure 9; Supplementary Table S2). For drug sensitivity experiments, we selected Huh7 and HepG2 cell lines to represent the high-risk score and low-risk score subgroups of HCC patients, respectively. Consistent with the drug sensitivity prediction results, our CCK-8 assay data showed that Huh7 cells with a high-risk score were more sensitive to Etoposide than HepG2 cells (Figure 10A), supporting the notion that this chemotherapy drug may be a promising candidate for precision therapy in HCC patients (Figure 10B).

4 Discussion

While the rise in the incidence of HCC has shown a decelerating trend in recent times, the morbidity and mortality associated with this disease remain significant (Yang et al., 2019b; Sung et al., 2021). As per current estimates, over 70% of patients who undergo radical resection experience recurrence of the disease within 5 years (Xu et al., 2019; Zhou et al., 2020c). Given these challenges, developing an accurate predictive model for postoperative recurrence and identification of HCC patients with a reduced overall survival is crucial for optimal clinical decision-making and prognostic outcomes.

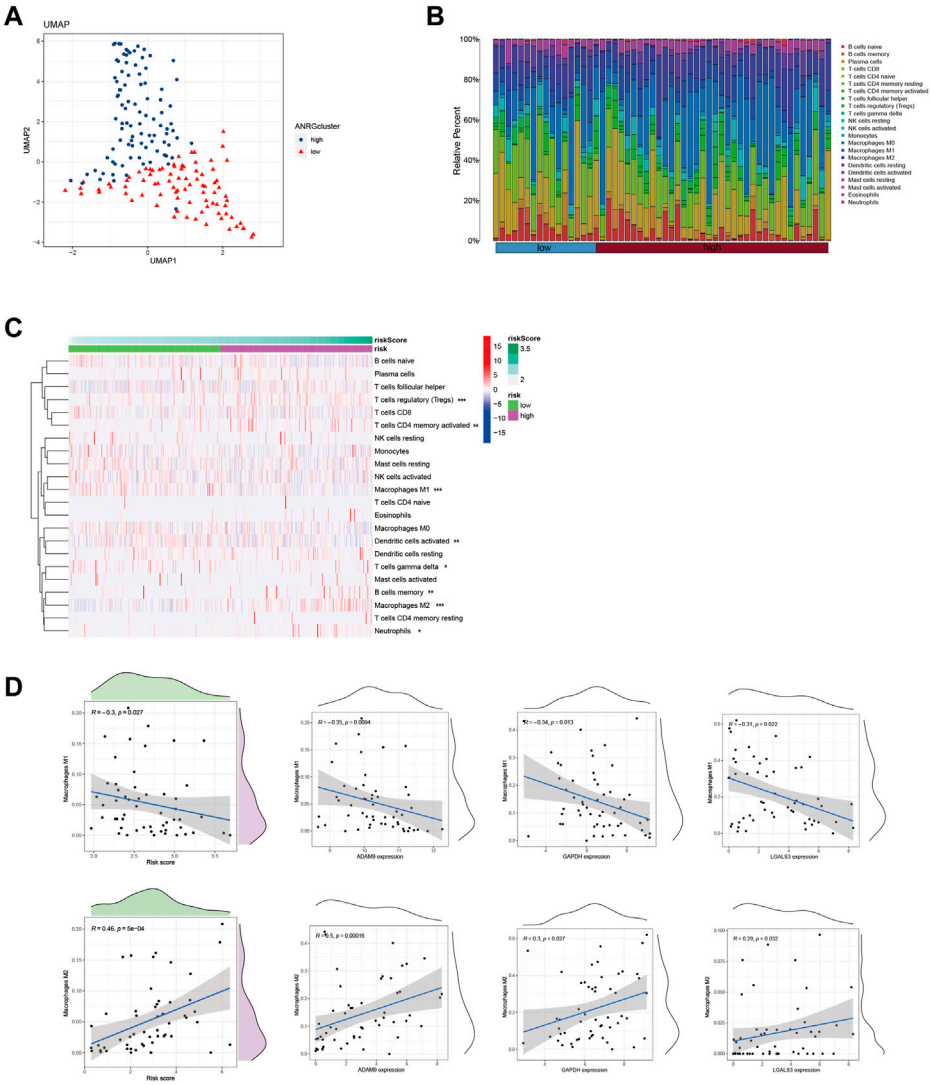


FIGURE 4
Identify immune landscape of HCC based on efferocytosis-associated signature. **(A)** UMAP demonstrates different immune profiles among HCC subgroups. **(B)** Proportion of immune cells in HCC tissues. **(C)** Differences in immune infiltration between HCC subgroups. **(D)** Correlation between immune cells and 6-ERGs.

Efferocytosis is an important process in the immune system, which maintains tissue health by clearing apoptotic cells. It plays a crucial role in both normal physiology and pathological conditions, particularly in tumor progression (Morioka et al., 2019; Myers et al., 2019; Yang et al., 2022). In HCC tissue, tumor cells continuously proliferate, die, and undergo apoptosis, leading to the release of various cytokines and chemicals that trigger inflammatory reactions and immune responses. These reactions lead to the accumulation of macrophages, dendritic cells, and NK cells, which control inflammation and immune responses by clearing apoptotic cells (Yang et al., 2019a; Garcia-Pras et al., 2021; Leone et al., 2021). However, tumor cells can exploit the mechanism of efferocytosis to evade immune system attack (Werfel and Cook, 2018). Studies have shown that tumor cells can express efferocytosis-related receptors and ligands, which attract immune cells to clear apoptotic cells around them and evade immune system attack by interfering with the activation of M1 macrophages and increasing the number of M2 macrophages (Graham et al., 2014; Poon et al., 2014; Soki et al., 2014). Therefore, efferocytosis plays an important role in tumor progression. It has been demonstrated regulation of efferocytosis processes enhances immune cell attack on tumors and increases apoptosis of tumor cells, thereby delaying tumor growth and metastasis (Vaught et al., 2015). Therefore, understanding the relationship between efferocytosis and tumors is of great significance for the development of more effective tumor treatment methods.

Utilizing machine learning to construct a reliable prognostic model based on known efferocytosis-related genes is essential for improving the accuracy of personalized diagnosis and treatment prediction for patients with HCC. Such an approach holds great potential for enhancing the clinical management of HCC patients. We conducted an investigation to identify differential genes from

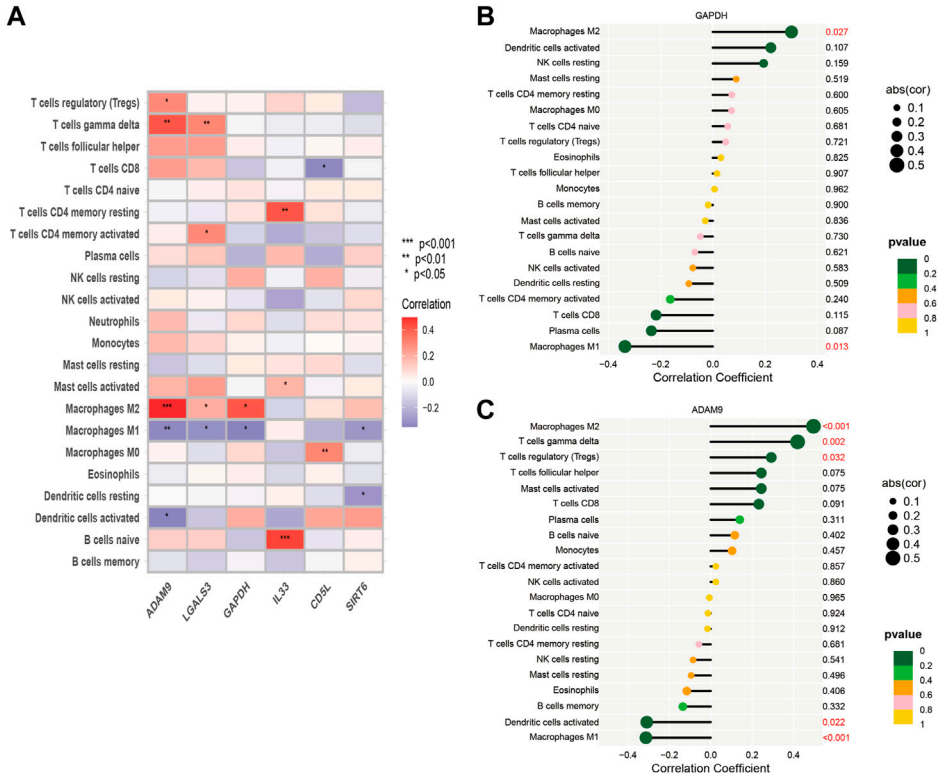


FIGURE 5
Correlation between immune cells and 6-ERGs. **(A)** Heatmap was used to show the correlation between immune cells and 6-efferocytosis genes (6-ERGs). **(B,C)** Bar plots were used to illustrate the relationship between GAPDH and ADAM9 with immune cell infiltration.

111 ERGs and subsequently explored their potential prognostic relevance (Figures 1A, B). The identified differential genes were regarded as promising markers with the potential to influence the survival outcomes of HCC patients. Six ERGs, specifically ADAM9, GAPDH, SIRT6, LGALS3, CD5L, and IL33, were selected for the development of an HCC prognostic model (Figure 1C). Our model accurately predicted overall survival of HCC patients at 1, 3, and 5 years, demonstrating its robust predictive ability (Figures 1D–F).

Cell death frequently occurs in solid tumors during malignant progression, and is influenced by the tumor microenvironment (TME), which plays a crucial role in tumor heterogeneity and tumorigenesis (Roy et al., 2018; Gadiyar et al., 2020; Lahey et al., 2022; Li et al., 2023a; Sandri et al., 2023). The nature of cell death and the mechanisms involved in corpse clearance can significantly impact the immune phenotype within the TME (Werfel et al., 2019). The process of efferocytosis, which clears dying cell corpses in the TME, has conventionally been viewed as immunosuppressive (Poon et al., 2014). Our investigation revealed that individuals exhibiting high ERG levels exhibited elevated infiltration of Tregs, activated CD4 memory T cells, Macrophages M2, and neutrophils, in comparison to those with low ERG levels (Figure 4C). Notably, Tregs have been implicated in regulating immune response during the development of HCC, from the early stages to advanced disease. Additionally, Tregs may exert a suppressive influence on liver function, thereby contributing to the emergence of primary liver cancer. The role of M2 macrophages in promoting tumor progression has been widely explored in HCC. The polarization of TAMs towards the M2 macrophage activation state plays a crucial role in anti-inflammatory and pro-tumor activity during tumor progression, which is in stark contrast to the cytocidal and tumoricidal properties of M1 activation states. These fuctions of TAMs partly explain the pronounced enrichment of M2 macrophages in patients with high ERGs, while a divergent trend was observed for M1 macrophages (Figure 4D). The results provide insight into the potential role of macrophage polarization in the prognosis of HCC patients and suggest that targeting ERGs could be a promising therapeutic strategy for the treatment of HCC. However, additional research is needed to elucidate the underlying mechanisms of this relationship and to further evaluate the clinical relevance of these findings.

Immunotherapy has become an essential therapeutic strategy for cancer and has been extensively investigated (Gong et al., 2022; Jin et al., 2022; Zhao et al., 2022; Wang et al., 2023c). This approach involves leveraging the immune system to recognize and eradicate cancerous cells. Numerous types of immune therapies, such as checkpoint inhibitors, adoptive cell transfer, and cancer vaccines, have been developed (Llovet et al., 2022; Peng et al., 2022; Liu et al., 2023b). PD-1 and PD-L1 have been closely associated with macrophages (Liu et al., 2018). Moreover, Abrogation of Efferocytosis leads to diminished immunosuppressive characteristics of macrophages, as evidenced by a reduction in the expression of M2-associated markers such as PD-L1 and PD-L2 (Cruz Cruz et al., 2023). Our study results demonstrate that high ERGs group exhibited elevated PD-1 and PD-L1 expression levels in

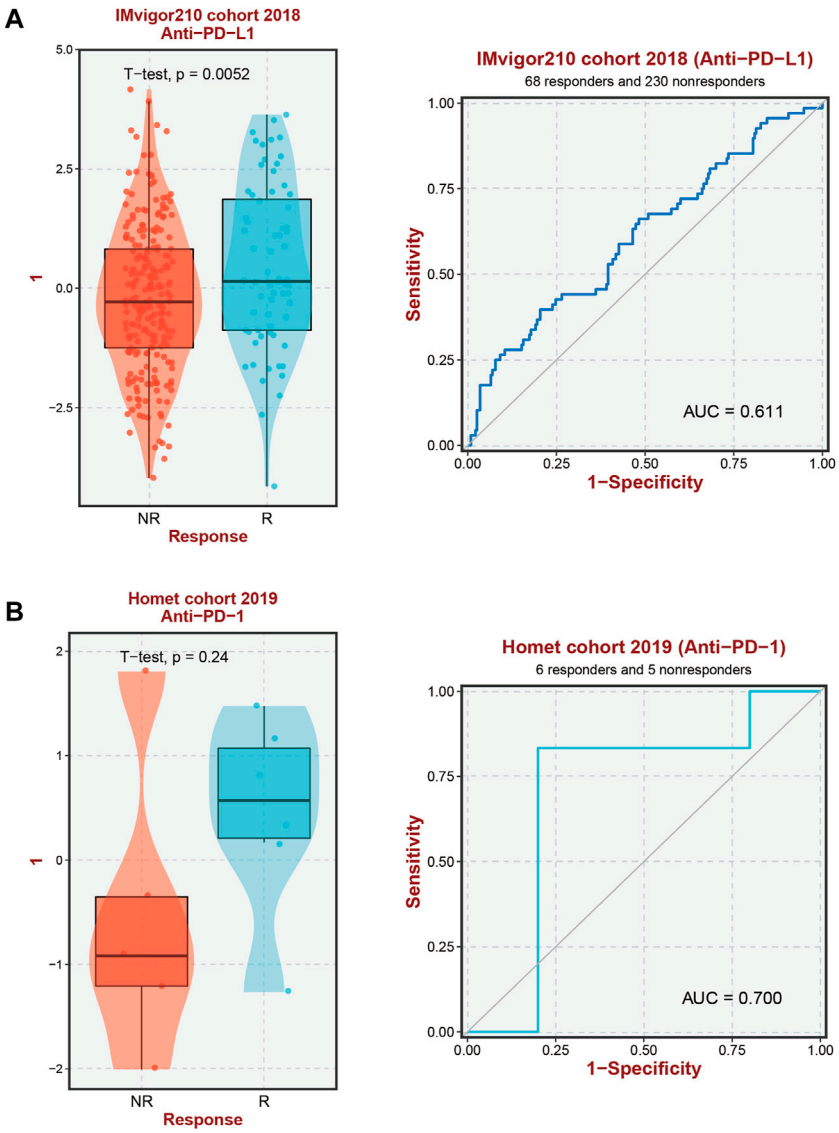


FIGURE 6
Immunotherapy response prediction. **(A)** Prediction of immune therapy response to anti-PD-L1 treatment in HCC patients based on 6-ERGs. **(B)** Prediction of immune therapy response to anti-PD-1 treatment in HCC patients based on 6-ERGs.

comparison with the low ERGs group, which could be linked to the macrophage infiltration phenomenon (Figures 7C, D). This observation provides a partial rationale for the superior efficacy of anti-PD-1 and anti-PD-L1 immunotherapy among high ERGs group patients with HCC, emphasizing the important value of the ERGs model in immunotherapy strategies (Figure 6). To further advance the clinical application of this model for predicting anti-PD-1 and PD-L1 response, our subsequent investigations will focus on the evidence from *in vivo* immunotherapy experiments. It is crucial and valuable to assess the accuracy of the ERGs-based model in predicting immunotherapy response by utilizing different risk-scored HCC cell lines or ERGs gene knockout mice. HBV, one of the etiological factors contributing to hepatocellular carcinoma (HCC), exerts a significant impact on the progression of this malignancy. However, in our study, HBV-positive HCC patients did not exhibit significant differences in immune therapy response scores compared to HBV-negative patients (Figure 7B). We postulate that this observation may be closely associated with the viral load of HBV. Despite both groups being HBV-positive HCC patients, variations in viral load could lead to divergent responses to immune therapy. Therefore, overall, HBV-positive HCC patients may demonstrate immune response outcomes comparable to those of HCC patients without HBV infection.

In addition to the interplay between tumor-immune cells, disrupted pathways within tumor cells can also affect the advancement of HCC (Llovet et al., 2018). To gain deeper insights into the differences in pathway enrichment among HCC patients with varying risks, we performed GO and KEGG analyses (Figure 3C). Our findings revealed significant differences in the enrichment of immune response pathways

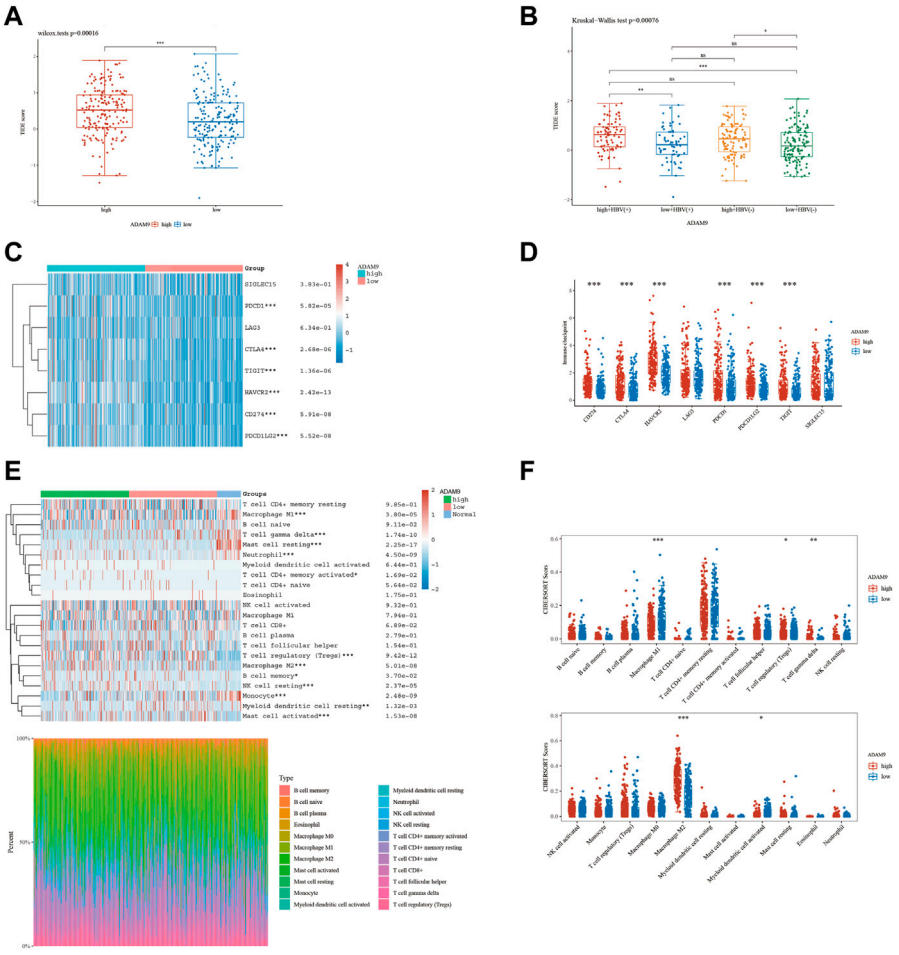


FIGURE 7
The level of immune checkpoint in HCC subtypes. **(A)** The TIDE score between ADAM9 subgroups. **(B)** HBV infection does not affect the effectiveness of immunotherapy. **(C,D)** There are differences in the expression of immune checkpoint markers between the high-risk and low-risk groups of HCC. **(E,F)** CIBERSORT analysis revealed differences in immune infiltration between the subgroups.

within HCC subgroups classified based on ERGs. These results suggest that ERGs may have an impact on the immune response outcomes of HCC patients (Chang et al., 2013; Fenutria et al., 2014; Braga et al., 2017; Kornberg et al., 2018; Fu et al., 2020; Nie et al., 2020).Our pathway enrichment analysis revealed significant enrichment in cytoplasmic ribosomal proteins, eukaryotic translation elongation, humoral immune response, and cytosolic large ribosomal subunit pathways in HCC patients with high expression of ERGs (Figures 3B–E), which could potentially influence the response of HCC patients to chemotherapy. Drug resistance is also an important factor in persistent tumor progression. Leveraging the six ERGs, we identified sixteen potential clinical drugs that could be tailored to specific subtypes of HCC (Figure 8). We then tested the reliability of our predictions by selecting etoposide as a representative drug. Based on the expression levels of ERGs, HCC cell lines were classified into high and low ERG expression groups, with Huh7 and HepG2 representing the high and low groups, respectively (Figure 9). Our results demonstrated that Huh7 cells exhibited greater sensitivity to etoposide, with a lower IC50, compared to HepG2 cells when treated with various concentrations of the drug (Figure 10A). Furthermore, under identical treatment conditions with the same concentration of etoposide, the drug exhibited greater cytotoxicity to Huh7 cells (Figure 10B). Our findings are in agreement with our initial predictions of drug sensitivity and underscore the reliability of ERGs in predicting chemotherapy response. Based on the drug sensitivity list provided in Supplementary Table S1, the implementation of combination therapeutic strategies involving specific drugs in conjunction with first-line treatments may potentially enhance the anti-tumor therapeutic efficacy for high-risk or low-risk hepatocellular carcinoma patients. It is important to note that prior to implementation, rigorous *in vivo* experiments are required to validate these approaches adequately.

In recent years, the connection between efferocytosis and tumors has garnered notable interest. The expression of glycolytic metabolic genes is known to influence the TME and thus the susceptibility of HCC cells to immunotherapy or chemotherapy. As such, personalized therapeutic approaches should be

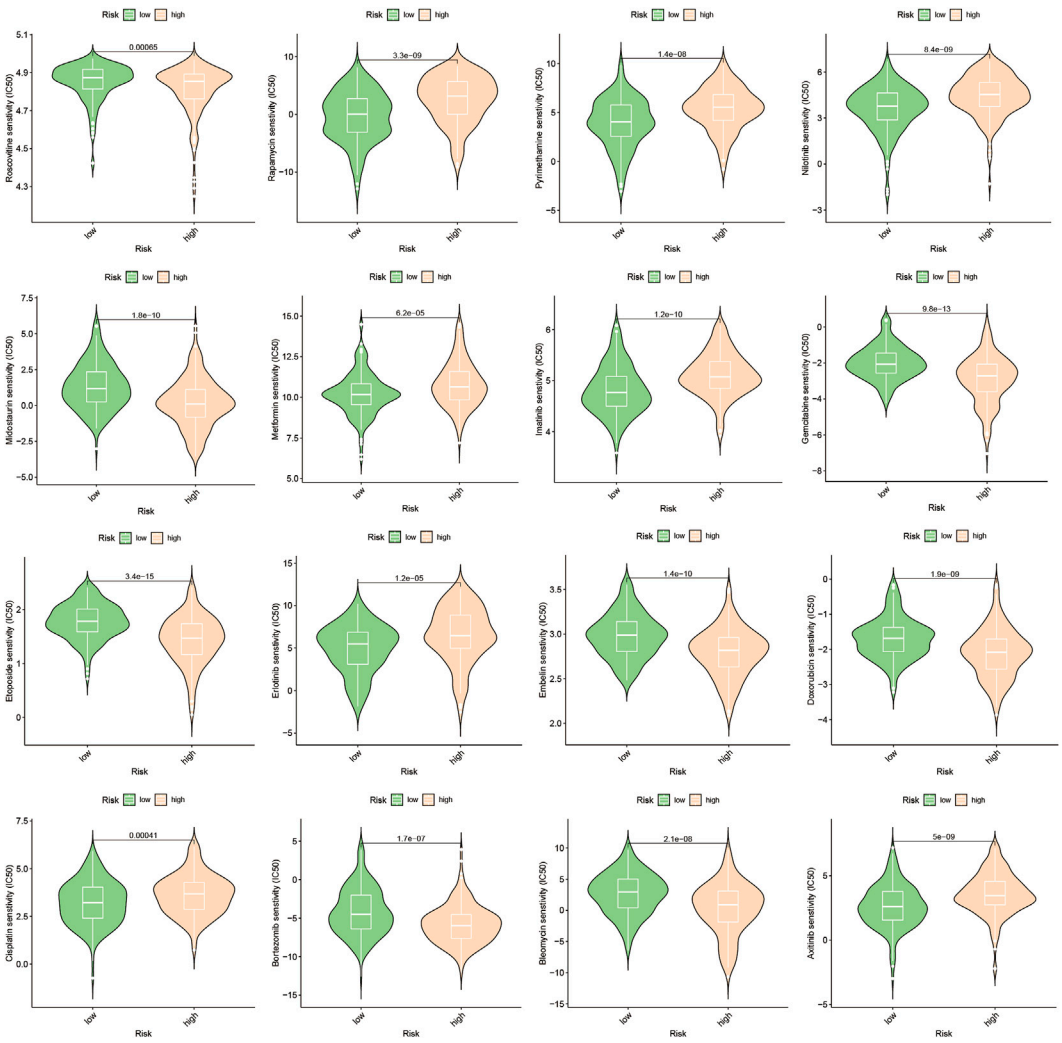


FIGURE 8
Drug sensitivity prediction.

implemented for cancer patients based on their specific degree of efferocytosis. The classification of tumor samples using gene expression profiling has been well-established as a reliable technique (Werfel et al., 2019; Gong et al., 2022; Zhao et al., 2022). In the present investigation, we classified HCC patients based on the expression levels of efferocytosis-related genes, revealing significant differences in prognostic outcomes and immune infiltration profiles among patients with varying ERG expression levels. Our findings support the use of a six-gene efferocytosis-related model to accurately predict patient prognosis. Furthermore, our cell toxicity assays have confirmed the efficacy of our chemotherapy sensitivity predictions, which could aid clinicians in selecting optimal treatment regimens (Jin et al., 2021; Liu et al., 2022b; Zhong et al., 2022). These results emphasize the capability of our 6-gene model to serve as a reliable prognostic indicator for overall survival in individuals with HCC. Furthermore, our findings suggest that this model could be instrumental in pinpointing novel therapeutic targets for high-risk patient cohorts.

Despite the valuable clinical implications of our investigation regarding prognostic assessment and treatment selection for patients diagnosed with HCC, it is essential to acknowledge the limitations present in our study. Primarily, our research is retrospective in nature, necessitating validation through prospective studies in the future. Due to the unavailability of mRNA expression profile data for HCC patients undergoing immunotherapy, an indirect assessment was conducted to explore the predictive capability of this signature in terms of immunotherapy response. It is important to note that this approach may deviate from the actual circumstances, introducing a potential limitation in the analysis. Consequently, it is imperative to conduct further validation studies that incorporate data obtained from HCC patients undergoing immunotherapy. Besides, the migratory capacity of tumor cells is closely associated with unfavorable prognosis and recurrence (Wu et al., 2021; Li et al., 2023b). However, the relationship between efferocytosis and the migratory potential remains understudied in our investigation. Ultimately, the

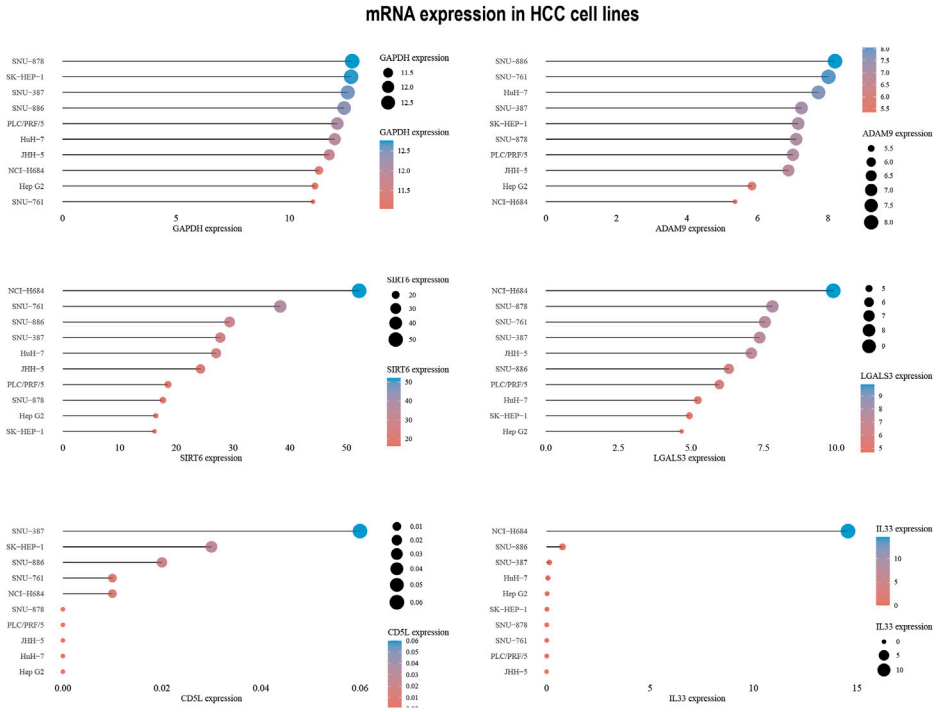


FIGURE 9
6-ERGs expression levels in HCC cell lines.

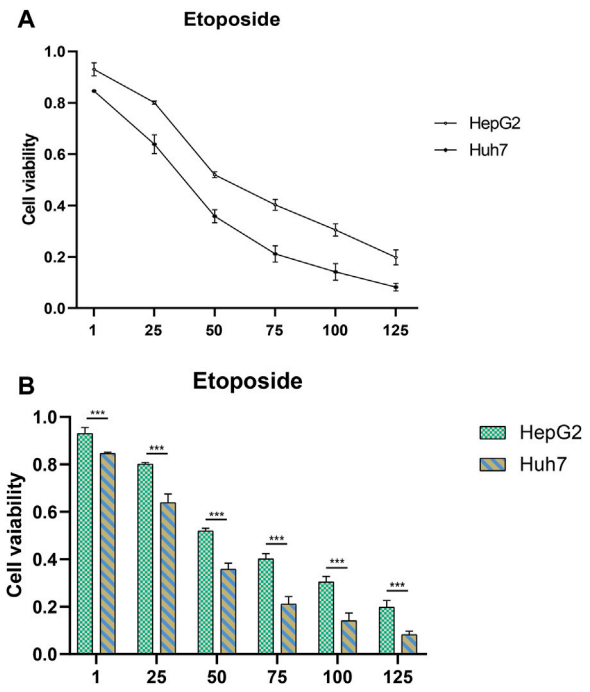


FIGURE 10
Drug sensitivity testing. **(A,B)** The CCK-8 assay revealed the cytotoxic effects of Etoposide on Huh7 and HepG2 cells at different concentrations.

current study lacks sufficient *in vivo* experiments to enhance the reliability of drug prediction outcomes, thus impacting their potential for further clinical applications. These limitations warrant the need for future investigations aimed at refining and expanding upon these aspects.

5 Conclusion

Efferocytosis plays a critical role in both normal physiological processes and pathological conditions, particularly in tumor progression, within the immune system. Despite this, the function of efferocytosis-related genes in HCC progression and prognosis remains largely unexplored. To address this gap, we developed a risk model based on six efferocytosis-related genes and identified two subtypes associated with ERG that demonstrate significant differences in both tumor immune landscape and prognostic stratification. Our results underscore the importance of efferocytosis status in HCC, and have significant implications for predicting patterns of immune therapy and chemotherapy response. Furthermore, *in vitro* validation confirms the model's predictive ability for drug response, offering important insights into potential therapeutic strategies for patients with different efferocytosis statuses. Overall, our study highlights the crucial role of efferocytosis in HCC and serves as a valuable foundation for further research into HCC progression and treatment response, as well as for guiding clinical decision-making in HCC management.

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.

Author contributions

Conceptualization, TW; writing original draft preparation, KX; visualization, KX and YL; data resources, KX; supervision, validation, and funding acquisition, TW; review and editing, TW; KX, HL and TW helped with the final revision of this manuscript. All authors reviewed and approved the final manuscript.

Acknowledgments

All authors acknowledge the contributions from the TCGA project.

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/fphar.2023.1218244/full#supplementary-material

References

Banerjee, H. N., Bartlett, V., Krauss, C., Aurelius, C., Johnston, K., Hedley, J., et al. (2021). Efferocytosis and the story of "find me," "eat me," and "don't eat me" signaling in the tumor microenvironment. *Adv. Exp. Med. Biol.* 1329, 153–162. doi:10.1007/978-3-030-73119-9_8

Braga, D., Barcella, M., D'Avila, F., Lupoli, S., Tagliaferri, F., Santamaria, M. H., et al. (2017). Preliminary profiling of blood transcriptome in a rat model of hemorrhagic shock. *Exp. Biol. Med. (Maywood)* 242, 1462–1470. doi:10.1177/1535370217717978

Brightwell, R. M., Grzankowski, K. S., Lele, S., Eng, K., Arshad, M., Chen, H., et al. (2016). The CD47 "don't eat me signal" is highly expressed in human ovarian cancer. *Gynecol. Oncol.* 143, 393–397. doi:10.1016/j.ygyno.2016.08.325

Bukong, T. N., Cho, Y., Iracheta-Vellve, A., Saha, B., Lowe, P., Adejumo, A., et al. (2018). Abnormal neutrophil traps and impaired efferocytosis contribute to liver injury and sepsis severity after binge alcohol use. *J. Hepatol.* 69, 1145–1154. doi:10.1016/j.jhep.2018.07.005

Chang, C. H., Curtis, J. D., Maggi, L. B., Faubert, B., Villarino, A. V., O'Sullivan, D., et al. (2013). Posttranscriptional control of T cell effector function by aerobic glycolysis. *Cell* 153, 1239–1251. doi:10.1016/j.cell.2013.05.016

Chi, H., Jiang, P., Xu, K., Zhao, Y., Song, B., Peng, G., et al. (2022a). A novel anoikis-related gene signature predicts prognosis in patients with head and neck squamous cell carcinoma and reveals immune infiltration. *Front. Genet.* 13, 984273. doi:10.3389/fgene.2022.984273

Chi, H., Xie, X., Yan, Y., Peng, G., Strohmer, D. F., Lai, G., et al. (2022b). Natural killer cell-related prognosis signature characterizes immune landscape and predicts prognosis of HNSCC. *Front. Immunol.* 13, 1018685. doi:10.3389/fimmu.2022.1018685

Chi, H., Yang, J., Peng, G., Zhang, J., Song, G., Xie, X., et al. (2023a). Circadian rhythm-related genes index: A predictor for HNSCC prognosis, immunotherapy efficacy, and chemosensitivity. *Front. Immunol.* 14, 1091218. doi:10.3389/fimmu.2023.1091218

Chi, H., Zhao, S., Yang, J., Gao, X., Peng, G., Zhang, J., et al. (2023b). T-cell exhaustion signatures characterize the immune landscape and predict HCC prognosis via integrating single-cell RNA-seq and bulk RNA-sequencing. *Front. Immunol.* 14, 1137025. doi:10.3389/fimmu.2023.1137025

Conche, C., Finkelmeier, F., Pesic, M., Nicolas, A. M., Bottger, T. W., Kennel, K. B., et al. (2023). Combining ferroptosis induction with MDSC blockade renders primary tumours and metastases in liver sensitive to immune checkpoint blockade. *Gut* 2022, 327909. doi:10.1136/gutjnl-2022-327909

Cruz Cruz, J., Allison, K. C., Page, L. S., Jenkins, A. J., Wang, X., Earp, H. S., et al. (2023). Inhibiting efferocytosis reverses macrophage-mediated immunosuppression in the leukemia microenvironment. *Front. Immunol.* 14, 1146721. doi:10.3389/fimmu.2023.1146721

Engebretsen, S. (2019). Bohlin J: Statistical predictions with glmnet. *Clin. Epigenetics* 11, 123. doi:10.1186/s13148-019-0730-1

Fenutria, R., Martinez, V. G., Simoes, I., Postigo, J., Gil, V., Martinez-Florensa, M., et al. (2014). Transgenic expression of soluble human CD5 enhances experimentally-induced autoimmune and anti-tumoral immune responses. *PLoS One* 9, e84895. doi:10.1371/journal.pone.0084895

Fu, W., Li, H., Fu, H., Zhao, S., Shi, W., Sun, M., et al. (2020). The SIRT3 and SIRT6 promote prostate cancer progression by inhibiting necroptosis-mediated innate immune response. *J. Immunol. Res.* 2020, 8820355. doi:10.1155/2020/8820355

Gadiyar, V., Lahey, K. C., Calianese, D., Devoe, C., Mehta, D., Bono, K., et al. (2020). Cell death in the tumor microenvironment: Implications for cancer immunotherapy. *Cells* 9, 2207. doi:10.3390/cells9102207

Garcia-Pras, E., Fernandez-Iglesias, A., Gracia-Sancho, J., and Perez-Del-Pulgar, S. (2021). Cell death in hepatocellular carcinoma: Pathogenesis and therapeutic opportunities. *Cancers (Basel)* 14, 48. doi:10.3390/cancers14010048

Geeleher, P., Cox, N., and Huang, R. S. (2014). pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* 9, e107468. doi:10.1371/journal.pone.0107468

Gong, X., Chi, H., Strohmer, D. F., Teichmann, A. T., Xia, Z., and Wang, Q. (2022). Exosomes: A potential tool for immunotherapy of ovarian cancer. *Front. Immunol.* 13, 1089410. doi:10.3389/fimmu.2022.1089410

Graham, D. K., DeRyckere, D., Davies, K. D., and Earp, H. S. (2014). The TAM family: Phosphatidylserine sensing receptor tyrosine kinases gone awry in cancer. *Nat. Rev. Cancer* 14, 769–785. doi:10.1038/nrc3847

Hanzelmann, S., Castelo, R., and Guinney, J. (2013). Gsva: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Icard, P., Simula, L., Wu, Z., Berzan, D., Sogni, P., Dohan, A., et al. (2021). Why may citrate sodium significantly increase the effectiveness of transarterial chemoembolization in hepatocellular carcinoma? *Drug Resist Updat* 59, 100790. doi:10.1016/j.drup.2021.100790

Jensen, K. N., Omarsdottir, S. Y., Reinhardsdottir, M. S., Hardardottir, I., and Freysdottir, J. (2020). Docosahexaenoic acid modulates NK cell effects on neutrophils and their crosstalk. *Front. Immunol.* 11, 570380. doi:10.3389/fimmu.2020.570380

Jin, W., Yang, Q., Chi, H., Wei, K., Zhang, P., Zhao, G., et al. (2022). Ensemble deep learning enhanced with self-attention for predicting immunotherapeutic responses to cancers. *Front. Immunol.* 13, 1025330. doi:10.3389/fimmu.2022.1025330

Jin, W., Zhang, Y., Liu, Z., Che, Z., Gao, M., and Peng, H. (2021). Exploration of the molecular characteristics of the tumor-immune interaction and the development of an individualized immune prognostic signature for neuroblastoma. *J. Cell Physiol.* 236, 294–308. doi:10.1002/jcp.29842

Kornberg, M. D., Bhargava, P., Kim, P. M., Putluri, V., Snowman, A. M., Putluri, N., et al. (2018). Dimethyl fumarate targets GAPDH and aerobic glycolysis to modulate immunity. *Science* 360, 449–453. doi:10.1126/science.aan4665

Lahey, K. C., Gadiyar, V., Hill, A., Desind, S., Wang, Z., Davra, V., et al. (2022). Mertk: An emerging target in cancer biology and immuno-oncology. *Int. Rev. Cell Mol. Biol.* 368, 35–59. doi:10.1016/bs.ircmb.2022.04.004

Lee, M., Lee, S. Y., and Bae, Y. S. (2022). Emerging roles of neutrophils in immune homeostasis. *BMB Rep.* 55, 473–480. doi:10.5483/BMBRep.2022.55.10.115

Leone, V., Ali, A., Weber, A., Tschaharganeh, D. F., and Heikenwalder, M. (2021). Liver inflammation and hepatobiliary cancers. *Trends Cancer* 7, 606–623. doi:10.1016/j.trecan.2021.01.012

Li, Z., Wang, Q., Huang, X., Yang, M., Zhou, S., Li, Z., et al. (2023a). Lactate in the tumor microenvironment: A rising star for targeted tumor therapy. *Front. Nutr.* 10, 1113739. doi:10.3389/fnut.2023.1113739

Li, Z., Zhou, H., Xia, Z., Xia, T., Du, G., Franziska, S. D., et al. (2023b). HMGA1 augments palbociclib efficacy via PI3K/mTOR signaling in intrahepatic cholangiocarcinoma. *Biomark. Res.* 11, 33. doi:10.1186/s40364-023-00473-w

Lin, J., Xu, A., Jin, J., Zhang, M., Lou, J., Qian, C., et al. (2022). MerTK-mediated efferocytosis promotes immune tolerance and tumor progression in osteosarcoma through enhancing M2 polarization and PD-L1 expression. *Oncoimmunology* 11, 2024941. doi:10.1080/2162402X.2021.2024941

Liu, C. Q., Xu, J., Zhou, Z. G., Jin, L. L., Yu, X. J., Xiao, G., et al. (2018). Expression patterns of programmed death ligand 1 correlate with different microenvironments and patient prognosis in hepatocellular carcinoma. *Br. J. Cancer* 119, 80–88. doi:10.1038/s41416-018-0144-4

Liu, G., Xiong, D., Che, Z., Chen, H., and Jin, W. (2022b). A novel inflammation-associated prognostic signature for clear cell renal cell carcinoma. *Oncol. Lett.* 24, 307. doi:10.3892/ol.2022.13427

Liu, G., Zhang, P., Yang, F., Jiang, K., Sun, S., Xia, Z., et al. (2023a). Integrating single-cell analysis and machine learning to create glycosylation-based gene signature for prognostic prediction of uveal melanoma. *Front. Endocrinol. (Lausanne)* 14, 1163046. doi:10.3389/fendo.2023.1163046

Liu, N., Wang, X., Steer, C. J., and Song, G. (2022a). MicroRNA-206 promotes the recruitment of CD8(+) T cells by driving M1 polarisation of Kupffer cells. *Gut* 71, 1642–1655. doi:10.1136/gutjnl-2021-324170

Liu, N., Xun, Z., Ma, K., Liang, S., Li, X., Zhou, S., et al. (2023b). Identification of a tumour immune barrier in the HCC microenvironment that determines the efficacy of immunotherapy. *J. Hepatol.* 78, 770–782. doi:10.1016/j.jhep.2023.01.011

Llovet, J. M., Castet, F., Heikenwalder, M., Maini, M. K., Mazzaferro, V., Pinato, D. J., et al. (2022). Immunotherapies for hepatocellular carcinoma. *Nat. Rev. Clin. Oncol.* 19, 151–172. doi:10.1038/s41571-021-00573-2

Llovet, J. M., Montal, R., Sia, D., and Finn, R. S. (2018). Molecular therapies and precision medicine for hepatocellular carcinoma. *Nat. Rev. Clin. Oncol.* 15, 599–616. doi:10.1038/s41571-018-0073-4

Lu, L. G., Zhou, Z. L., Wang, X. Y., Liu, B. Y., Lu, J. Y., Liu, S., et al. (2022). PD-L1 blockade liberates intrinsic antitumourigenic properties of glycolytic macrophages in hepatocellular carcinoma. *Gut* 71, 2551–2560. doi:10.1136/gutjnl-2021-326350

Maschalidi, S., Mehrotra, P., Keceli, B. N., De Cleene, H. K. L., Lecomte, K., Van der Cruyssen, R., et al. (2022). Targeting SLC7A11 improves efferocytosis by dendritic cells and wound healing in diabetes. *Nature* 606, 776–784. doi:10.1038/s41586-022-04754-6

Mehrotra, P., and Ravichandran, K. S. (2022). Drugging the efferocytosis process: Concepts and opportunities. *Nat. Rev. Drug Discov.* 21, 601–620. doi:10.1038/s41573-022-00470-y

Morioka, S., Maueroder, C., and Ravichandran, K. S. (2019). Living on the edge: Efferocytosis at the interface of homeostasis and pathology. *Immunity* 50, 1149–1162. doi:10.1016/j.immuni.2019.04.018

Myers, K. V., Amend, S. R., and Pienta, K. J. (2019). Targeting Tyro3, axl and MerTK (TAM receptors): Implications for macrophages in the tumor microenvironment. *Mol. Cancer* 18, 94. doi:10.1186/s12943-019-1022-2

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457. doi:10.1038/nmeth.3337

Ni, L., Tang, C., Wang, Y., Wan, J., Charles, M. G., Zhang, Z., et al. (2022). Construction of a miRNA-based nomogram model to predict the prognosis of endometrial cancer. *J. Pers. Med.* 12, 1154. doi:10.3390/jpm12071154

Nie, J., Yu, Z., Yao, D., Wang, F., Zhu, C., Sun, K., et al. (2020). Litopenaeus vannamei sirtuin 6 homolog (LvSIRT6) is involved in immune response by modulating hemocytes ROS production and apoptosis. *Fish. Shellfish Immunol.* 98, 271–284. doi:10.1016/j.fsi.2020.01.029

Nino-Castano, V. E., Penteado, L. A., Silva-Pereira, L., Bazzano, J. M. R., Orlando, A. B., Salina, A. C. G., et al. (2022). RIP2 contributes to expanded CD4(+) T cell IFN-gamma production during efferocytosis of Streptococcus pneumoniae-infected apoptotic cells. *Immunohorizons* 6, 559–568. doi:10.4049/immunohorizons.2200001

Peng, S., Chen, S., Hu, W., Mei, J., Zeng, X., Su, T., et al. (2022). Combination neoantigen-based dendritic cell vaccination and adoptive T-cell transfer induces antitumor responses against recurrence of hepatocellular carcinoma. *Cancer Immunol. Res.* 10, 728–744. doi:10.1158/2326-6066.CIR-21-0931

Poon, I. K., Lucas, C. D., Rossi, A. G., and Ravichandran, K. S. (2014). Apoptotic cell clearance: Basic biology and therapeutic potential. *Nat. Rev. Immunol.* 14, 166–180. doi:10.1038/nri3607

Roy, S., Bag, A. K., Dutta, S., Polavaram, N. S., Islam, R., Schellenburg, S., et al. (2018). Macrophage-derived neuropilin-2 exhibits novel tumor-promoting functions. *Cancer Res.* 78, 5600–5617. doi:10.1158/0008-5472.CAN-18-0562

Safran, M., Rosen, N., Twik, M., BarShir, R., Stein, T. I., Dahary, D., et al. (2021). *Lancet DJPgtlsd: The genecards suite*. Berlin, Germany: Springer, 27–56.

Sandri, S., Hebeda, C. B., Broering, M. F., de Paula Silva, M., Moredo, L. F., de Barros, E. S. M. J., et al. (2023). Role of annexin A1 secreted by neutrophils in melanoma metastasis. *Cells* 12, 425. doi:10.3390/cells12030425

Soki, F. N., Koh, A. J., Jones, J. D., Kim, Y. W., Dai, J., Keller, E. T., et al. (2014). Polarization of prostate cancer-associated macrophages is induced by milk fat globule-EGF factor 8 (MFG-E8)-mediated efferocytosis. *J. Biol. Chem.* 289, 24560–24572. doi:10.1074/jbc.M114.571620

Sun, P., Wang, X., Zhong, J., Yu, D., Xuan, H., Xu, T., et al. (2023). Development and validation of a pyroptosis-related genes signature for risk stratification in gliomas. *Front. Genet.* 14, 1087563. doi:10.3389/fgene.2023.1087563

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Tajbakhsh, A., Gheibi Hayat, S. M., Movahedpour, A., Savardashtaki, A., Loveless, R., Barreto, G. E., et al. (2021). The complex roles of efferocytosis in cancer development, metastasis, and treatment. *Biomed. Pharmacother.* 140, 111776. doi:10.1016/j.biopha.2021.111776

Vaught, D. B., Stanford, J. C., and Cook, R. S. (2015). Efferocytosis creates a tumor microenvironment supportive of tumor survival and metastasis. *Cancer Cell Microenviron.* 2, e666. doi:10.14800/ccm.666

Wang, J., Yu, H., Dong, W., Zhang, C., Hu, M., Ma, W., et al. (2023a). N6-Methyladenosine-Mediated up-regulation of FZD10 regulates liver cancer stem cells' properties and lenvatinib resistance through WNT/β-Catenin and hippo signaling pathways. *Gastroenterology* 164, 990–1005. doi:10.1053/j.gastro.2023.01.041

Wang, J., Liu, Y., Li, Z., Tang, Y., Long, W., Xin, H., et al. (2023c). Establishment of a novel lysosomal signature for the diagnosis of gastric cancer with *in-vitro* and *in-situ* validation. *Front. Immunol.* 14, 1182277. doi:10.3389/fimmu.2023.1182277

Wang, X., Zhao, Y., Strohmer, D. F., Yang, W., Xia, Z., and Yu, C. (2022). The prognostic value of MicroRNAs associated with fatty acid metabolism in head and neck squamous cell carcinoma. *Front. Genet.* 13, 983672. doi:10.3389/fgene.2022.983672

Wang, J., Trzeciak, A. J., Rojas, W. S., Saavedra, P., Chen, Y. T., Chirayil, R., et al. (2023b). Metabolic adaptation supports enhanced macrophage efferocytosis in limited-oxygen environments. *Cell Metab.* 35, 316–331.e6. doi:10.1016/j.cmet.2022.12.005

Wang, Z., Jensen, M. A., and Zenklusen, J. C. (2016). A practical guide to the cancer Genome Atlas (TCGA). *Methods Mol. Biol.* 1418, 111–141. doi:10.1007/978-1-4939-3578-9_6

Werfel, T. A., and Cook, R. S. (2018). Efferocytosis in the tumor microenvironment. *Semin. Immunopathol.* 40, 545–554. doi:10.1007/s00281-018-0698-5

Werfel, T. A., Elion, D. L., Rahman, B., Hicks, D. J., Sanchez, V., Gonzales-Ericsson, P. I., et al. (2019). Treatment-Induced tumor cell apoptosis and secondary necrosis drive tumor progression in the residual tumor microenvironment through MerTK and Ido1. *Cancer Res.* 79, 171–182. doi:10.1158/0008-5472.CAN-18-1106

Wu, J., Gao, W., Tang, Q., Yu, Y., You, W., Wu, Z., et al. (2021). M2 macrophage-derived exosomes facilitate HCC metastasis by transferring α_M β_2 integrin to tumor cells. *Hepatology* 73, 1365–1380. doi:10.1002/hep.31432

Xu, C., Song, L., Peng, H., Yang, Y., Liu, Y., Pei, D., et al. (2022). Clinical eosinophil-associated genes can serve as a reliable predictor of bladder urothelial cancer. *Front. Mol. Biosci.* 9, 963455. doi:10.3389/fmolb.2022.963455

Xu, X. F., Xing, H., Han, J., Li, Z. L., Lau, W. Y., Zhou, Y. H., et al. (2019). Risk factors, patterns, and outcomes of late recurrence after liver resection for hepatocellular carcinoma: A multicenter study from China. *JAMA Surg.* 154, 209–217. doi:10.1001/jamasurg.2018.4334

Yang, J. D., Hainaut, P., Gores, G. J., Amadou, A., Plymoth, A., and Roberts, L. R. (2019a). A global view of hepatocellular carcinoma: Trends, risk, prevention and management. *Nat. Rev. Gastroenterol. Hepatol.* 16, 589–604. doi:10.1038/s41575-019-0186-y

Yang, Y., Wang, Y., Guo, L., Gao, W., Tang, T. L., and Yan, M. (2022). Interaction between macrophages and ferroptosis. *Cell Death Dis.* 13, 355. doi:10.1038/s41419-022-04775-z

Yang, J. D., Kim, S. Y., and Seki, E. (2019b). Inflammation and liver cancer: Molecular mechanisms and therapeutic targets. *Semin. Liver Dis.* 39, 26–42. doi:10.1055/s-0038-1676806

Zhai, X., Xia, Z., Du, G., Zhang, X., Xia, T., Ma, D., et al. (2022). LRP1B suppresses HCC progression through the NCSTN/PI3K/AKT signaling axis and affects doxorubicin resistance. *Genes Dis.* 2022, 21. doi:10.1016/j.gendis.2022.10.021

Zhang, M., Liu, K., Zhang, Q., Xu, J., Liu, J., Lin, H., et al. (2023a). Alpha fetoprotein promotes polarization of macrophages towards M2-like phenotype and inhibits macrophages to phagocytize hepatoma cells. *Front. Immunol.* 14, 1081572. doi:10.3389/fimmu.2023.1081572

Zhang, M., Zhuge, J., Liu, J., Xia, Z., Wang, H., Gao, Q., et al. (2023b). Prognostic signatures of sphingolipids: Understanding the immune landscape and predictive role in immunotherapy response and outcomes of hepatocellular carcinoma. *Front. Immunol.* 14, 1153423. doi:10.3389/fimmu.2023.1153423

Zhang, Y., Wang, Y., Ding, J., and Liu, P. (2022). Efferocytosis in multisystem diseases (Review). *Mol. Med. Rep.* 25, 13. doi:10.3892/mmr.2021.12529

Zhao, S., Chi, H., Yang, Q., Chen, S., Wu, C., Lai, G., et al. (2023b). Identification and validation of neurotrophic factor-related gene signatures in glioblastoma and Parkinson's disease. *Front. Immunol.* 14, 1090040. doi:10.3389/fimmu.2023.1090040

Zhao, S., Zhang, X., Gao, F., Chi, H., Zhang, J., Xia, Z., et al. (2023a). Identification of copper metabolism-related subtypes and establishment of the prognostic model in ovarian cancer. *Front. Endocrinol. (Lausanne)* 14, 1145797. doi:10.3389/fendo.2023.1145797

Zhao, Y., Wei, K., Chi, H., Xia, Z., and Li, X. (2022). IL-7: A promising adjuvant ensuring effective T cell responses and memory in combination with cancer vaccines? *Front. Immunol.* 13, 1022808. doi:10.3389/fimmu.2022.1022808

Zhong, H., Yang, L., Zeng, Q., Chen, W., Zhao, H., Wu, L., et al. (2023). Machine learning predicts the oxidative stress subtypes provide an innovative insight into colorectal cancer. *Oxid. Med. Cell Longev.* 2023, 1737501. doi:10.1155/2023/1737501

Zhong, Y., Zhang, Y., Wei, S., Chen, J., Zhong, C., Cai, W., et al. (2022). Dissecting the effect of sphingolipid metabolism gene in progression and microenvironment of osteosarcoma to develop a prognostic signature. *Front. Endocrinol. (Lausanne)* 13, 1030655. doi:10.3389/fendo.2022.1030655

Zhou, C., Weng, J., Liu, C., Liu, S., Hu, Z., Xie, X., et al. (2023). Disruption of SLFN11 deficiency-induced CCL2 signaling and macrophage M2 polarization potentiates anti-PD-1 therapy efficacy in hepatocellular carcinoma. *Gastroenterology* 164, 1261–1278. doi:10.1053/j.gastro.2023.02.005

Zhou, Y., Cai, Z., Ma, N., Xie, W., Gao, C., Huang, M., et al. (2020c). A novel ten-gene signature predicting prognosis in hepatocellular carcinoma. *Front. Cell Dev. Biol.* 8, 629. doi:10.3389/fcell.2020.00629

Zhou, Y., Fei, M., Zhang, G., Liang, W. C., Lin, W., Wu, Y., et al. (2020a). Blockade of the phagocytic receptor MerTK on tumor-associated macrophages enhances P2x7r-dependent STING activation by tumor-derived cGAMP. *Immunity* 52, 357–373. doi:10.1016/j.immuni.2020.01.014

Zhou, Y., Yao, Y., Deng, Y., and Shao, A. (2020b). Regulation of efferocytosis as a novel cancer therapy. *Cell Commun. Signal* 18, 71. doi:10.1186/s12964-020-00542-9

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 23 August 2023
DOI 10.3389/fphar.2023.1213891

OPEN ACCESS

EDITED BY
Linhui Wang,
Second Military Medical University, China

REVIEWED BY
Min Tang,
Nanjing Medical University, China
Ying Luo,
East China Normal University, China
Jiao Hu,
Central South University, China

*CORRESPONDENCE
Mingchao Wang,
3316019@zju.edu.cn
Lei Gao,
drgaolei0417@zju.edu.cn

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

RECEIVED 28 April 2023
ACCEPTED 07 August 2023
PUBLISHED 23 August 2023

CITATION
Li L, Li F, Xu Z, Li L, Hu H, Li Y, Yu S, Wang M and Gao L (2023), Identification and validation of *SERPINE1* as a prognostic and immunological biomarker in pan-cancer and in ccRCC.
Front. Pharmacol. 14:1213891.
doi: 10.3389/fphar.2023.1213891

Identification and validation of *SERPINE1* as a prognostic and immunological biomarker in pan-cancer and in ccRCC

Lingqin Li[1†], Fan Li[2†], Zhehao Xu[2], Liyang Li[3], Haiyi Hu[2], Yang Li[2], Shicheng Yu[2], Mingchao Wang[2*] and Lei Gao[2*]

[1]Department of Operating Room, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, HangZhou, China, [2]Department of Urology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [3]University of New South Wales, School of Medicine, Sydney, NSW, Australia

Background: *SERPINE1*, a serine protease inhibitor involved in the regulation of the plasminogen activation system, was recently identified as a cancer-related gene. However, its clinical significance and potential mechanisms in pan-cancer remain obscure.

Methods: In pan-cancer multi-omics data from public datasets, including The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx), and online web tools were used to analyze the expression of *SERPINE1* in different cancers and its correlation with prognosis, genetic alteration, DNA promoter methylation, biological processes, immunoregulator expression levels, immune cell infiltration into tumor, tumor mutation burden (TMB), microsatellite instability (MSI), immunotherapy response and drug sensitivity. Further, two single-cell databases, Tumor Immune Single-cell Hub 2 (TISCH2) and CancerSEA, were used to explore the expression and potential roles of *SERPINE1* at a single-cell level. The aberrant expression of *SERPINE1* was further verified in clear cell renal cell carcinoma (ccRCC) through qRT-PCR of clinical patient samples, validation in independent cohorts using The Gene Expression Omnibus (GEO) database, and proteomic validation using the Clinical Proteomic Tumor Analysis Consortium (CPTAC) database.

Results: The expression of *SERPINE1* was dysregulated in cancers and enriched in endothelial cells and fibroblasts. Copy number amplification and low DNA promoter methylation could be partly responsible for high *SERPINE1* expression. High *SERPINE1* expression was associated with poor prognosis in 21 cancers. The results of gene set enrichment analysis (GSEA) indicated *SERPINE1* involvement in the immune response and tumor malignancy. *SERPINE1* expression was also associated with the expression of several immunoregulators and immune cell infiltration and could play an immunosuppression role. Besides, *SERPINE1* was found to be related with TMB, MSI, immunotherapy response and sensitivity to several drugs in cancers. Finally, the high expression of *SERPINE1* in ccRCC was verified using qRT-PCR performed on patient samples, six independent GEO cohorts, and proteomic data from the CPTAC database.

Conclusion: The findings of the present study revealed that *SERPINE1* exhibits aberrant expression in various types of cancers and is associated with cancer

immunity and tumor malignancy, providing novel insights for individualized cancer treatment.

KEYWORDS

***SERPINE1*, pan-cancer, multi-omics, clear cell renal cell carcinoma, cancer immunity**

1 Introduction

As one of the leading causes of death worldwide, cancer imposes an immense burden on the human society every year (Sung et al., 2021; Siegel et al., 2023). Despite significant advances in cancer treatment and early screening over the recent years, the overall survival prognosis for patients with cancer remains unsatisfactory, especially in certain cancer types (Sung et al., 2021; Hu et al., 2022). As such, there is an urgent need to explore new targets for cancer diagnosis and treatment.

Serine protease inhibitor clade E member 1 (*SERPINE1*, also known as PAI-1) is a serine protease inhibitor that plays key roles in regulating the plasminogen activation system (Placencio and DeClerck, 2015). By binding to and inactivating tissue-type plasminogen activator (tPA) and urokinase-type plasminogen activator (uPA), *SERPINE1* exerts antifibrinolytic effects (Declerck and Gils, 2013). The inhibition of PAI-1 leads to increased thrombolysis in artery disease (Kohler and Grant, 2000). *SERPINE1* consists of 379 amino acids and is synthesized and secreted primarily by vascular endothelial cells, adipocytes, and platelets (Chen et al., 2021).

In addition to regulating the plasminogen/plasminase system, *SERPINE1* has been found to be involved in a variety of other processes, such as pericellular proteolysis, tissue remodeling, cell migration, inflammation, angiogenesis, and apoptosis, implying its involvement in various diseases (Placencio and DeClerck, 2015; Sillen and Declerck, 2021). In recent years, the abnormal expression of *SERPINE1* has been found in various cancer types. Specifically, *SERPINE1* overexpression has been observed in breast cancer (Duffy et al., 2014; Jevrić et al., 2019), melanoma (Hanekom et al., 2002), non-small cell lung cancer (Sotiropoulos et al., 2019), bladder cancer (Becker et al., 2010), and ovarian cancer (Nakatsuka et al., 2017). The inhibition of *SERPINE1* expression has been shown to impede tumor progression and angiogenesis in several cancer types (Gomes-Giacoia et al., 2013; Masuda et al., 2013; Mashiko et al., 2015; Placencio et al., 2015; Takayama et al., 2016). *SERPINE1*-deficient mice also exhibited delayed tumor development, cancer invasion, and vascularization (Bajou et al., 1998; Gutierrez et al., 2000). Therefore, *SERPINE1* is expected to be a promising novel target for the diagnosis and treatment of cancers. However, the detailed mechanisms underlying the involvement of *SERPINE1* in cancers remain unclear.

Currently, there is no comprehensive study on the role of *SERPINE1* in pan-cancer. In the present study, we performed a multi-omics analysis of *SERPINE1* in pan-cancer. Our results confirmed the aberrant expression of *SERPINE1* in multiple cancers and the relationship of *SERPINE1* expression with the tumor microenvironment and cancer immunity. Furthermore, qRT-PCR was performed to validate the abnormal expression of *SERPINE1* in clear cell renal cell carcinoma samples.

2 Materials and methods

2.1 The workflow of the study

Based on several public datasets, we analyzed the expression of *SERPINE1* in pan-cancer and explored the possible causes of its abnormal expression. Then we evaluated the prognostic and diagnostic value of *SERPINE1* in pan-cancer to assess its clinical value. Next, we explored the biological functions associated with *SERPINE1* expression and found that it was associated with tumor malignancy and cancer immunity. Co-expression analysis, immune cell infiltration analysis, and cancer immunity cycle analysis were performed as well to validate its association with cancer immunity. Then we analyzed the correlation of *SERPINE1* expression with response of cancer treatments to further explore its clinical value. At last, validation analyses were performed in ccRCC. A summary of the workflow of this study is shown in Figure 1.

2.2 Tissue samples and sample size calculation

The Ethics Committee of the Sir Run Run Shaw Hospital, Zhejiang University approved this study, and written informed consent was obtained from all participating patients. The research procedures adhered to the guidelines of the Declaration of Helsinki. For assessing *SERPINE1* expression, 26 tissue samples (13 tumor tissues and 13 paired adjacent normal tissues) were randomly selected from clear cell renal cell carcinoma patients. The power analysis was performed using G*Power (G*Power, version 3.1 for MAC, Dusseldorf, North Rhine-Westphalia). According to the TCGA-KIRC project, the *SERPINE1* expression of tumor samples (with available paired adjacent normal tissue data) were 5.8746 ± 1.6083 (log2FPKM) and the expression of adjacent normal tissues were 3.9290 ± 1.7263 (log2FPKM), indicating a effect size of 0.8615 (Cancer Genome Atlas Research Network, 2013). With power (1-β) of 0.85 and α error of 0.05, the sample size was calculated to be 13 pairs of samples.

2.3 Gene expression analysis of *SERPINE1* in pan-cancer

Transcriptional data of tumor and normal samples were collected from the UCSC Xena (https://xenabrowser.net/

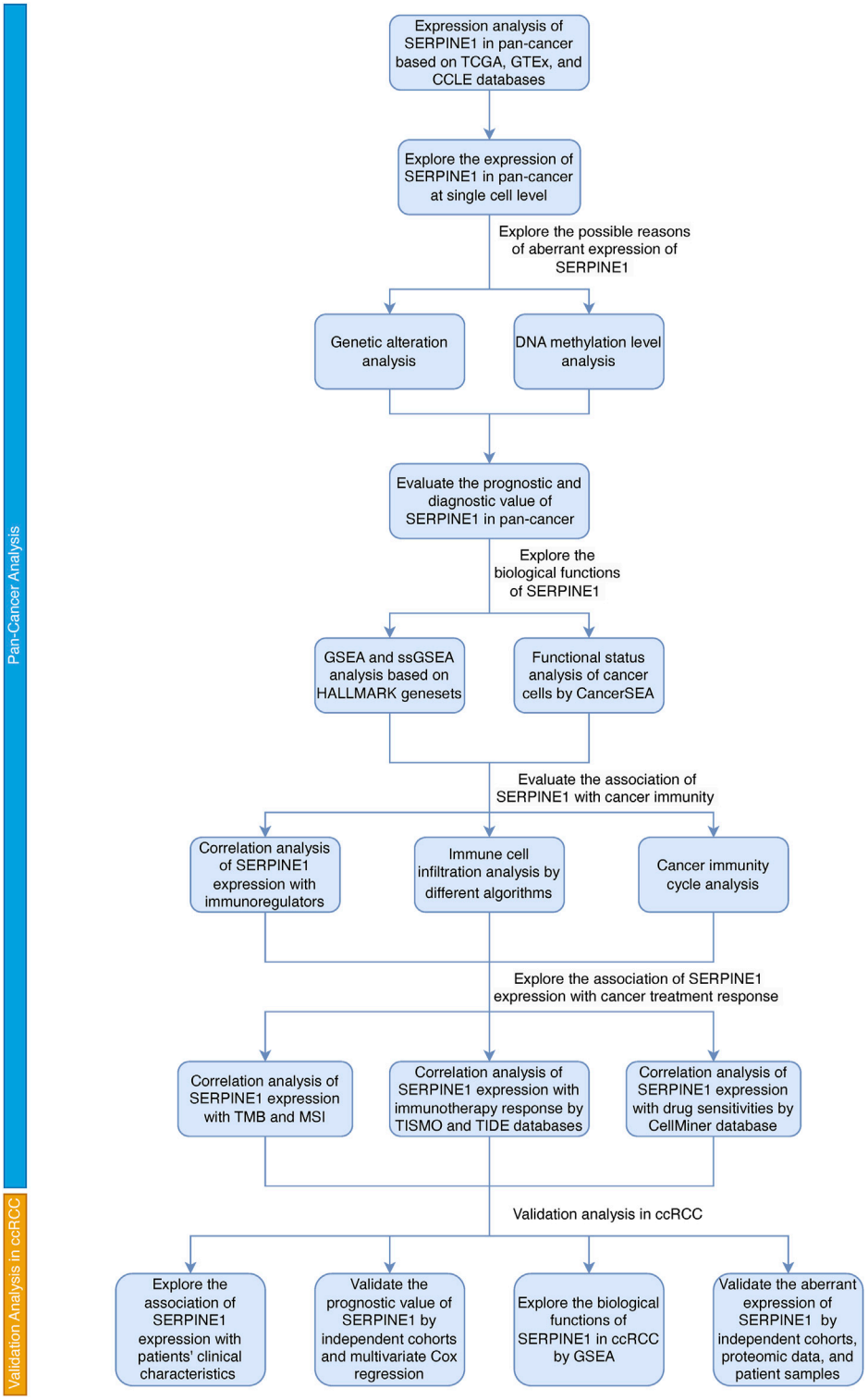


FIGURE 1
The flow chart of this study.

datapages/) dataset (Goldman et al., 2020). Expression data of cancer cell lines were downloaded from the Cancer Cell Line Encyclopedia (CCLE, https://sites.broadinstitute.org/ccle) (Barretina et al., 2012). The abbreviations of cancer names are presented in Supplementary Table S1. Wilcox rank sum test was used to compare gene expression levels. R (version 4.2.3) and R package "ggplot2" were used for statistical analysis and visualization.

2.4 Single-cell analysis of *SERPINE1* expression

Expression of *SERPINE1* at a single-cell level was analyzed using Tumor Immune Single-cell Hub 2 (TISCH2), based on the MAESTRO workflow (Han et al., 2022). The following parameters were used for analysis: *SERPINE1* (Gene), major lineages (Cell-type annotation), and all cancers (Cancer type). The R Package "ComplexHeatmap" (version 2.14.0) was used for data visualization (Gu et al., 2016).

2.5 Genetic alteration and DNA methylation analysis

Genetic alteration and DNA methylation data from the TCGA database were download from The cBioPortal (http://cbioportal.org) and GSCA (http://bioinfo.life.hust.edu.cn/GSCA/#/) (Cerami et al., 2012; Liu et al., 2023). Spearman's rank correlation coefficient was used to evaluate the correlations between copy number variation, the DNA methylation level, and *SERPINE1* expression.

2.6 Clinical and prognostic value analysis of *SERPINE1*

Clinical data and different types of prognostic data were extracted from UCSC Xena (https://xenabrowser.net/datapages/) (Goldman et al., 2020). Both Cox regression and Kaplan–Meier analysis were conducted to assess the correlation between *SERPINE1* expression and patient prognosis. The cutoff value of the Kaplan–Meier estimator was determined using the "surv-cutpoint" function of the R package "survminer" (version 0.4.9). Information regarding immune subtypes was obtained from a previous report (Thorsson et al., 2018). Moreover, six independent datasets from BEST (https://rookieutopia.com/) were analyzed to validate the prognostic value of *SERPINE1* expression in clear cell renal cell carcinoma, including the GSE167573, GSE29609, and GSE22541 cohorts from GEO (https://www.ncbi.nlm.nih.gov/), the E-MTAB-1980 cohort from EMBL's European Bioinformatics Institute (EMBL-EBI, https://www.ebi.ac.uk/), and Renal Cell Cancer-European Union (RECA-EU) project data from the International Cancer Genome Consortium (ICGC, https://dcc.icgc.org/). The diagnostic value of *SERPINE1* was assessed by receiver operating characteristic curve (ROC) using R package "pROC" (version 1.18.4).

2.7 Functional enrichment analysis and gene set enrichment analysis of *SERPINE1*

The associations between *SERPINE1* expression and several biological processes were examined using GSEA and ssGSEA. Gene sets were downloaded from the Molecular Signatures Database (MSigDB, https://www.gsea-msigdb.org/gsea/index.jsp). Samples were grouped according to the median *SERPINE1* expression for each cancer in GSEA. The R package "GSVA" (version 1.46.0) was used for the GSEA and ssGSEA analyses.

2.8 Single cell-level analysis of *SERPINE1*

CancerSEA, which is based on Gene Set Variance Analysis (GSVA), was used to assess the functional status of cancer cells at the single-cell level and Spearman's rank correlation test were used to calculate their relationship to *SERPINE1* expression (Yuan et al., 2019).

2.9 Correlation analysis of *SERPINE1* with immune-associated genes, immune cell infiltration, and cancer-immunity cycle

Based on TCGA in pan-cancer data, the correlation of *SERPINE1* expression with several immunoregulators, infiltration scores of different cells in the tumor microenvironment, and cancer-immunity cycle were evaluated using Pearson's correlation coefficients. A $\log_2$ (TPM+1) transformation was performed before analysis. There were four types of immune-associated genes analyzed, including immune checkpoints, chemokines, chemokine receptors, and MHC-related genes. The infiltration scores of different cells in tumor microenvironment were evaluated by seven algorithms, including ESTIMATE (Yoshihara et al., 2013), TIMER (Li et al., 2020), MCP-counter (Becht et al., 2016), CIBERSORT(Newman et al., 2015), quanTIseq (Finotello et al., 2019), xCell (Aran et al., 2017), and EPIC(Racle et al., 2017), using R package "IOBR" (version 0.99.9) (Zeng et al., 2021). The correlation between the expression of *SERPINE1*and marker genes of immune cells, which was obtained from a previous study, was assessed using the "Gene_Corr" module of TIMER2.0 tool (Li et al., 2020; Li et al., 2023). The activity scores of each step of cancer-immunity cycle, which reflects the stepwise events of immune systems' response to cancer (Chen and Mellman, 2013), were calculated by Tracking Tumor Immunophenotype (TIP, http://biocc.hrbmu.edu.cn/TIP/) (Xu et al., 2018).

2.10 Analysis of the relationships of *SERPINE1* expression with TMB, MSI, immunotherapy response, and drug sensitivities

The simple nucleotide variation in pan-cancer data of TCGA annotated by MuTect2 were downloaded from the TCGA GDC (https://portal.gdc.cancer.gov/), and the R package "Maftools" (version 2.8.05) was used to calculate the TMB (Beroukhim et al., 2010; Mayakonda et al., 2018). The MSI in pan-cancer data from TCGA were obtained from a previous study (Bonneville et al., 2017). Cancers with a sample size less than three were eliminated from the TMB and MSI analyses. The relationship between *SERPINE1* expression and immune checkpoint blockade (ICB) therapy response was evaluated by Tumor Immune Syngeneic MOuse (TISMO, http://tismo.cistrome.org/), a syngeneic mouse tumor database for investigation of tumor immunity and

10.3389/fphar.2023.1213891

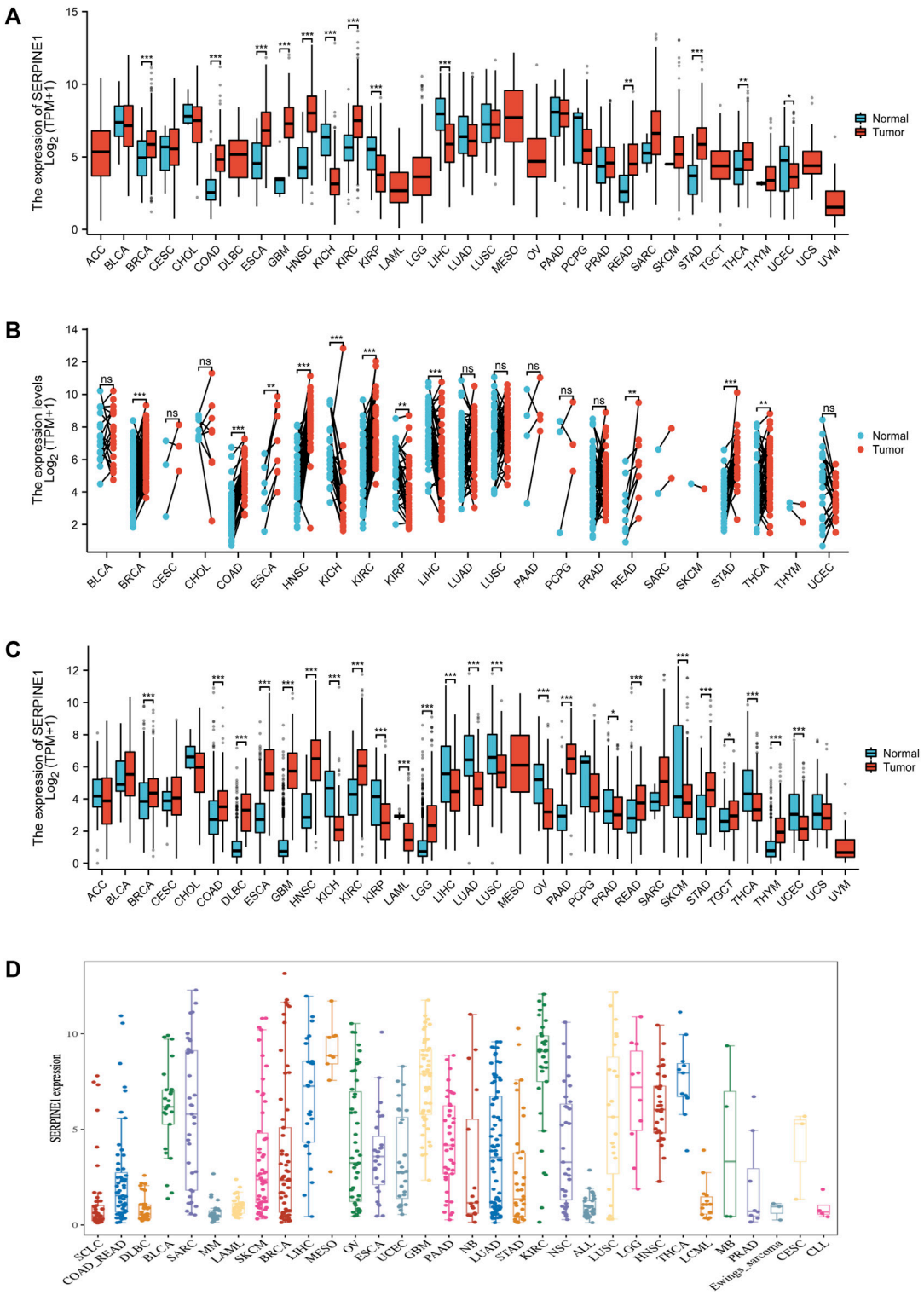


FIGURE 2
SERPINE1 expression in pan-cancer. **(A)** The expression of *SERPINE1* in human cancers and normal tissues based on TCGA database. **(B)** The paired analysis of *SERPINE1* expression in human cancers with adjacent normal tissues. **(C)** The differential expression of *SERPINE1* between tumor and normal tissues among cancers based on the integrated data from TCGA and GTEx datasets. **(D)** The expression levels of different cancer cell lines according to CCLE database. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

immunotherapy response (Zeng et al., 2022). The accuracy of *SERPINE1* in predicting ICB therapy response was further verified by comparing it with other well-known biomarkers in human cohorts based on data from Tumor Immune Dysfunction and Exclusion (TIDE, http://tide.dfci.harvard.edu/) (Fu et al., 2020). Drug sensitivity and mRNA expression data of the NCI-60 cell lines were downloaded from CellMiner (http://discover.nci.nih.gov/cellminer/) (Luna et al., 2021). The relationships of *SERPINE1* expression with TMB, MSI, and drug sensitivities were evaluated using Pearson's correlation test and the differences between groups are statistically evaluated by Wald test using DESeq2 (Love et al., 2014).

2.11 Validating the expression of *SREPINE1* in tissue samples using qRT-PCR

Total RNA was isolated using the TRIzol reagent (Invitrogen). This was followed by its reverse transcription into cDNA using the HiFiScript cDNA Synthesis Kit (CWBio). Quantitative PCR was then conducted using the SYBR Green method in the LightCycler® 480 System (Roche). The relative expression levels of genes were calculated using the $2^{-\Delta\Delta CT}$ method, with β-actin as the internal reference gene. The forward and reverse primer sequences for *SERPINE1* were 5′-CTCATCAGCCACTGGAAAGGCA-3′ and 5′-GACTCGTGAAGTCAGCCTGAAAC-3′, respectively.

2.12 Independent cohorts and proteomic level validation of *SERPINE1* abnormal expression in ccRCC

Six ccRCC cohorts were downloaded from the GEO database to verify the abnormal mRNA expression of *SERPINE1* in ccRCC: GSE14994, GSE17895, GSE53000, GSE53757, GSE68417, and GSE71963. Proteomic data and corresponding clinical data from the CPTAC ccRCC discovery Study (PDC000127) were downloaded from CPTAC (https://pdc.cancer.gov/pdc/) to further explore its expression at the protein level (Clark et al., 2019).

3 Results

3.1 Aberrant expression of *SERPINE1* in pan-cancer

To explore the expression of *SERPINE1* among cancers, mRNA expression data from the TCGA, GTEx, and CCLE databases were analyzed. Based on the TCGA database alone, the mRNA expression of *SERPINE1* was significantly increased in BRCA, COAD, ESCA, GBM, HNSC, KIRC, READ, STAD, and THCA, but significantly decreased in KICH, KIRP, LIHC, and UCEC (Figure 2A). The paired analysis results were consistent with those of unpaired analysis, except that the result of UCEC was no longer significant (Figure 2B). When analyzed in conjunction with data from the GTEx database, *SERPINE1* expression was also increased in DLBC, PAAD, TGCT, and THYM and decreased in LAML, LUAD, LUSC, OV, PRAD, SKCM, and THYM (Figure 1C). Analysis of data from the CCLE database also indicated that the expression of *SERPINE1* was increased in various cancer types, including BLCA, MESO, GBM, and KIRC, similar to analysis of TCGA data (Figures 2A, D).

3.2 *SERPINE1* is enriched in endothelial cells and fibroblasts

Considering the boundedness of bulk RNA-seq, we further investigated the expression of *SERPINE1* at a single-cell level using TISCH2. The results indicated that the expression of *SERPINE1* was enriched in endothelial cells and fibroblasts in most cancer types (Figure 3A). The results for GSE11360 and GSE172301 are shown as examples (Figures 3B–E).

3.3 Genetic alteration and DNA methylation analysis of *SERPINE1* in pan-cancer

Considering that the mRNA expression level of a gene can be influenced by epigenetics, we investigated the genetic alteration and DNA methylation level of *SERPINE1* using cBioPortal and GSCA datasets. Amplification is the most common type of genetic alteration in most cancers, dominated by heterozygous amplification, followed by mutation, dominated by missense mutation (Figures 4A–C). Correlation analysis showed that the expression of *SERPINE1* was negatively correlated with the DNA methylation level and positively correlated with the DNA copy number in most cancer types (Figure 4D).

3.4 *SERPINE1* is a prognostic and diagnostic factor for various cancers

To evaluate the prognostic capacity of *SERPINE1* in cancers, both univariate Cox regression and Kaplan–Meier analysis were performed. *SERPINE1* was found to significantly reduce overall survival in ACC, BLCA, BRCA, CESC, COAD, GBM, HNSC, KIRC, KIRP, LGG, LIHC, LUAD, LUSC, MESO, OV, PAAD, SARC, STAD, THCA, UCEC, and UVM, while it was found to play a protective role for overall survival in PCPG and SKCM (Figure 5A). Besides, it is noteworthy that *SERPINE1* was associated with poor prognosis for all four types of prognostic outcomes in LUAD and PAAD (Figure 5A). Univariate Cox regression results for the OS of all cancers are shown in Figure 5B. These findings suggested the general prognostic value of *SERPINE1* expression in various cancers. The diagnostic value of *SERPINE1* was examined using ROC curve as well. As shown in Supplementary Figure S1, SERPINE1 showed high diagnostic value in 11 kinds of cancers (AUC > 0.7), including CHOL, COAD, ESCA, GBM, HNSC, KICH, KIRC, KIRP, LIHC, READ, and STAD, indicating its crucial role in cancer diagnosis.

10.3389/fphar.2023.1213891

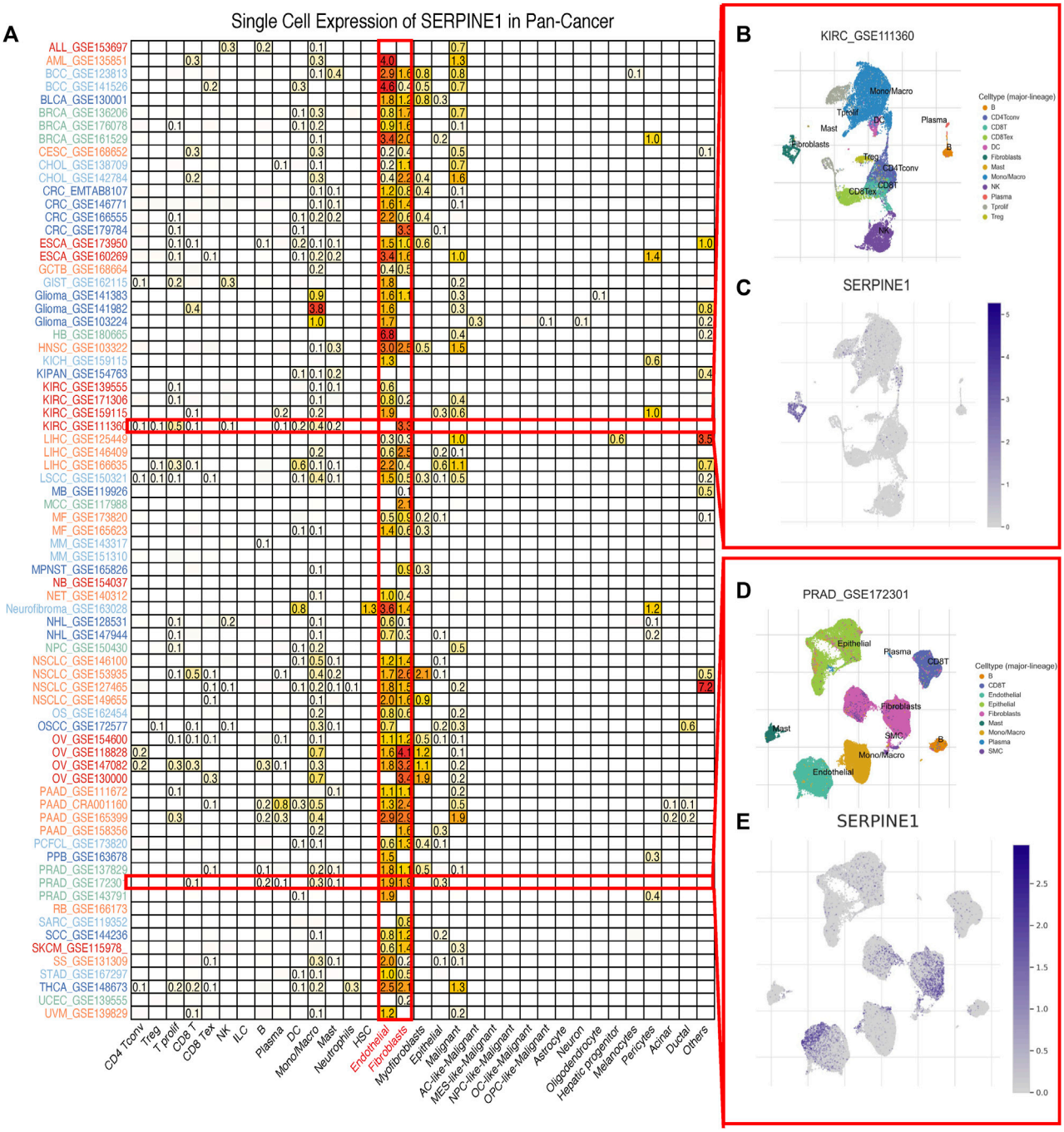


FIGURE 3
Single cell expression of *SERPINE1* in pan-cancer. **(A)** The expression of *SERPINE1* in different cell types across cancers. **(B)** The distribution of cell types in KIRC GSE111360 cohort. **(C)** The single cell expression of *SERPINE1* in GSE111360 cohort. **(D)** The distribution of cell types in PRAD GSE172301 cohort. **(E)** The single cell expression of *SERPINE1* in GSE 172301 cohort.

3.5 *SERPINE1* is associated with the immune response and tumor malignancy in pan-cancer

To better understand the biological roles of *SERPINE1*, we evaluated the enrichment of pathways associated with *SERPINE1* expression using GSEA at the bulk-RNA level and CancerSEA at the single-cell level. The GSEA results revealed that a number of immune-related pathways were enriched in samples with high *SERPINE1* expression in most cancers, including TNF-α signaling via NF-κB, IFN-γ response, IFN-β response, inflammation response, IL-6-JAK-STAT3 signaling, IL-2-STATA5 signaling, and complement and allograft rejection, indicating a potential relationship between *SERPINE1* and cancer immunity (Figure 6A). The results of ssGSEA for selected gene sets were consistent with those of GSEA (Figure 6B). The results from

10.3389/fphar.2023.1213891

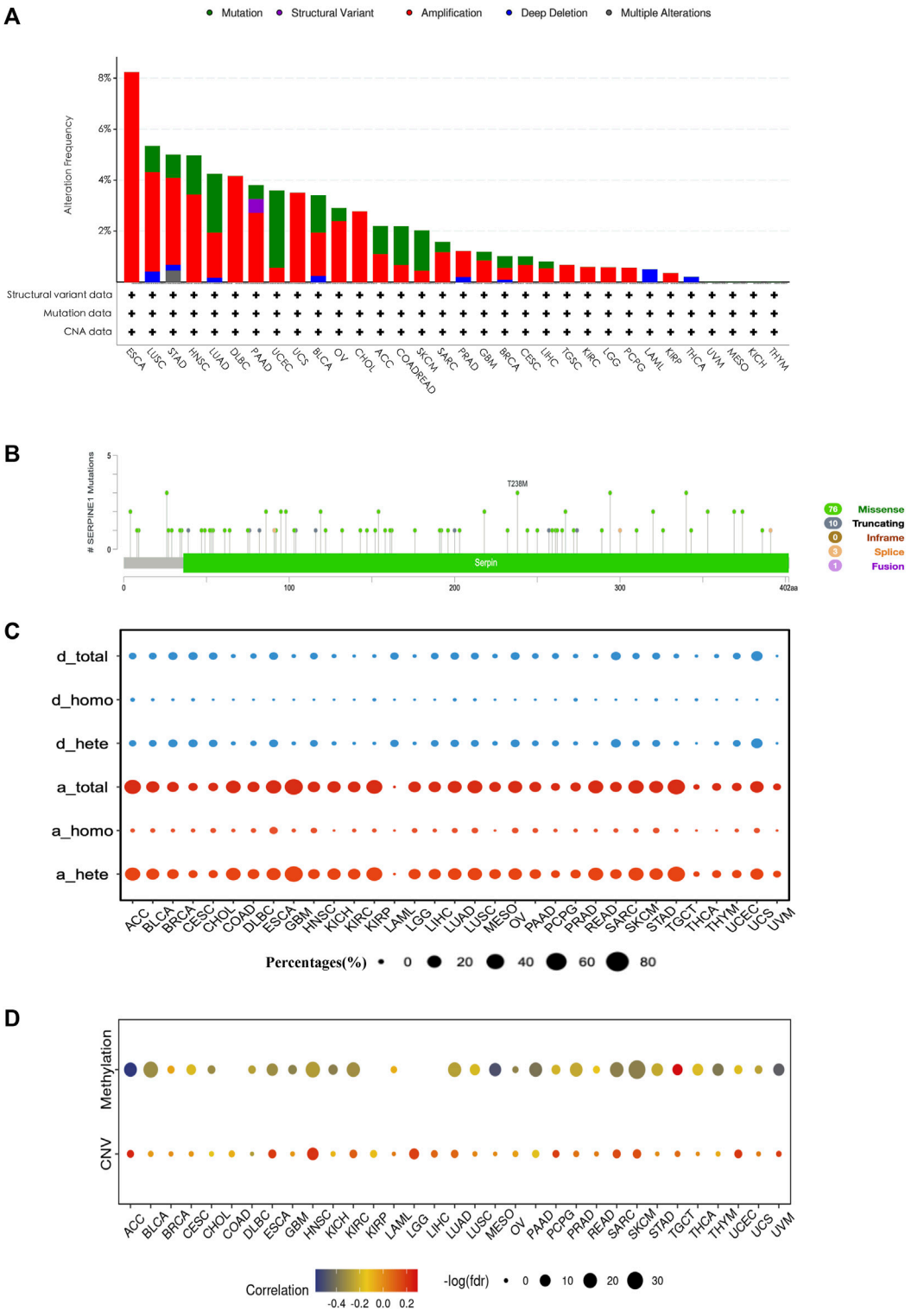


FIGURE 4
The genetic alteration and DNA methylation profile of *SERPINE1* in pan-cancer. **(A)** The genetic alteration condition of *SERPINE1* among cancers based on cBioPortal database. **(B)** The composition of *SERPINE1* mutation in pan-cancer according to cBioPortal database. **(C)** The composition of *SERPINE1* copy number variation according to GSCA database. d_total, total copy number deletion percentage; d_homo, homozygous copy number deletion percentage; d_hete, heterozygous copy number deletion percentage; a_total, total copy number amplification percentages percentage; a_homo, homozygous copy number amplification percentage; a_hete, heterozygous copy number amplification percentage. **(D)** Spearman's rank correlations of *SERPINE1* expression with the DNA methylation level and copy number variance.

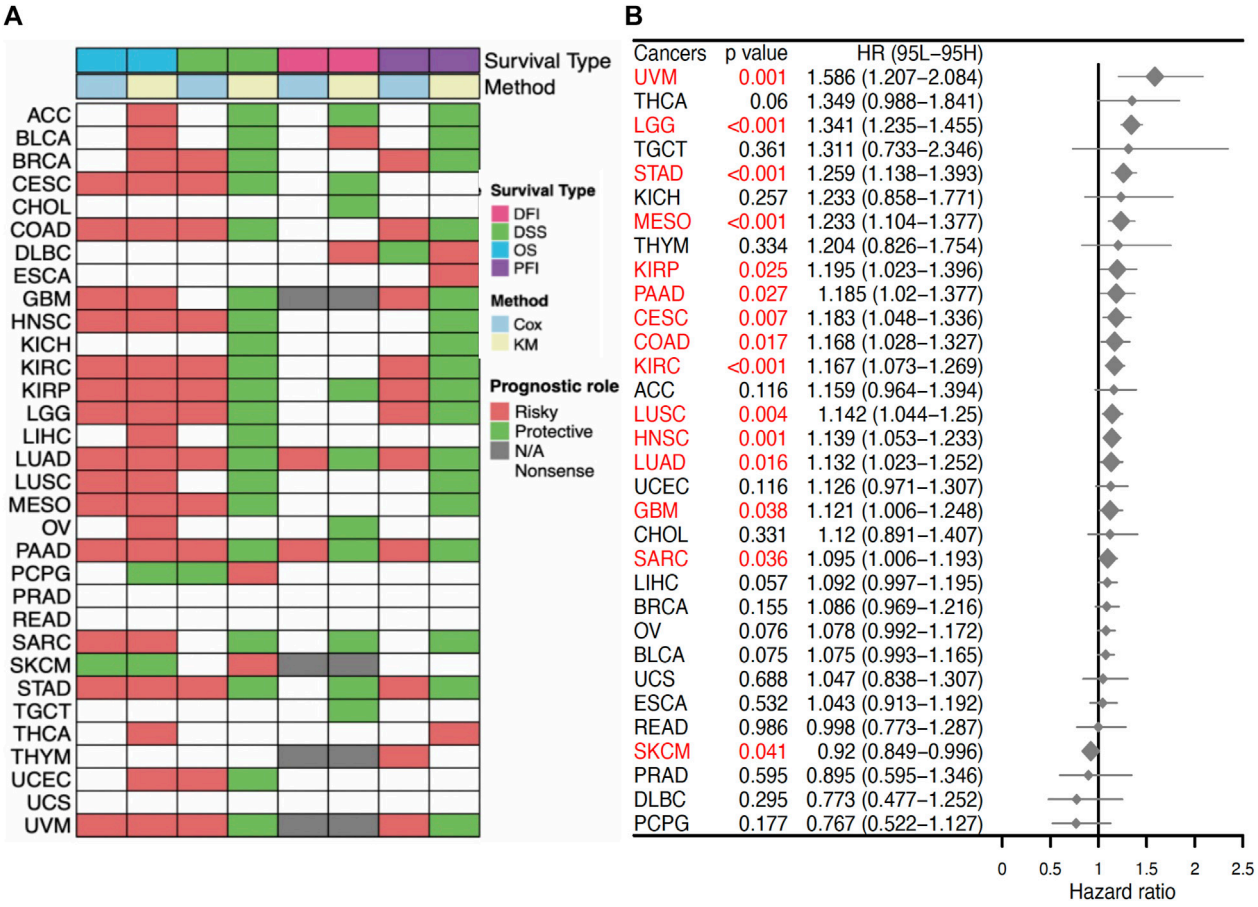


FIGURE 5
Prognostic value of *SERPINE1* in pan-cancer. **(A)** Correlations of *SERPINE1* expression with overall survival (OS), disease-specific survival (DSS), disease-free interval (DFI) and progression-free interval (PFI) based on univariate Cox regression and Kaplan-Meier method. Red indicates a risky role of *SERPINE1* for prognosis, while green represents a protective role. **(B)** The prognostic role of *SERPINE1* expression to patients' overall survival (OS) in cancers by univariate Cox regression. Significant results ($p < 0.05$) are highlighted in red.

CancerSEA demonstrated that *SERPINE1* was positively related with angiogenesis, hypoxia, inflammation, and metastasis in most cancers and negatively related with DNA damage, DNA repair, and stemness in some cancers (Figures 7A, B), which were consistent with the abovementioned results. Based on these results, we speculated that *SERPINE1* may regulate tumor progression by influencing immune-related processes within tumors.

3.6 *SERPINE1* expression is correlated with the expression of various immunoregulators in pan-cancer

Given the well-known roles of immunoregulators in cancer immunity, the relationships between *SERPINE1* and immune checkpoints, chemokines, chemokine receptors, and MHC-related genes were analyzed. The results demonstrated that *SERPINE1* expression was positively correlated with the expression of most inhibitory immune checkpoints, including *CD274* (PD-L1), *PDCD1* (PD-1), *CTLA4,* and *HAVCR2* (TIM-3) in most cancers, especially GBM-LGG, KIPAN, UVM, PAAD, and COAD-READ (Figure 8A). Moreover, expression of *SERPINE1* was found to be significantly associated with expressions of various immunostimulators, chemokines, chemokine receptors and MHC-related genes in pan cancer as well (Figures 8B–E).

3.7 *SERPINE1* expression is correlated with immune cell infiltration and cancer-immunity cycle in pan-cancer

To gain further insight into the relationship between *SERPINE1* and cancer immunity, we conducted immune cell infiltration analysis using different algorithms across a range of cancers. All algorithms showed that *SERPINE1* expression was significantly correlated with the infiltration scores of multiple immune cells, although there was some variation in the results for specific cell types. The results from ESTIMATE indicated that *SERPINE1* expression was positively correlated with both the immune score and stromal score in most cancers, especially KIPAN, GBM-LGG, and COAD-READ. Notably, no significantly negative relationship was found between *SERPINE1* expression and stromal score in all cancers (Supplementary Figure S2). According to the CIBERSORT algorithm, *SERPINE1* expression was negatively correlated with the

10.3389/fphar.2023.1213891

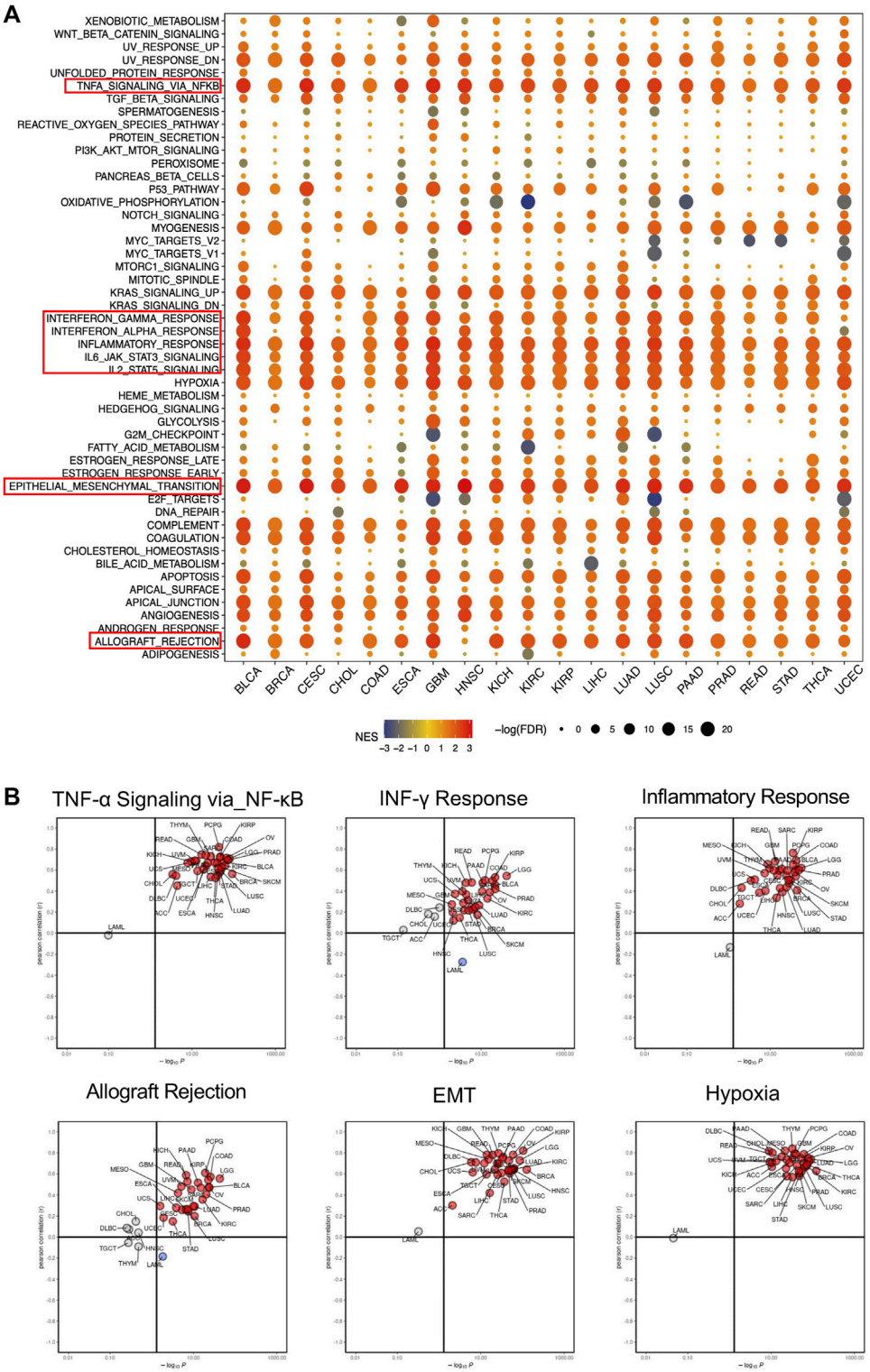


FIGURE 6
Gene function analysis of *SERPINE1* in pan-cancer based on GSEA and ssGSEA. **(A)** The hallmarks gene set enrichment analysis (GSEA) of *SERPINE1* in pan-cancer. The size of circle represents the FDR value of enrich term in each cancer, and the color indicates the normalized enrichment score (NES) of each term. **(B)** The single sample gene set enrichment analysis (ssGSEA) of *SERPINE1* in pan-cancer. Cancer types with significant positive correlations ($p < 0.05$) with ssGSEA scores are highlighted in red, while those with significant negative correlations ($p < 0.05$) are highlighted in blue.

10.3389/fphar.2023.1213891

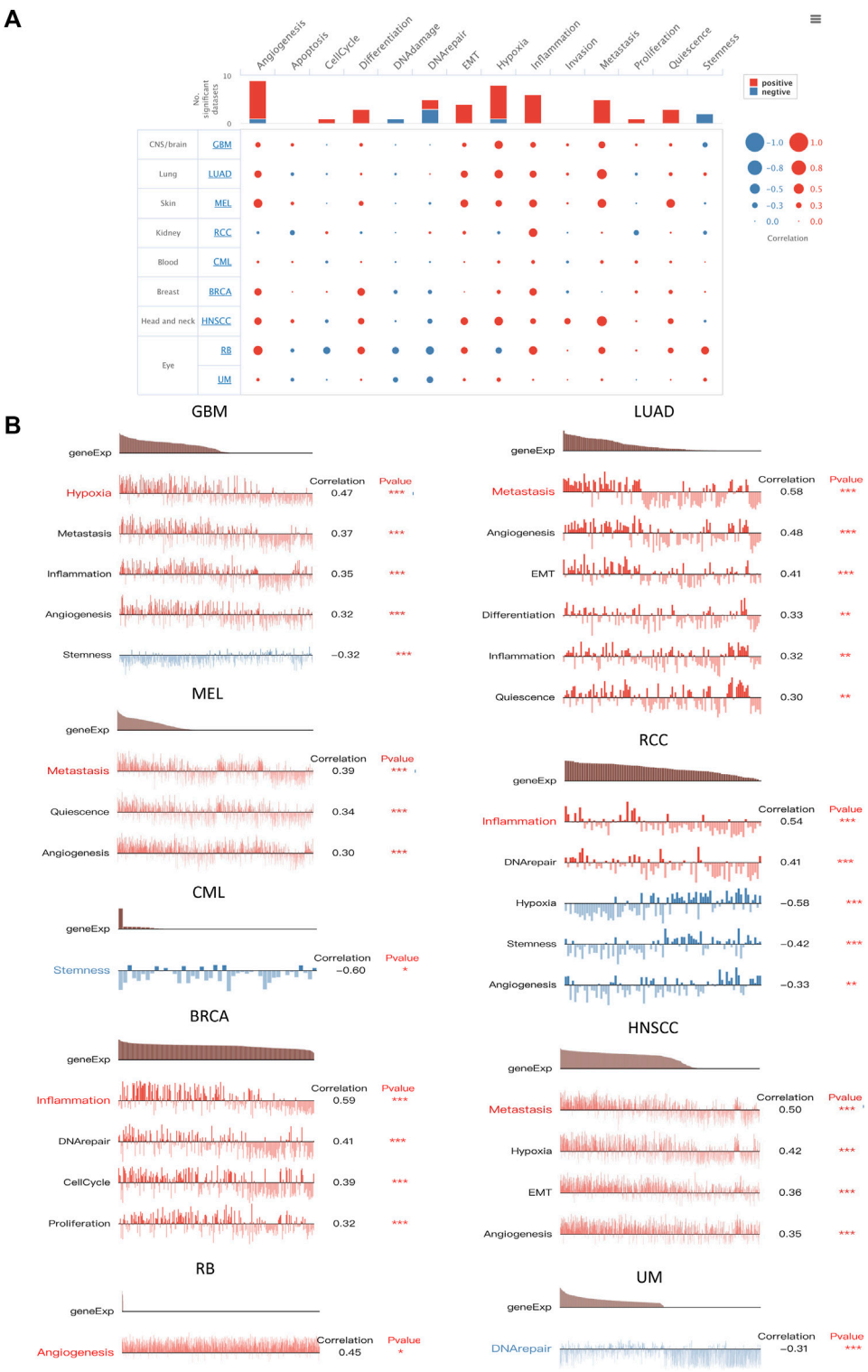


FIGURE 7
Functional relevance analysis of *SERPINE1* in single-cell resolution by CancerSEA. **(A)** The association between *SERPINE1* expression and single-cell states in cancers. **(B)** Single-cell states that significantly associated with each cancer. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

infiltration of plasma cells, memory B cells, naïve $CD4^+$ T cells, $CD8^+$ T cells, follicular helper T cells, regulatory T cells, activated NK cells, and resting mast cells and positively correlated with the infiltration of activated memory $CD4^+$ T cells, macrophages, activated mast cells, and neutrophils in most cancers (Figure 9A). As for the xCell algorithm, its results indicated that *SERPINE1* expression was negatively correlated with the infiltration of $CD4^+$ Tcm cells, NKT cells, plasma cells, and Th1 cells and positively correlated

10.3389/fphar.2023.1213891

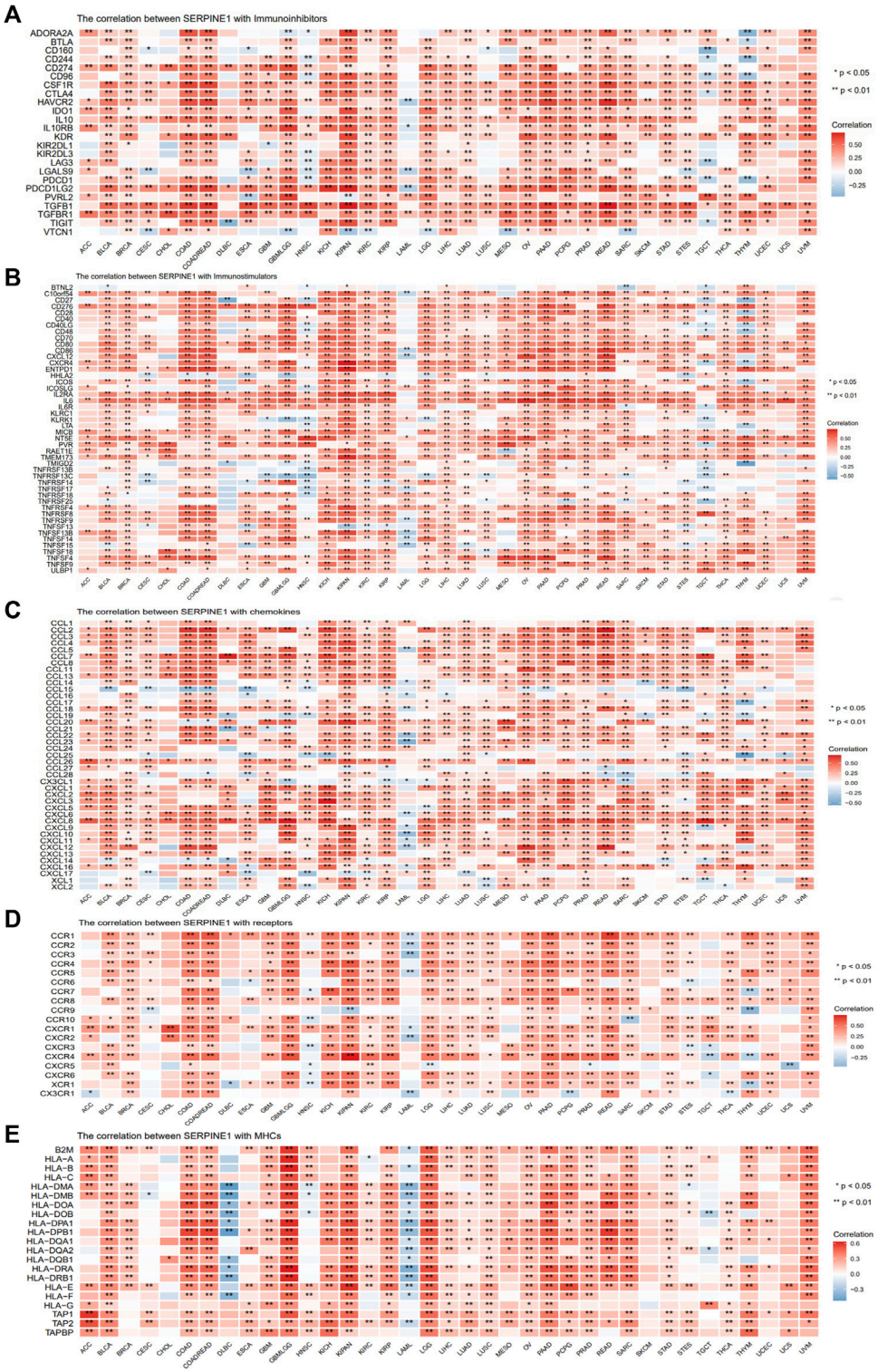


FIGURE 8
Co-expression analysis of *SERPINE1* with immunoregulators in pan-cancer. **(A)** The Pearson's correlation of *SERPINE1* with inhibitory immune-checkpoints genes in cancers. **(B)** The Pearson's correlation of *SERPINE1* with stimulatory immune-checkpoints genes in cancers. **(C)** The Pearson's correlation of *SERPINE1* with chemokines. **(D)** The Pearson's correlation of *SERPINE1* with chemokine receptors **(E)** The Pearson's correlation of *SERPINE1* with MHC-related genes in cancers. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

10.3389/fphar.2023.1213891

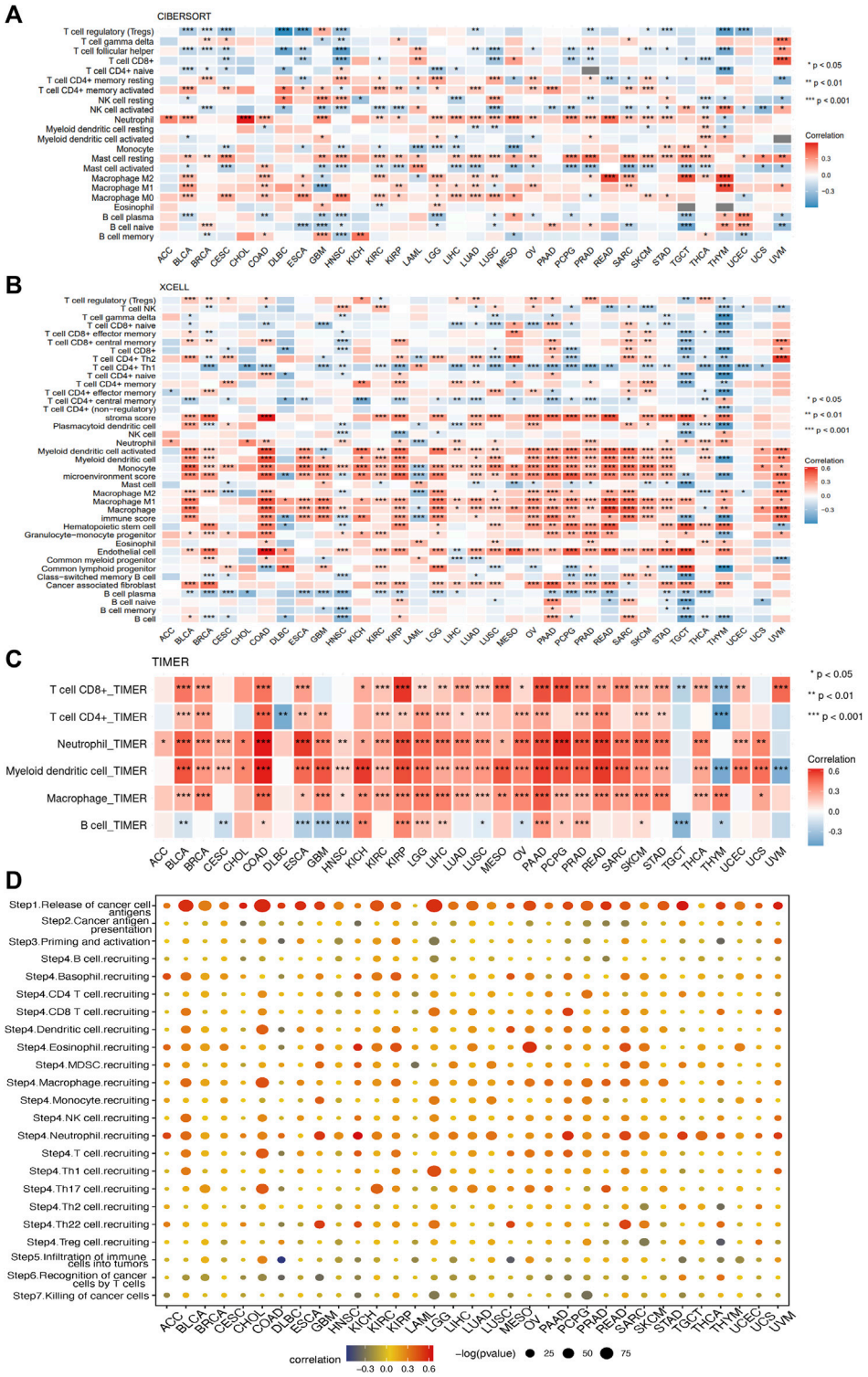


FIGURE 9
Correlations of SERPINE1 expression with immune cells infiltrations in pan-cancer. **(A)** The correlations of *SERPINE1* expression with infiltration of 22 types of immune cells in cancers based on CIBERSORT algorithm. **(B)** The correlations of *SERPINE1* expression with infiltration of 64 types of immune and stromal cell in cancers based on xCell algorithm. **(C)** The correlations of *SERPINE1* expression with infiltration of six types of immune cell in cancers based on TIMER algorithm. **(D)** Correlation between SERPINE1 expression and cancer-immunity cycle. The size of circle represents the *p*-value of correlation test, and the color indicates the Pearson's correlation coefficient. $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

10.3389/fphar.2023.1213891

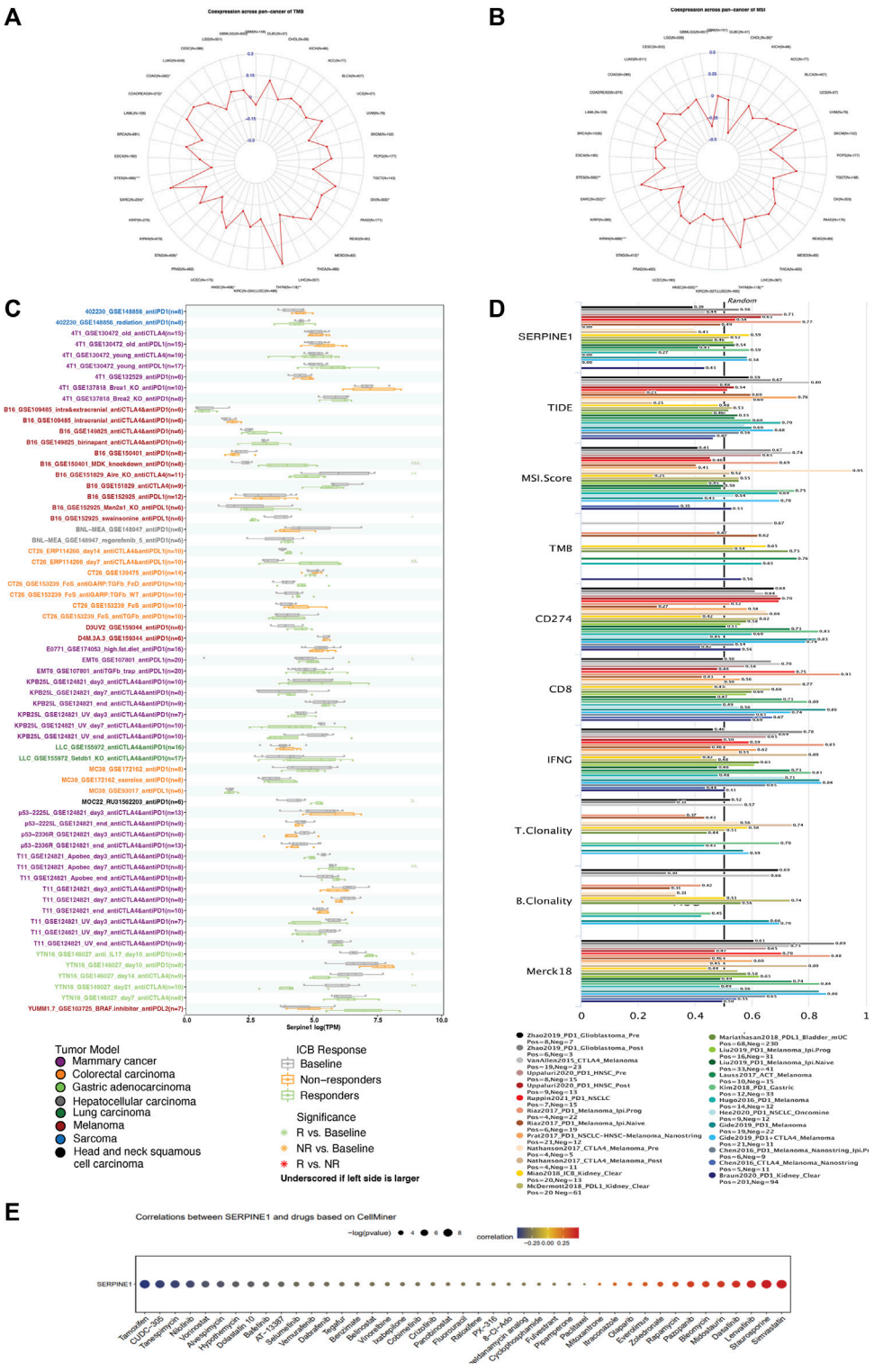


FIGURE 10
Relevance analysis of *SERPINE1* expression with tumor mutation burden (TMB), microsatellite instability (MSI), immunotherapy response, and drug sensitivity. **(A)** The correlation of *SERPINE1* with TMB in pan-cancer. **(B)** The correlation of *SERPINE1* with MSI in pan-cancer. **(C)** Expression of *SERPINE1* in mice with different responses to immune checkpoint blockade (ICB) therapy based on data from TISMO database. **(D)** The area under the receiver operating characteristic curve (AUC) of *SERPINE1* and other biomarkers in predicting ICB therapy response. **(E)** The association of *SERPINE1* expression with drug sensitivity based on CellMiner database. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

with the infiltration of fibroblasts, macrophages, neutrophils and regulatory T cells (Figure 9B). The results from TIMER algorithm revealed that *SERPINE1* expression was positively correlated with the infiltration of all cell types, except B cells which did not reach significance in most cancers (Figure 9C). The results of MCP-counter algorithm were similar with that of TIMER algorithm, except that the infiltration of T cells showed significant negative correlation with *SERPINE1* expression in 10 types of cancers (Supplementary Figure S3A). The results of the quanTIseq algorithm indicated that *SERPINE1* expression was positively correlated with M1 macrophages, M2 macrophages, neutrophils and regulatory T cells (Supplementary Figure S3B). The results of the EPIC algorithm showed that *SERPINE1* expression was positively correlated with the infiltration of cancer associated fibroblasts (CAFs), endothelial cells, and macrophages, and negatively correlated with the infiltration of $CD8^+$ T cells (Supplementary Figure S3C). We also analyzed the correlation between the expression of *SERPINE1* and immune cells' marker genes. The results showed the expression of SERPINE1 was positively correlated with that of most immune cells' marker genes in most cancer, especially the maker genes of neutrophils, monocyte, tumor-associated macrophages (TAMs), M2 macrophages, and regulatory T cells (Supplementary Figure S4). These results suggested that *SERPINE1* may regulate the immune response of cancers.

The relationship between *SERPINE1* expression and the cancer-immunity cycle, which can reflect the stepwise events in anticancer immune response, was analyzed as well. As shown in Figure 8D, high expression of *SERPINE1* was positively correlated with the activities of step 1 and step 4, which represents the release of cancer cell antigen and the recruiting of immune cells respectively, but negatively correlated with the activities of step 5, 6 and 7, which represents the infiltration of immune cells into tumors, the recognition of cancer cells by T cells, and the killing of cancer cells, in lots of cancer types. These results provide further evidence that *SERPINE1* may be involved in the regulation of the immune response to cancer.

3.8 *SERPINE1* is associated with TMB, MSI, immunotherapy response, and drug sensitivity in pan-cancer

Due to the predictive value of TMB and MSI for immunotherapy response, the relationships between TMB, MSI, and *SERPINE1* expression were evaluated. *SERPINE1* expression was found to be positively associated with TMB in THYM, COAD, COADREAD, SARC, and OV and negatively correlated with STES, STAD, and HNSC (Figure 10A). In addition, *SERPINE1* expression levels in seven cancers were also significantly correlated with MSI in GBM-LGG, KIPAN, STES, HNSC, STAD, CHOL, SARC and THYM (Figure 10B). Then, we directly analyzed the relationship between *SERPINE1* expression and immunotherapy response. As shown in Figure 10C, *SERPINE1* expression could significantly predict the immunotherapy response in nine murine cohorts, with responders showing elevated *SERPINE1* expression levels in six cohorts and decreased *SERPINE1* expression levels in three cohorts. We also compared the predictive power of *SERPINE1* to ICB response with other biomarkers in human cohorts. The area under curve (AUC) of *SERPINE1* was above 0.5 in 12 cohorts and above 0.7 in two cohorts, which is similar to T clonality but lower than other classical biomarkers such as MSI, TMB, and CD274 (Figure 10D). These results indicated that *SERPINE1* could predict the response of cancer patients to immunotherapy to some extent, though its predictive power might be lower than that of some classical markers. Furthermore, the correlations between *SERPINE1* expression and drug sensitivities were explored using CellMiner. We found that *SERPINE1* expression was positively correlated with sensitivities towards 13 drugs, including simvastatin, staurosporine, and pazopanib, and negatively correlated with sensitivities towards 30 drugs, including tamoxifen, tanespimycin, and nilotinib (Figure 10E).

3.9 Elevated expression of *SERPINE1* is associated with unfavorable prognosis in patients with ccRCC

Given the limited availability of reports on the role of *SERPINE1* in ccRCC, we investigated the role of *SERPINE1* in ccRCC. Our findings revealed that *SERPINE1* expression is associated with male sex, lymph node metastasis, higher T stage, higher histological grade, higher pathological stage, and different immune subtypes (Figures 11A–H). Cox regression analyses revealed upregulated *SERPINE1* expression as a risk factor for OS, DSS, and PFS using data from the TCGA database and as a risk factor for OS using GSE167573 cohort data from the GEO database (Figure 11I). However, the result of Cox regression using RECA-EU project data from ICGC database indicated upregulated *SERPINE1* expression as a protective factor for OS (Figure 11I). The multivariate Cox regression was performed based on TCGA database, and the results verified the independence of *SERPINE1*'s prognostic value for patient with ccRCC (Supplementary Figure S5).

3.10 *SERPINE1* expression is associated with the tumor microenvironment in clear cell renal cell carcinoma

To further investigate the role of *SERPINE1* in ccRCC, gene set enrichment analysis was performed. The results revealed that *SERPINE1* expression was significantly related with various biological processes. Further, GO analysis showed that *SERPINE1* expression was positively correlated with collagen fibril organization, collagen biosynthetic process, and the regulation of T helper 1 type immune response, while it was negatively correlated with metabolism-related processes, such as tricarboxylic acid cycle, fatty acid beta oxidation, and the alpha amino acid catabolic process (Figure 12A). In the KEGG pathway analysis, several malignancy-related pathways, such as the P53 signaling pathway, cell cycle and DNA replication pathways, as well as immunity-related pathways, such as natural killer cell mediated cytotoxicity and T cell receptor signaling pathways, were found to be significantly positively correlated with *SERPINE1* expression (Figure 12B). Similar results were obtained through the hallmark analysis (Figures 12C, D). These findings revealed that *SERPINE1* may be

10.3389/fphar.2023.1213891

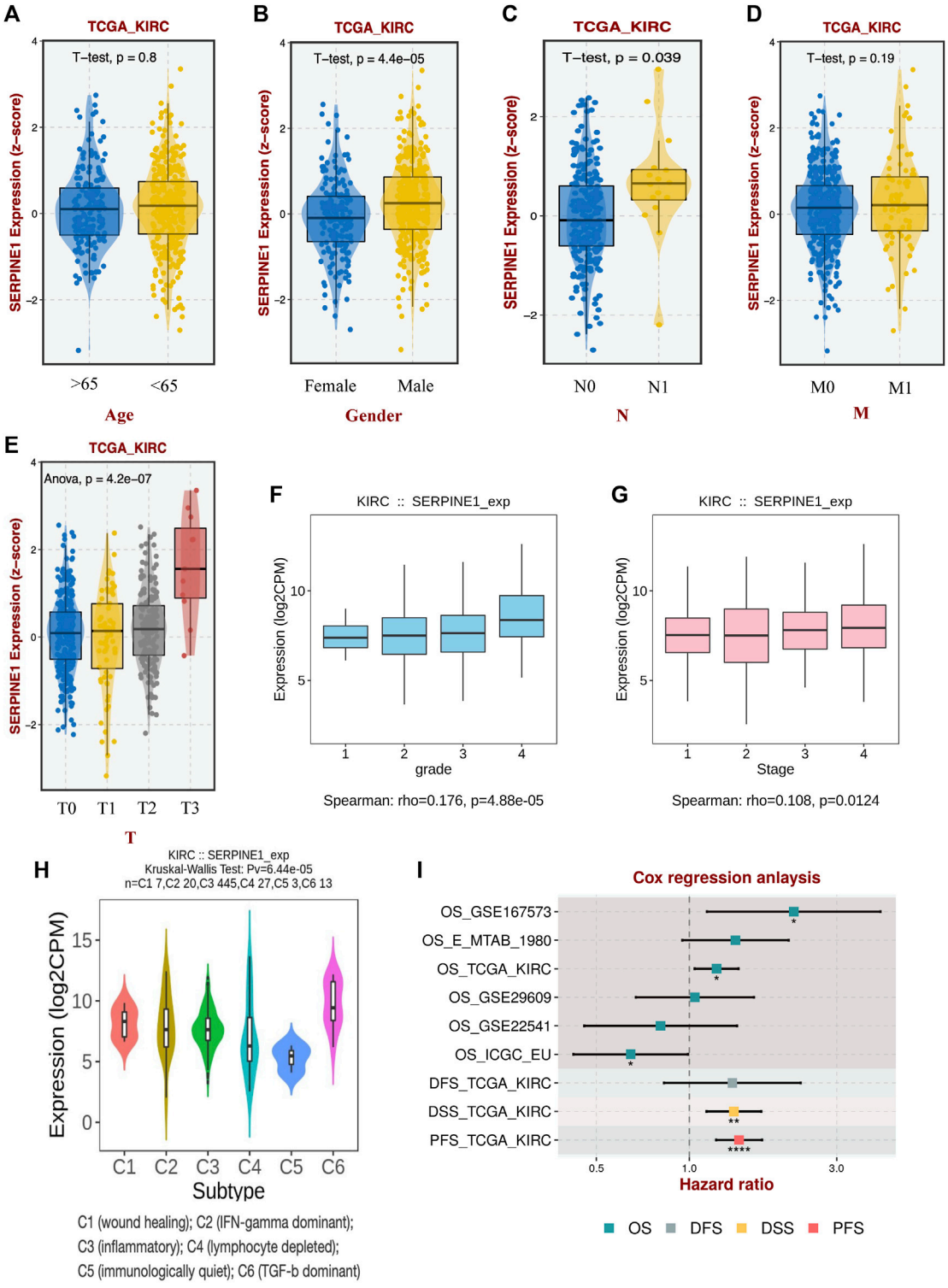


FIGURE 11
Association between *SERPINE1* and clinicopathological characteristics of clear cell renal cell carcinoma. **(A)** The expression of *SERPINE1* in patients of different ages. **(B)** The expression of *SERPINE1* in patients of different genders. **(C)** The expression of *SERPINE1* in patients of different lymph node metastasis statuses. **(D)** The expression of *SERPINE1* in patients of different remote metastasis statuses. **(E)** The expression of *SERPINE1* in patients of different primary tumor stages. **(F)** The expression of *SERPINE1* in different histological grades. **(G)** The expression of *SERPINE1* in different cancer stages. **(H)** The expression of *SERPINE1* in different immune subtypes. **(I)** The correlation of *SERPINE1* with prognosis based on different datasets. OS, overall survival; DFS, disease-free survival; DSS, disease-specific survival; PFS, progression-free survival; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

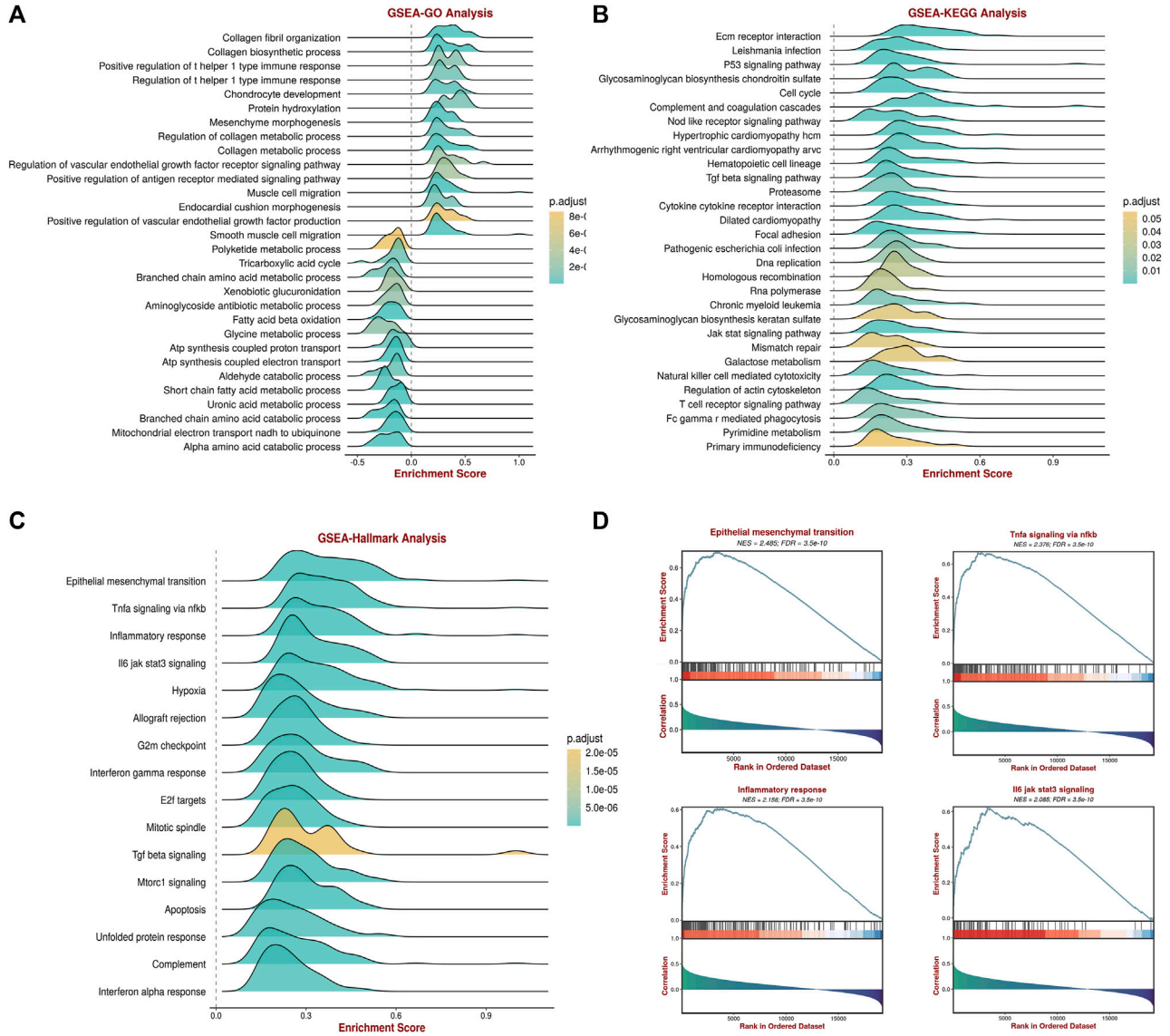


FIGURE 12
Gene set enrichment analysis (GSEA) of *SERPINE1* in clear cell renal cell carcinoma (ccRCC). **(A)** The associations of *SERPINE1* expression with several Gene Ontology (GO) terms. **(B)** The associations of *SERPINE1* expression with several Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. **(C)** The associations of *SERPINE1* expression with several Hallmark terms. **(D)** The GSEA results of certain Hallmark terms.

involved in the formation of the tumor microenvironment in ccRCC and were consistent with the results in pan-cancer.

3.11 Validation of *SERPINE1* expression in tumor tissues using qRT-PCR, independent cohorts, and proteomic data

Finally, we confirmed the aberrant expression of *SERPINE1* in ccRCC samples. Our ccRCC patient samples and six independent GEO cohorts exhibited significantly increased levels of *SERPINE1* expression, as expected (Figures 13A, B). Additionally, analysis of CPTAC data revealed that *SERPINE1* expression was also elevated at the protein level in ccRCC samples (Figures 13C–F), which further enhanced the credibility of our study findings.

4 Discussion

SERPINE1, a regulator of the fibrinolytic system, was found to be associated with tumor progression and metastasis in several cancers; however, its detailed mechanisms of action in various cancers remain obscure (Hanekom et al., 2002; Becker et al., 2010; Duffy et al., 2014; Mashiko et al., 2015; Jevrić et al., 2019; Sotiropoulos et al., 2019). In the present research, we performed a multi-omics integrated analysis to explore the expression, prognostic value, and possible underlying mechanisms of action of *SERPINE1* in pan-cancer.

Expression analysis based on the TCGA and GTEx databases indicated the aberrant expression of *SERPINE1* in several cancers. Overexpression of *SERPINE1* and its correlation with poor prognosis has been reported in several cancers (Hanekom et al.,

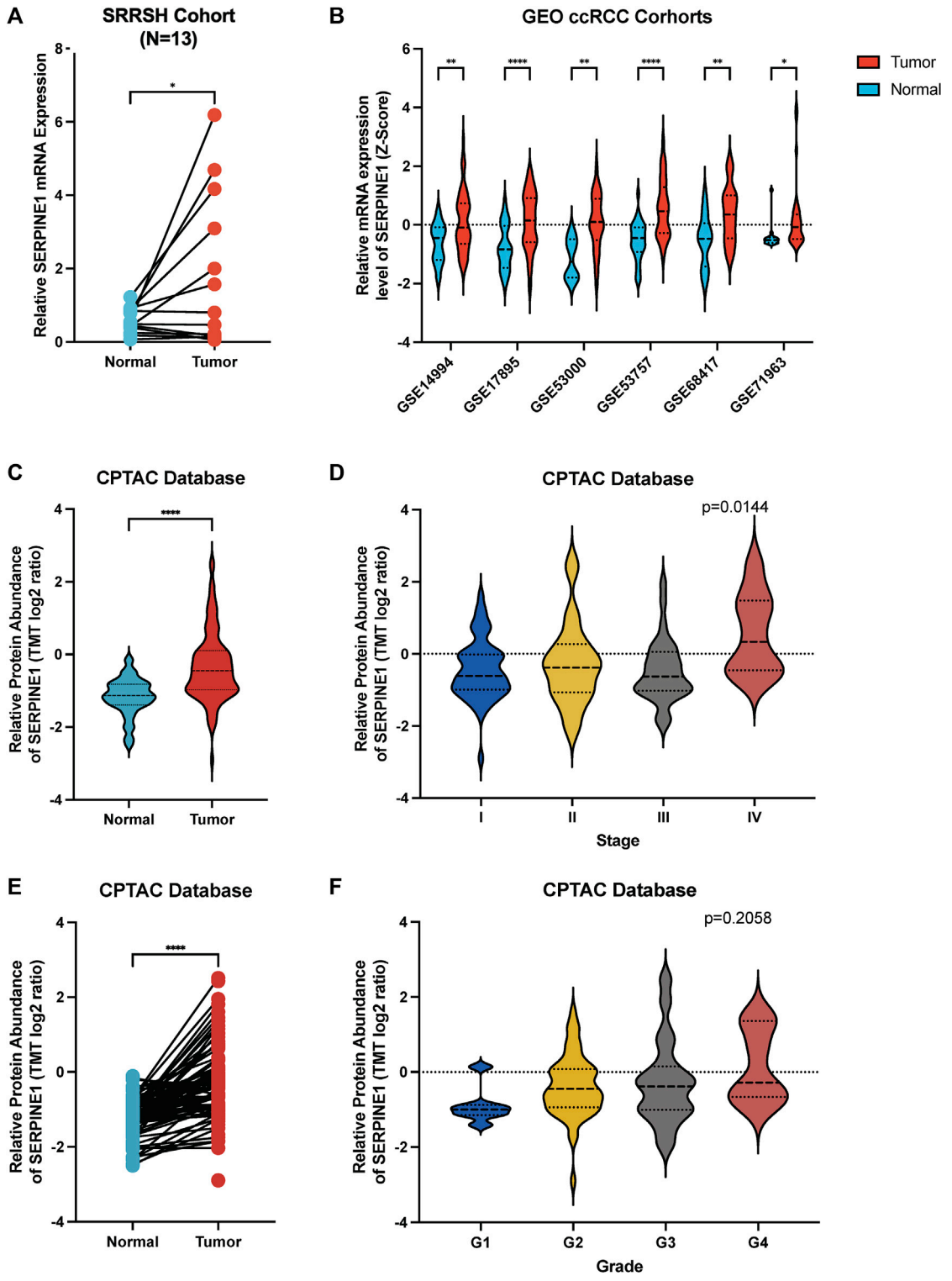


FIGURE 13
Validation of aberrant expression of *SERPINE1* in ccRCC by qRT-PCR, independent cohorts and proteome. **(A)** qRT-PCR analysis of *SERPINE1* expression in clear cell renal cell carcinoma and paired adjacent normal tissues based on patient samples from Sir Run Run Shaw Hospital. **(B)** Analysis of 6 GEO cohorts regarding *SERPINE1* expression in ccRCC and normal samples. **(C)** Unpaired analysis of *SERPINE1* protein abundance in ccRCC and adjacent normal tissue based on CPTAC database. **(D)** The SERPINE1 protein abundance in patients with different stages of ccRCC. **(E)** Paired analysis of SERPINE1 protein abundance in ccRCC and adjacent normal tissue based on CPTAC database. **(F)** The SERPINE1 protein abundance in patients with different histological grades of ccRCC $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

2002; Becker et al., 2010; Duffy et al., 2014; Nakatsuka et al., 2017; Jevrić et al., 2019; Sotiropoulos et al., 2019). The results of the present study confirmed the prevalence of aberrant *SERPINE1* expression among cancers. Given the bulk RNA sequencing data is the average expression of different cells, which will lead to the loss of information about cells heterogeneity, we analyzed the expression of *SERPINE1* at single-cell level as well, and the results showed that *SERPINE1* expression was enriched in the endothelial cells and fibroblasts, consistent with the findings of a previous study, indicating the functions of *SERPINE1* might be related with these two cell types, such as angiogenesis and regulating TME (Placencio and DeClerck, 2015; Clark et al., 2019; Chen et al., 2021b; de Visser and Joyce, 2023). Subsequently, we analyzed possible reasons for the aberrant expression of *SERPINE1* in cancers. DNA methylation level and DNA copy number variation are well-known ways to influence the gene expression, thus their correlation with *SERPINE1* was analyzed (Morgan et al., 2018; Pös et al., 2021). *SERPINE1* expression showed a positive correlation with copy number variance and a negative correlation with the DNA promoter methylation level in most cancers, suggesting its potential role in CNV and the DNA methylation level. Some studies have indicated that some microRNAs and long noncoding RNAs are involved in the regulation of *SERPINE1* expression (Tan et al., 2021; Teng et al., 2021; Zhao and Liu, 2021). Moreover, transforming growth factor β (TGF-β) has been implicated in the regulation of *SERPINE1* expression (Kutz et al., 2001; Ma et al., 2002), and a positive correlation of TGF-β expression with *SERPINE1* expression was observed through the GSEA analysis conducted in the present study.

Next, we explored the relationship between *SERPINE1* expression and the prognosis of cancer patients. High *SERPINE1* expression was found to be a risk factor for overall survival in several cancers, consistent with the findings of previous studies (Hanekom et al., 2002; Becker et al., 2010; Duffy et al., 2014; Nakatsuka et al., 2017; Jevrić et al., 2019; Sotiropoulos et al., 2019). Hence, we speculated that *SERPINE1* may be a prognostic biomarker for various cancers.

To further explore the mechanisms underlying the role of *SERPINE1* in cancers, we performed GSEA on pan-cancer data from TCGA. Several immune-related pathways were found to be significantly associated with high *SERPINE1* expression in most cancers, including TNF-α signaling via NF-κβ, INF-γ response, and inflammation response. Similar results were also found in the single cell-level analysis performed by us using CancerSEA. Therefore, we believe that *SERPINE1* is involved in cancer immunity and tumor malignancy (Iwaki et al., 2012; Declerck and Gils, 2013; Placencio and DeClerck, 2015; Chen et al., 2021; Sillen and Declerck, 2021).

Since immunoregulators are known to be important for the immune response, we next performed a co-expression analysis of *SERPINE1* to further understand its roles in cancer immunity. The results demonstrated that *SERPINE1* expression was positively and significantly correlated with the expression of immune-checkpoints, chemokines, chemokine receptors, and MHC-related genes in most cancers, indicating the remarkable effect of *SERPINE1* on the immune system and consistent with previous studies about the pro-inflammatory and pro-angiogenesis roles of *SERPINE1* (Iwaki et al., 2012; Declerck and Gils, 2013; Placencio and DeClerck, 2015). Traditionally, $CD8^+$ T cells, memory B cells, plasma cells, follicular helper T cells, activated NK cells, NKT cells, and M1 macrophages are thought to be anti-cancer cells in the tumor microenvironment (TME), while regulatory T cells, M2 macrophages, and cancer associated fibroblasts are considered to be pro-cancer cells (Fridman et al., 2012; Sica and Mantovani, 2012; Bae et al., 2019; Chen and Song, 2019; Togashi et al., 2019; St Paul and Ohashi, 2020; Laskowski et al., 2022; Laumont et al., 2022; Li H. et al., 2023; Cai et al., 2023; Gutiérrez-Melo and Baumjohann, 2023). Therefore, we speculated that high *SERPINE1* expression may play an immunosuppressive role in the tumor microenvironment due to its inverse correlation with several anti-cancer cells and positive correlation with several pro-cancer cells. The result of cancer-immunity cycle analysis further validated our speculation. Although *SERPINE1* expression was positive correlated with the recruiting of most immune cells, the infiltration of immune cells into tumors, recognition of cancer cells by T cells, and killing of cancer cells were found to be negatively correlated with *SERPINE1* expression in lots of cancers. However, given that our bioinformatic analysis based on bulk-RNA sequencing data has several limitations, further investigations are warranted.

Next, we validated the abnormal expression of *SERPINE1* and its potential biological functions in clear cell renal cell carcinoma. Upregulated *SERPINE1* expression was found to be associated with several clinical features of ccRCC, such as lymph node metastasis, high T stage, high histological grade, and high pathological stage. As for the gene function analysis performed using GSEA, collagen-associated processes, immune-associated processes, and malignancy-related pathways were found to be positively correlated with *SERPINE1* expression in ccRCC. These findings are consistent with our speculation that *SERPINE1* expression is involved in the regulation of the tumor microenvironment. Next, we further explored *SERPINE1* expression in ccRCC in clinical patient samples and independent datasets.

Given the role of abnormal *SERPINE1* expression in cancer development, it is important to consider how to interfere its expression to benefit cancer patients. As PAI-1 is a well-known regulator of the plasminogen activation system, many efforts have been devoted to the development of selective PAI-1 inhibitors (Fortenberry, 2013; Placencio et al., 2015; Sillen and Declerck, 2020; 2021). Some marketed drugs, including insulin-sensitizing agents, angiotensin-converting enzyme inhibitors (ACEI), and statins, have shown the ability to attenuate the synthesis or secretion of *SERPINE1* (Brown et al., 2002b; Ersoy et al., 2008; Baluta and Vintila, 2015). Specifically, insulin resistance has been found to be associated with elevated plasma PAI-1 levels (Juhan-Vague and Alessi, 1997), and both proinsulin and insulin can stimulate PAI-1 expression (Sakamoto et al., 1999; Nordt et al., 2001), thus insulin-sensitizing agents such as metformin may have independent effects in decreasing PAI-1 levels in patients with type 2 diabetic (Ersoy et al., 2008). Besides, activation of renin-angiotensin-aldosterone system (RASS) has also been found to be involved in the regulation of PAI-1 levels, and ACEI, such as quinapril, ramipril, and perindopril, have shown the ability to reduce PAI-1 level in both healthy people and hypertensive patients (Brown et al., 1998; 2002a; 2002b; Erdem et al., 1999). Statins can inhibit the production of PAI-1 by regulating a variety of signaling pathways as well (Ma et al., 2002; Laumen et al., 2008; Dunoyer-Geindre et al., 2011; Ni et al., 2013). However, the role of

these drugs in cancer remains unclear. Apart from these traditional drugs, there are many more novel drugs under development. Due to the crucial role of reactive center loop (RCL) in inhibitory mechanism of PAI-1, several synthetic peptides that mimic various parts of the RCL of PAI-1 were developed (Eitzman et al., 1995; Kvassman et al., 1995; Xue et al., 1998; D'Amico et al., 2012). In general, peptides mimicking the C-terminal part of the loop can accelerate the irreversible transition of PAI-1 to its latent form, while peptides mimicking the N-terminal part can induce it to be cleaved (D'Amico et al., 2012). In addition, several RNA aptamers designed to block the interaction of pai-1 with its partner have been developed as well, and have shown the ability to reduce cancer migration, invasion, and angiogenesis (Blake et al., 2009; Damare et al., 2014; Fortenberry et al., 2016). Another class of PAI-1 inhibitors is small molecules. These compounds work by binding a common binding pocket within the flexible joint region of PAI-1 or by interfering structural elements within that region through interactions at the surface of PAI-1, thereby inducing the substrate behavior of PAI-1 and its conversion to an inert form (Egelund et al., 2001; Fjellström et al., 2013; Lin et al., 2013; Sillen et al., 2021). There are also lots of antibody based PAI-1 inhibitors, including antibodies and antibody derivatives. Their target sits and mechanisms are more extensive and complex than drugs mentioned above (Sillen and Declerck, 2020). However, although these different types of novel PAI-1inhibitors have been shown to be efficient *in vivo* or *in vitro*, their role and safety in cancers remain unclear and require more research and clinical trials to understand them, but they still promise a bright future for cancer therapies based on *SERPINE1*

Admittedly, there are several limitations to our study. First, there are some contradictory findings in our study. For example, in the Cox regression analysis, the result based on RECA-EU project data from ICGC database indicated a protective role of *SERPINE1* in renal cell carcinoma, which is contrary to the results of other datasets. We speculate that this may be related to the heterogeneity of samples from different datasets, and the sample size of this dataset (n = 91) is smaller than that of TCGA database, but further studies and follow-up are still needed to verify it. In addition, the results of different cancers are not always consistent in pan-cancer analysis. Therefore, further research focused on the differences in *SERPINE1* roles among cancers is needed as well. Second, although we have identified the prognostic value and possible action mechanisms of *SERPINE1* in cancers through correlation analysis in the present study, direct evidence supporting these conclusions are required. Finally, our research is mainly based on public databases, which may have inevitably introduced systemic bias; further experimental verification is therefore needed.

5 Conclusion

In the present study, we conducted a comprehensive multi-omics analysis of *SERPINE1* in pan-cancer, revealing its prognostic value and potential action mechanisms in cancers. The aberrant expression of *SERPINE1* is common in cancers and is associated with patient prognosis, cancer immunity, immunotherapy response and drug sensitivities. *SERPINE1* may thus be a promising new target for cancer diagnosis and treatment.

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 The Ethics Committee of the Sir Run Run Shaw Hospital, Zhejiang University. 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

FL, LqL, LyL, and MW performed bioinformatic analysis. ZX, HH, and YL performed and analyzed experiments. SY and LG were involved in data analysis and interpretation. FL, LG, and MW designed the experiment, interpreted the data, and wrote the manuscript. All authors contributed to the article and approved the submitted version.

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/fphar.2023.1213891/full#supplementary-material

References

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

Bae, E. A., Seo, H., Kim, I. K., Jeon, I., and Kang, C. Y. (2019). Roles of NKT cells in cancer immunotherapy. *Arch. Pharm. Res.* 42, 543–548. doi:10.1007/s12272-019-01139-8

Bajou, K., Noël, A., Gerard, R. D., Masson, V., Brunner, N., Holst-Hansen, C., et al. (1998). Absence of host plasminogen activator inhibitor 1 prevents cancer invasion and vascularization. *Nat. Med.* 4, 923–928. doi:10.1038/nm0898-923

Baluta, M. M., and Vintila, M. M. (2015). PAI-1 inhibition - another therapeutic option for cardiovascular protection. *Maedica (Bucur)* 10, 147–152.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607. doi:10.1038/nature11003

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17, 218. doi:10.1186/s13059-016-1070-5

Becker, M., Szarvas, T., Wittschier, M., vom Dorp, F., Tötsch, M., Schmid, K. W., et al. (2010). Prognostic impact of plasminogen activator inhibitor type 1 expression in bladder cancer. *Cancer* 116, 4502–4512. doi:10.1002/cncr.25326

Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., et al. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature* 463, 899–905. doi:10.1038/nature08822

Blake, C. M., Sullenger, B. A., Lawrence, D. A., and Fortenberry, Y. M. (2009). Antimetastatic potential of PAI-1-specific RNA aptamers. *Oligonucleotides* 19, 117–128. doi:10.1089/oli.2008.0177

Bonneville, R., Krook, M. A., Kautto, E. A., Miya, J., Wing, M. R., Chen, H. Z., et al. (2017). Landscape of microsatellite instability across 39 cancer types. *JCO Precis. Oncol.* 1, 1–15. doi:10.1200/PO.17.00073

Brown, N. J., Abbas, A., Byrne, D., Schoenhard, J. A., and Vaughan, D. E. (2002a). Comparative effects of estrogen and angiotensin-converting enzyme inhibition on plasminogen activator inhibitor-1 in healthy postmenopausal women. *Circulation* 105, 304–309. doi:10.1161/hc0302.102570

Brown, N. J., Agirbasli, M. A., Williams, G. H., Litchfield, W. R., and Vaughan, D. E. (1998). Effect of activation and inhibition of the renin-angiotensin system on plasma PAI-1. *Hypertension* 32, 965–971. doi:10.1161/01.hyp.32.6.965

Brown, N. J., Kumar, S., Painter, C. A., and Vaughan, D. E. (2002b). ACE inhibition versus angiotensin type 1 receptor antagonism: differential effects on PAI-1 over time. *Hypertension* 40, 859–865. doi:10.1161/01.hyp.0000040264.15961.48

Cai, Z., Chen, J., Yu, Z., Li, H., Liu, Z., Deng, D., et al. (2023). BCAT2 shapes a noninflamed tumor microenvironment and induces resistance to anti-PD-1/PD-L1 immunotherapy by negatively regulating proinflammatory chemokines and anticancer immunity. *Adv. Sci. (Weinh)* 10, e2207155. doi:10.1002/advs.202207155

Cancer Genome Atlas Research Network (2013). Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 499, 43–49. doi:10.1038/nature12222

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2, 401–404. doi:10.1158/2159-8290.CD-12-0095

Chen, D. S., and Mellman, I. (2013). Oncology meets immunology: the cancer-immunity cycle. *Immunity* 39, 1–10. doi:10.1016/j.immuni.2013.07.012

Chen, T. Y., Zhou, M., Lin, M. Q., Liang, S. T., Yan, Y., Wang, S. M., et al. (2021a). Research progress on the SERPINE1 protein and chronic inflammatory diseases of the upper respiratory tract: a literature review. *Int. Arch. Allergy Immunol.* 182, 1097–1102. doi:10.1159/000516195

Chen, X., and Song, E. (2019). Turning foes to friends: targeting cancer-associated fibroblasts. *Nat. Rev. Drug Discov.* 18, 99–115. doi:10.1038/s41573-018-0004-1

Chen, Y., McAndrews, K. M., and Kalluri, R. (2021b). Clinical and therapeutic relevance of cancer-associated fibroblasts. *Nat. Rev. Clin. Oncol.* 18, 792–804. doi:10.1038/s41571-021-00546-5

Clark, D. J., Dhanasekaran, S. M., Petralia, F., Pan, J., Song, X., Hu, Y., et al. (2019). Integrated proteogenomic characterization of clear cell renal cell carcinoma. *Cell.* 179, 964–983.e31. doi:10.1016/j.cell.2019.10.007

Damare, J., Brandal, S., and Fortenberry, Y. M. (2014). Inhibition of PAI-1 antiproteolytic activity against tPA by RNA aptamers. *Nucleic Acid. Ther.* 24, 239–249. doi:10.1089/nat.2013.0475

D'Amico, S., Martial, J. A., and Struman, I. (2012). A peptide mimicking the C-terminal part of the reactive center loop induces the transition to the latent form of plasminogen activator inhibitor type-1. *FEBS Lett.* 586, 686–692. doi:10.1016/j.febslet.2012.02.013

de Visser, K. E., and Joyce, J. A. (2023). The evolving tumor microenvironment: from cancer initiation to metastatic outgrowth. *Cancer Cell.* 41, 374–403. doi:10.1016/j.ccell.2023.02.016

Declerck, P. J., and Gils, A. (2013). Three decades of research on plasminogen activator inhibitor-1: a multifaceted serpin. *Semin. Thromb. Hemost.* 39, 356–364. doi:10.1055/s-0033-1334487

Duffy, M. J., McGowan, P. M., Harbeck, N., Thomssen, C., and Schmitt, M. (2014). uPA and PAI-1 as biomarkers in breast cancer: validated for clinical use in level-of-evidence-1 studies. *Breast Cancer Res.* 16, 428. doi:10.1186/s13058-014-0428-4

Dunoyer-Geindre, S., Fish, R. J., and Kruithof, E. K. O. (2011). Regulation of the endothelial plasminogen activator system by fluvastatin. Role of Rho family proteins, actin polymerisation and p38 MAP kinase. *Thromb. Haemost.* 105, 461–472. doi:10.1160/TH10-07-0444

Egelund, R., Einholm, A. P., Pedersen, K. E., Nielsen, R. W., Christensen, A., Deinum, J., et al. (2001). A regulatory hydrophobic area in the flexible joint region of plasminogen activator inhibitor-1, defined with fluorescent activity-neutralizing ligands. Ligand-induced serpin polymerization. *J. Biol. Chem.* 276, 13077–13086. doi:10.1074/jbc.M009024200

Eitzman, D. T., Fay, W. P., Lawrence, D. A., Francis-Chmura, A. M., Shore, J. D., Olson, S. T., et al. (1995). Peptide-mediated inactivation of recombinant and platelet plasminogen activator inhibitor-1 *in vitro*. *J. Clin. Invest.* 95, 2416–2420. doi:10.1172/JCI117937

Erdem, Y., Usalan, C., Haznedaroğlu, I. C., Altun, B., Arici, M., Yasavul, U., et al. (1999). Effects of angiotensin converting enzyme and angiotensin II receptor inhibition on impaired fibrinolysis in systemic hypertension. *Am. J. Hypertens.* 12, 1071–1076. doi:10.1016/s0895-7061(99)00145-4

Ersoy, C., Kiyici, S., Budak, F., Oral, B., Guclu, M., Duran, C., et al. (2008). The effect of metformin treatment on VEGF and PAI-1 levels in obese type 2 diabetic patients. *Diabetes Res. Clin. Pract.* 81, 56–60. doi:10.1016/j.diabres.2008.02.006

Finotello, F., Mayer, C., Plattner, C., Laschober, G., Rieder, D., Hackl, H., et al. (2019). Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* 11, 34. doi:10.1186/s13073-019-0638-6

Fjellström, O., Deinum, J., Sjögren, T., Johansson, C., Geschwindner, S., Nerme, V., et al. (2013). Characterization of a small molecule inhibitor of plasminogen activator inhibitor type 1 that accelerates the transition into the latent conformation. *J. Biol. Chem.* 288, 873–885. doi:10.1074/jbc.M112.371732

Fortenberry, Y. M., Brandal, S. M., Carpentier, G., Hemani, M., and Pathak, A. P. (2016). Intracellular expression of PAI-1 specific aptamers alters breast cancer cell migration, invasion and angiogenesis. *PLoS One* 11, e0164288. doi:10.1371/journal.pone.0164288

Fortenberry, Y. M. (2013). Plasminogen activator inhibitor-1 inhibitors: a patent review (2006-present). *Expert Opin. Ther. Pat.* 23, 801–815. doi:10.1517/13543776.2013.782393

Fridman, W. H., Pagès, F., Sautès-Fridman, C., and Galon, J. (2012). The immune contexture in human tumours: impact on clinical outcome. *Nat. Rev. Cancer* 12, 298–306. doi:10.1038/nrc3245

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

Goldman, M. J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* 38, 675–678. doi:10.1038/s41587-020-0546-8

Gomes-Giacoia, E., Miyake, M., Goodison, S., and Rosser, C. J. (2013). Targeting plasminogen activator inhibitor-1 inhibits angiogenesis and tumor growth in a human cancer xenograft model. *Mol. Cancer Ther.* 12, 2697–2708. doi:10.1158/1535-7163.MCT-13-0500

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849. doi:10.1093/bioinformatics/btw313

Gutierrez, L. S., Schulman, A., Brito-Robinson, T., Noria, F., Ploplis, V. A., and Castellino, F. J. (2000). Tumor development is retarded in mice lacking the gene for urokinase-type plasminogen activator or its inhibitor, plasminogen activator inhibitor-1. *Cancer Res.* 60, 5839–5847.

Gutiérrez-Melo, N., and Baumjohann, D. (2023). T follicular helper cells in cancer. *Trends Cancer* 9, 309–325. doi:10.1016/j.trecan.2022.12.007

Han, Y., Wang, Y., Dong, X., Sun, D., Liu, Z., Yue, J., et al. (2022). TISCH2: expanded datasets and new tools for single-cell transcriptome analyses of the tumor microenvironment. *Nucleic Acids Res.* 51, D1425–D1431. doi:10.1093/nar/gkac959

Hanekom, G. S., Stubbings, H. M., and Kidson, S. H. (2002). The active fraction of plasmatic plasminogen activator inhibitor type 1 as a possible indicator of increased risk for metastatic melanoma. *Cancer Detect Prev.* 26, 50–59. doi:10.1016/s0361-090x(02)00002-8

Hu, J., Chen, J., Ou, Z., Chen, H., Liu, Z., Chen, M., et al. (2022). Neoadjuvant immunotherapy, chemotherapy, and combination therapy in muscle-invasive bladder cancer: a multi-center real-world retrospective study. *Cell. Rep. Med.* 3, 100785. doi:10.1016/j.xcrm.2022.100785

Iwaki, T., Urano, T., and Umemura, K. (2012). PAI-1, progress in understanding the clinical problem and its aetiology. *Br. J. Haematol.* 157, 291–298. doi:10.1111/j.1365-2141.2012.09074.x

Jevrić, M., Matić, I. Z., Krivokuća, A., Đorđić Crnogorac, M., Besu, I., Damjanović, A., et al. (2019). Association of uPA and PAI-1 tumor levels and 4G/5G variants of PAI-1 gene with disease outcome in luminal HER2-negative node-negative breast cancer patients treated with adjuvant endocrine therapy. *BMC Cancer* 19, 71. doi:10.1186/s12885-018-5255-z

Juhan-Vague, I., and Alessi, M. C. (1997). PAI-1, obesity, insulin resistance and risk of cardiovascular events. *Thromb. Haemost.* 78, 656–660. doi:10.1055/s-0038-1657607

Kohler, H. P., and Grant, P. J. (2000). Plasminogen-activator inhibitor type 1 and coronary artery disease. *N. Engl. J. Med.* 342, 1792–1801. doi:10.1056/NEJM200006153422406

Kutz, S. M., Hordines, J., McKeown-Longo, P. J., and Higgins, P. J. (2001). TGF-beta1-induced PAI-1 gene expression requires MEK activity and cell-to-substrate adhesion. *J. Cell. Sci.* 114, 3905–3914. doi:10.1242/jcs.114.21.3905

Kvassman, J. O., Lawrence, D. A., and Shore, J. D. (1995). The acid stabilization of plasminogen activator inhibitor-1 depends on protonation of a single group that affects loop insertion into beta-sheet A. *J. Biol. Chem.* 270, 27942–27947. doi:10.1074/jbc.270.46.27942

Laskowski, T. J., Biederstädt, A., and Rezvani, K. (2022). Natural killer cells in antitumour adoptive cell immunotherapy. *Nat. Rev. Cancer* 22, 557–575. doi:10.1038/s41568-022-00491-0

Laumen, H., Skurk, T., and Hauner, H. (2008). The HMG-CoA reductase inhibitor rosuvastatin inhibits plasminogen activator inhibitor-1 expression and secretion in human adipocytes. *Atherosclerosis* 196, 565–573. doi:10.1016/j.atherosclerosis.2007.06.005

Laumont, C. M., Banville, A. C., Gilardi, M., Hollern, D. P., and Nelson, B. H. (2022). Tumour-infiltrating B cells: immunological mechanisms, clinical impact and therapeutic opportunities. *Nat. Rev. Cancer* 22, 414–430. doi:10.1038/s41568-022-00466-1

Li, H., Chen, J., Li, Z., Chen, M., Ou, Z., Mo, M., et al. (2023a). S100A5 attenuates efficiency of anti-PD-L1/PD-1 immunotherapy by inhibiting CD8+ T cell-mediated anti-cancer immunity in bladder carcinoma. *Adv. Sci. (Weinh)* 2023, e2300110. doi:10.1002/advs.202300110

Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., et al. (2020). TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 48, W509–W514. doi:10.1093/nar/gkaa407

Li, Y. Z., Xie, J., Wang, R. Q., Gao, X. Q., Liu, P. J., and Liu, J. (2023b). KLF2 is a clinical diagnostic and treatment biomarker of breast cancer. *Front. Cell. Dev. Biol.* 11, 1182123. doi:10.3389/fcell.2023.1182123

Lin, Z., Jensen, J. K., Hong, Z., Shi, X., Hu, L., Andreasen, P. A., et al. (2013). Structural insight into inactivation of plasminogen activator inhibitor-1 by a small-molecule antagonist. *Chem. Biol.* 20, 253–261. doi:10.1016/j.chembiol.2013.01.002

Liu, C. J., Hu, F. F., Xie, G. Y., Miao, Y. R., Li, X. W., Zeng, Y., et al. (2023). Gsca: an integrated platform for gene set cancer analysis at genomic, pharmacogenomic and immunogenomic levels. *Brief. Bioinform* 24, bbac558. doi:10.1093/bib/bbac558

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

Luna, A., Elloumi, F., Varma, S., Wang, Y., Rajapakse, V. N., Aladjem, M. I., et al. (2021). CellMiner Cross-Database (CellMinerCDB) version 1.2: exploration of patient-derived cancer cell line pharmacogenomics. *Nucleic Acids Res.* 49, D1083–D1093. doi:10.1093/nar/gkaa968

Ma, Y., Ryu, J. S., Dulay, A., Segal, M., and Guller, S. (2002). Regulation of plasminogen activator inhibitor (PAI)-1 expression in a human trophoblast cell line by glucocorticoid (GC) and transforming growth factor (TGF)-beta. *Placenta* 23, 727–734. doi:10.1016/s0143-4004(02)90863-5

Mashiko, S., Kitatani, K., Toyoshima, M., Ichimura, A., Dan, T., Usui, T., et al. (2015). Inhibition of plasminogen activator inhibitor-1 is a potential therapeutic strategy in ovarian cancer. *Cancer Biol. Ther.* 16, 253–260. doi:10.1080/15384047.2014.1001271

Masuda, T., Hattori, N., Senoo, T., Akita, S., Ishikawa, N., Fujitaka, K., et al. (2013). SK-216, an inhibitor of plasminogen activator inhibitor-1, limits tumor progression and angiogenesis. *Mol. Cancer Ther.* 12, 2378–2388. doi:10.1158/1535-7163.MCT-13-0041

Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28, 1747–1756. doi:10.1101/gr.239244.118

Morgan, A. E., Davies, T. J., and Mc Auley, M. T. (2018). The role of DNA methylation in ageing and cancer. *Proc. Nutr. Soc.* 77, 412–422. doi:10.1017/S0029665118000150

Nakatsuka, E., Sawada, K., Nakamura, K., Yoshimura, A., Kinose, Y., Kodama, M., et al. (2017). Plasminogen activator inhibitor-1 is an independent prognostic factor of ovarian cancer and IMD-4482, a novel plasminogen activator inhibitor-1 inhibitor, inhibits ovarian cancer peritoneal dissemination. *Oncotarget* 8, 89887–89902. doi:10.18632/oncotarget.20834

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457. doi:10.1038/nmeth.3337

Ni, X. Q., Zhu, J. H., Yao, N. H., Qian, J., and Yang, X. J. (2013). Statins suppress glucose-induced plasminogen activator inhibitor-1 expression by regulating RhoA and nuclear factor-κB activities in cardiac microvascular endothelial cells. *Exp. Biol. Med. (Maywood)* 238, 37–46. doi:10.1258/ebm.2012.012127

Nordt, T. K., Bode, C., and Sobel, B. E. (2001). Stimulation *in vivo* of expression of intra-abdominal adipose tissue plasminogen activator inhibitor Type I by proinsulin. *Diabetologia* 44, 1121–1124. doi:10.1007/s001250100618

Placencio, V. R., and DeClerck, Y. A. (2015). Plasminogen activator inhibitor-1 in cancer: rationale and insight for future therapeutic testing. *Cancer Res.* 75, 2969–2974. doi:10.1158/0008-5472.CAN-15-0876

Placencio, V. R., Ichimura, A., Miyata, T., and DeClerck, Y. A. (2015). Small molecule inhibitors of plasminogen activator inhibitor-1 elicit anti-tumorigenic and anti-angiogenic activity. *PLoS One* 10, e0133786. doi:10.1371/journal.pone.0133786

Pös, O., Radvanszky, J., Buglyó, G., Pös, Z., Rusnakova, D., Nagy, B., et al. (2021). DNA copy number variation: main characteristics, evolutionary significance, and pathological aspects. *Biomed. J.* 44, 548–559. doi:10.1016/j.bj.2021.02.003

Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E., and Gfeller, D. (2017). Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife* 6, e26476. doi:10.7554/eLife.26476

Sakamoto, T., Woodcock-Mitchell, J., Marutsuka, K., Mitchell, J. J., Sobel, B. E., and Fujii, S. (1999). TNF-alpha and insulin, alone and synergistically, induce plasminogen activator inhibitor-1 expression in adipocytes. *Am. J. Physiol.* 276, C1391–C1397. doi:10.1152/ajpcell.1999.276.6.C1391

Sica, A., and Mantovani, A. (2012). Macrophage plasticity and polarization: *in vivo* veritas. *J. Clin. Invest.* 122, 787–795. doi:10.1172/JCI59643

Siegel, R. L., Miller, K. D., Wagle, N. S., and Jemal, A. (2023). Cancer statistics, 2023. *CA Cancer J. Clin.* 73, 17–48. doi:10.3322/caac.21763

Sillen, M., and Declerck, P. J. (2021). A narrative review on plasminogen activator inhibitor-1 and its (Patho)Physiological role: to target or not to target? *Int. J. Mol. Sci.* 22, 2721. doi:10.3390/ijms22052721

Sillen, M., and Declerck, P. J. (2020). Targeting PAI-1 in cardiovascular disease: structural insights into PAI-1 functionality and inhibition. *Front. Cardiovasc Med.* 7, 622473. doi:10.3389/fcvm.2020.622473

Sillen, M., Miyata, T., Vaughan, D. E., Strelkov, S. V., and Declerck, P. J. (2021). Structural insight into the two-step mechanism of PAI-1 inhibition by small molecule TM5484. *Int. J. Mol. Sci.* 22, 1482. doi:10.3390/ijms22031482

Sotiropoulos, G., Kotopouli, M., Karampela, I., Christodoulatos, G. S., Antonakos, G., Marinou, I., et al. (2019). Circulating plasminogen activator inhibitor-1 activity: a biomarker for resectable non-small cell lung cancer? *J. BUON* 24, 943–954.

St Paul, M., and Ohashi, P. S. (2020). The roles of CD8+ T cell subsets in antitumor immunity. *Trends Cell. Biol.* 30, 695–704. doi:10.1016/j.tcb.2020.06.003

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Takayama, Y., Hattori, N., Hamada, H., Masuda, T., Omori, K., Akita, S., et al. (2016). Inhibition of PAI-1 limits tumor angiogenesis regardless of angiogenic stimuli in malignant pleural mesothelioma. *Cancer Res.* 76, 3285–3294. doi:10.1158/0008-5472.CAN-15-1796

Tan, P., Chen, H., Huang, Z., Huang, M., Du, Y., Li, T., et al. (2021). MMP25-AS1/hsa-miR-10a-5p/SERPINE1 axis as a novel prognostic biomarker associated with immune cell infiltration in KIRC. *Mol. Ther. Oncolytics* 22, 307–325. doi:10.1016/j.omto.2021.07.008

Teng, F., Zhang, J., Chen, Y., Shen, X., Su, C., Guo, Y., et al. (2021). LncRNA NKX2-1-AS1 promotes tumor progression and angiogenesis via upregulation of SERPINE1 expression and activation of the VEGFR-2 signaling pathway in gastric cancer. *Mol. Oncol.* 15, 1234–1255. doi:10.1002/1878-0261.12911

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The immune landscape of cancer. *Immunity* 48, 812–830. doi:10.1016/j.immuni.2018.03.023

Togashi, Y., Shitara, K., and Nishikawa, H. (2019). Regulatory T cells in cancer immunosuppression - implications for anticancer therapy. *Nat. Rev. Clin. Oncol.* 16, 356–371. doi:10.1038/s41571-019-0175-7

Xu, L., Deng, C., Pang, B., Zhang, X., Liu, W., Liao, G., et al. (2018). Tip: a web server for resolving tumor Immunophenotype profiling. *Cancer Res.* 78, 6575–6580. doi:10.1158/0008-5472.CAN-18-0689

Xue, Y., Björquist, P., Inghardt, T., Linschoten, M., Musil, D., Sjölin, L., et al. (1998). Interfering with the inhibitory mechanism of serpins: crystal structure of a complex formed between cleaved plasminogen activator inhibitor type 1 and a reactive-centre loop peptide. *Structure* 6, 627–636. doi:10.1016/s0969-2126(98)00064-1

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

Yuan, H., Yan, M., Zhang, G., Liu, W., Deng, C., Liao, G., et al. (2019). CancerSEA: a cancer single-cell state atlas. *Nucleic Acids Res.* 47, D900–D908. doi:10.1093/nar/gky939

Zeng, D., Ye, Z., Shen, R., Yu, G., Wu, J., Xiong, Y., et al. (2021). Iobr: multi-omics immuno-oncology biological research to decode tumor microenvironment and Signatures. *Front. Immunol.* 12, 687975. doi:10.3389/fimmu.2021.687975

Zeng, Z., Wong, C. J., Yang, L., Ouardaoui, N., Li, D., Zhang, W., et al. (2022). Tismo: syngeneic mouse tumor database to model tumor immunity and immunotherapy response. *Nucleic Acids Res.* 50, D1391–D1397. doi:10.1093/nar/gkab804

Zhao, C., and Liu, Z. (2021). MicroRNA 617 targeting SERPINE1 inhibited the progression of oral squamous cell carcinoma. *Mol. Cell. Biol.* 41, e0056520–20. doi:10.1128/MCB.00565-20

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 30 August 2023
DOI 10.3389/fphar.2023.1260697

OPEN ACCESS

EDITED BY
Hongzhou Cai,
Nanjing Medical University, China

REVIEWED BY
Dechao Feng,
Sichuan University, China
Ying Wang,
Sun Yat-Sen University Cancer Center (SYSUCC), China
Tao Zhennan,
Nanjing Drum Tower Hospital, China

*CORRESPONDENCE
Hao Chen,
chenhao6938@163.com,
qiluchenhao@email.sdu.edu.cn
Ming Lu,
lvming@sdu.edu.cn

RECEIVED 18 July 2023
ACCEPTED 14 August 2023
PUBLISHED 30 August 2023

CITATION
Wang Z, Ma C, Teng Q, Man J, Zhang X, Liu X, Zhang T, Chong W, Chen H and Lu M (2023), Identification of a ferroptosis-related gene signature predicting recurrence in stage II/III colorectal cancer based on machine learning algorithms.
Front. Pharmacol. 14:1260697.
doi: 10.3389/fphar.2023.1260697

Identification of a ferroptosis-related gene signature predicting recurrence in stage II/III colorectal cancer based on machine learning algorithms

Ze Wang[1,2], Chenghao Ma[3], Qiong Teng[3], Jinyu Man[1,2], Xuening Zhang[1,2], Xinjie Liu[1,2], Tongchao Zhang[2,4], Wei Chong[5], Hao Chen[2,4]* and Ming Lu[1,2,4]*

[1]Department of Epidemiology and Health Statistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, China, [2]Clinical Epidemiology Unit, Qilu Hospital of Shandong University, Jinan, China, [3]Department of Gastroenterological Surgery, Shandong Provincial Hospital, Shandong University, Jinan, Shandong, China, [4]Clinical Research Center of Shandong University, Jinan, China, [5]Department of Gastrointestinal Surgery, Key Laboratory of Engineering of Shandong Province, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Medical Science and Technology Innovation Center, Shandong First Medical University & Shandong Academy of Medical Sciences, Jinan, Shandong, China

Background: Colorectal cancer (CRC) is one of the most prevalent cancer types globally. A survival paradox exists due to the inherent heterogeneity in stage II/III CRC tumor biology. Ferroptosis is closely related to the progression of tumors, and ferroptosis-related genes can be used as a novel biomarker in predicting cancer prognosis.

Methods: Ferroptosis-related genes were retrieved from the FerrDb and KEGG databases. A total of 1,397 samples were enrolled in our study from nine independent datasets, four of which were integrated as the training dataset to train and construct the model, and validated in the remaining datasets. We developed a machine learning framework with 83 combinations of 10 algorithms based on 10-fold cross-validation (CV) or bootstrap resampling algorithm to identify the most robust and stable model. C-indice and ROC analysis were performed to gauge its predictive accuracy and discrimination capabilities. Survival analysis was conducted followed by univariate and multivariate Cox regression analyses to evaluate the performance of identified signature.

Results: The ferroptosis-related gene (FRG) signature was identified by the combination of Lasso and plsRcox and composed of 23 genes. The FRG signature presented better performance than common clinicopathological features (e.g., age and stage), molecular characteristics (e.g., BRAF mutation and microsatellite instability) and several published signatures in predicting the prognosis of the CRC. The signature was further stratified into a high-risk group and low-risk subgroup, where a high FRG signature indicated poor prognosis among all collected datasets. Sensitivity analysis showed the FRG signature remained a significant prognostic factor. Finally, we have developed a nomogram and a decision tree to enhance prognosis evaluation.

Conclusion: The FRG signature enabled the accurate selection of high-risk stage II/III CRC population and helped optimize precision treatment to improve their clinical outcomes.

KEYWORDS

stage II/III colorectal cancer, machine learning, prognosis, ferroptosis-related gene, tumor heterogeneity

Introduction

Colorectal cancer (CRC) is a common and deadly disease with 147,950 new cases estimated in 2020 (Siegel et al., 2020; Mi et al., 2023). Early detection through regular screening, effective treatment options such as adjuvant chemotherapy and targeted therapies, and promoting healthy lifestyle choices can all help reduce the CRC recurrence risk and improve survival rates (Dekker et al., 2019). The pathological staging at the time of diagnosis is a crucial determinant of both the recurrence risk and survival (Jeffery et al., 2019). Meanwhile, stage II/III CRC represents a significant proportion, accounting for about 70% of all CRC cases (Liu et al., 2022). Unfortunately, even with curative resection, 30–40% of the patients will experience recurrence, which can significantly impact their survival rates (Jeffery et al., 2019). More importantly, a survival paradox exists for patients with stage IIB/IIC and IIIA CRC, which cannot be well explained by traditional clinicopathological features or molecular signatures (Kim et al., 2015; Kim et al., 2019). Meanwhile, evidence in previous studies showed that patients who routinely received adjuvant chemotherapy after surgery did not respond equally even with the same stage (Shiovitz et al., 2014). Therefore, there is still a need to establish a novel recurrence-related prognostic model to identify the high-risk population of stage II/III CRC for clinical decision-making.

Recent studies have brought light to various molecular features in CRC that have been highly correlated with the prognosis and therapy response. Notably, these features include CMS classification, genomic alterations such as TP53, KRAS, and BRAF mutation, microsatellite instability (MSI), and tumor mutational burden (TMB), which have been recognized as relatively reliable biomarkers (Dienstmann et al., 2017; Dienstmann et al., 2019; Chong et al., 2022). The immunohistochemistry technique, particularly the multiplex immunohistochemistry or immunofluorescence (mIHC/IF) method, is frequently used to aid in pathology diagnosis as it reduces inter-observer variability and has the ability to label multiple markers per tissue section. However, it is important to point out that one potential disadvantage of mIHC is that the number of markers that can be simultaneously labeled is typically limited, usually between 3 and 7, which may not capture the full complexity of the biomolecular interactions underlying the disease pathology (Tan et al., 2020). It is worth mentioning that the CMS classification, which employs bulk RNA-seq data to stratify CRC patients into four subtypes, has emerged as a highly effective tool for identifying strong prognostic effects for both recurrence and survival, warranting further attention and analysis (Guinney et al., 2015; Stintzing et al., 2019).

Several molecular models have been developed to predict the recurrence and survival of stage II/III CRC, including lncRNA, hallmark-based, immune-based, methylation-based, and epithelial–mesenchymal transition (EMT)-related signatures, among others (Li et al., 2020a; Chong et al., 2021; Liu et al., 2022; Ren et al., 2022; Li et al., 2023). Ferroptosis plays a critical role in the development of CRC through several mechanisms, such as the build-up of lipid peroxides, disruption of the balance between glutathione and glutathione peroxidase 4, and disturbances in iron homeostasis (Song et al., 2023). Accumulating evidence has shown that the induction of ferroptosis in CRC successfully eliminates cancer cells resistant to other modes of cell death (Wang et al., 2022a). Several studies also established ferroptosis-related gene (FRG) or lncRNA signatures to predict stage II/III CRCs recurrence or prognosis (Wu et al., 2021; Yu et al., 2021; Du et al., 2022). However, the performance of these molecular models in prediction was different, and several studies did not emphasize sufficient validation with multiple datasets and attempt at the multiple modeling algorithm. The modeling algorithm combination should be further fine-tuned, and the validation procedure for the signature should be intensified to improve the credibility of the model.

Accordingly, the aim of the present study was to construct an mRNA expression signature using FRGs to identify patients at risk of relapse via a 10-fold and bootstrap machine learning framework. The constructed signature was also validated in five independent datasets and compared with clinical traits and molecular features and published signatures. Sensitivity analysis was performed to test the performance of the signature. A nomogram and decision tree, which integrated clinical and molecular features with the signature, were established to improve clinical outcomes and guide clinical practice.

Materials and methods

Data collection and preprocessing

The overall workflow of the study is shown in Figure 1. Gene expression data and corresponding clinical features of stage II/III CRC samples were collected from publicly available datasets of the NCBI GEO (https://www.ncbi.nlm.nih.gov/geo/), cBioPortal (https://www.cbioportal.org/), and TCGA (https://cancergenome.nih.gov/) database. Seven microarray datasets from GEO (GSE14333, GSE37892, GSE39582, GSE103479, GSE29621, GSE92921, and GSE12945) were sequenced by using Affymetrix HG-U133 Plus 2.0 Array, and TCGA (the combination of TCGA-COAD/READ) and MSK-READ datasets were produced from Illumina high-throughput sequencing platform. A total of 1,397 samples enrolled in the establishment and validation of the model met the following criteria: 1) primary tumor in colorectum; 2) with clinical information and gene expression data; and 3) stage II/III in the AJCC staging system.

Four cohorts (GSE14333, GSE37892, GSE39582, and TCGA-CRC, 1,000 samples) of the total patients were combined as the training dataset. The other five cohorts were used as independent validation datasets. The combined gene expression data in the training dataset was collected from the study of Guinney et al. (2015). Careful data preprocessing before the merge was implemented to address the batch effect among datasets due to the different platforms, labs, and time points. The raw data of the four GEO validation datasets were processed by the robust multiarray averaging (RMA) algorithm with 'affy' package, and the duplicate probes were merged via the median number. TCGA RNA FPKM format sequencing data were curated from UCSC Xena (https://xenabrowser.net/datapages/). TCGA-COAD/READ data were merged after removing the batch effect by Combat algorithm via "sva" package. Then, they were converted into

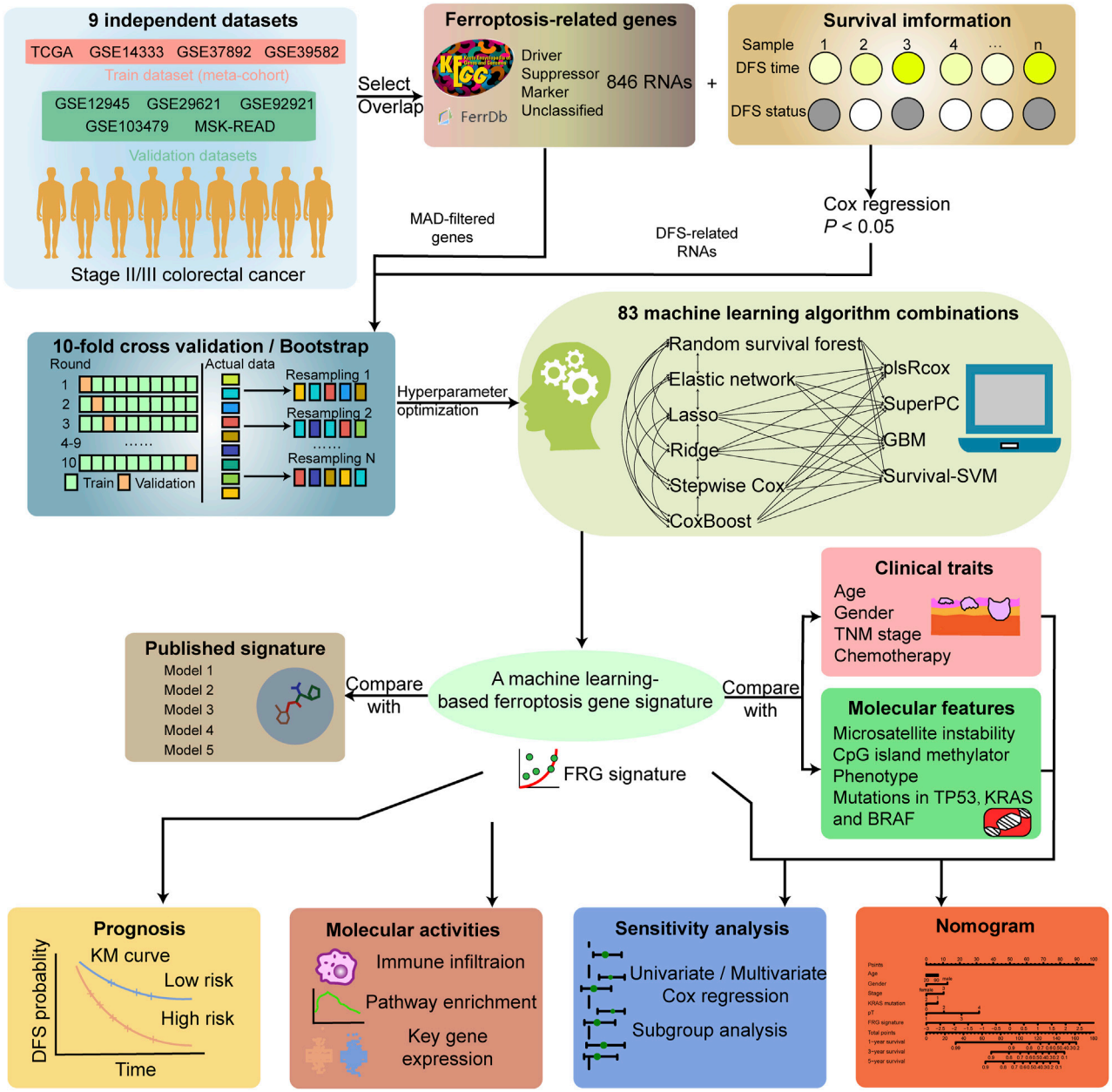


FIGURE 1
Workflow of the study.

TPM format and further log-2 transformed. The log-2 transformed RNA-seq data from the MSK dataset were collected from cBioPortal. All gene expressions were transformed into Z-score among samples when training the model.

Acquisition and screening of ferroptosis-related genes

We obtained FRGs from the public databases FerrDb (http://www.zhounan.org/ferrdb/current/) and KEGG. A total of 846 RNAs were obtained, including those from the driver, suppressor, marker, and unclassified categories (Supplementary Table S1). After removing duplicates, a total of 546 RNAs remained. Since low-expressed or non-varying genes usually represent noise, the transcriptome data in the training dataset were downloaded with nearly 6,000 genes by the largest median absolute deviation (MAD). These genes were measured by at least one probeset in all datasets, and each gene was represented by the probeset with the largest MAD (Guinney et al., 2015). We took the intersection of the curated FRGs and MAD-filtered genes as the variable features in the training meta-cohort and performed univariate Cox regression analysis to screen out disease-free survival (DFS)-associated genes. A total of 80 representative recurrence-related ferroptosis genes were enrolled in the machine learning framework as initial variables (Supplementary Table S2).

Construction of a prognostic gene signature

A total of 10 separate machine learning algorithms and their combinations composed the machine learning framework. The 10 algorithms included random survival forest (RSF), elastic network (ENet), least absolute shrinkage and selection operator (Lasso), ridge, stepwise Cox proportional hazards regression (Stepwise Cox), CoxBoost (Boosting Cox's proportional hazards regression), generalized boosted regression modeling (GBM), supervised principal components (SuperPC), partial least squares regression for Cox (plsRcox), and survival support vector machine (survival-SVM). Six algorithms, including Lasso, RSF, CoxBoost, ridge, ENet, and stepwise Cox, could be used to perform feature selection. Within this framework, we utilize six specific algorithms for feature selection: Lasso, RSF, CoxBoost, ridge, ENet, and stepwise Cox. These algorithms play a fundamental role in the preliminary phase of gene screening, working synergistically with other algorithms. Both individual algorithms and combinations of two algorithms contribute to this comprehensive framework. To optimize the performance of each algorithm, we employed either 10-fold cross-validation (CV) or bootstrap resampling techniques. These approaches assist us in evaluating and honing the models. Finally, a total of combined 83 algorithms based on 10-fold CV or bootstrap resampling were utilized to select the optimistic performance model. The implementation of machine learning algorithms framework and corresponding hyperparameter optimization function are shown in Supplementary Table S3. The Harrell's concordance index (C-index) was calculated among all datasets. The algorithm with the highest average C-index across all validation datasets was regarded as the optimal model to generate the FRG signature. The samples were categorized into high and low risk based on the optimal thresholds for signature scores determined by the surv_cutpoint function of the R package "survminer".

Collection of published signatures

To further assess our identified model's performance, we curated five previously published mRNA signatures (Supplementary Table S4). These signatures were constructed by using diverse computational algorithms and curated from various biological processes, including hypoxia and the tumor microenvironment. To evaluate the performance of each signature, we employed univariate Cox regression analysis and computed the C-index across all cohorts.

Construction and validation of the nomogram

A novel nomogram for predicting relapse in patients with stage II/III CRC was established by the "rms" package. We integrated common clinical and molecular features in the nomogram with our signature to compose a comprehensive model applying Cox proportional hazards regression. The samples in the training meta-cohort, which had these relevant variables (632 samples), were enrolled in the analysis. The calibration curve was used to visualize the relationship between the predicted probability generated by the nomogram and the actual observations. The decision curve analysis (DCA) results could be performed to obtain the clinical net benefit of different models, and all and none strategies (Van Calster et al., 2018). Finally, we employed recursive partitioning analysis with the R package "rpart" to construct a decision tree model for DFS, aiming to refine risk stratification.

Additional bioinformatic analysis

The deconvolution approach xCell algorithm was selected to exhibit molecular features regarding immunology between risk groups by 'xCell' packages (Newman et al., 2015; Aran et al., 2017). xCell could utilize bulk RNA-seq data to infer infiltrating immune and stromal cell subsets. The correlation coefficients between the signature scores and each gene expression acquired were calculated. The sorted correlation coefficients were used as the ranked gene list input to perform gene set enrichment analysis (GSEA) via the "clusterProfiler" package against KEGG and REACTOME reference gene set (Subramanian et al., 2005).

Statistical analysis

The data processing, statistical analysis, and plotting were generated in the R 4.2.2 software. The heatmap of genes enrolled in the signature with clinical annotations was generated using the R package "ComplexHeatmap." Correlations between two continuous variables were evaluated via Spearman correlation coefficients. The Wilcoxon rank-sum test or *t*-test was applied to compare the difference between two groups for quantitative data. Two-sided Fisher exact tests were used to analyze categorical variables. The Cox proportional hazards model and Kaplan–Meier analysis were performed with the 'survival' or 'rms' package. Receiver operating characteristic curves (ROCs) were used to evaluate the prognostic classification performance of the signature with the 'timeROC' package. The C-index comparisons between clinical and molecular traits and the risk score were implemented by the "compareC" package. All statistical tests were two-sided. $p < 0.05$ was considered as statistically significant. The length of error bars represented 95% confidence intervals. The Benjamini–Hochberg method was applied to control the false discovery rate (FDR) for multiple hypothesis testing in appropriate conditions.

Results

Development of machine learning-based ferroptosis-related gene signature

The meta-cohort of the four datasets (GSE14333, GSE37892, GSE39582, and TCGA-CRC) was regarded as the training dataset, and the principal component analysis showed no

10.3389/fphar.2023.1260697

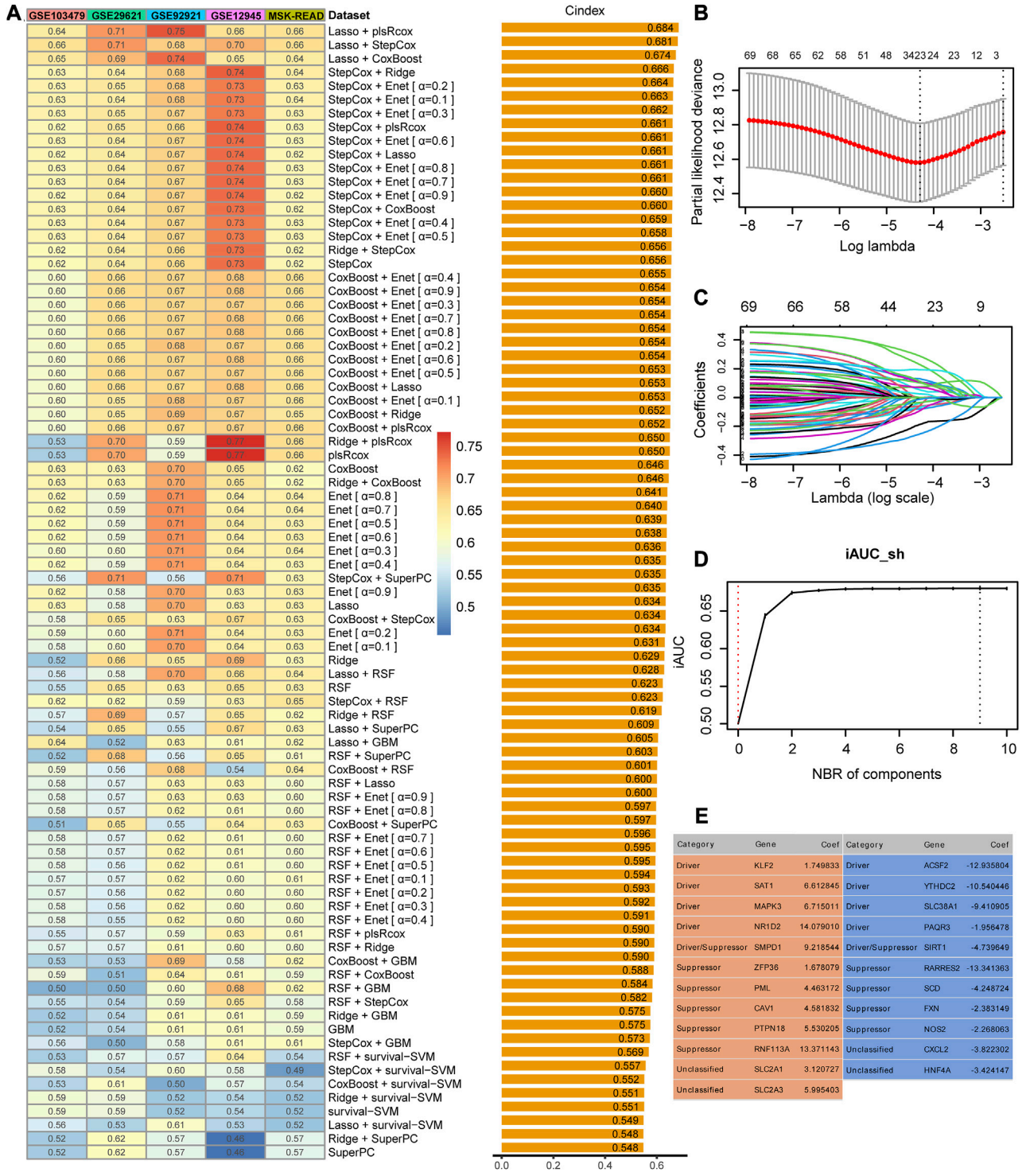


FIGURE 2
Identification and construction of the best performance signature. **(A)** C-indices of 83 combinations of machine learning prediction models in five validation cohorts. **(B)** Determination of the optimal lambda was obtained when the partial likelihood deviance reached the minimum value and further generated the gene features with non-zero coefficients. **(C)** Lasso coefficient profiles of the candidate genes for FRG signature construction. **(D)** Determination of the optimal number of components when the iAUC reached the maximum value. **(E)** Categories and coefficients of 23 genes finally obtained in plsRcox regression.

significant batch effects within the meta dataset (Supplementary Figure S1A). The expression profiles of the 80 ferroptosis-related prognostic genes were subjected to the machine learning-based modeling framework. We fitted 83 kinds of prediction models via 10 machine learning algorithms based on the 10-fold cross-validation or bootstrap resampling algorithm to optimize the

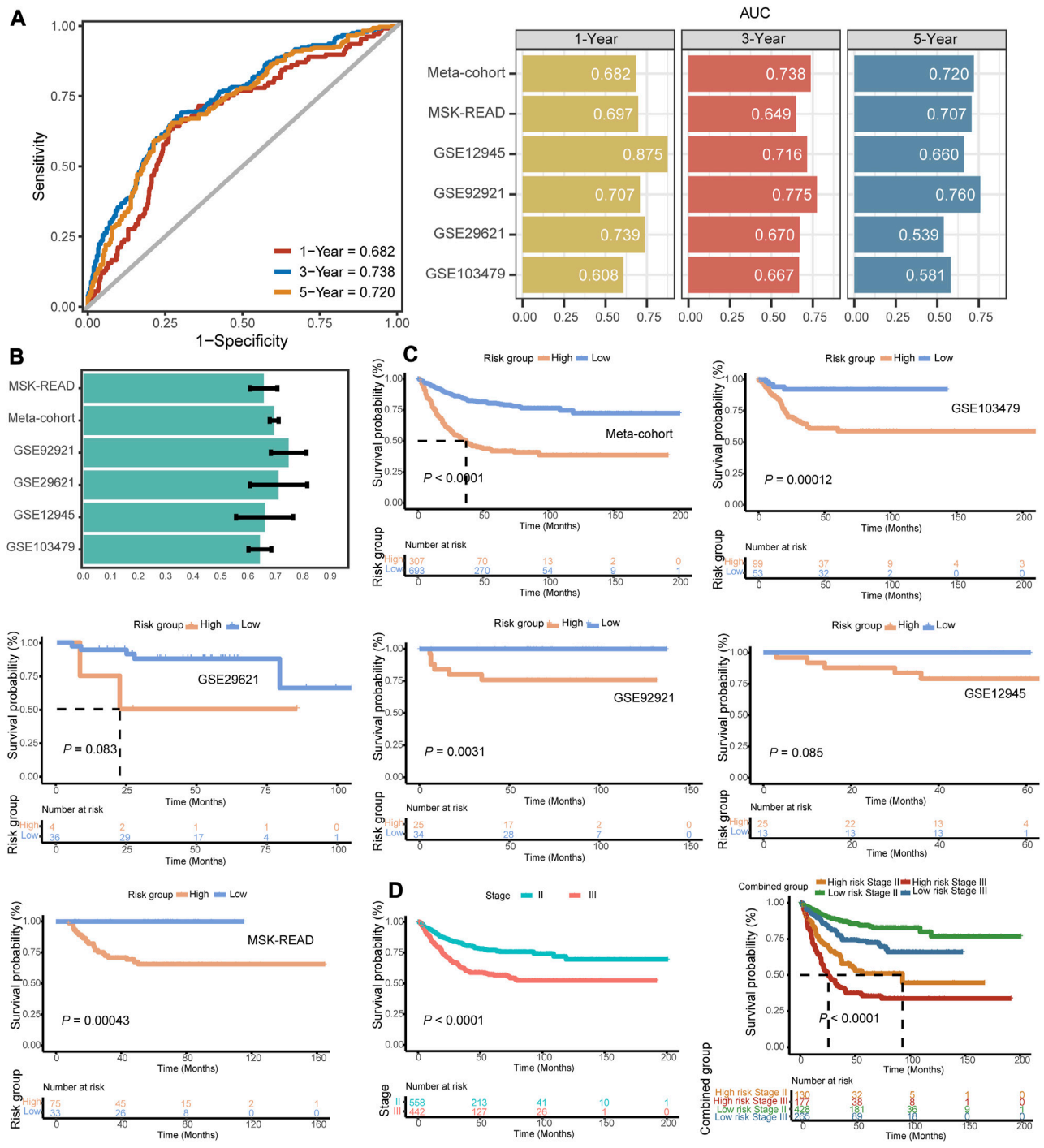


FIGURE 3
Evaluation indicators and prognostic value of the FRG signature. **(A)** Time-dependent ROC analysis for predicting DFS at 1-, 3-, and 5-year across the training meta-cohort and all validation datasets. **(B)** C-indices of the signature across all datasets. **(C)** Kaplan–Meier survival curve of DFS between patients with a high-signature score and with a low-signature score in the training meta-cohort and five validation datasets. **(D)** Kaplan–Meier survival curve of DFS between patients in stage II vs. stage III patients and with respect to the stage and the identified gene signature of the meta-cohort.

model parameter in the training meta-cohort. The C-indices were calculated exclusively in the five validation cohorts of all models, and the model that exhibited the highest average C-index was deemed the optimal solution. The most robust model with the highest mean C-index in the five validation datasets was the combination of Lasso and plsRcox (Figure 2A). Using Lasso regression and 10-fold CV, we found that 23 FRGs had non-zero Lasso coefficients and were associated with recurrence in stage II/III CRC. The regression partial likelihood deviance reached its minimum value, indicating that these FRGs are important predictive features for recurrence (Figures 2B, C). The chosen features underwent a 10-fold cross-validation

plsRcox to build a predictive model with the optimal number of components. The incremental area under the curve (iAUC) value reached its maximum at nine components (Figure 2D), so the model used components 0–9 to obtain fit statistics using both the Akaike information criterion (AIC) and Bayesian information criterion (BIC) (Supplementary Figure S1B). Finally, the risk score for each patient was calculated using the expression of 23 gene features multiplied by their corresponding coefficients to generate the FRG signature (Figure 2E).

The predictive and prognostic value of FRG signature

We used ROC analysis to measure the DFS discrimination of the signature, with 1-, 3-, and 5-year AUCs of 0.682, 0.738, and 0.720 in the training meta-cohort; 0.697, 0.649, and 0.707 in GSE103479; 0.875, 0.716, and 0.660 in GSE12945; and 0.707, 0.775, and 0.760 in GSE92921, respectively (Figures 3A, B). The 3-year AUC of GSE103479 and the 1-year AUC of GSE29621 were 0.670 and 0.739, respectively. The model had an overall certain decent degree of 3-year AUCs across all independent datasets. Both the training meta-cohort and validation datasets also showed stable C-indices around 0.7 (Figure 3C).

In our independent datasets, we divided the samples into high- and low-risk groups using the optimal cutoff for the predicted signature risk score. By analyzing the expression patterns of the 23 identified genes, we observed clear differences between the high- and low-risk groups (Supplementary Figure S1C). Specifically, genes such as ZFP36, KLF2, PML, PTPN18, MAPK3, SMPD1, SLC2A3, RARRES2, CAV1, and SAT1 were found to be highly expressed in the high-risk group. On the other hand, other genes showed predominant expression in the low-risk group. These findings suggest that these genes may play a significant role in distinguishing different risk groups in stage II/III CRC. Patients in the high-risk group had significantly shorter DFS compared to the low-risk group in the training meta-cohort ($p < 0.0001$), and similar trends were also observed in the validation datasets such as GSE103479 ($p = 0.00012$), GSE92921 ($p = 0.0031$), and MSK-READ ($p = 0.00043$) (Figure 3D). The results of Kaplan–Meier survival analysis in two cohorts, GSE29621 and GSE12945, reached a marginal statistical significance ($p = 0.083$ and 0.085, respectively), considering the smaller sample size. Meanwhile, the discriminatory power of the FRG signature scores was similar to the hazard obtained through pathological staging. In both stage II and stage III subgroups, patients with high signature scores had significantly shorter DFS ($p < 0.0001$) (Figure 3E). In addition, we integrated patients into a pooled cohort containing the training and validation cohorts to revalidate the prognostic value. The pooled cohort still showed a significant difference in DFS between the high- and low-risk groups ($p < 0.0001$), and this difference was also observed within both stage II and stage III subgroups ($p < 0.0001$) (Supplementary Figures S2A–C). Totally, the FRG-based model provided the promising potential in predicting the recurrence risk of stage II/III CRC.

The comparisons with other features and collected signatures

Apart from selecting the most suitable model combination, we also compared the C-index of the signature with clinical characteristics and other molecular features in all the datasets included in our study. The clinical characteristics included demographic information, such as age, race, and gender, as well as tumor histology data such as AJCC stage, pT, pN, grade, tumor size, and chemotherapy response. Molecular features were specific to several mutational alterations (KRAS mutations, BRAF mutations, and microsatellite state) and molecular classification (such as CMS classification). These characteristics are commonly used in clinical evaluation of patients (Ng et al., 2019; Timar and Kashofer, 2020; Battistuzzi et al., 2021; Chen et al., 2021). The FRG signature score basically had the highest C-index compared to these clinical and molecular features in the training and validation cohorts, demonstrating the survival prediction capability of our signature (Figure 4A). Furthermore, our FRG signature achieved the highest C-index among five published molecular signatures in four datasets (Figure 4B). The multifaceted evaluation demonstrated that the FRG signature performed well in identifying stage II/III CRC patients with distinct clinical outcomes.

Sensitivity analysis

Table 1 summarized the distribution of demographics, tumor-related clinical characteristics, and molecular features of the four cohorts in the training meta-cohort, which included 1,000 patients identified as the high-risk group (N = 307) or low-risk group (N = 693) with the FRG signature. We recognized that stage, pN status, KRAS mutation, and CMS classification were significantly differentially distributed between high- and low-risk groups. To further inspect the robustness of the model, clinicopathological and molecular features, together with the identified signature, were assessed in univariate and multivariate models (Figures 5A, B). Importantly, only pT stage, KRAS mutation, and signature scores were independent recurrence prognostic factors in multivariable models. Gender, AJCC stage, and pN stage did not significantly improve prognosis prediction over pT stage and KRAS mutation when the FRG signature was considered. Sensitivity analysis showed that the FRG signature was still robust within the subgroups of clinicopathological and molecular annotation variables of interest, including pT4 and KRAS mutation (Figure 5C).

Potential molecular processes associated with FRG signature

To explore possible underlying molecular mechanisms for the FRG signature, we utilized the xCell method to analyze the immune infiltration landscape with our signature. A total of 14 immune infiltrating cells were correlated with the FRG signature

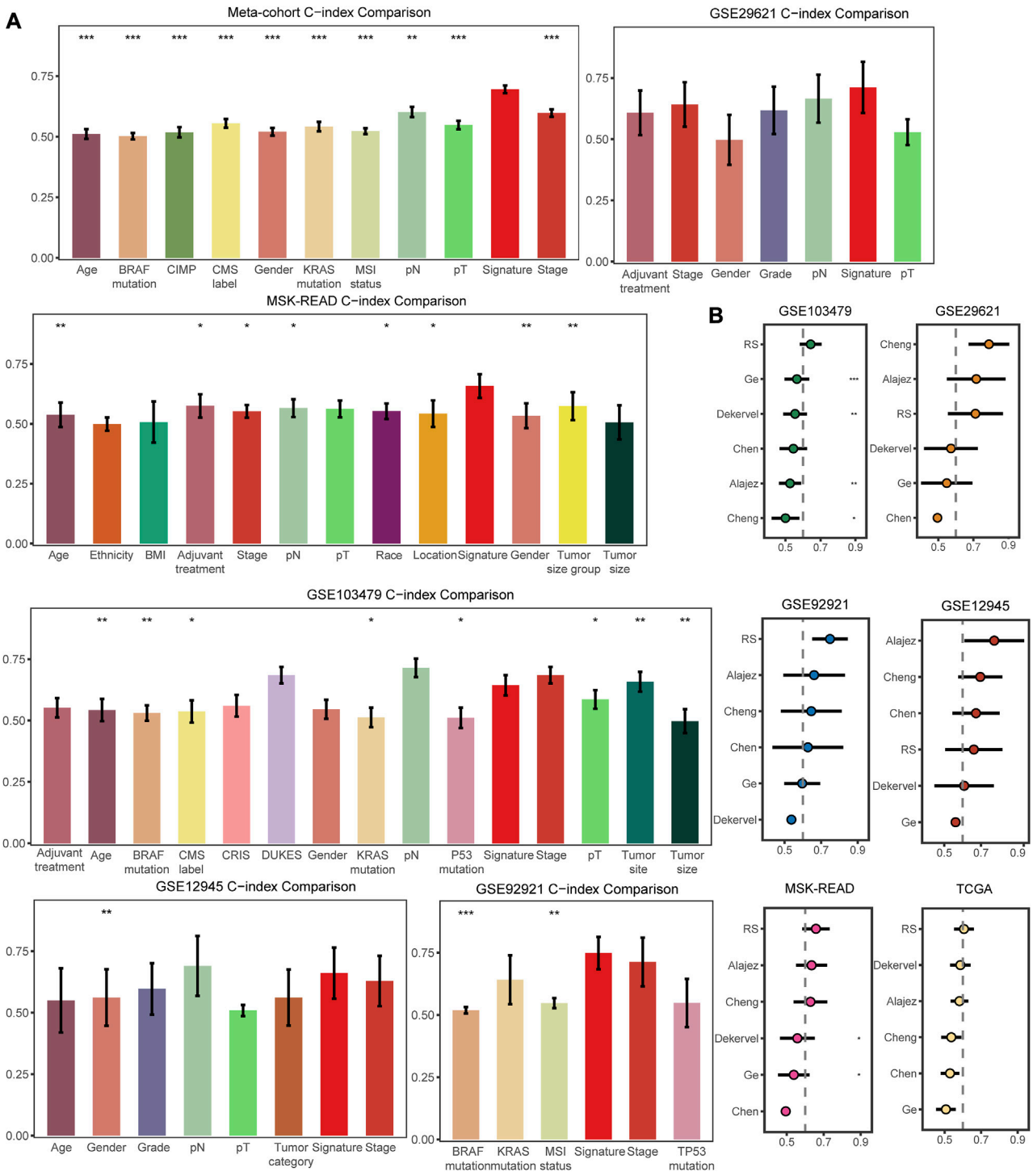


FIGURE 4
Comparisons of clinical and molecular characteristics, and published signatures with the FRG signature. **(A)** C-index comparisons between clinical and molecular variables and signature in the training meta-cohort and validation datasets. **(B)** C-index comparisons between signature and five published signatures. * means $p < 0.05$, ** means $p < 0.01$, and *** means $p < 0.001$.

(Supplementary Figure S3A). Representative cells, T cells and B cells, scored significantly higher in the low-risk group than in the high-risk group, while the levels of endothelial cells and epithelial cells were significantly higher in the high-risk group. To gain the comprehensive biological mechanisms of the FRG signature, we used the correlation coefficients between two major gene sets (KEGG and REACTOME) and signatures to perform GSEA. The extracellular matrix, cell adhesion processes, and elastic fiber formation were found positively correlated with the signature score. On the contrary, base excision repair, cell cycle, and RNA

TABLE 1 Distribution of clinicopathological characteristics with low- and high-risk groups in the identified signature.

Variables	Level	Low	High	*p*-value
N		693	307	
Age (Sd)		67.380 (12.934)	66.897 (12.906)	0.6036
Gender (%)				
	Female	318 (45.89)	138 (44.95)	0.8373
	Male	375 (54.11)	169 (55.05)	
Stage (%)				
	II	428 (61.76)	130 (42.35)	<0.0001
	III	265 (38.24)	177 (57.65)	
pT (%)				
	1/2	22 (4.16)	5 (2.23)	0.232
	3	431 (81.47)	179 (79.91)	
	4	76 (14.37)	40 (17.86)	
pN (%)				
	0	320 (60.61)	99 (45.21)	<0.0001
	1	141 (26.70)	55 (25.11)	
	2/3	67 (12.69)	65 (29.68)	
MSI status (%)				
	MSI	85 (17.31)	27 (11.84)	0.0765
	MSS	406 (82.69)	201 (88.16)	
CIMP (%)				
	High	90 (19.48)	34 (17.71)	0.6528
	Low	117 (25.32)	55 (28.65)	
	Negative	255 (55.19)	103 (53.65)	
KRAS mutation (%)				
	No	310 (68.43)	110 (55.84)	0.0027
	Yes	143 (31.57)	87 (44.16)	
BRAF mutation (%)				
	No	388 (88.58)	159 (88.83)	1
	Yes	50 (11.42)	20 (11.17)	
CMS label (%)				
	CMS1	113 (17.94)	42 (15.11)	<0.0001
	CMS2	297 (47.14)	84 (30.22)	
	CMS3	90 (14.29)	35 (12.59)	
	CMS4	130 (20.63)	117 (42.09)	

processes were found negatively correlated with the signature (Supplementary Figure S3B). In addition, the expression of key ferroptosis-related genes was also compared. We found DHFR and CYB5R1 were highly expressed in low- and high-risk groups, respectively (Supplementary Figure S3C).

Establishment of predictive nomogram and decision tree

To provide potential clinical tool for recurrence prediction, we attempted to establish a nomogram together with the clinical and molecular features, and our signature. Age, gender, and stage status are considered primary characteristics that are routinely obtained in clinical practice. In addition, the results of a multivariate Cox regression analysis demonstrated that the FRG signature, pT, and KRAS mutation remained statistically significant even after adjusting for other factors. As a result, these specific characteristics were utilized in the development of a nomogram model (Figure 5D). In the calibration analysis of the nomogram, the prediction lines for 1-, 3-, and 5-year survival probability were closely aligned with the ideal reference line, indicating the favorable performance of the nomogram (Figure 5E). The DCA showed that the predictive model with the FRG signature yielded a higher net benefit compared to traditional prognostic factors enrolled in the nomogram (Figure 5F). In order to refine risk stratification, a recursive partitioning analysis was conducted on the 1,000 patients with the known tumor stage and FRG signature, resulting in a classifier decision tree (Figure 5G). The corresponding complexity parameter (CP) value with a minimum CV error at the first level was used to prune the decision tree (Supplementary Figure S4). In this analysis, the stage and FRG signature were identified as key determinants. Specifically, patients with stage II and a low FRG signature were categorized as the new low-risk group, while those with stage III and a high FRG signature were labeled as the new high-risk group. Patients who failed to align with these specific classifications were assigned to the intermediate-risk group. Overall, this approach helped to optimize risk stratification for each patient based on their unique characteristics.

Discussion

The tumor AJCC stage is still the most widely used biomarker in clinical practice to provide guidance for treatment (Yoshihara et al., 2013). CRC shows apparent tumor heterogeneity in prognosis and therapy response, even with the same stage (Srdjan et al., 2016). In addition, the five-year postoperative recurrence rate for patients with stage II/III CRC is approximately 10–30% (Osterman et al., 2020; Xu et al., 2020; Benson et al., 2021). It is controversial to give all of them identical adjuvant therapies, regardless of the tumor genetic and molecular heterogeneity. Ferroptosis is a regulated form of cell death that is driven by iron-dependent lipid peroxidation. It plays a critical role in various physiological and pathological processes (Cui et al., 2020). There is emerging evidence suggesting that ferroptosis may be involved in cancer progression and treatment response, which has led to the interest in exploring its potential as a prognostic biomarker (Zuo et al., 2022). Accordingly, it is reasonable for us to use comprehensive FRG signature to develop the prognosis model and recognize high-risk subpopulations.

Several ferroptosis signatures for CRC have been developed, demonstrating the prognostic significance. However, these signatures primarily rely on Lasso and multivariable Cox regression analyses, without considering the combination of

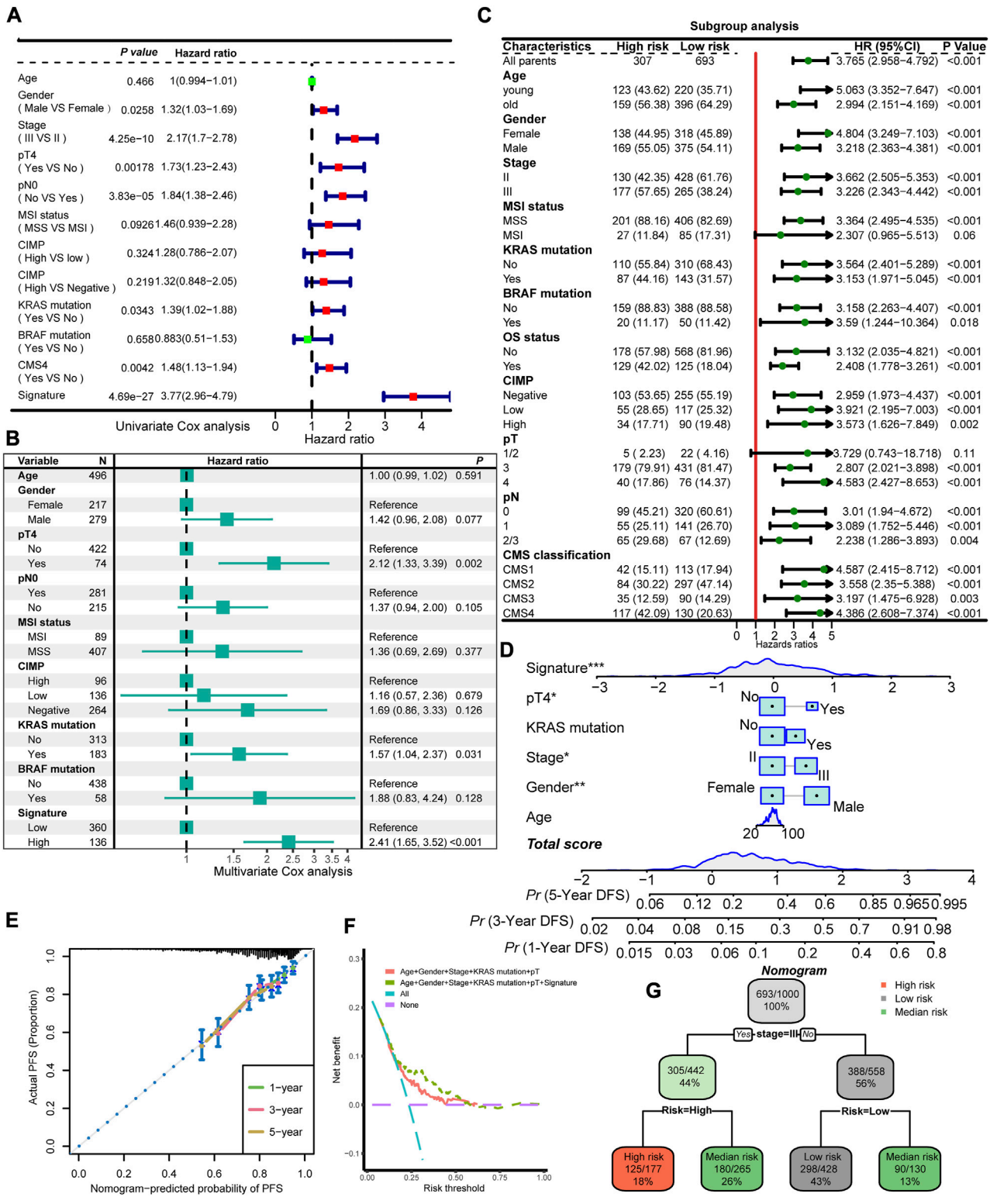


FIGURE 5
Interaction and combinations of the FRG signature with clinical and molecular features. Univariate Cox regression analysis **(A)** and multivariate Cox regression analysis **(B)** of prognostic factors for DFS for the training meta-cohort. **(C)** Subgroup analysis of the identified signature in clinical and molecular markers. **(D)** Prognostic nomogram predicting the probability of 1-, 3-, and 5-year DFS. **(E)** Calibration plot of the nomogram for 1-, 3-, and 5-year DFS prediction. Model performance is shown by the plot, relative to the 45-degree line, which represents perfect prediction. **(F)** DCA curve of the FRG signature and established risk factors in terms of DFS in the training cohort. The x-axis indicates the threshold probability, and the y-axis represents the net benefit. **(G)** A decision tree classifies patients into low-risk, intermediate-risk, and high-risk according to the probability of recurrent disease. * means $p < 0.05$, ** means $p < 0.01$, and *** means $p < 0.001$.

multiple algorithms to determine the optimal solution and performing thorough evaluations of the models on the validation set to assess their performance (Shao et al., 2021; Wang et al., 2022b; Du et al., 2022; Feng et al., 2022). To address this limitation, it is important to consider the combination of multiple algorithms and perform thorough evaluations of the models on the validation sets to assess their performance. This can help identifying the optimal solution and improve the robustness and generalizability of the prognostic signature. Machine learning algorithms have the advantages in making accurate predictions based on bulk data and using these predictions to guide future research efforts (Greener et al., 2022). Therefore, we were able to maximize the predictive accuracy of our model while also ensuring rigor and reproducibility. Additionally, the use of multiple independent validation datasets allowed us to evaluate the generalizability of the model and its performance in diverse population groups. However, while machine learning algorithms have shown great potential in making accurate predictions, selecting the optimal algorithm for model fitting can be challenging. Simply relying on researcher preference may not yield the best results and can lead to inefficiencies. One approach to address this issue is to use standardized methods for algorithm selection and model fitting. In our study, we took advantage of the strengths of machine learning and curated 10 different algorithms commonly used in survival analysis to generate an FRG signature for predicting stage II/III CRC prognosis. Rather than relying on a single algorithm or researcher preference, we utilized a combination of algorithms to create a framework and highlighted the importance of careful validation in this process. We calculated the C-index in multiple independent validation datasets to identify the best-performing model for predicting CRC recurrence.

In our study, we recognized 23 ferroptosis-related prognostic genes determined by the combination of Lasso and plsRcox with the highest average C-indices in validation datasets. The identified gene signature includes several FRGs associated with different aspects of cancer development. SAT1 plays a key role in immune regulation and metabolic signaling pathways, and it has been closely associated with chemoradiotherapy resistance and disease recurrence (Mou et al., 2022). MAPK3, a component of the RAS/MAPK pathway, may promote ferroptosis while potentially inhibiting antitumor immunity (Sun et al., 2022). NR1D2 is a transcriptional repressor that has been implicated in the epithelial–mesenchymal transition (EMT), a process that is crucial for cancer metastasis (Tong et al., 2020). Its knockdown could potentially slow down cancer progression by inhibiting EMT. CAV1 is a protein that plays a role in various cellular processes, including endocytosis and signal transduction. It has been identified as a suppressor of ferroptosis, a form of regulated cell death, and its high levels have been associated with the poor prognosis in cancer patients. Therefore, targeting CAV1 could potentially enhance the effect of ferroptosis and inhibit cancer progression (Lu et al., 2022). PTPN18 stabilizes the MYC protein level, leading to the activation of the MYC-CDK4 axis and promoting CRC development (Li et al., 2021). YTHDC2 is a tumor suppressor gene that is typically expressed at high levels in normal tissues and at low levels in tumor tissues. It has been associated with immune infiltrating levels, suggesting a role in the immune response to cancer (Li et al., 2020b). PAQR3 has been shown to induce apoptosis and inhibit proliferation and invasiveness of cancer cells when its expression is restored (Yu et al., 2015). Therefore, strategies to restore PAQR3 expression could potentially have therapeutic benefits in cancer treatment. Overall, these findings highlight the complex interplay between identified FRG genes and proteins in cancer progression.

In addition to the AJCC stage, several current and emerging clinically relevant biomarkers, such as BRAF mutations, HER2 overexpression and microsatellite state, were utilized to guide therapy in stage II/III CRCs (Sveen et al., 2020). To further validate the performance of the model, we compared C-index between the common clinical and molecular features (e.g., age, gender, T, N, AJCC stage, TMB, microsatellite state, and TP53, KRAS, or BRAF mutations) and our signature. Aging, male gender, and late stages are considered risk factors of CRC (Baraibar et al., 2023). TMB, KRAS, BRAF, and TP53 mutations were associated with worse prognosis, while high microsatellite instability (MSI-H) is a favorable prognostic biomarker and has been suggested as the predictors of immunotherapy response (Ganesh et al., 2019; Koncina et al., 2020). Our signature had high C-index levels in our comparisons with these common features, and the AUC of the 1- or 3-year relapses survival was robust across all datasets. Importantly, these markers of interest were found to be significant for recurrence in the univariate test, and the signature remained strongly predictive even after adjusting for them in the multivariate model. Moreover, our model demonstrated consistent predictive performance across subgroups with different clinical and molecular characteristics, further highlighting the robustness of the FRG signature as a prognostic tool. Collectively, these findings demonstrated that our signature could be a promising biomarker for predicting the high-risk recurrence population of stage II/III CRC in clinical practice.

The remarkable potential of the FRG signature score lies in its ability to precisely stratify patients into distinct high- and low-risk subgroups. This consequential stratification provides tailored and timely guidance regarding adjuvant therapy, specifically for stage II/III CRC patients. The FRG signature emerges as an effective tool in identifying stage II/III patients who are particularly susceptible to recurrence. Patients classified as stage III with a high FRG signature exhibited markedly decreased prognosis in comparison to those in stage II with a low FRG signature, while the intermediate-risk group comprising high-risk stage II and low-risk stage III patients shows similar outcomes.

The relevant biological activities of the extracellular matrix and cell adhesion and endothelial and epithelial cells were enriched in the high-risk group. They have been implicated in fibrosis, inflammation, thrombosis, cell division, and metastasis (Lin et al., 2021; Matthews et al., 2022). Several tumor suppressor pathways (e.g., base excision repair and cell cycle) were also enriched in the low-risk group. Alterations in iron metabolism and oxidative stress, key processes involved in ferroptosis, can be influenced by ECM remodeling and cell adhesion. Changes in iron import, export, and storage, as well as the presence of reactive oxygen species, can be regulated by ECM-related signals and cell adhesion molecules, thereby affecting the occurrence and progression of ferroptosis (He et al., 2023). Among the FRGs that we have identified in the signature, CAV1 plays a crucial role in the efficient deposition of ECM by fibroblast-derived exosomes, ultimately promoting tumor invasion. The activation of SIRT1 has a positive impact on the

expression of the major ECM components and helps to regulate ECM organization (Albacete-Albacete et al., 2020; Smith et al., 2022). T cells and B cells were found highly infiltrated in the low-risk group. T cells have emerged as powerful allies in the fight against cancer, while B cells play a crucial role by presenting tumor-associated antigens to T cells. Recent studies have revealed that activated $CD8^+$ T cells can enhance ferroptosis-specific lipid peroxidation in tumor cells, and this increased ferroptosis contributes to the antitumor efficacy of immunotherapy (Wang et al., 2019). KLF2 in the FRG signature is a transcription factor that has been demonstrated to play a crucial role in regulating the quiescence and trafficking of T lymphocytes. SATB1 directs lineage-specific transcriptional programs in the thymus, thereby influencing the development of the primary T-cell pool (Kakugawa et al., 2017; Wittner and Schuh, 2021). In addition, B cells can produce antibodies that enhance antigen presentation to T cells or directly target and kill tumor cells. This dynamic cooperation between T cells and B cells has a positive clinical impact (Waldman et al., 2020; Fridman et al., 2021). We found DHFR and CYB5R1, critical genes in ferroptosis, were highly expressed in low- and high-risk groups, respectively. Blockade of DHFR, either genetically or pharmacologically, enhances the effectiveness of GPX4 inhibition in triggering ferroptosis (Zheng and Conrad, 2020). Ferroptosis can also be induced by incidental electron transfer facilitated by POR/CYB5R1 oxidoreductase (Yan et al., 2021). This suggests that therapeutic approaches targeting ferroptosis induction may achieve favorable outcomes in the high-risk group. Overall, the underlying molecular mechanism suggested the biological plausibility and reliability in predicting the prognosis of the signature.

The nomogram we finally built present excellent performance. The capability of the FRG signature was validated with the calibration curve and DCA. The DCA curve demonstrated that incorporating the FRG signature yielded greater net benefit improvement compared to the conventional prognostic evaluation system. The prediction lines of the calibration curve for 1-, 3-, and 5-year survival probability were also fitted with the ideal reference line. In decision tree analysis, the intermediate-risk group regrouped stage II patients with a high signature and stage III patients with a low signature, thus enhancing the rationalization of risk groupings for stage II/III CRC patients. These results reinforced the potential for the FRG signature to guide personalized treatment decisions, improve outcomes for patients with CRC, and exhibit usability in daily routine practice.

Some limitations must be underscored with the current study even though the results of our investigation were profound. First, although a total of 1,397 patients were included with both microarray and RNA-seq platforms, they were all from retrospective cohorts. The signature should be further validated in a prospective study. Second, our study concentrated on the scope of ferroptosis-related mRNAs. Using the combination of lncRNA and mRNA might generate a more robust signature, which could be explored in future research. Last, the nomogram has been developed and validated in a computational model. However, it requires further clinical trials to confirm its effectiveness in real-world clinical settings and to evaluate its cost-effectiveness.

Identifying specific molecular targets involved in ferroptosis opens up avenues for developing novel therapeutic interventions. By targeting these genes or biological pathways in the FRG signature, it may be possible to modulate or inhibit ferroptosis, leading to improved treatment strategies for CRC. This could include developing small-molecule inhibitors or therapeutic agents that selectively target ferroptosis-related pathways. Moreover, integrating these FRGs and pathways with other omics data, such as proteomics and metabolomics, has the potential to uncover novel biomarkers and therapeutic targets. By combining multiple layers of molecular information, we can further gain insights into the complex interplay between different biological processes and identify key molecular players that can be targeted for therapeutic intervention.

In conclusion, our analysis established a stable and powerful ferroptosis-related signature based on consensus machine learning algorithms by sequencing the data of genes. The performance of the signature has been validated in multiple independent datasets and in comparison with the common clinical and molecular features. Furthermore, the model had great implications in the prognosis, even in subgroup analyses and after adjusting for common clinical and molecular markers. Finally, the developed nomogram, utilizing the common features and the signature, can potentially be a valuable tool to categorize high-risk patients. These findings indicate that the FRG signature shows promise in aiding clinical decision-making and facilitating personalized therapy for stage II/III CRC patients.

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.

Author contributions

ZW: data curation, formal analysis, investigation, methodology, software, validation, visualization, writing—original draft, and writing—review and editing. CM: project administration and writing—original draft. QT, XZ, and XL: formal analysis and writing—original draft. JM: validation and writing—original draft. TZ and WC: supervision and writing—review and editing. HC: conceptualization, funding acquisition, investigation, methodology, project administration, software, supervision, and writing—review and editing. ML: funding acquisition, resources, software, supervision, and writing—review and editing.

Funding

This study was supported by grants from the National Scientific Foundation of China (82103322, 81973116, 82173591, and 82102702), Natural Science Foundation of Shandong Province of China (ZR2020QH180 and ZR2021QH141), and National Key Research and Development Program of China (2021YFF1201101).

Acknowledgments

The authors thank all the researchers who took part in the databases mentioned in this study.

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

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2023.1260697/full#supplementary-material

References

Albacete-Albacete, L., Navarro-Lérida, I., López, J. A., Martín-Padura, I., Astudillo, A. M., Ferrarini, A., et al. (2020). ECM deposition is driven by caveolin-1-dependent regulation of exosomal biogenesis and cargo sorting. *J. Cell Biol.* 219, e202006178. doi:10.1083/jcb.202006178

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

Baraibar, I., Ros, J., Saoudi, N., Salvà, F., García, A., Castells, M. R., et al. (2023). Sex and gender perspectives in colorectal cancer. *ESMO Open* 8, 101204. doi:10.1016/j.esmoop.2023.101204

Battistuzzi, L., Puccini, A., and Sciallero, S. (2021). Microsatellite-instability-high advanced colorectal cancer. *N. Engl. J. Med.* 384, 971–972. doi:10.1056/NEJMc2036233

Benson, A. B., Venook, A. P., Al-Hawary, M. M., Arain, M. A., Chen, Y. J., Ciombor, K. K., et al. (2021). Colon cancer, version 2.2021, NCCN clinical practice guidelines in oncology. *J. Natl. Compr. Canc Netw.* 19, 329–359. doi:10.6004/jnccn.2021.0012

Chen, K., Collins, G., Wang, H., and Toh, J. W. T. (2021). Pathological features and prognostication in colorectal cancer. *Curr. Oncol.* 28, 5356–5383. doi:10.3390/curroncol28060447

Chong, W., Shang, L., Liu, J., Fang, Z., Du, F., Wu, H., et al. (2021). m(6 A regulator-based methylation modification patterns characterized by distinct tumor microenvironment immune profiles in colon cancer. *Theranostics* 11, 2201–2217. doi:10.7150/thno.52717

Chong, W., Zhu, X., Ren, H., Ye, C., Xu, K., Wang, Z., et al. (2022). Integrated multi-omics characterization of KRAS mutant colorectal cancer. *Theranostics* 12, 5138–5154. doi:10.7150/thno.73089

Cui, Y., Chen, H., Xi, R., Cui, H., Zhao, Y., Xu, E., et al. (2020). Whole-genome sequencing of 508 patients identifies key molecular features associated with poor prognosis in esophageal squamous cell carcinoma. *Cell Res.* 30, 902–913. doi:10.1038/s41422-020-0333-6

Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M., and Wallace, M. B. (2019). Colorectal cancer. *Lancet* 394, 1467–1480. doi:10.1016/S0140-6736(19)32319-0

Dienstmann, R., Mason, M. J., Sinicrope, F. A., Phipps, A. I., Tejpar, S., Nesbakken, A., et al. (2017). Prediction of overall survival in stage II and III colon cancer beyond TNM system: a retrospective, pooled biomarker study. *Ann. Oncol.* 28, 1023–1031. doi:10.1093/annonc/mdx052

Dienstmann, R., Villacampa, G., Sveen, A., Mason, M. J., Niedzwiecki, D., Nesbakken, A., et al. (2019). Relative contribution of clinicopathological variables, genomic markers, transcriptomic subtyping and microenvironment features for outcome prediction in stage II/III colorectal cancer. *Ann. Oncol.* 30, 1622–1629. doi:10.1093/annonc/mdz287

Du, S., Zeng, F., Sun, H., Liu, Y., Han, P., Zhang, B., et al. (2022). Prognostic and therapeutic significance of a novel ferroptosis related signature in colorectal cancer patients. *Bioengineered* 13, 2498–2512. doi:10.1080/21655979.2021.2017627

Feng, D., Shi, X., Xiong, Q., Zhang, F., Li, D., Wei, W., et al. (2022). A ferroptosis-related gene prognostic index associated with biochemical recurrence and radiation resistance for patients with prostate cancer undergoing radical radiotherapy. *Front. Cell Dev. Biol.* 10, 803766. doi:10.3389/fcell.2022.803766

Fridman, W. H., Petitprez, F., Meylan, M., Chen, T. W., Sun, C. M., Roumenina, L. T., et al. (2021). B cells and cancer: to B or not to B? *J. Exp. Med.* 218, 218. doi:10.1084/jem.20200851

Ganesh, K., Stadler, Z. K., Cercek, A., Mendelsohn, R. B., Shia, J., Segal, N. H., et al. (2019). Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat. Rev. Gastroenterol. Hepatol.* 16, 361–375. doi:10.1038/s41575-019-0126-x

Greener, J. G., Kandathil, S. M., Moffat, L., and Jones, D. T. (2022). A guide to machine learning for biologists. *Nat. Rev. Mol. Cell Biol.* 23, 40–55. doi:10.1038/s41580-021-00407-0

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., et al. (2015). The consensus molecular subtypes of colorectal cancer. *Nat. Med.* 21, 1350–1356. doi:10.1038/nm.3967

He, J., Abikoye, A. M., McLaughlin, B. P., Middleton, R. S., Sheldon, R., Jones, R. G., et al. (2023). Reprogramming of iron metabolism confers ferroptosis resistance in ECM-detached cells. *iScience* 26, 106827. doi:10.1016/j.isci.2023.106827

Jeffery, M., Hickey, B. E., and Hider, P. N. (2019). Follow-up strategies for patients treated for non-metastatic colorectal cancer. *Cochrane Database Syst. Rev.* 9, Cd002200. doi:10.1002/14651858.CD002200.pub2

Kakugawa, K., Kojo, S., Tanaka, H., Seo, W., Endo, T. A., Kitagawa, Y., et al. (2017). Essential roles of SATB1 in specifying T lymphocyte subsets. *Cell Rep.* 19, 1176–1188. doi:10.1016/j.celrep.2017.04.038

Kim, H. S., Kim, K. M., Lee, S. B., Kim, G. R., Han, Y. D., Cho, M. S., et al. (2019). Clinicopathological and biomolecular characteristics of stage IIB/IIC and stage IIIA colon cancer: insight into the survival paradox. *J. Surg. Oncol.* 120, 423–430. doi:10.1002/jso.25515

Kim, M. J., Jeong, S. Y., Choi, S. J., Ryoo, S. B., Park, J. W., Park, K. J., et al. (2015). Survival paradox between stage IIB/C (T4N0) and stage IIIA (T1-2N1) colon cancer. *Ann. Surg. Oncol.* 22, 505–512. doi:10.1245/s10434-014-3982-1

Koncina, E., Haan, S., Rauh, S., and Letellier, E. (2020). Prognostic and predictive molecular biomarkers for colorectal cancer: updates and challenges. *Cancers (Basel)* 12, 319. doi:10.3390/cancers12020319

Li, C., Li, S. Z., Huang, X. C., Chen, J., Liu, W., Zhang, X. D., et al. (2021). PTPN18 promotes colorectal cancer progression by regulating the c-MYC-CDK4 axis. *Genes Dis.* 8, 838–848. doi:10.1016/j.gendis.2020.08.001

Li, M., Zhu, C., Xue, Y., Miao, C., He, R., Li, W., et al. (2023). A DNA methylation signature for the prediction of tumour recurrence in stage II colorectal cancer. *Br. J. Cancer* 128, 1681–1689. doi:10.1038/s41416-023-02155-8

Li, X., Wen, D., Li, X., Yao, C., Chong, W., and Chen, H. (2020a). Identification of an immune signature predicting prognosis risk and lymphocyte infiltration in colon cancer. *Front. Immunol.* 11, 1678. doi:10.3389/fimmu.2020.01678

Li, Y., Zheng, J. N., Wang, E. H., Gong, C. J., Lan, K. F., and Ding, X. (2020b). The m6A reader protein YTHDC2 is a potential biomarker and associated with immune infiltration in head and neck squamous cell carcinoma. *PeerJ* 8, e10385. doi:10.7717/peerj.10385

Lin, D., Fan, W., Zhang, R., Zhao, E., Li, P., Zhou, W., et al. (2021). Molecular subtype identification and prognosis stratification by a metabolism-related gene expression signature in colorectal cancer. *J. Transl. Med.* 19, 279. doi:10.1186/s12967-021-02952-w

Liu, Z., Guo, C., Dang, Q., Wang, L., Liu, L., Weng, S., et al. (2022). Integrative analysis from multi-center studies identities a consensus machine learning-derived lncRNA signature for stage II/III colorectal cancer. *EBioMedicine* 75, 103750. doi:10.1016/j.ebiom.2021.103750

Lu, T., Zhang, Z., Pan, X., Zhang, J., Wang, X., Wang, M., et al. (2022). Caveolin-1 promotes cancer progression via inhibiting ferroptosis in head and neck squamous cell carcinoma. *J. Oral Pathol. Med.* 51, 52–62. doi:10.1111/jop.13267

Matthews, H. K., Bertoli, C., and de Bruin, R. A. M. (2022). Cell cycle control in cancer. *Nat. Rev. Mol. Cell Biol.* 23, 74–88. doi:10.1038/s41580-021-00404-3

Mi, M., Weng, S., Xu, Z., Hu, H., Wang, Y., and Yuan, Y. (2023). CSCO guidelines for colorectal cancer version 2023: updates and insights. *Chin. J. Cancer Res.* 35, 233–238. doi:10.21147/j.issn.1000-9604.2023.03.02

Mou, Y., Zhang, L., Liu, Z., and Song, X. (2022). Abundant expression of ferroptosis-related SAT1 is related to unfavorable outcome and immune cell infiltration in low-grade glioma. *BMC Cancer* 22, 215. doi:10.1186/s12885-022-09313-w

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457. doi:10.1038/nmeth.3337

Ng, J. Y., Lu, C. T., and Lam, A. K. (2019). BRAF mutation: current and future clinical pathological applications in colorectal carcinoma. *Histol. Histopathol.* 34, 469–477. doi:10.14670/HH-18-079

Osterman, E., Hammarström, K., Imam, I., Osterlund, E., Sjöblom, T., and Glimelius, B. (2020). Recurrence risk after radical colorectal cancer surgery-less than before, but how high is it? Cancers. *(Basel)* 12, 3308. doi:10.3390/cancers12113308

Ren, H., Bösch, F., Pretzsch, E., Jacob, S., Westphalen, C. B., Walter Holch, J., et al. (2022). Identification of an EMT-related gene signature predicting recurrence in stage II/III colorectal cancer: a retrospective study in 1780 patients. *Ann. Surg.* 276, 897–904. doi:10.1097/SLA.0000000000005644

Shao, Y., Jia, H., Huang, L., Li, S., Wang, C., Aikemu, B., et al. (2021). An original ferroptosis-related gene signature effectively predicts the prognosis and clinical status for colorectal cancer patients. *Front. Oncol.* 11, 711776. doi:10.3389/fonc.2021.711776

Shiovitz, S., Bertagnolli, M. M., Renfro, L. A., Nam, E., Foster, N. R., Dzieciatkowski, S., et al. (2014). CpG island methylator phenotype is associated with response to adjuvant irinotecan-based therapy for stage III colon cancer. *Gastroenterology* 147, 637–645. doi:10.1053/j.gastro.2014.05.009

Siegel, R. L., Miller, K. D., Goding Sauer, A., Fedewa, S. A., Butterly, L. F., Anderson, J. C., et al. (2020). Colorectal cancer statistics, 2020. *CA Cancer J. Clin.* 70, 145–164. doi:10.3322/caac.21601

Smith, C. A., Humphreys, P. A., Bates, N., Naven, M. A., Cain, S. A., Dvir-Ginzberg, M., et al. (2022). SIRT1 activity orchestrates ECM expression during hESC-chondrogenic differentiation. *Faseb J.* 36, e22314. doi:10.1096/fj.202200169R

Song, Y. Q., Yan, X. D., Wang, Y., Wang, Z. Z., Mao, X. L., Ye, L. P., et al. (2023). Role of ferroptosis in colorectal cancer. *World J. Gastrointest. Oncol.* 15, 225–239. doi:10.4251/wjgo.v15.i2.225

Srdjan, M., Jadranka, A., Ivan, D., Branimir, Z., Daniela, B., Petar, S., et al. (2016). Microsatellite instability & survival in patients with stage II/III colorectal carcinoma. *Indian J. Med. Res.* 143, S104–s111. doi:10.4103/0971-5916.191801

Stintzing, S., Wirapati, P., Lenz, H. J., Neureiter, D., Fischer von Weikersthal, L., Decker, T., et al. (2019). Consensus molecular subgroups (CMS) of colorectal cancer (CRC) and first-line efficacy of FOLFIRI plus cetuximab or bevacizumab in the FIRE3 (AIO KRK-0306) trial. *Ann. Oncol.* 30, 1796–1803. doi:10.1093/annonc/mdz387

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15545–15550. doi:10.1073/pnas.0506580102

Sun, H., Qian, X., Yang, W., Zhou, W., Zhou, C., Liu, S., et al. (2022). Novel prognostic signature based on HRAS, MAPK3 and TFRC identified to be associated with ferroptosis and the immune microenvironment in hepatocellular carcinoma. *Am. J. Transl. Res.* 14, 6924–6940.

Sveen, A., Kopetz, S., and Lothe, R. A. (2020). Biomarker-guided therapy for colorectal cancer: strength in complexity. *Nat. Rev. Clin. Oncol.* 17, 11–32. doi:10.1038/s41571-019-0241-1

Tan, W. C. C., Nerurkar, S. N., Cai, H. Y., Ng, H. H. M., Wu, D., Wee, Y. T. F., et al. (2020). Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. *Cancer Commun. (Lond)* 40, 135–153. doi:10.1002/cac2.12023

Timar, J., and Kashofer, K. (2020). Molecular epidemiology and diagnostics of KRAS mutations in human cancer. *Cancer Metastasis Rev.* 39, 1029–1038. doi:10.1007/s10555-020-09915-5

Tong, H., Liu, X., Li, T., Qiu, W., Peng, C., Shen, B., et al. (2020). NR1D2 accelerates hepatocellular carcinoma progression by driving the epithelial-to-mesenchymal transition. *Onco Targets Ther.* 13, 3931–3942. doi:10.2147/OTT.S237804

Van Calster, B., Wynants, L., Verbeek, J. F. M., Verbakel, J. Y., Christodoulou, E., Vickers, A. J., et al. (2018). Reporting and interpreting decision curve analysis: a guide for investigators. *Eur. Urol.* 74, 796–804. doi:10.1016/j.eururo.2018.08.038

Waldman, A. D., Fritz, J. M., and Lenardo, M. J. (2020). A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat. Rev. Immunol.* 20, 651–668. doi:10.1038/s41577-020-0306-5

Wang, W., Green, M., Choi, J. E., Gijón, M., Kennedy, P. D., Johnson, J. K., et al. (2019). CD8(+) T cells regulate tumour ferroptosis during cancer immunotherapy. *Nature* 569, 270–274. doi:10.1038/s41586-019-1170-y

Wang, X., Xu, Y., Dai, L., Yu, Z., Wang, M., Chan, S., et al. (2022b). A novel oxidative stress- and ferroptosis-related gene prognostic signature for distinguishing cold and hot tumors in colorectal cancer. *Front. Immunol.* 13, 1043738. doi:10.3389/fimmu.2022.1043738

Wang, Y., Zhang, Z., Sun, W., Zhang, J., Xu, Q., Zhou, X., et al. (2022a). Ferroptosis in colorectal cancer: potential mechanisms and effective therapeutic targets. *Biomed. Pharmacother.* 153, 113524. doi:10.1016/j.biopha.2022.113524

Wittner, J., and Schuh, W. (2021). Krüppel-like factor 2 (KLF2) in immune cell migration. *Vaccines (Basel)* 9, 1171. doi:10.3390/vaccines9101171

Wu, Z., Lu, Z., Li, L., Ma, M., Long, F., Wu, R., et al. (2021). Identification and validation of ferroptosis-related LncRNA signatures as a novel prognostic model for colon cancer. *Front. Immunol.* 12, 783362. doi:10.3389/fimmu.2021.783362

Xu, W., He, Y., Wang, Y., Li, X., Young, J., Ioannidis, J. P. A., et al. (2020). Risk factors and risk prediction models for colorectal cancer metastasis and recurrence: an umbrella review of systematic reviews and meta-analyses of observational studies. *BMC Med.* 18, 172. doi:10.1186/s12916-020-01618-6

Yan, B., Ai, Y., Sun, Q., Ma, Y., Cao, Y., Wang, J., et al. (2021). Membrane damage during ferroptosis is caused by oxidation of phospholipids catalyzed by the oxidoreductases POR and CYB5R1. *Mol. Cell* 81, 355–369.e10. doi:10.1016/j.molcel.2020.11.024

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

Yu, T. J., Ma, D., Liu, Y. Y., Xiao, Y., Gong, Y., Jiang, Y. Z., et al. (2021). Bulk and single-cell transcriptome profiling reveal the metabolic heterogeneity in human breast cancers. *Mol. Ther.* 29, 2350–2365. doi:10.1016/j.ymthe.2021.03.003

Yu, X., Li, Z., Chan, M. T., and Wu, W. K. (2015). PAQR3: a novel tumor suppressor gene. *Am. J. Cancer Res.* 5, 2562–2568.

Zheng, J., and Conrad, M. (2020). The metabolic underpinnings of ferroptosis. *Cell Metab.* 32, 920–937. doi:10.1016/j.cmet.2020.10.011

Zuo, Y. B., Zhang, Y. F., Zhang, R., Tian, J. W., Lv, X. B., Li, R., et al. (2022). Ferroptosis in cancer progression: role of noncoding RNAs. *Int. J. Biol. Sci.* 18, 1829–1843. doi:10.7150/ijbs.66917

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 01 September 2023
DOI 10.3389/fphar.2023.1224828

OPEN ACCESS

EDITED BY
Lin Qi,
Central South University, China

REVIEWED BY
Wenjian Hu,
Southwest Medical University, China
Bl Gan,
Guangxi Medical University, China

*CORRESPONDENCE
Juntian Liu,
ljt23340123@163.com
Ming Gao,
headandneck2008@126.com

†These authors have contributed equally to this work

RECEIVED 18 May 2023
ACCEPTED 21 August 2023
PUBLISHED 01 September 2023

CITATION
Li L, Wu N, Zhuang G, Geng L, Zeng Y, Wang X, Wang S, Ruan X, Zheng X, Liu J and Gao M (2023), Heterogeneity and potential therapeutic insights for triple-negative breast cancer based on metabolic-associated molecular subtypes and genomic mutations. *Front. Pharmacol.* 14:1224828. doi: 10.3389/fphar.2023.1224828

Heterogeneity and potential therapeutic insights for triple-negative breast cancer based on metabolic-associated molecular subtypes and genomic mutations

Lijuan Li[1†], Nan Wu[1†], Gaojian Zhuang[2†], Lin Geng[3], Yu Zeng[3], Xuan Wang[4], Shuang Wang[1], Xianhui Ruan[3], Xiangqian Zheng[3], Juntian Liu[1*] and Ming Gao[3,5*]

[1]Department of Cancer Prevention Center, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Key Laboratory of Breast Cancer Prevention and Therapy, Tianjin Medical University, Ministry of Education, Key Laboratory of Cancer Prevention and Therapy, Tianjin, China, [2]The Sixth Affiliated Hospital of Guangzhou Medical University, Qingyuan People's Hospital, Qingyuan, China, [3]Department of Thyroid and Neck Tumor, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Key Laboratory of Breast Cancer Prevention and Therapy, Tianjin Medical University, Ministry of Education, Key Laboratory of Cancer Prevention and Therapy, Tianjin, China, [4]Department of Phase I Clinical Trial, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Key Laboratory of Breast Cancer Prevention and Therapy, Tianjin Medical University, Ministry of Education, Key Laboratory of Cancer Prevention and Therapy, Tianjin, China, [5]Department of Thyroid and Breast Surgery, Tianjin Union Medical Center, Tianjin Key Laboratory of General Surgery in construction, Tianjin Union Medical Center, Tianjin, China

Objective: Due to a lack of effective therapy, triple-negative breast cancer (TNBC) is extremely poor prognosis. Metabolic reprogramming is an important hallmark in tumorigenesis, cancer diagnosis, prognosis, and treatment. Categorizing metabolic patterns in TNBC is critical to combat heterogeneity and targeted therapeutics.

Methods: 115 TNBC patients from TCGA were combined into a virtual cohort and verified by other verification sets, discovering differentially expressed genes (DEGs). To identify reliable metabolic features, we applied the same procedures to five independent datasets to verify the identified TNBC subtypes, which differed in terms of prognosis, metabolic characteristics, immune infiltration, clinical features, somatic mutation, and drug sensitivity.

Abbreviations: TNBC, Triple-negative breast cancer; TCGA, The Cancer Genome Atlas; NMF, Non-negative matrix factorization; DEGs, Differentially expressed genes; PIK3CA, Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; KMT2D, Lysine methyltransferase 2D; KMT2C, Lysine methyltransferase 2C; GEO, Gene Expression Omnibus; GDC, Genome Data Commons; OS, overall survival; DFS, Disease-free survival; MSigDB, Molecular Signature Database; GSEA, Gene Set Enrichment Analysis; GSVA, Genomic variation analysis; ssGSEA, Single-sample gene-set enrichment analysis; TMB, Tumor mutation burden; CNV, Copy number variant; GDSC, Genomics of Drug Sensitivity in Cancer; PD-1, Programmed cell death-1; PD-L1, Programmed cell death ligand-1; CTLA-4, Cytotoxic T-lymphocyte antigen-4; AKR1B10, Aldo-Keto reductase family 1 member B10.

Results: In general, TNBC could be classified into two metabolically distinct subtypes. C1 had high immune checkpoint genes expression and immune and stromal scores, demonstrating sensitivity to the treatment of PD-1 inhibitors. On the other hand, C2 displayed a high variation in metabolism pathways involved in carbohydrate, lipid, and amino acid metabolism. More importantly, C2 was a lack of immune signatures, with late pathological stage, low immune infiltration and poor prognosis. Interestingly, C2 had a high mutation frequency in PIK3CA, KMT2D, and KMT2C and displayed significant activation of the PI3K and angiogenesis pathways. As a final output, we created a 100-gene classifier to reliably differentiate the TNBC subtypes and AKR1B10 was a potential biomarker for C2 subtypes.

Conclusion: In conclusion, we identified two subtypes with distinct metabolic phenotypes, provided novel insights into TNBC heterogeneity, and provided a theoretical foundation for therapeutic strategies.

KEYWORDS

metabolic subtypes, triple-negative breast cancer, metabolic pathway, immune signature, immunotherapy response, mutation landscape

1 Introduction

In 2020, breast cancer become the main cause of malignant tumors and the fifth leading cause of death. Three million new patients and 685,000 deaths (Sung et al., 2021). As a disease with high heterogeneity, the treatment and prognosis of patients are greatly different. With the definition of breast cancer molecular subtypes being proposed, triple-negative breast cancer (TNBC) is classified as a type of breast cancer. This type has no expression of estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) (also known as ERBB2) (Goldhirsch et al., 2013; Waks and Winer, 2019). TNBC accounts for 10%–20%, with being prone to recurrence and metastasis. Due to the high early recurrence rate and limited treatment, the prognosis is very poor (Denkert et al., 2017; Garrido-Castro et al., 2019). Much effort has been devoted to classifying TNBC into subtypes of several molecular with different mutational characteristics and genomic changes (Bareche et al., 2018; Garrido-Castro et al., 2019; Jiang et al., 2021). Previous studies showed that cluster analysis identified TNBC subtypes, which provided new ideas for the treatment of TNBC (Lehmann et al., 2011; Jiang et al., 2019; Xiao et al., 2022).

Metabolic reprogramming, as an emerging hallmark, is a new tumor biomarker that plays a major role in the occurrence, progression, diagnosis, treatment, and prognosis (Martinez-Outschoorn et al., 2017; Xia et al., 2021). Due to the heterogeneous metabolic dependencies existing across different tumor types and even the same tissue (Hensley et al., 2016; Kim and DeBerardinis, 2019), we know little about the impact of tumor metabolic reprogramming on TNBC. In addition to some previous pan-cancer analysis (Rosario et al., 2018), the understanding of TNBC metabolic heterogeneity is still insufficient. Thanks to advancements in bioinformatics, we are now equipped to analyze high-throughput genetic data to gain insights into diseases, such as autoimmune disorders (Li et al., 2022; Li et al., 2023a; Li et al., 2023b) or cancers (Cheng et al., 2023; Tu et al., 2023). Building on this, our study delves into classifying Triple-Negative Breast Cancer (TNBC) from a metabolic perspective, shedding light on its underlying heterogeneity.

We used the screened metabolic genes to systematically check the diverse metabolic signatures of TNBC and identify two distinct metabolic subtypes. Differentially expressed genes (DEGs) were revealed by comparing transcriptome levels of patients with different subtypes. Subtyping TNBC Prognosis, metabolic characteristics, immune infiltration, clinical features, *in vivo* cell mutation characteristics, and drug sensitivity vary. Finally, a 100-gene classifier was designed and preliminarily verified to determine the classification of TNBC. This investigation may also provide insightful information into tumor-immune cell interactions, which retains tremendous potential for clinical therapeutic interventions in TNBC patients.

2 Materials and methods

2.1 Patients and samples

BRCA gene expression profiles were downloaded from five independent cohorts of patients, including TCGA-BRCA, GSE25066, GSE21653, GSE103091and METABRIC. Only samples from TNBC were reserved in all cohorts. Survival analysis only considered overall survival (OS) and disease-free survival (DFS). In the above five cohorts, METABRIC had no patient prognostic information. The remaining histological data were obtained from the TCGA-BRCA cohort, including copy number variant data obtained via firehose, and mutation MAF files obtained from the cBioPortal Pancancer Project. TCGA-BRCA partial samples of the predicted neoantigen numbers were obtained from published literature (Rooney et al., 2015). The metabolic gene file used for clustering (Possemato et al., 2011), the metabolism signatures (Rosario et al., 2018), the immune pathway signatures (Bindea et al., 2013) and the oncogenetic signature.txt (Sanchez-Vega et al., 2018) from different published literature. The drug information is from the GDSC database involving the drug's R package "pRRophetic" for use in predicting the drug's IC50. The external datasets were used to determine whether the defined subtypes are likely to respond to immunotherapy (Roh et al., 2017).

Limma package used for identifying DEGs (|log2FC| > 1 and $p < 0.01$). Genetic feature set files "c2.cp.kegg.v6.2.symbols.gmt" and "h,all,v60.2.symbols" were obtained from the Molecular Signature Database (MSigDB). Then, Clusterprofiler R package was used for Gene Set Enrichment Analysis (GSEA) (Yu et al., 2012).

To identify Aldo-Keto reductase family 1 member B10 (AKR1B10), we collected peripheral blood samples from 30 TNBC patients and 30 healthy individuals as controls from Tianjin Medical University Cancer Institute and Hospital (Tianjin, China) in 2022 for RT-qPCR, and their paraffin-embedded tissues for IHC. All patients were female patients who were recently admitted and had not undergone radiotherapy, chemotherapy, or surgery. Control group was determined to be free from TNBC and other malignant tumors.

2.2 Identification of TNBC subtypes through non-negative matrix factorization clustering

Because all data used in this study were derived from five platforms, and some of the data were normalized, we combined the data after normalizing each data using z-score to eliminate potential batch effects. We performed consensus NMF (Possemato et al., 2011) with 2-5 cluster numbers using TCGA data expression profiles and calculated the covariance coefficients for each decomposition. The MOVICS package (Lu et al., 2021) was used for differential expression analysis of these two subtypes, while the top 50 most significantly upregulated genes in each subtype were used as biomarkers for the different subtypes ($p < 0.05$, FDR < 0.25). In addition, we constructed a template using MOVICS.

2.3 Gene mutation analysis and single-sample gene-set enrichment analysis (ssGSEA)

Genomic variation analysis (GSVA) is a method genome augmentation, which calculates the characteristics of certain pathways or different populations based on expression spectra. The differences of gene sets between samples were investigated by GSVA R software package from relevant metabolic pathway gene sets (Rosario et al., 2018). Then, the limma package (Liu et al., 2019) was used to obtain the substitution gene scores, analyze the differences, and screen for DEGs features.

To identify the extent to which genes are up or downregulated within a single sample, ssGSEA is used for quantifying the immune composition of tumors. Here, we assessed the enrichment fraction of gene sets representing biological processes as well as biological pathways in bulk tumors or individual cancer cells by ssGSEA.

2.4 Detection of tumor microenvironment characteristics

The ESTIMATE algorithm (Yoshihara et al., 2013) can be applied to calculate the permeability and matrix content of immune cells (immune fraction) and stromal content (stromal fraction) of different subtypes, thus reflecting the microenvironmental characteristics of tumors. Microenvironmental Cell Population counter (MCPcounter) (Becht et al., 2016) was used for evaluating the penetration frequency of immune and non-immune cell populations in two subtypes.

2.5 Evaluation of genomic changes, number of new antigens, tumor mutation burden (TMB) and copy number variant (CNV) in different groups

The detection of co-occurrence and mutually exclusive mutations mainly relied on the CoMEt algorithm. Next, we predicted the different genotypes between different subtypes, including the number of neoantigens, TMB, copy number amplification, and the frequency of copy number deletions. We also performed an online analysis using GISTIC2 (Cibulskis et al., 2013) to obtain the number of amplifications and deletions for all samples and to calculate arm- and focal level somatic copy-number alterations (SCNAs) and G-scores in tumors with the input of "SNP6" files.

2.6 Prediction of treatment for each subgroup of immune checkpoint

MD-Anderson melanoma cohort treated with anti-CTLA-4 or anti-PD-1 is considered to be used to predict immunotherapy response (Roh et al., 2017). And then, we analyzed the sub map from the Genomics of Drug Sensitivity in Cancer (GDSC) database (Roh et al., 2017) and studied the sensitivity differences between the C1 and C2 groups after multiple drug treatments.

2.7 IHC staining

IHC staining was used to slice the dewaxed tissue portion of TNBC samples and cure with 3% hydrogen peroxide for a period of time. Block endogenous peroxidase for 30 min, then solidify with appropriate horseradish AKR1B10 antibody. The IHC fraction is calculated by multiplying the dyeing intensity by the percentage of cells. Definition of intensity: 0 (unstained), 1 (soft), 2 (medium), 3 (strong). Definition of percentage of cells: 1 (25%), 2 (26%–50%), 3 (51%–75%), and 4 (>75%). More than 3 was defined as positive, while less than or equal to 3 as negative.

2.8 RNA isolation and RT-qPCR

Triazole solution (AC0101-B; SparkJade, China) was used for extracting RNA from blood and tissues samples. 2×HQ SYBR qPCR Mix (ZF501; ZOMANBIO; Beijing, China) was used for PCR reaction. Primer sequences were listed in Supplementary Table S1. The levels of AKR1B10 expression were calculated by the method of $2^{-\Delta\Delta Cq}$.

10.3389/fphar.2023.1224828

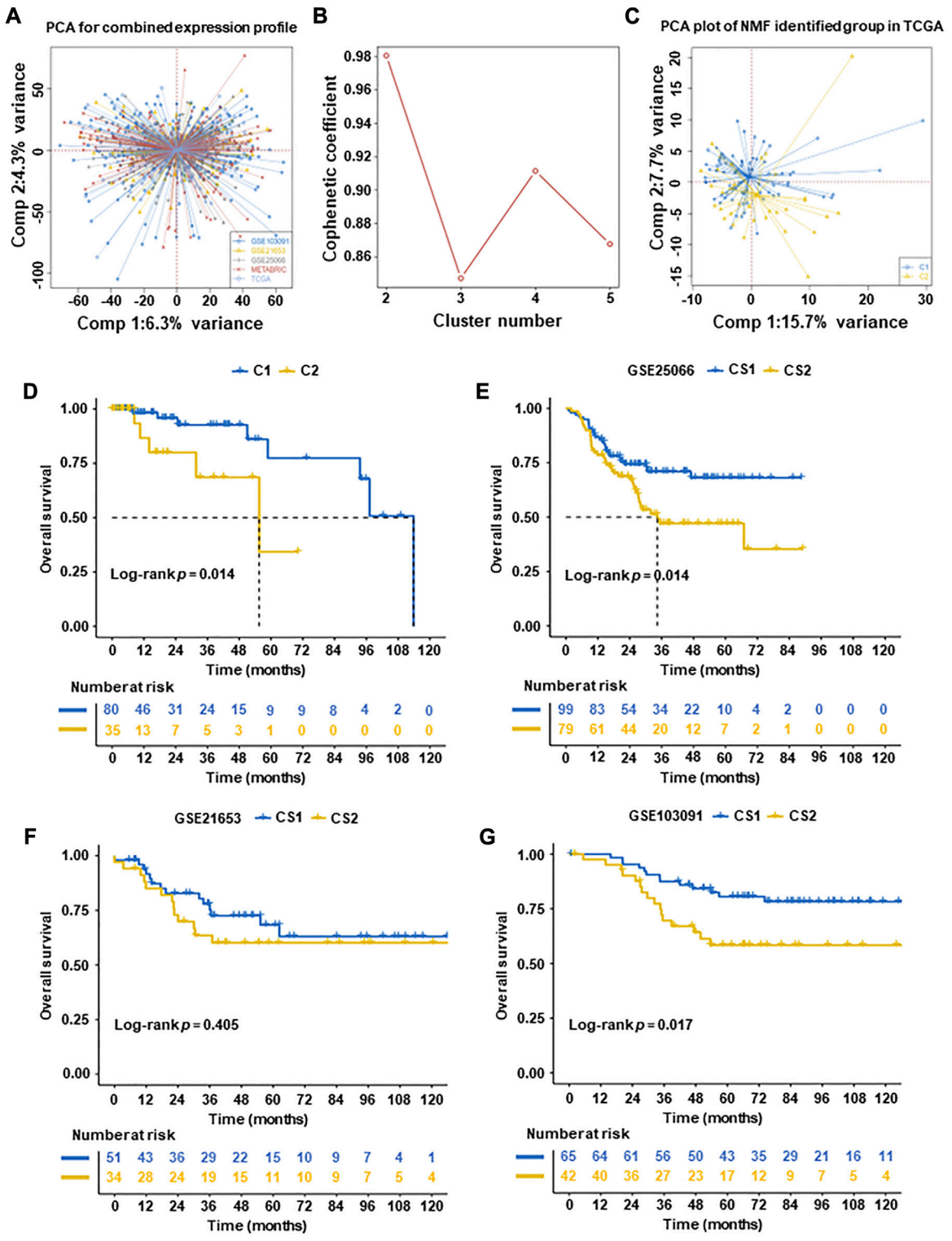


FIGURE 1
Subtyping of TNBC tumors according to non-negative matrix factorization (NMF) in five datasets. **(A)** Principal component analysis (PCA) of integrated expression profiles based on five TNBC datasets. **(B)** After comprehensive consideration, the optimal clustering number (k value) was 2. **(C)** PCA dimension reduction analysis was used to support the classification into two TNBC-subtypes. **(D)** Overall survival (OS) analysis of two subtypes in TCGA datasets. **(E–G)** Overall survival (OS) analysis of validation datasets (GSE25066, GSE21653, and GSE103091, excluding METABRIC with a lack of patient OS). The results of OS revealed that C1 had significantly better than C2 in TCGA datasets and validation datasets (GSE25066, GSE21653, and GSE103091) (p = 0.014, p = 0.014, and p = 0.017, respectively).

10.3389/fphar.2023.1224828

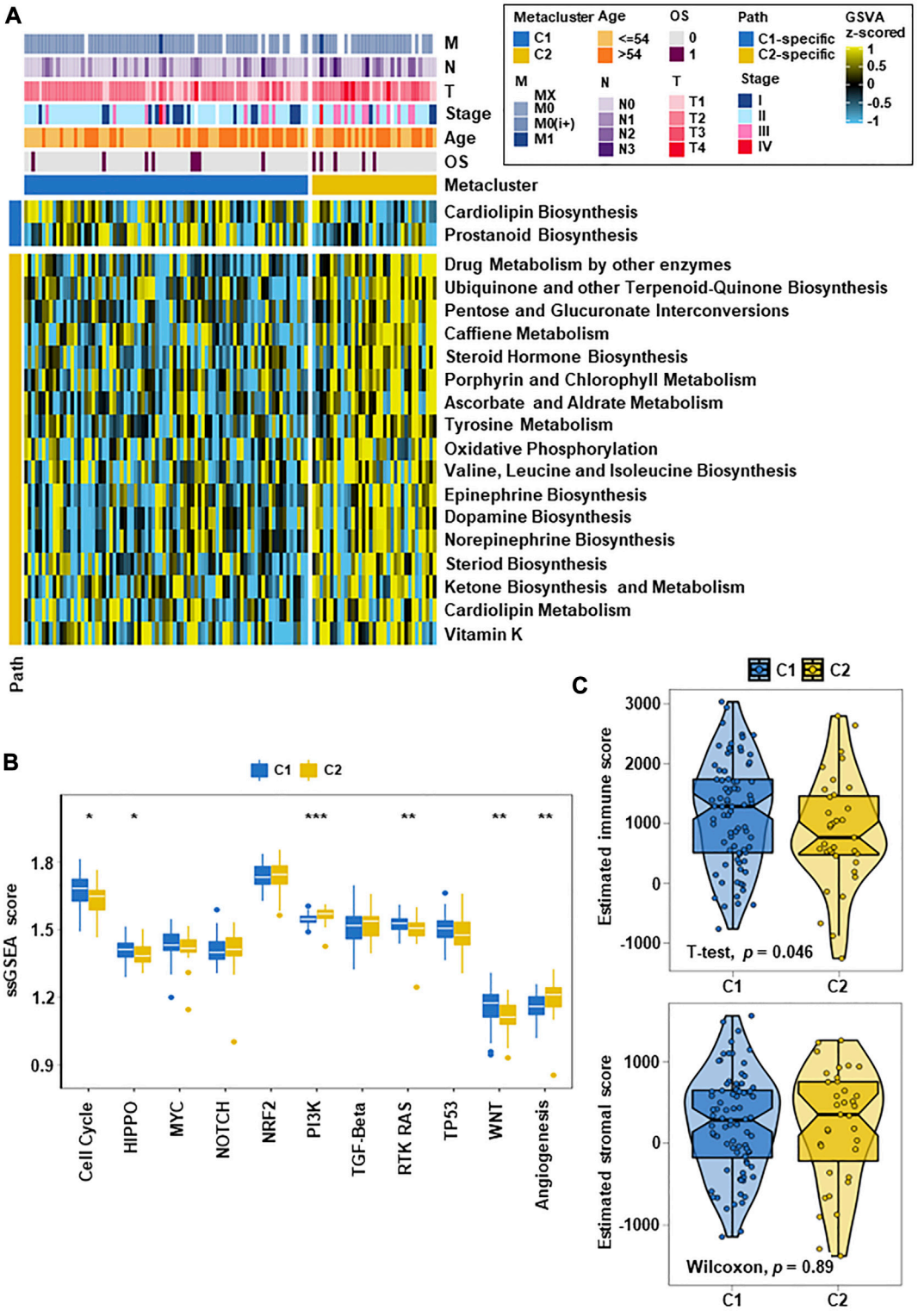


FIGURE 2
TNBC metabolic subtypes and tumor progression-related features. **(A)** Heatmap of metabolism-related features in the two subtypes. C1 and C2 had specific metabolic characteristics, with 2 metabolism-related pathways upregulated in C1 and 17 metabolism-related pathways significantly upregulated in C2. **(B)** Box plots of tumor progression-related signaling pathways in the two subtypes. After quantifying 11 carcinogenic pathways, the results showed differences between the them on multiple classic carcinogenic pathways. **(C)** Box line plots of the immune fraction and matrix fraction of ESTIMATE in the two subtypes (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

2.9 Statistical analysis

R software (version 4.0.2) was used to process all data. Contingency table ($\chi 2$) variables used the chi-square test and Fisher's precision probability test for statistical significance. Kaplan-Meier method was used for survival analysis and compared the results by the log-rank test. Z test was used to assess whether there was a significant difference between the two groups. Univariate Cox proportional hazards regression models were used to assess the risk ratio for univariate analysis. A two-tailed p-value < 0.05 was considered statistically significant.

3 Results

3.1 NMF distinguishes two subtypes of TNBC

Based on TCGA database and NMF algorithm analysis, we divided TNBC into two subtypes with different metabolic characteristics. In this study, 115 cases of TCGA-TNBC patients were screened. Supplementary Table S2 showed clinical characteristics of TNBC patients. Before analyzing the TNBC NMF algorithm, we used the ComBat algorithm to eliminate batch processing effects in the TNBC queue. And after deleting the batch processing effect, draw a key element analysis diagram (Figure 1A). Previously, a total of 2,752 reported metabolic related genes (Possemato et al., 2011) were screened and downloaded as the basis for analyzing metabolomics in our study (Supplementary Table S3).

To identify subtypes in TNBC, Cox regression was used. A total of 637 prognostic genes were obtained (Supplementary Table S4). After a further adjusted p-value ($p < 0.05$), 277 candidate genes were identified (Supplementary Table S5). We then used the NMF algorithm to cluster the 277 candidate genes and drew the NMF with two to five sets (Figure 1B). After comprehensive consideration, the optimal clustering number (k value) was 2, defining two subtypes C1 ($n = 80$) and C2 ($n = 35$). To verify the consistency between subtype designations and two-dimensional distribution patterns, we reduced the PCA dimension (Figure 1C). Subsequently, the same conclusion was validated in the validation set (GSE25066, GSE21653, and GSE103091, excluding METABRIC with a lack of patient OS).

Finally, two TNBC molecular subtypes were established. We also used the survival information in the four queues to analyze the subtype survival of TNBC subsets. The OS of C1 was verified better than that of C2 in TCGA-TNBC patients ($p = 0.014$, Figure 1D) and other datasets patients (GSE25066 and GSE103091) ($p = 0.014$ and $p = 0.017$, respectively) (Figures 1E, G). No significant difference was observed in datasets (GSE21653) ($p = 0.15$) (Figure 1F).

3.2 Association of TNBC subtypes with metabolism-related signatures

In this study, we analyzed whether different TNBC subtypes have their own characteristics in distinct metabolic pathways. Firstly, we used the GSVA R package to score metabolic pathways (Rosario et al., 2018) (Supplementary Table S6). Limma difference test cross group was performed to confirm subtype-specific differential metabolic pathways, and heatmaps were constructed for visualization (Supplementary Figure S1).

Furthermore, the DEGs between two groups were detected by GSVA enrichment again, and it was found that C1 and C2 had specific metabolic characteristics (Figure 2A). There were 17 metabolism-related pathways in C2 that were significantly upregulated, mainly involving pentose and glucuronate interconversion, steroid hormone biosynthesis, tyrosine metabolism, oxidative phosphorylation, ketone biosynthesis and metabolism. Similar outcomes that 17 metabolism-related pathways were activated in C2 were observed in the validation datasets (GSE25066, GSE21653, GSE103091 and METABRIC) (Supplementary Figures S2, S3). To determine the different activities of metabolic pathways, we represented the two subtypes in the TCGA-TNBC cohorts and validation cohorts and revealed that C2 contained the highest activation of metabolic pathways in all cohorts (Supplementary Figure S4).

In analyze the differences between the two subtypes in carcinogenesis-related pathways, we counted the GSVA enrichment points and plotted box-line plots. Eleven carcinogenesis-related pathways were selected and quantified. The results showed that different subgroups were closely related to the activation of different carcinogenic signaling pathways, which mainly involved cell cycle, PI3K, RTK-RAS, and angiogenesis (Figure 2B). C2 displayed significant activation of the PI3K and angiogenesis pathways. C1 had a stronger cell cycle, HIPPO, RTK-RAS and WNT signature than C2. These differences in carcinogenic pathway activity may affect their prognosis. After evaluating whether the subtype was related to the tumor microenvironment, it was found that the immune score of C1 was higher than that of C2 ($p = 0.046$), and the stromal score had no significant difference (Figure 2C).

3.3 Association of TNBC subtypes with immune infiltration

To evaluate the immune status of two subtypes, the MCP counter and ssGSEA algorithm were used to estimate the abundance of immune cells (Figure 3A). The results showed significant differences between different immune cell groups between the two subtypes (Figures 3B, C). Especially, the immune value of C2 in most immune cells was obvious lower than that of C1, except for neutrophils, fibroblasts and Th17 cells. According to this study, C1 was rich in more immune cells and had the highest immune score, which indicated that differences in the distribution of different immune cells may be the reason for the poorer prognosis of C2 than C1.

3.4 Association of TNBC subtypes with clinical features

To explore the relationship between these subtypes and clinical features, we analyzed the clinicopathological parameters between the two subtypes and constructed a clinical information heatmap of subtypes (Figure 3A). The results revealed that larger tumor size ($p = 0.007$) and advanced pathologic stage (TNM III/IV stage)

10.3389/fphar.2023.1224828

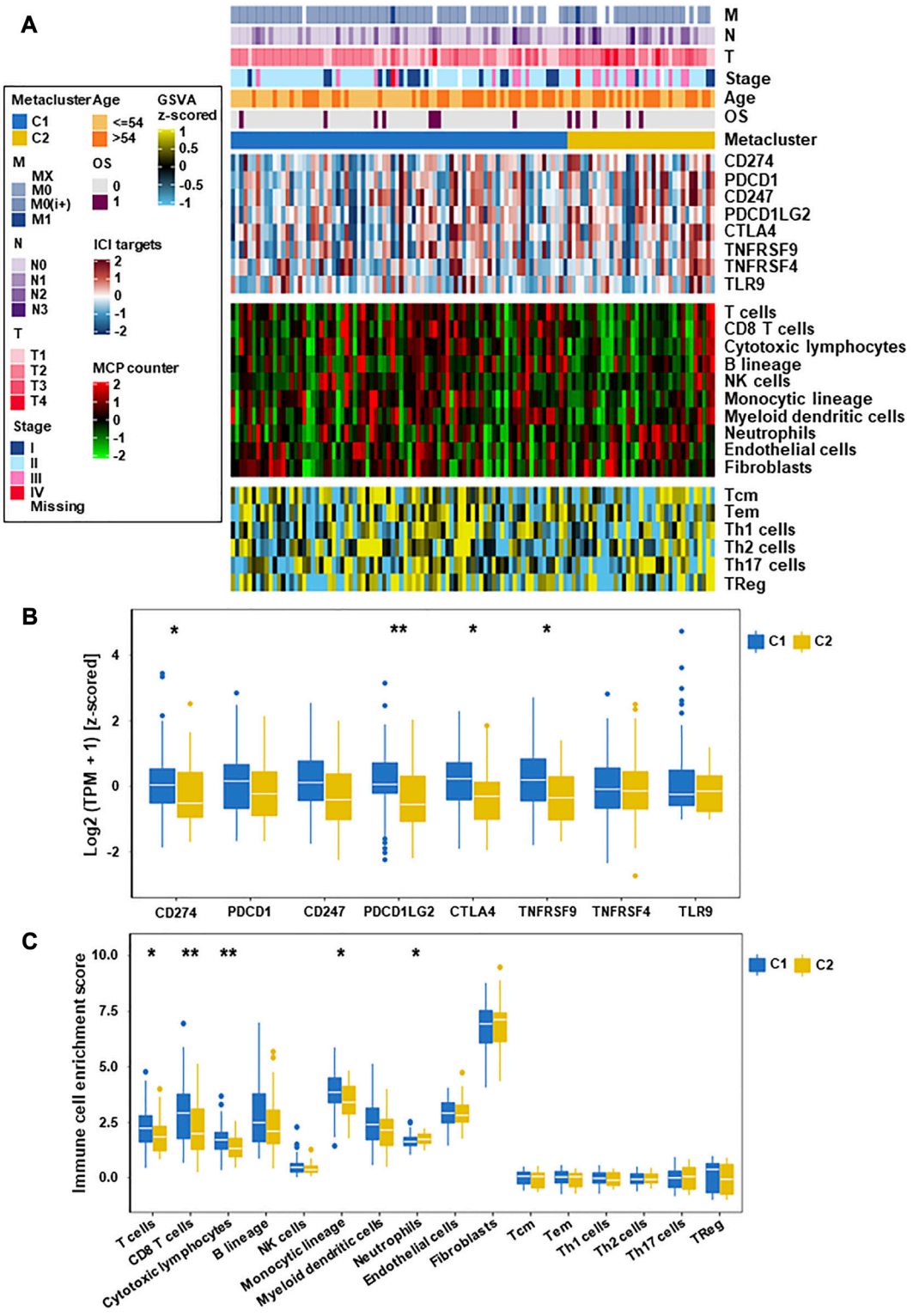


FIGURE 3
Immune characteristics of the two subtypes in the TCGA datasets. **(A)** Expression heatmap of immune cell and stromal cell populations in two TNBC subtypes. **(B)** Eight immune checkpoint genes in two TNBC subtypes. **(C)** Expression of different immune cells and stromal cells in C1 and C2 subtypes (*$p < 0.05$, **$p < 0.01$).

($p = 0.001$) were related to the C2 subtype (Supplementary Table S7). We also constructed a clinicopathological variables heatmap of subtypes in the validation cohorts and presented detailed data (Supplementary Figure S2; Supplementary Table S2). It is well known that larger tumor size and advanced TNM stage represent shorter survival in TNBC (Johansson et al., 2021).

10.3389/fphar.2023.1224828

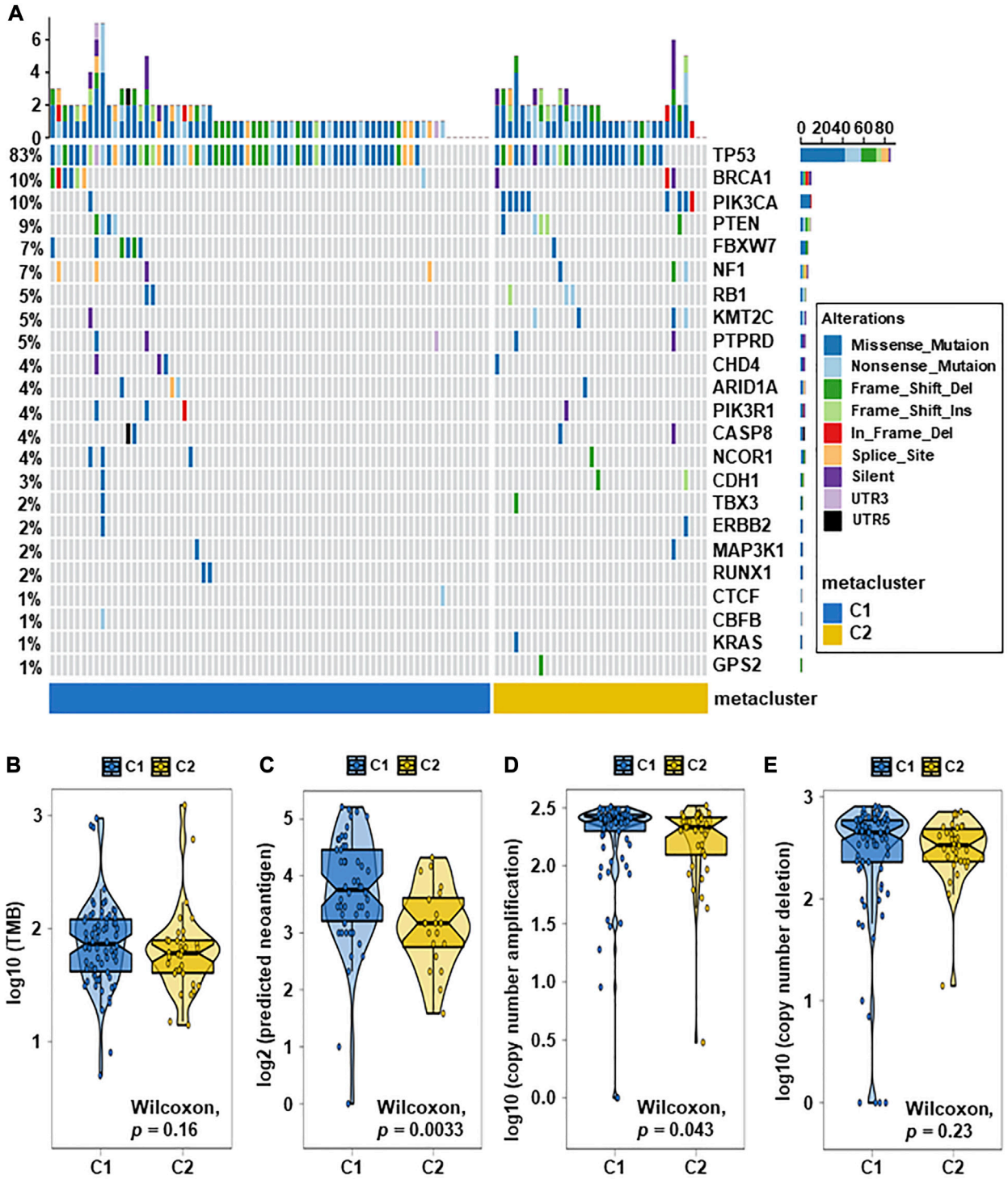


FIGURE 4
Relationship between TNBC subtypes and tumor mutation-related features. **(A)** Driver-type oncogenic mutations according to TCGA-TNBC typing with intragroup aggregation waterfall plots (see detailed statistical analysis in Supplementary Table S8). **(B)** Violin plots of gene mutations. There was a trend to show that the TMB of C1 was higher than that of C2, however there was no difference (p = 0.16). **(C)** Violin plots of predicted neoantigens. The quantity between two subtypes were significantly different (p = 0.0033). **(D,E)** Violin plots of copy number amplification and copy number deletion in TNBC subtypes. Patients within C1 only showed higher amplification than C2 (p = 0.043).

10.3389/fphar.2023.1224828

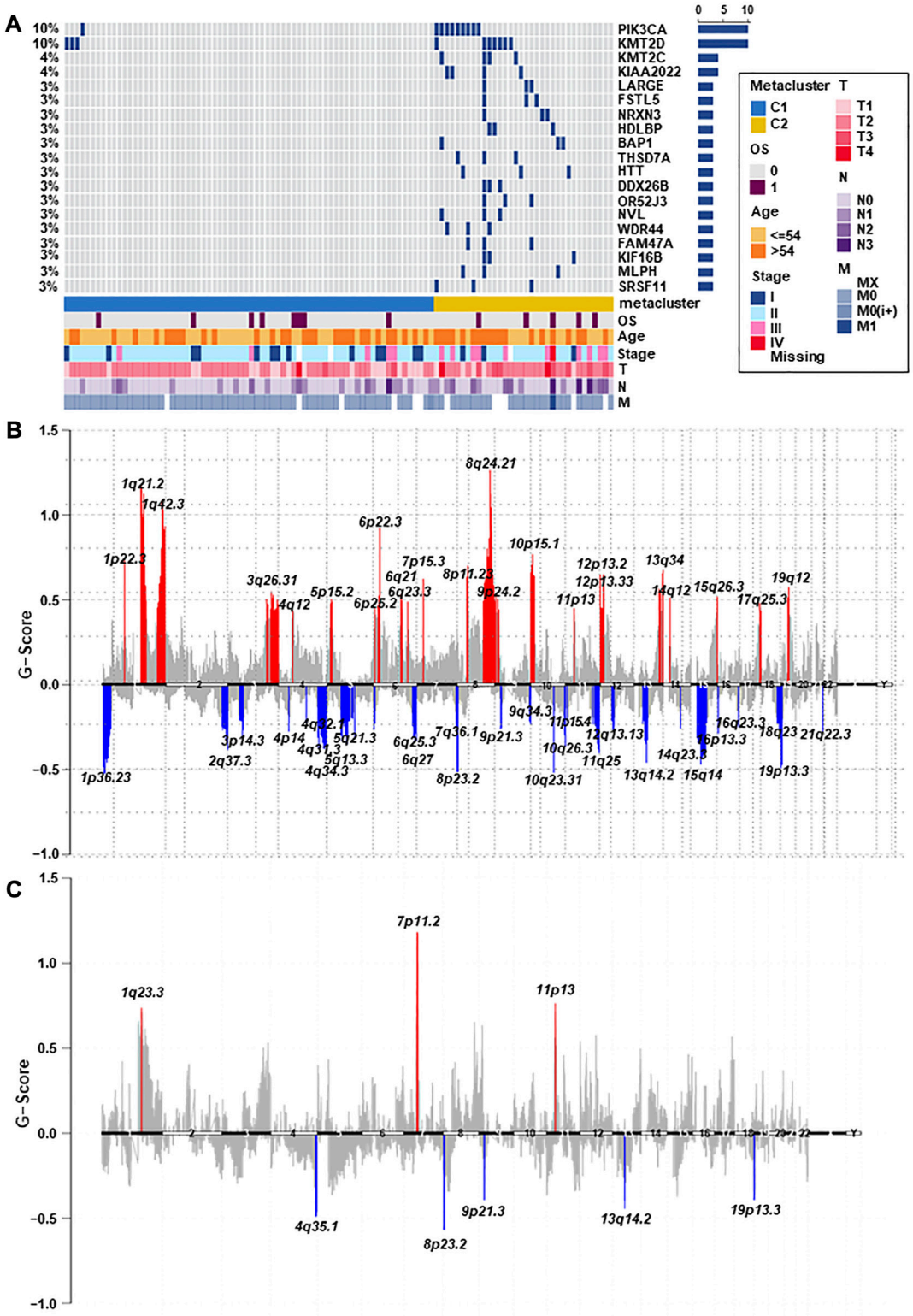


FIGURE 5
The landscape of somatic copy-number alterations in the two subtypes. **(A)** After adjusting $p < 0.05$, the genes with the most significant mutation frequencies between C1 and C2 groups were displayed. Specifically, C2 had a significantly higher mutation frequency of PIK3CA, KMT2D, KMT2C, and so on (see detailed statistical analysis in Supplementary Table S9). **(B,C)** Cytoband indicated differences in genomic copy-numbers between the two subtypes, with red representing amplification and blue representing deletion.

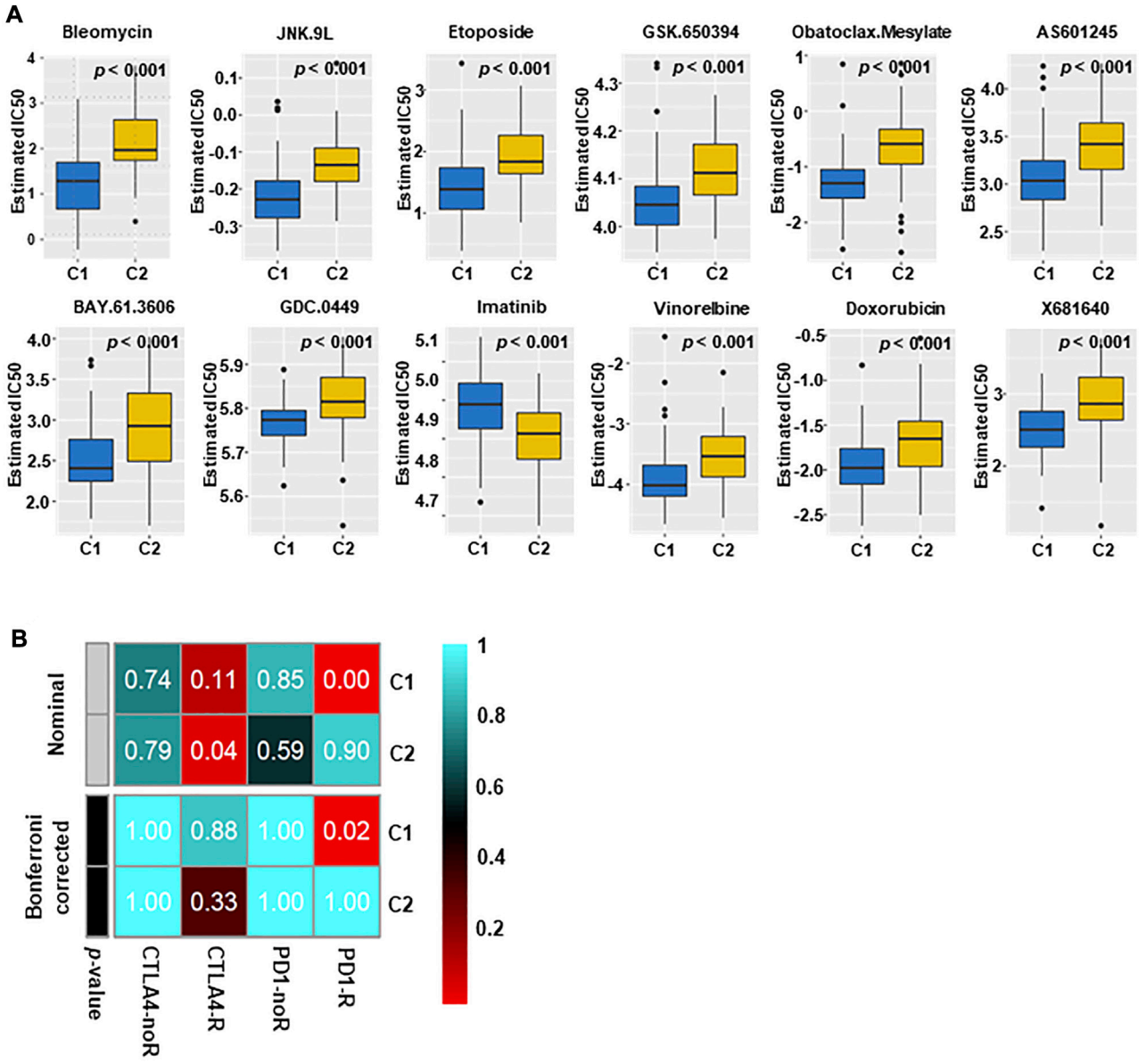


FIGURE 6
Immunotherapy and targeted therapy sensitivity of different subtypes. **(A)** The box plots of sensitivity to chemotherapy drugs in the two subtypes. The results indicated that C2 may not be sensitive to chemotherapy (All $p < 0.001$). **(B)** C1 may be more effective to PD1 inhibitors (Bonferroni correction, $p = 0.02$), and C2 may be more effective to CTLA4 inhibitors.

3.5 Association of TNBC subtypes with mutations and created heatmaps for visualization

Breast cancer has been closely related to many genomic mutations in the body (Kim et al., 2021). To investigate the difference of somatic mutations frequency between TNBC subtypes, we applied specific driver mutations for breast cancer (Bailey et al., 2018) to estimate gene mutations and draw waterfall map. High mutation frequencies of TP53, BRCA1, PIK3CA, PTEN, FBXW7, NF1, RB1, KMT2C, and PTPRD in both TNBC subtypes were observed (Figure 4A; Supplementary Table S8). We found that C2 exhibited different mutation characteristics from C1. Specifically, C2 has a higher mutation frequency, such as PIK3CA, KMT2D, KMT2C, and so on (Figure 5A; Supplementary Table S9). We calculated the TMB for each metabolic subtype (Figure 4B). Although there was no difference ($p = 0.16$), a trend showed that the TMB of C1 was higher than that of C2. We also analyzed the total number of mutations and expected neoantigens (Figure 4C) and observed a significant difference between them ($p = 0.0033$). Subsequently, the frequency of amplification (Figure 4D) and deletion (Figure 4E) was showed and found that patients within C1 only showed higher amplification than C2 ($p = 0.043$).

Finally, we mapped a cell column to change the number of copies of each group by performing online GISTIC2.0 analysis, in which red represented gains and blue represented losses (Figures 5B,

10.3389/fphar.2023.1224828

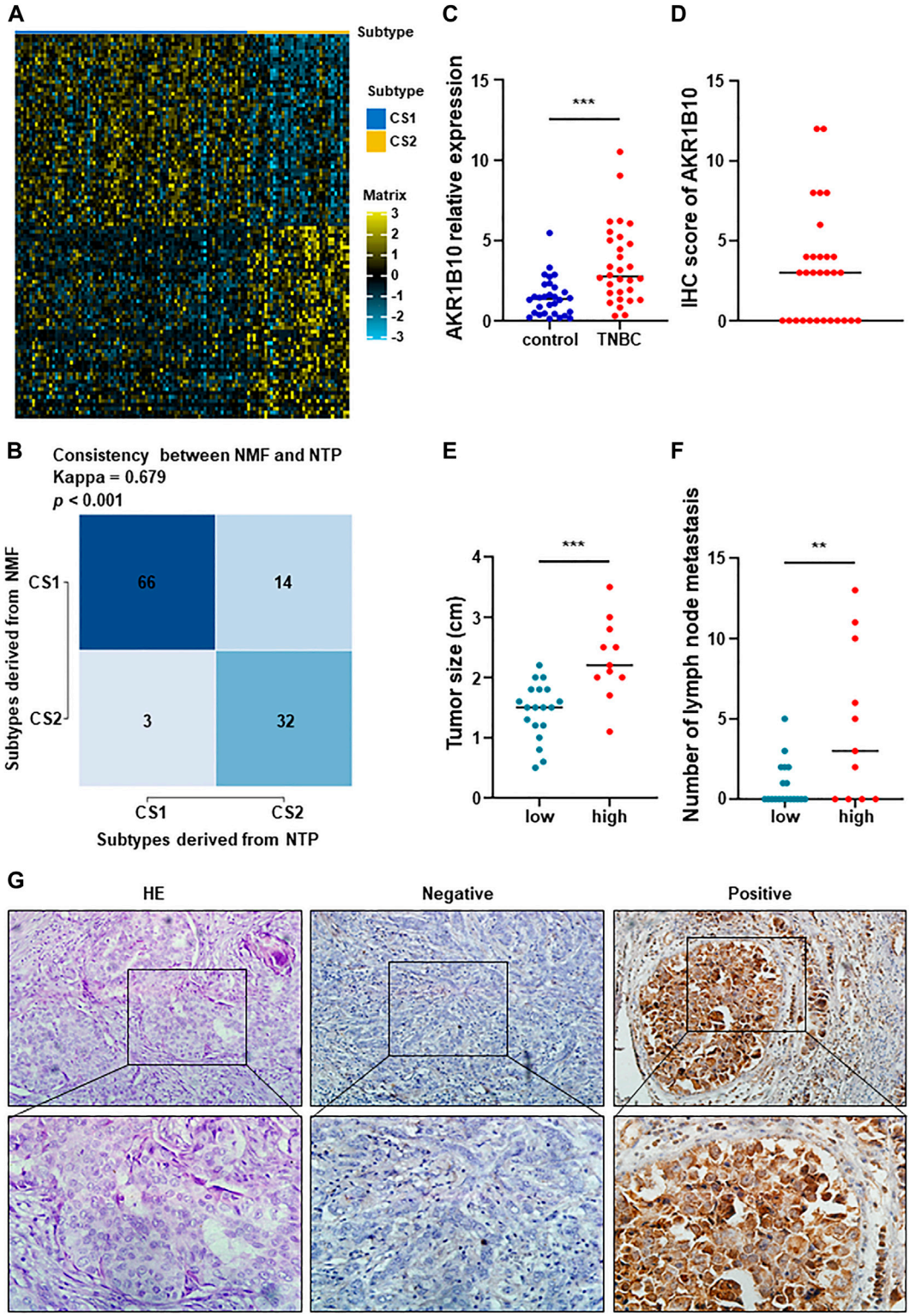


FIGURE 7
Performance validation of predictive metabolic-genes, and expression signature and preliminary validation of AKR1B10. **(A)** A 100-gene classifier was composed of the top 50 genes with significant differences in each TNBC subtype, and visualized by a heatmap. **(B)** Constructing a 100 gene classifier for identifying TNBC classification. **(C)** Expression of AKR1B10 was significantly increased in peripheral blood of TNBC patients (***$p < 0.0001$). **(D)** IHC score of AKR1B10. **(E,F)** The difference of tumor size and number of lymph node metastasis between AKR1B10 positive group and AKR1B10 negative group (**$p < 0.005$). **(G)** AKR1B10 was significantly overexpressed in part of TNBC tissues. From left to right, they were HE staining, negative and positive respectively (×200 in the upper section, ×400 in the lower section).

C). Both C1 and C2 observed copy number alterations in chromosome regions, including amplification at 11p13 and deletion at 8p23.2, 9p21.3, 13q14.2, and 19p13.3. In contrast to C1, C2 has significant amplification at 1q23.3 and 7p11.2. These differences could also explain that C2 has a better prognosis than C1. Therefore, changes in copy quantity might be the main mechanism behind the differences in metabolism and prognosis between the two group.

3.6 Specific sensitivity of TNBC subtypes for potential therapy

The difference in sensitivity to chemotherapeutic drugs and targeted drugs between two groups was analyzed by using the GDSC drug sensitivity database. The top 12 drugs with differential responses were plotted and listed (Figure 6A). After estimating the IC50 value, we found that C2 may be less sensitive to chemotherapy, including bleomycin, vinorelbine and doxorubicin (all $p < 0.001$).

The different immune infiltration patterns among TNBC subtypes suggested that further research on the response of immunotherapy was needed. To this aim, we matched the expression spectra of two subspecies to determine the similarity of the TCGA reaction spectra (Figure 6B). The results indicate that C1 may be more sensitive to PD1 treatment ($p = 0.02$), and C2 may have a better therapeutic effect on anti-CTLA4.

3.7 Performance validation of one hundred-gene classifier, and expression signature of Aldo-Keto reductase family 1 member B10 (AKR1B10)

And then, we extracted the top 50 genes of each metabolic specificity as biomarkers and constructed clinical models, and plotted correlation heatmaps using MOVICS (Bailey et al., 2018) package analysis. The classifier based on 100 genes was generated and visualized by heatmap (Figure 7A; Supplementary Table S10). In order to predict the identification of metabolic subtypes in each sample, we conducted consistency testing on the results of the two subtypes using the NTP algorithm and indicated that the characteristics of these genes can be replicated to determine the TNBC type (Figure 7B).

To better distinguish the two subtypes, we assume that AKR1B10 was an effective biomarker for C2. RT-qPCR and IHC staining were used to preliminarily verify this hypothesis. AKR1B10 were overexpressed in peripheral blood of TNBC patients than in healthy control (Figure 7C). IHC showed that AKR1B10 were positive in 11 cases of TNBC, with a positive rate of 36.7% (Figures 7D, G). The average tumor size of AKR1B10 positive group was 2.2 cm from 0.8 to 3.5 cm, which was higher than that of negative group (Figure 7E). In addition, the number of lymph node metastasis in AKR1B10 positive group were more than that in negative group (Figure 7F). Large tumor and many lymph node metastases often indicate poor prognosis of TNBC, which was consistent with C2 subtype. These results were listed in Supplementary Table S11.

4 Discussion

It is well known that the overall prognosis of TNBC is poor (Bianchini et al., 2016). With the increased understanding of metabolic reprogramming in breast cancer, traditional molecular characterization is no longer sufficient to fully elucidate tumor heterogeneity. As an important hallmark of tumors (Pavlova and Thompson, 2016; Pavlova et al., 2022), metabolic reprogramming may be beneficial to targeted therapy of TNBC. Recently, many TNBC classifications methods have been proposed, but a consensus on molecular taxonomy has not been reached. Thus, deeply exploring the metabolic characteristics and heterogeneity of TNBC is the key to providing provide precise treatment.

In this study, TNBC could be divided into two different metabolic related subtypes. Each subtype had different metabolic characteristics, prognoses, clinical features, tumor microenvironment characteristics, and so an. For C1, it was rich in immune signals and hardly involved in metabolic signals, gene expression was relatively high at immune monitoring points and scoring points. The increase of immunity and matrix indicated that these patients were allergic to drug allergy containing PD-1 inhibitor. In contrast, the C2 subtype displayed high variation in metabolism pathways involved in carbohydrate, lipid, and amino acid metabolism and a lack of immune signatures, with late pathological stage, weakened immunity and poor prognosis.

Our study indicated that C1 had abundant immune signatures and that C2 had overactivated metabolic related pathways. Considering the above results, we named C1 as the immune-related subtype and C2 as the metabolically active subtype. Subsequently, Patients in C2 had larger tumor size and later pathological stages, which implied that their overall prognosis were poor. The difference in metabolic characteristics and immune infiltration might be the important reason for the different prognoses of them. In this study, 17 associated metabolic pathways were significantly upregulated in C2, including pentose and glucuronate interconversions, oxidative phosphorylation, amino acid metabolism, steroid hormone and so on.

Previous studies have shown that glucose, amino acids and free fatty acids are important energy sources for tumor growth (Pavlova and Thompson, 2016), and metabolic disorders have a crucial impact on cancer (Micalizzi et al., 2021). Oxidative phosphorylation can promote distant metastasis and even induce chemotherapy resistance in TNBC (Davis et al., 2020; Evans et al., 2021). Tyrosine phosphorylation is an important mechanism for regulating signal transduction pathways and is also a common feature in oncogenic activation in cancer (Ostman et al., 2006; Taddei et al., 2020). Hence, the relationship between TNBC molecular subtypes may reveal the determining factors for TNBC metabolic differentiation classification.

Recently, more and more studies have confirmed that the tumor microenvironment (TME) plays an important role in the formation of breast cancer (Reis et al., 2018). Neoantigens can regulate the interaction between breast tumor cells and immune cells. This effect is presented by antigen-presenting cells (APCs) (Harbeck et al., 2019; Lhuillier et al., 2021). Although immune checkpoint inhibitors have achieved great advances in TNBC treatment, it is necessary to clearly distinguish which patients can benefit the most from this

treatment (Li et al., 2021). Therefore, we compared the response to immune checkpoints of two TNBC-subtypes to obtain the potential significance of immunotherapy. Our results showed that C1 was significantly superior to C2 in both immune cell infiltration and neoantigens, which indicated that C1 has a higher response to treatment targeting immune checkpoints. Due to the inconsistent in carcinogenic signaling pathways, C1 may benefit from RAS inhibitors and WNT inhibitors in the future, while C2 may benefit from targeting PI3K and anti-angiogenesis. A series of studies have shown that targeting RAS, WNT, and PIK3 signaling pathways and angiogenesis are potential strategies to enhance the efficacy of cancer therapy (Verret et al., 2019; Xu et al., 2020).

In order to identify the molecular driving factors between two groups, we noticed that C2 had significant mutation frequencies in PIK3CA, KMT2D, KMT2C, and so on. Notably, C2 was accompanied by special chromosome copy number alterations, such as amplification at 7p11.2 and deletion at 9p21.3 and 13q14.2. Amplification at chromosome 7p11.2 (EGFR) can promote the invasion and metastasis of breast tumors (Chen et al., 2022). EGFR was overexpressed in metaplastic breast cancer, and EGFR inhibitor was potential therapeutic agent for metaplastic breast cancer with 7p11.2 amplification (Reis-Filho et al., 2006). Patients with 9p21.3 deletion and concomitant PIK3CA mutation were prone to recurrence and distant metastasis (Bartels et al., 2018). The mutation frequency of PIK3CA is only second to TP53 (Pascual and Turner, 2019). The same characteristics were obtained in our study.

Multiple studies have clarified that PI3K inhibitors are beneficial in enhancing the sensitivity of PIK3CA mutant TNBC to CDK4/6 inhibitors (Asghar et al., 2017), and have a good effect on HR+ breast cancer carrying PIK3CA mutations (Di Leo et al., 2018), which indicates the potential of combined targeted therapy. In this study, C2 was not sensitive to a variety of chemotherapies and immunotherapies, with high PI3K mutations and amplification at 7p11.2 (EGFR), suggesting that these TNBC patients may receive good treatment outcomes after receiving PI3K inhibitors or EGFR inhibitors. Previous studies have shown that When PIK3CA mutates, the glutamate pyruvate transaminase 2 in colorectal cancer (CRC) cells is significantly upregulated, thereby affecting the reprogramming of glutamine metabolism (Hao et al., 2016). The metabolites of glutamine can be used not only to produce ATP, but also to synthesize certain macromolecules to promote tumor formation. For example, the ATP concentration and ATP/ADP ratio in PIK3CA mutant cells were higher. Mutations in PIK3CA in adipose tissue can lead cells to acquire many characteristic changes of cancer cells, such as increased glucose uptake, enhanced Warburg effect activity, and increased synthesis of oncogenic macromolecules (Ladraa et al., 2022). KMT2D mutations can significantly alter the biosynthesis of various metabolic products within cells, such as aerobic glycolysis and β-Oxidation, degradation, and uptake of lipids (Koutsioumpa et al., 2019). The above results indicate that these gene mutations can promote the differentiation of C2 subtypes by affecting metabolic reprogramming. Meanwhile, once these genes undergo mutations, they will further promote tumor progression by altering the activity of glucose and lipid metabolism in C2 patients. This may be the root cause of poor prognosis in C2 patients.

Our study had some limitations. First, bioinformatic analysis of metabolic and genomic alterations failed to pinpoint the precise cause of the difference in prognosis between the two subtypes. Second, the two subtypes classified according to immune and metabolic conditions need to be functionally validated further. Furthermore, the sensitivity of different subtypes of drugs must also be validated through clinical trials to explore the feasibility of translating these results into clinical practice. Finally, although it was preliminarily verified to identify the subtypes of TNBC, data from multiple centers and large samples will be needed to support this conclusion in the future.

5 Conclusion

In summary, this study revealed differences in TNBC metabolism and identified two subtypes. Subtype C1 was abundant in immune signatures but barely active in metabolic signatures, with higher gene expression at immune checkpoints and higher immune and matrix scores. This indicated that the C1 was allergic to PD-1 inhibitors. Subtype C2, on the other hand, had a high variation in metabolic pathways and a lack of immune signatures, as well as late pathological stage, low immune infiltration and poor prognosis. By dividing TNBC into two clusters, this study elucidated the reasons for the differences in prognosis of TNBC from the perspectives of metabolism and immune response. For the first time, we proved that C1 may be more sensitive to immunosuppressive drugs. RAS inhibitors and WNT inhibitors, whereas C2 may benefit from targeting PI3K and anti-angiogenesis. Furthermore, AKR1B10 based on the one hundred-gene classifier was a potential biomarker for identifying C2 subtypes. This provides a theoretical basis for further rationalizing TNBC subtypes to provide precise therapeutic strategies.

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 study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Tianjin Medical University Cancer Institute and Hospital (bc2022142). All subjects obtained informed consent before participating in the study.

Author contributions

LL, NW, and GZ provided ideas to the article and contributed to the verification and data analysis. LL, LG, and YZ contributed to clinical information collection and statistical analysis. XW and SW participated in the specimen collection and provided the experimental materials. XR and XZ contributed the experimental

platform, funding, and review of manuscript. XR, XZ, JL, and MG were involved in supervising this work, providing resources and revising the manuscript critically. JL and MG agreed to be accountable for all aspects of this work. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the National Natural Science Foundation of China (82103386 and 82172821); Tianjin Municipal Science and Technology Project (21JCZDJC00360); Beijing-Tianjin-Hebei Basic Research Cooperation Project (20JCZXJC00120); the Science and Technology Development Fund of Tianjin Education Commission for Higher Education (2021ZD033); Tianjin Medical Key Discipline (Specialty) Construction Project (TJYXZDXK-058B); Tianjin Key Medical Discipline (Specialty) Construction Project (TJYXZDXK-009A) and Tianjin Health Research Project (TJWJ2022XK024).

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/fphar.2023.1224828/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Heatmap of metabolic related characteristics in TCGA cohorts. Univariate Cox regression was used to identify metabolic related genes related to OS and visualized with heatmap.

SUPPLEMENTARY FIGURE S2
TNBC metabolic subtypes and tumor progression-related features in validation datasets. Heatmap of all metabolism-related genes were structured between two subtypes in the GSE25066 cohorts **(A)**, GSE21653 cohorts **(B)**, GSE103091 cohorts **(C)** and METABRIC cohorts **(D)** to verify the initial classification in TCGA cohorts. The results showed that the classification was consistent.

SUPPLEMENTARY FIGURE S3
Association between 17 metabolism-related signatures and the TNBC subtypes in validation datasets. There were 17 metabolism-related pathways in C2 that were significantly upregulated in GSE25066 cohorts **(A)**, GSE21653 cohorts **(B)**, GSE103091 cohorts **(C)** and METABRIC cohorts **(D)**.

SUPPLEMENTARY FIGURE S4
Validation of the differential activity of metabolic pathways between the two subtypes. The nearest template prediction (NTP) indicated two subtypes in TCGA-TNBC cohorts and validation cohorts, and demonstrated that C2 subtypes from five cohorts had stronger metabolic activity. Representing the two subtypes in the TCGA-TNBC cohorts **(A)**, GSE25066 cohorts **(B)**, GSE21653 cohorts **(C)**, GSE103091 cohorts **(D)** and METABRIC cohorts **(E)**.

References

Asghar, U. S., Barr, A. R., Cutts, R., Beaney, M., Babina, I., Sampath, D., et al. (2017). Single-cell dynamics determines response to CDK4/6 inhibition in triple-negative breast cancer. *Clin. Cancer Res.* 23 (18), 5561–5572. doi:10.1158/1078-0432.CCR-17-0369

Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., et al. (2018). Comprehensive characterization of cancer driver genes and mutations. *Cell.* 173 (2), 371–385.e18. doi:10.1016/j.cell.2018.02.060

Bareche, Y., Venet, D., Ignatiadis, M., Aftimos, P., Piccart, M., Rothe, F., et al. (2018). Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis. *Ann. Oncol.* 29 (4), 895–902. doi:10.1093/annonc/mdy024

Bartels, S., van Luttikhuizen, J. L., Christgen, M., Mägel, L., Luft, A., Hänzelmann, S., et al. (2018). CDKN2A loss and PIK3CA mutation in myoepithelial-like metaplastic breast cancer. *J. Pathol.* 245 (3), 373–383. doi:10.1002/path.5091

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17 (1), 218. doi:10.1186/s13059-016-1070-5

Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E., and Gianni, L. (2016). Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat. Rev. Clin. Oncol.* 13 (11), 674–690. doi:10.1038/nrclinonc.2016.66

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39 (4), 782–795. doi:10.1016/j.immuni.2013.10.003

Chen, H., Fan, S., Stone, J., Thompson, D. J., Douglas, J., Li, S., et al. (2022). Genome-wide and transcriptome-wide association studies of mammographic density phenotypes reveal novel loci. *Breast Cancer Res.* 24 (1), 27. doi:10.1186/s13058-022-01524-0

Cheng, L., Xiong, W., Li, S., Wang, G., Zhou, J., and Li, H. (2023). CRISPR-Cas9 screening identified lethal genes enriched in necroptosis pathway and of prognosis significance in osteosarcoma. *J. Gene Med.* 2023, e3563. doi:10.1002/jgm.3563

Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., et al. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31 (3), 213–219. doi:10.1038/nbt.2514

Davis, R. T., Blake, K., Ma, D., Gabra, M. B. I., Hernandez, G. A., Phung, A. T., et al. (2020). Transcriptional diversity and bioenergetic shift in human breast cancer metastasis revealed by single-cell RNA sequencing. *Nat. Cell. Biol.* 22 (3), 310–320. doi:10.1038/s41556-020-0477-0

Denkert, C., Liedtke, C., Tutt, A., and von Minckwitz, G. (2017). Molecular alterations in triple-negative breast cancer-the road to new treatment strategies. *Lancet* 389 (10087), 2430–2442. doi:10.1016/S0140-6736(16)32454-0

Di Leo, A., Johnston, S., Lee, K. S., Ciruelos, E., Lønning, P. E., Janni, W., et al. (2018). Buparlisib plus fulvestrant in postmenopausal women with hormone-receptor-positive, HER2-negative, advanced breast cancer progressing on or after mTOR inhibition (BELLE-3): A randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* 19 (1), 87–100. doi:10.1016/S1470-2045(17)30688-5

Evans, K. W., Yuca, E., Scott, S. S., Zhao, M., Paez Arango, N., Cruz Pico, C. X., et al. (2021). Oxidative phosphorylation is a metabolic vulnerability in chemotherapy-resistant triple-negative breast cancer. *Cancer Res.* 81 (21), 5572–5581. doi:10.1158/0008-5472.CAN-20-3242

Garrido-Castro, A. C., Lin, N. U., and Polyak, K. (2019). Insights into molecular classifications of triple-negative breast cancer: improving patient selection for treatment. *Cancer Discov.* 9 (2), 176–198. doi:10.1158/2159-8290.CD-18-1177

Goldhirsch, A., Winer, E. P., Coates, A. S., Gelber, R. D., Piccart-Gebhart, M., Thürlimann, B., et al. (2013). Personalizing the treatment of women with early breast cancer: highlights of the st gallen international expert consensus on the primary therapy of early breast cancer 2013. *Ann. Oncol.* 24 (9), 2206–2223. doi:10.1093/annonc/mdt303

Hao, Y., Samuels, Y., Li, Q., Krokowski, D., Guan, B. J., Wang, C., et al. (2016). Oncogenic PIK3CA mutations reprogram glutamine metabolism in colorectal cancer. *Nat. Commun.* 7, 11971. doi:10.1038/ncomms11971

Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., et al. (2019). Breast cancer. *Nat. Rev. Dis. Prim.* 5 (1), 66. doi:10.1038/s41572-019-0111-2

Hensley, C. T., Faubert, B., Yuan, Q., Lev-Cohain, N., Jin, E., Kim, J., et al. (2016). Metabolic heterogeneity in human lung tumors. *Cell.* 164 (4), 681–694. doi:10.1016/j.cell.2015.12.034

Jiang, Y. Z., Liu, Y., Xiao, Y., Hu, X., Jiang, L., Zuo, W. J., et al. (2021). Molecular subtyping and genomic profiling expand precision medicine in refractory metastatic triple-negative breast cancer: the FUTURE trial. *Cell. Res.* 31 (2), 178–186. doi:10.1038/s41422-020-0375-9

Jiang, Y. Z., Ma, D., Suo, C., Shi, J., Xue, M., Hu, X., et al. (2019). Genomic and transcriptomic landscape of triple-negative breast cancers: subtypes and treatment strategies. *Cancer Cell.* 35 (3), 428–440. doi:10.1016/j.ccell.2019.02.001

Johansson, A. L. V., Trewin, C. B., Fredriksson, I., Reinertsen, K. V., Russnes, H., and Ursin, G. (2021). In modern times, how important are breast cancer stage, grade and receptor subtype for survival: A population-based cohort study. *Breast Cancer Res.* 23 (1), 17. doi:10.1186/s13058-021-01393-z

Kim, J., and DeBerardinis, R. J. (2019). Mechanisms and implications of metabolic heterogeneity in cancer. *Cell. Metab.* 30 (3), 434–446. doi:10.1016/j.cmet.2019.08.013

Kim, M., Park, J., Bouhaddou, M., Kim, K., Rojc, A., Modak, M., et al. (2021). A protein interaction landscape of breast cancer. *Science* 374 (6563), eabf3066. doi:10.1126/science.abf3066

Koutsioumpa, M., Hatziapostolou, M., Polytarchou, C., Tolosa, E. J., Almada, L. L., Mahurkar-Joshi, S., et al. (2019). Lysine methyltransferase 2D regulates pancreatic carcinogenesis through metabolic reprogramming. *Gut* 68 (7), 1271–1286. doi:10.1136/gutjnl-2017-315690

Ladraa, S., Zerbib, L., Bayard, C., Fraissenon, A., Venot, Q., Morin, G., et al. (2022). PIK3CA gain-of-function mutation in adipose tissue induces metabolic reprogramming with Warburg-like effect and severe endocrine disruption. *Sci. Adv.* 8 (49), eade7823. doi:10.1126/sciadv.ade7823

Lehmann, B. D., Bauer, J. A., Chen, X., Sanders, M. E., Chakravarthy, A. B., Shyr, Y., et al. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* 121 (7), 2750–2767. doi:10.1172/JCI45014

Lhuillier, C., Rudqvist, N. P., Yamazaki, T., Zhang, T., Charpentier, M., Galluzzi, L., et al. (2021). Radiotherapy-exposed CD8+ and CD4+ neoantigens enhance tumor control. *J. Clin. Invest.* 131 (5), e138740. doi:10.1172/JCI138740

Li, H., Yu, L., Zhang, X., Shang, J., and Duan, X. (2022). Exploring the molecular mechanisms and shared gene signatures between rheumatoid arthritis and diffuse large B cell lymphoma. *Front. Immunol.* 13, 1036239. doi:10.3389/fimmu.2022.1036239

Li, H., Zhang, X., Shang, J., Feng, X., Yu, L., Fan, J., et al. (2023b). Identification of NETs-related biomarkers and molecular clusters in systemic lupus erythematosus. *Front. Immunol.* 14, 1150828. doi:10.3389/fimmu.2023.1150828

Li, H., Zhou, J., Zhou, L., Zhang, X., Shang, J., Feng, X., et al. (2023a). Identification of the shared gene signatures and molecular pathways in systemic lupus erythematosus and diffuse large B-cell lymphoma. *J. Gene Med.* 2023, e3558. doi:10.1002/jgm.3558

Li, Z., Sun, C., and Qin, Z. (2021). Metabolic reprogramming of cancer-associated fibroblasts and its effect on cancer cell reprogramming. *Theranostics* 11 (17), 8322–8336. doi:10.7150/thno.62378

Liu, J., Zhou, S., Li, S., Jiang, Y., Wan, Y., Ma, X., et al. (2019). Eleven genes associated with progression and prognosis of endometrial cancer (EC) identified by comprehensive bioinformatics analysis. *Cancer Cell. Int.* 19, 136. doi:10.1186/s12935-019-0859-1

Lu, X., Meng, J., Zhou, Y., Jiang, L., and Yan, F. (2021). Movics: an R package for multi-omics integration and visualization in cancer subtyping. *Bioinformatics* 36 (22-23), 5539–5541. doi:10.1093/bioinformatics/btaa1018

Martinez-Outschoorn, U. E., Peiris-Pagés, M., Pestell, R. G., Sotgia, F., and Lisanti, M. P. (2017). Cancer metabolism: A therapeutic perspective. *Nat. Rev. Clin. Oncol.* 14 (2), 11–31. doi:10.1038/nrclinonc.2016.60

Micalizzi, D. S., Ebright, R. Y., Haber, D. A., and Maheswaran, S. (2021). Translational regulation of cancer metastasis. *Cancer Res.* 81 (3), 517–524. doi:10.1158/0008-5472.CAN-20-2720

Ostman, A., Hellberg, C., and Böhmer, F. D. (2006). Protein-tyrosine phosphatases and cancer. *Nat. Rev. Cancer* 6 (4), 307–320. doi:10.1038/nrc1837

Pascual, J., and Turner, N. C. (2019). Targeting the PI3-kinase pathway in triple-negative breast cancer. *Ann. Oncol.* 30 (7), 1051–1060. doi:10.1093/annonc/mdz133

Pavlova, N. N., and Thompson, C. B. (2016). The emerging hallmarks of cancer metabolism. *Cell. Metab.* 23 (1), 27–47. doi:10.1016/j.cmet.2015.12.006

Pavlova, N. N., Zhu, J., and Thompson, C. B. (2022). The hallmarks of cancer metabolism: still emerging. *Cell. Metab.* 34 (3), 355–377. doi:10.1016/j.cmet.2022.01.007

Possemato, R., Marks, K. M., Shaul, Y. D., Pacold, M. E., Kim, D., Birsoy, K., et al. (2011). Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* 476 (7360), 346–350. doi:10.1038/nature10350

Reis, E. S., Mastellos, D. C., Ricklin, D., Mantovani, A., and Lambris, J. D. (2018). Complement in cancer: untangling an intricate relationship. *Nat. Rev. Immunol.* 18 (1), 5–18. doi:10.1038/nri.2017.97

Reis-Filho, J. S., Pinheiro, C., Lambros, M. B., Milanezi, F., Carvalho, S., Savage, K., et al. (2006). EGFR amplification and lack of activating mutations in metaplastic breast carcinomas. *J. Pathol.* 209 (4), 445–453. doi:10.1002/path.2004

Roh, W., Chen, P. L., Reuben, A., Spencer, C. N., Prieto, P. A., Miller, J. P., et al. (2017). Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci. Transl. Med.* 9 (379), eaah3560. doi:10.1126/scitranslmed.aah3560

Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell.* 160 (1-2), 48–61. doi:10.1016/j.cell.2014.12.033

Rosario, S. R., Long, M. D., Affronti, H. C., Rowsam, A. M., Eng, K. H., and Smiraglia, D. J. (2018). Pan-cancer analysis of transcriptional metabolic dysregulation using the Cancer Genome Atlas. *Nat. Commun.* 9 (1), 5330. doi:10.1038/s41467-018-07232-8

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W. K., Luna, A., La, K. C., et al. (2018). Oncogenic signaling pathways in the cancer genome atlas. *Cell.* 173 (2), 321–337.e10. doi:10.1016/j.cell.2018.03.035

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Taddei, M. L., Pardella, E., Pranzini, E., Raugei, G., and Paoli, P. (2020). Role of tyrosine phosphorylation in modulating cancer cell metabolism. *Biochim. Biophys. Acta Rev. Cancer* 1874 (2), 188442. doi:10.1016/j.bbcan.2020.188442

Tu, W., Tu, Y., Tan, C., Zhong, H., Xu, S., Wang, J., et al. (2023). Elucidating the role of T-cell exhaustion-related genes in colorectal cancer: A single-cell bioinformatics perspective. *Funct. Integr. Genomics* 23 (3), 259. doi:10.1007/s10142-023-01188-9

Verret, B., Cortes, J., Bachelot, T., Andre, F., and Arnedos, M. (2019). Efficacy of PI3K inhibitors in advanced breast cancer. *Ann. Oncol.* 30 (10), x12–x20. doi:10.1093/annonc/mdz381

Waks, A. G., and Winer, E. P. (2019). Breast cancer treatment: A review. *Jama* 321 (3), 288–300. doi:10.1001/jama.2018.19323

Xia, L., Oyang, L., Lin, J., Tan, S., Han, Y., Wu, N., et al. (2021). The cancer metabolic reprogramming and immune response. *Mol. Cancer* 20 (1), 28. doi:10.1186/s12943-021-01316-8

Xiao, Y., Ma, D., Yang, Y. S., Yang, F., Ding, J. H., Gong, Y., et al. (2022). Comprehensive metabolomics expands precision medicine for triple-negative breast cancer. *Cell. Res.* 32 (5), 477–490. doi:10.1038/s41422-022-00614-0

Xu, X., Zhang, M., Xu, F., and Jiang, S. (2020). Wnt signaling in breast cancer: biological mechanisms, challenges and opportunities. *Mol. Cancer* 19 (1), 165. doi:10.1186/s12943-020-01276-5

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

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 16 (5), 284–287. doi:10.1089/omi.2011.0118

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 07 September 2023
DOI 10.3389/fphar.2023.1244752

OPEN ACCESS

EDITED BY
Linhui Wang,
Second Military Medical University, China

REVIEWED BY
Wenhong Deng,
Renmin Hospital of Wuhan University, China
Feng Zhu,
Jincheng People's Hospital, China
Qingshan Chen,
Shanghai First Maternity and Infant Hospital, China

*CORRESPONDENCE
Chongchan Bao,
903773508@qq.com
Yunfei Liu,
yunfei.liu@med.uni-muenchen.de
Zhizhou Wang,
wzz901102@126.com

†These authors have contributed equally to this work

RECEIVED 23 June 2023
ACCEPTED 23 August 2023
PUBLISHED 07 September 2023

CITATION
Zhang B, Liu J, Li H, Huang B, Zhang B, Song B, Bao C, Liu Y and Wang Z (2023), Integrated multi-omics identified the novel intratumor microbiome-derived subtypes and signature to predict the outcome, tumor microenvironment heterogeneity, and immunotherapy response for pancreatic cancer patients. *Front. Pharmacol.* 14:1244752. doi: 10.3389/fphar.2023.1244752

Integrated multi-omics identified the novel intratumor microbiome-derived subtypes and signature to predict the outcome, tumor microenvironment heterogeneity, and immunotherapy response for pancreatic cancer patients

Biao Zhang[1†], Jifeng Liu[1†], Han Li[2†], Bingqian Huang[1,3], Bolin Zhang[4], Binyu Song[5], Chongchan Bao[6*], Yunfei Liu[7*] and Zhizhou Wang[1*]

[1]Department of General Surgery, The First Affiliated Hospital of Dalian Medical University, Dalian, China, [2]Department of Oncology, Southwest Medical University, Luzhou, China, [3]Institute (College) of Integrative Medicine, Dalian Medical University, Dalian, China, [4]Department of Visceral, Martin-Luther-University Halle-Wittenberg, University Medical Center Halle, Halle, Germany, [5]Department of Plastic Surgery, Xijing Hospital, Xi'an, China, [6]Department of Breast and Thyroid Surgery, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, China, [7]Department of General, Visceral, and Transplant Surgery, Ludwig-Maximilians-University Munich, Munich, Germany

Background: The extremely malignant tumour known as pancreatic cancer (PC) lacks efficient prognostic markers and treatment strategies. The microbiome is crucial to how cancer develops and responds to treatment. Our study was conducted in order to better understand how PC patients' microbiomes influence their outcome, tumour microenvironment, and responsiveness to immunotherapy.

Methods: We integrated transcriptome and microbiome data of PC and used univariable Cox regression and Kaplan–Meier method for screening the prognostic microbes. Then intratumor microbiome-derived subtypes were identified using consensus clustering. We utilized LASSO and Cox regression to build the microbe-related model for predicting the prognosis of PC, and utilized eight algorithms to assess the immune microenvironment feature. The OncoPredict package was utilized to predict drug treatment response. We utilized qRT-PCR to verify gene expression and single-cell analysis to reveal the composition of PC tumour microenvironment.

Results: We obtained a total of 26 prognostic genera in PC. And PC samples were divided into two microbiome-related subtypes: Mcluster A and B. Compared with Mcluster A, patients in Mcluster B had a worse prognosis and higher TNM stage and pathological grade. Immune analysis revealed that neutrophils, regulatory T cell, $CD8^+$ T cell, macrophages M1 and M2, cancer associated fibroblasts, myeloid dendritic cell, and activated mast cell had remarkably higher infiltrated levels within the tumour microenvironment of Mcluster B. Patients in Mcluster A

were more likely to benefit from CTLA-4 blockers and were highly sensitive to 5-fluorouracil, cisplatin, gemcitabine, irinotecan, oxaliplatin, and epirubicin. Moreover, we built a microbe-derived model to assess the outcome. The ROC curves showed that the microbe-related model has good predictive performance. The expression of LAMA3 and LIPH was markedly increased within pancreatic tumour tissues and was linked to advanced stage and poor prognosis. Single-cell analysis indicated that besides cancer cells, the tumour microenvironment of PC was also rich in monocytes/macrophages, endothelial cells, and fibroblasts. LIPH and LAMA3 exhibited relatively higher expression in cancer cells and neutrophils.

Conclusion: The intratumor microbiome-derived subtypes and signature in PC were first established, and our study provided novel perspectives on PC prognostic indicators and treatment options.

KEYWORDS

microbiome, pancreatic cancer, prognosis, tumour microenvironment, immunotherapy, single-cell analysis

1 Introduction

An extremely dangerous tumour of the digestive tract, pancreatic cancer (PC) has a sneaky onset and quick progression. Clinical practise for PC lacks efficient therapeutic medications, and the prognosis is extremely poor (Chi et al., 2022a; Zhang B. et al., 2023). With over 459,000 new cases and 432,000 fatalities per year, epidemiological studies have shown that PC is the seventh greatest cause of cancer-related deaths globally (Ryan et al., 2014; Bray et al., 2018). Since most PCs are not discovered until they are advanced, the probability of surviving more than 5 years is low, at only 9% (Mizrahi et al., 2020; Siegel et al., 2020). Surgery combined with adjuvant chemotherapy is currently the standard treatment for PC. However, due to the complexity and heterogeneity of the tumour microenvironment of PC, it often leads to the generation of treatment resistance and the differential response of different patients to treatment. Therefore, the development of indicators for early detection as well as risk assessment is an important clinical problem to be solved urgently. So far, many studies have developed biomarkers for early diagnosis and risk assessment of PC from the perspectives of subcellular organelle function, tumour immune response, and gene modification (Romero et al., 2020; Xiao et al., 2022; Zhuo et al., 2022). Nevertheless, clinically effective early diagnostic markers, therapeutic targets, and risk assessment schemes in PC are still lacking.

Recent studies have shown that microbes have been considered to influence the occurrence, development, metastasis, as well as therapy response of different tumours, especially closely related to tumour microenvironment and immune response (Sepich-Poore et al., 2021). Many basic researches revealed that intratumor microbiome could affect the progression, metastases, prognosis, as well as immunotherapy of cancer patients by regulating oxidative stress, Toll-like receptors-mediated immune response, and tumour cell metabolism, involving mTOR, STAT3, Wnt, MAPK and other signaling pathways (Pushalkar et al., 2018; Wang et al., 2019; Wei et al., 2019). The diversity and composition of microbiome play crucial functions in the prognosis of PC, and can regulate the tumour immune microenvironment (Riquelme et al., 2019; Kartal et al., 2022). Mao et al. (Mao et al., 2022) have constructed the intratumor microbiome signature for breast cancer to predict the outcome. However, the correlation of intratumor microbiome with the clinicopathological features, prognosis, tumour microenvironment heterogeneity, and therapeutic response in PC is still not reported.

Our study first constructed intratumor microbiome-derived subtypes for PC by integrating microbiome and transcriptome data, and comprehensively analyzed the important role of microbiome in clinicopathological characteristics, prognosis, tumour immune microenvironment, and immunotherapy response of PC patients. Meanwhile, we also identified the microbiome-related differentially expressed genes and utilized them to build a prognostic model. Subsequently, we verified the LIPH and LAMA3 mRNA expression by real-time quantitative PCR (qRT-PCR). Finally, we used single-cell analysis to further reveal the cell subpopulation composition in pancreatic tumour microenvironment, as well as the relative expression of LIPH and LAMA3 in different cell subpopulations. This study can provide innovative ideas for the outcome assessment as well as therapy of PC.

2 Materials and methods

2.1 Data acquisition

Transcriptome data (containing 178 PC tissues and 4 paracancerous tissues) of PC, along with clinical data (containing 185 PC samples), were downloaded via The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/). Microbiome data of PC were obtained via the cBioPortal platform (https://www.cbioportal.org/) (Cerami et al., 2012; Mao et al., 2022). Gene expression data and survival information for GSE62452 (containing 69 PC tissues), GSE28735 (containing 45 PC tissues and paracancerous tissues), and GSE57495 (containing 69 PC tissues and paracancerous tissues) datasets were obtained via the Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/). Using "sva" R package to eliminate batch effects between different datasets (Chi et al., 2023a). Microbiomes associated with PC prognosis were identified by univariable Cox regression analysis as well as Kaplan-Meier (KM) method (Chi et al.,

2023b). These microbes associated with PC prognosis were used in subsequent analyses.

2.2 Clustering analysis

Consensus cluster was carried out utilizing "ConsensusClusterPlus" R package for PC samples based on the relative abundance of prognosis-related microbes (Zhang X. et al., 2023). The ideal clustering number was established based on the cumulative distribution function (CDF) curve as well as variations of CDF curve area. Using principal component analysis (PCA) as well as t-distributed stochastic neighbor embedding (t-SNE) analyses to demonstrate the accuracy of our clustering findings. Then the prognosis and clinical pathological features between different subtypes were further compared. Additionally, the differences in the relative abundance of prognosis-associated microbes among different subtypes were compared.

2.3 Gene set enrichment analysis (GSEA)

Gene set enrichment analysis (GSEA) was performed using the gene sets "c2.cp.kegg.v7.5.1.symbols.gmt" as well as "c5.go.v7.5.1.symbols.gmt" to compare the putative mechanisms behind the variations between the different intratumor microbiome subtypes (Subramanian et al., 2005). GSEA enrichment analysis was conducted using the R packages "limma", "org.Hs.e.g.,.db", "clusterProfiler", and "enrichplot". An adjusted p-value <0.05 were considered statistically significant.

2.4 Immune analysis

To analyze the differences in tumour immune microenvironment among different intratumor microbiome subtypes, the stromal, immune, and ESTIMATE score of every PC patient were evaluated utilizing "ESTIMATE" algorithm (Yoshihara et al., 2013). For evaluating the variations in infiltrated levels of immune cell subset between different subtypes, the infiltration scores calculated utilizing eight algorithms were downloaded from the Tumor Immune Estimation Resource database (TIMER, http://timer.cistrome.org/) (Yuan et al., 2022). The Cancer Immunome Atlas (TCIA, https://tcia.at/) is a database developed based on the TCGA database, which analyzes the tumour immune microenvironment and tumor antigen genes in 20 solid tumours (Charoentong et al., 2017). The immunophenoscore (IPS) of PC patients from TCIA database were downloaded. Then the differences in the responsiveness to cytotoxic T lymphocyte antigen-4 (CTLA-4) and programmed cell death protein 1 (PD-1) blockers between the different intratumor microbiome subtypes were further analyzed.

2.5 Drug sensitivity analysis

OncoPredict is an R package created via Maeser et al., which is used to predict drug response and biomarkers *in vivo* or in cancer patients based on cell line screening data (Maeser et al., 2021). OncoPredict was employed to assess the variations of drug sensitivity between the patients with different intratumor microbiome subtypes.

2.6 Differential expression analysis

In order to further analyze the differences between intratumor microbiome subtypes, we utilized "limma" package to find differentially expressed genes (DEGs), which were defined as intratumor microbe-related DEGs (Chi et al., 2022b). The filtering criteria were $|\log_2 FC| > 1$, the adjusted p-value <0.05. Additionally, PC patients of TCGA, GSE28735, GSE62452, and GSE57495 datasets were merged for identifying DEGs between PC tissues and paracancerous tissues. The DEGs obtained from both approaches were then combined, then we utilized Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses to assess the biological processes and functions they participate in (Ashburner et al., 2000; Kanehisa and Goto, 2000). For assessing the prognostic value of these DEGs in PC, consensus clustering was carried out. Survival times of the various subtypes were compared using KM curve.

2.7 Construction of the prognostic signature

For accurately assessing the prognosis of individual PC samples, we employed LASSO regression and Cox regression analysis to build a prognostic signature utilizing intratumor microbe-related DEGs. Samples from TCGA dataset were grouped into a training cohort as well as an internal validation cohort utilizing "caret" package in a 5: 5 ratio, while samples from GEO dataset were utilized as an external validation cohort. The score for every sample could be computed using risk score formula. Additionally, a comparison with the median score of the training cohort was used to categorise each sample into high- or low-risk score categories or groups. Utilizing KM curve to compare the prognosis between different risk categories. The performance of the signature was appraised by plotting time-dependent receiver operating characteristic (ROC) curve as well as figuring out the area under the curve (AUC).

2.8 Correlation of clinicopathological features, independent prognostic analysis, and construction of nomogram prediction model

Our study combined the clinicopathological information of PC samples with the risk scores and grouped them based on clinicopathological characteristics. Risk scores between different categories were compared utilizing Wilcoxon signed-rank test and Kruskal–Wallis rank sum test. Utilizing univariable as well as multivariable Cox regression to identify the independent prognostic factor of PC. Subsequently, the clinicopathological features and risk score were utilized for building a nomogram prediction model by "rms" R package (Park, 2018). Utilizing calibration curve to appraise the predictive accuracy of the nomogram.

2.9 Expression analysis of model genes and single-cell analysis

The GEPIA platform (http://gepia.cancer-pku.cn/) can allow for gene differential expression as well as survival analysis utilizing data of TCGA and GTEx databases (Tang et al., 2017). We utilized it to examine the expressed variations in LAMA3 and LIPH at RNA level between pancreatic cancer and normal tissues as well as the prognostic significance of LAMA3 and LIPH. Human Protein Atlas database (HPA, version22, https://www.proteinatlas.org/) aims at creating expressed patterns in protein of cells as well as tissues (Pontén et al., 2011). We can download immunohistochemistry images of pancreatic cancer and normal tissues via HPA platform. The tumor immune single-cell hub database (TISCH, http://tisch.comp-genomics.org) as a single-cell RNA-seq platform, focuses on the tumour microenvironment. It was utilized to reveal the composition of various cell subpopulations in the pancreatic tumour microenvironment and the relative expressed level of LAMA3 and LIPH within various cell subpopulations (Sun et al., 2021).

2.10 Real-time quantitative PCR

RNAs were extracted from cell lines, including a normal pancreatic epithelial cell line (HPDE6-C7) as well as three PC cell lines (CF-PAC1, PANC-1, and BxPC-3). The cDNAs were prepared using Reverse Transcription Reagent. Subsequently, PCR was performed. GAPDH served as the reference standard. Utilizing the ΔΔCt method to illustrate the relative expressed level of LAMA3 and LIPH. The primer sequences for human genes, including LAMA3 (Forward: 5′-ATTGAATTGAGCACCAGC GATAGC-3′, Reverse: 5′-CGATGAGAAGCCGTAGTCCAGAG-3′) as well as LIPH (Forward: 5′-TACGGGACTAAATGTGAG GC-3′, Reverse: 5′-CCTAGACTTACTCCGATCATG-3′).

2.11 Data analysis

Data analysis was performed utilizing R (Version 4.1.2) as well as GraphPad Prism 9. For normally distributed quantitative data, utilizing *t*-test to compare the differences. For non-normally distributed quantitative data, utilizing Wilcoxon signed-rank test to compare the differences between two groups, and utilizing Kruskal–Wallis rank sum test to compare the differences among multiple groups. KM curve was utilized to compare the prognosis between different subtypes or categories. The *p*-value <0.05 represented remarkable significance.

3 Results

3.1 Identification of intratumor microbiome-derived subtypes

The workflow of our research was depicted in Figure 1. Totally 1406 genera were obtained from the pancreatic tumour microenvironment through the cBioPortal platform (Supplementary Table S1). Univariate Cox regression analysis identified 63 genera associated with the prognosis of PC, with 24 genera associated with a favorable prognosis and 39 genera associated with a poor prognosis (Figure 2A). KM method identified 44 genera related to PC patients' prognosis (Supplementary Table S2). The intersection of genera obtained from univariate Cox regression analysis and Kaplan-Meier analysis yielded 26 genera: Alpharetrovirus, Azohydromonas, *Bacteroides*, Carlavirus, *Chlamydia*, Derxia, Domibacillus, Francisella, Gemmatimonas, Halothermothrix, Histophilus, Holospora, Hylemonella, Indibacter, Mesoplasma, Natronolimnobius, Paucibacter, Pseudarthrobacter, Puniceibacterium, Riemerella, Ruegeria, Runella, Silanimonas, Starkeya, Vagococcus, and Xanthobacter (Figure 2B). Correlation analysis revealed complicated relationships among the 26 genera. For example, Vagococcus had positive correlations with Puniceibacterium, Halothermothrix, Derxia, Starkeya, Pseudarthrobacter, Domibacillus, Gemmatimonas, and Silanimonas, while had negative correlations with Ruegeria, *Chlamydia*, Francisella, Carlavirus, and Alpharetrovirus (Supplementary Figure S1). Subsequently, consensus clustering was carried out utilizing the abundance of 26 genera. The CDF curve as well as the area variation under the curve were depicted in Supplementary Figures S2, S3, which indicated the ideal k value was 2. The consensus matrix at $k = 2$ was shown in Figure 2C. PC patients were divided into two intratumor microbiome-derived subtypes: Mcluster A and Mcluster B. PCA as well as t-SNE can clearly differentiate samples of Mcluster A and Mcluster B (Figures 2D, E). Survival analysis indicated a significantly better prognosis for Mcluster A compared to Mcluster B (Figure 2F). Compared to Mcluster B, a higher proportion of T1-2 stage, N0 stage, M0 stage, Stage I, and pathological grade G1 was observed in Mcluster A (Figures 2G–K). Furthermore, the abundance differences of 26 genera in different subtypes were analyzed. The results showed that Azohydromonas, Derxia, Holospora, Hylemonella, Paucibacter, Silanimonas, Starkeya, and Xanthobacter had remarkably higher abundance in Mcluster A, while Alpharetrovirus, Indibacter, Riemerella, and Ruegeria had remarkably higher abundance in Mcluster B (Figure 2L).

3.2 Gene set enrichment analysis

For investigating putative molecular mechanisms between different intratumor microbiome-derived subtypes, we performed GSEA analysis. In "c5.go.v7.5.1.symbols.gmt" gene set, we found that Mcluster A was mainly enriched in cell body, presynapse, gated channel activity, as well as potassium channel activity (Figure 3A). And Mcluster B was significantly more abundant in cytokine-mediated signaling pathway, immune response-regulating signaling pathway, immunoglobulin production, as well as membrane invagination (Figure 3B). In "c2.cp.kegg.v7.5.1.symbols.gmt" gene set, Mcluster A was found to be enriched in pathways containing oxidative phosphorylation, Parkinson's disease, ribosome, and steroid hormone biosynthesis (Figure 3C). Conversely, Mcluster B was enriched in pathways including cytokine-cytokine receptor interaction, ECM-receptor interaction, focal adhesion, JAK-STAT signaling pathway, etc (Figure 3D).

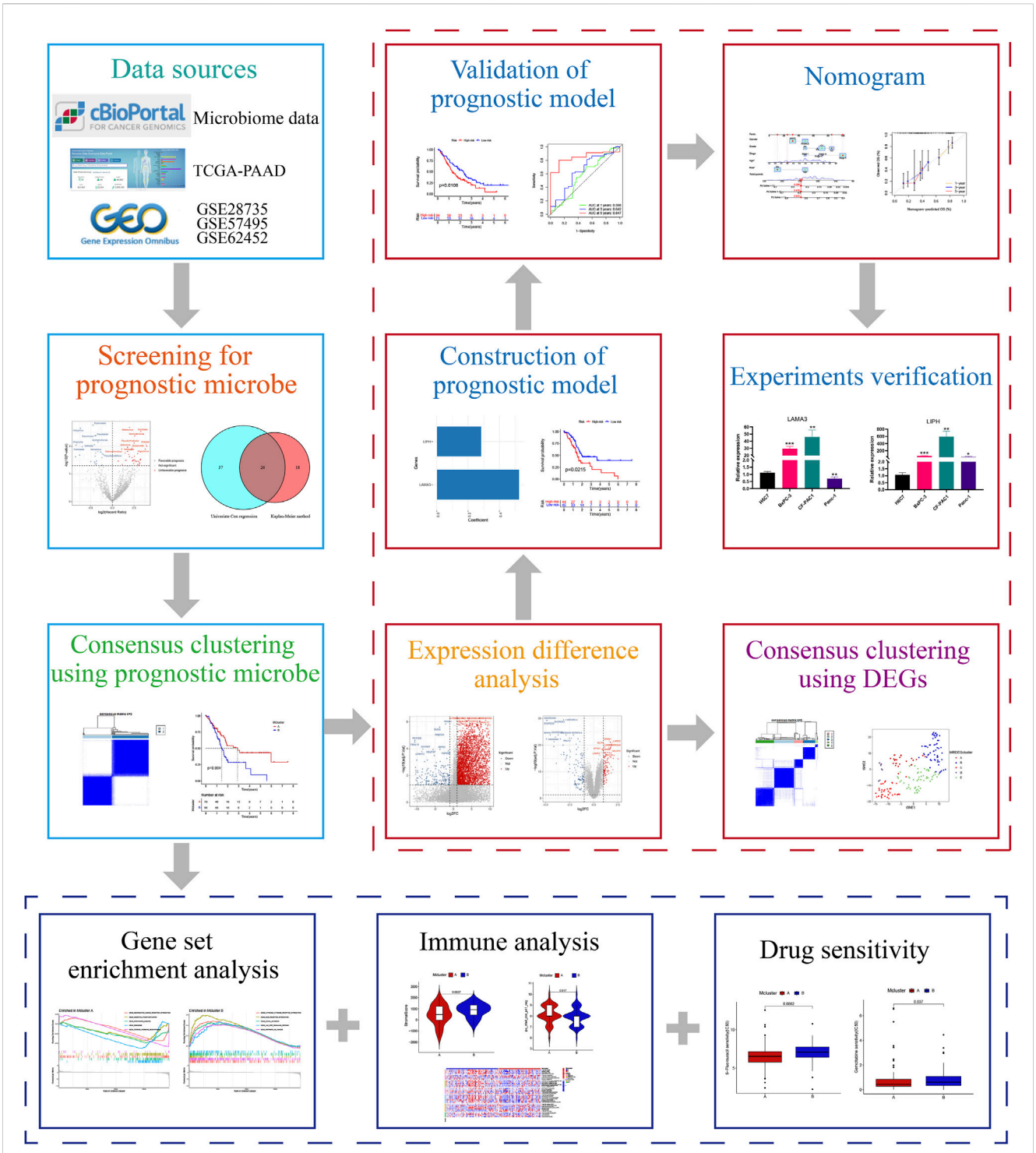


FIGURE 1
The study's general procedure.

3.3 Immune analysis

We evaluated the variations in immune infiltration between distinct subtypes using various algorithms to investigate the association between intratumor microbiome-derived subtypes and tumour microenvironment. The "ESTIMATE" algorithm showed that Mcluster B had higher stroma as well as ESTIMATE score (Figures 4A, C), whereas there was no discernible difference in immunological score between Mcluster A and Mcluster B (Figure 4B). For investigating the infiltrated variations in immune cell subpopulations between different subtypes, the infiltrated scores of immune cell subpopulations were obtained from the TIMER database. We found that neutrophils, regulatory T cell (Treg), $CD8^+$ T cell, macrophages M1 and M2, cancer-

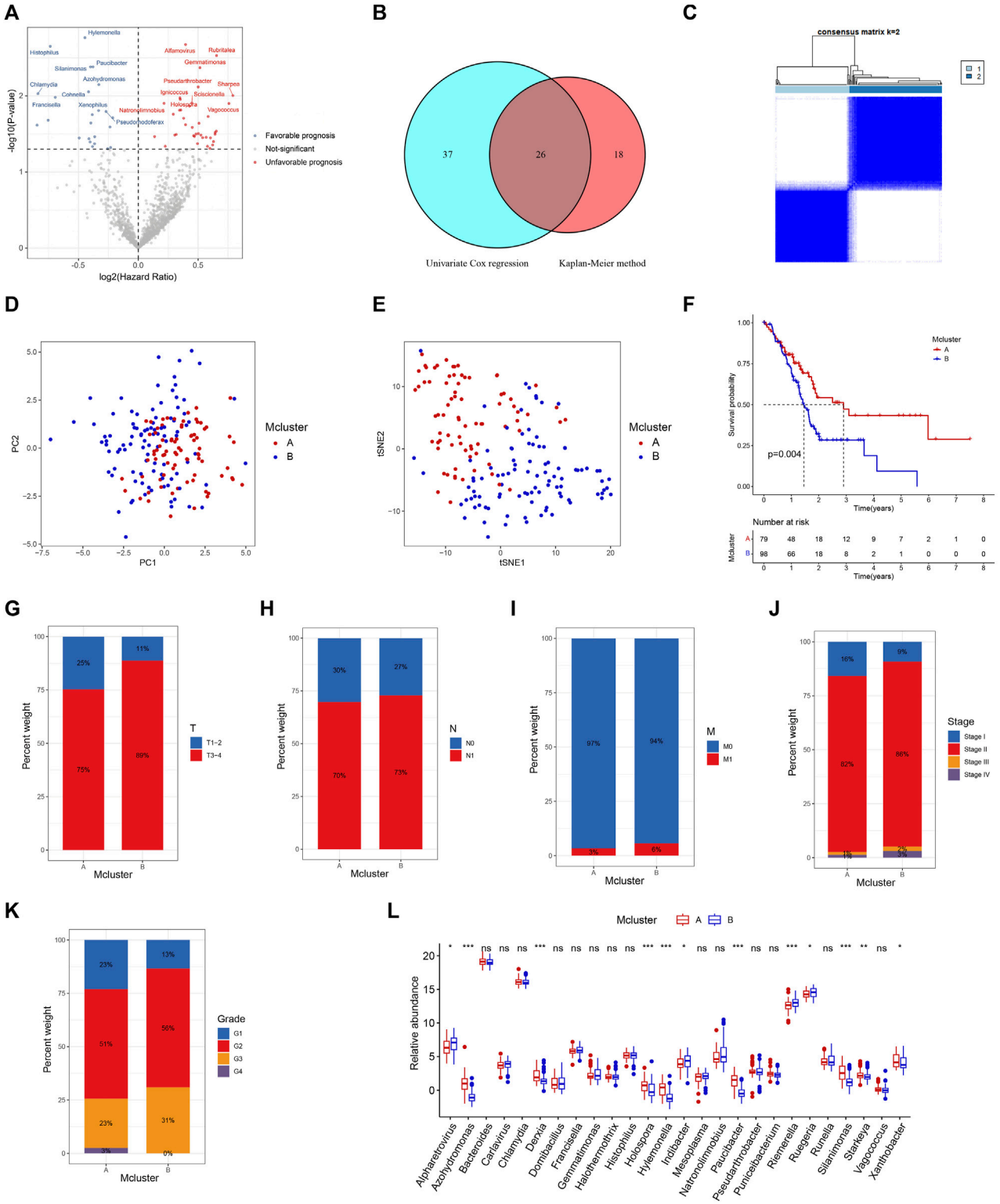


FIGURE 2
Identification of intratumor microbiome-derived subtypes. **(A)** Volcanic map of prognostic genera screened by univariable Cox regression. **(B)** Intersection of prognostic genera found by univariable Cox regression and Kaplan-Meier method. **(C)** Heatmap of consensus matrix when the cluster number was 2. PCA **(D)** and t-SNE **(E)** can clearly distinguish samples between Mcluster A and B. **(F)** Survival curves for Mcluster A and B. The proportion of T **(G)**, N **(H)**, M **(I)**, TNM **(J)** stage and pathological grade **(K)** between Mcluster A and B. **(L)** Differences in genera abundance between Mcluster A and B. (ns, no significant; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

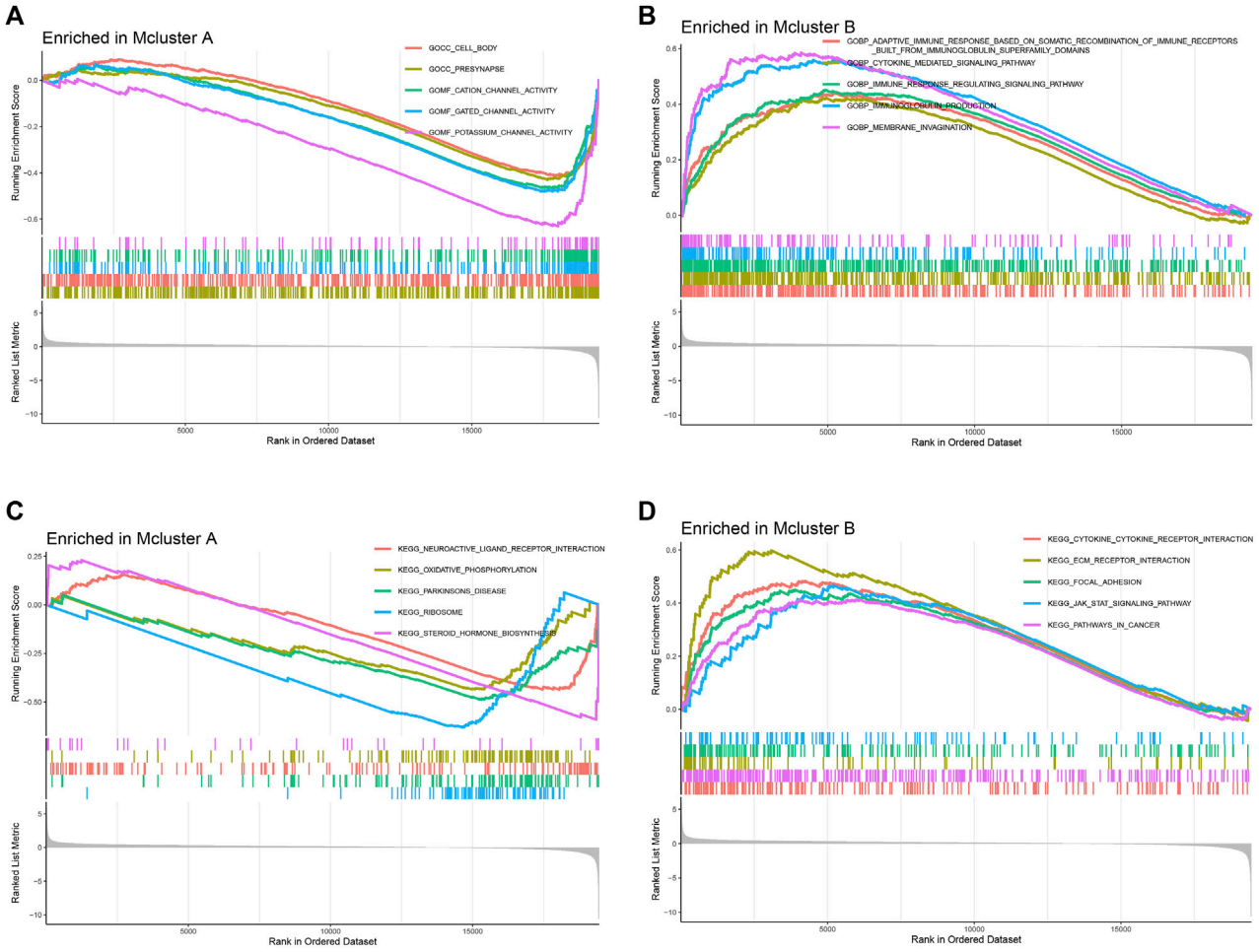


FIGURE 3
Gene set enrichment analysis. The remarkably enriched pathways of Mcluster A **(A)** and B **(B)** in "c5.go.v7.5.1.symbols.gmt" gene set. The remarkably enriched pathways of Mcluster A **(C)** and B **(D)** in "c2.cp.kegg.v7.5.1.symbols.gmt" gene set.

associated fibroblasts, myeloid dendritic cell, as well as activated mast cell exhibited remarkably higher infiltration levels within Mcluster B (Figure 4D). Then, we further utilized the TCIA platform to investigate the response of PC to immune checkpoint blockers. In the matter of overall immunophenoscore, immunophenoscore for PD-1 blocker, and immunophenoscore for CTLA-4 and PD-1 blocker, our investigation could not detect any remarkable variations between the two subtypes (Figures 4E, F, H). However, the immunophenoscore for CTLA-4 blocker had a remarkably higher score in Mcluster A (Figure 4G), suggesting that patients with PC in Mcluster A may have a better response to CTLA-4 blockers.

3.4 Drug sensitivity analysis

Drug adjuvant therapy is an important means for enhancing PC patients' prognoses. Nevertheless, the emergence of primary and secondary drug resistance often leads to treatment failure. To enhance the curative impact, it is crucial to choose medications with high sensitivity for various patients. Our results showed that patients in Mcluster A were more sensitive to 5-fluorouracil, cisplatin, gemcitabine, irinotecan, oxaliplatin, sorafenib, and epirubicin (Figures 5A–G), while patients in Mcluster B were more sensitive to sapitinib and osimertinib (Figures 5H, I). Thus, intratumor microbiome-derived subtypes can provide new strategies for personalized therapy in PC.

3.5 Identification and analysis of differentially expressed genes

To explore the transcriptional differences in genes among different intratumor microbiome-derived subtypes, we identified 4716 DEGs, with 281 genes having higher expression within Mcluster A as well as 4435 genes having higher expression within Mcluster B (Figure 6A). Additionally, between pancreatic tumour tissue and paracancerous tissues, we identified 230 DEGs, with 107 genes had higher expression within tumour tissues and 123 genes had higher expression within paracancerous tissues (Figure 6B). Further intersection analysis obtained 29 DEGs (Figure 6C). To analyze the biological processes associated with these 29 genes, we carried out GO as well as KEGG analyses. GO

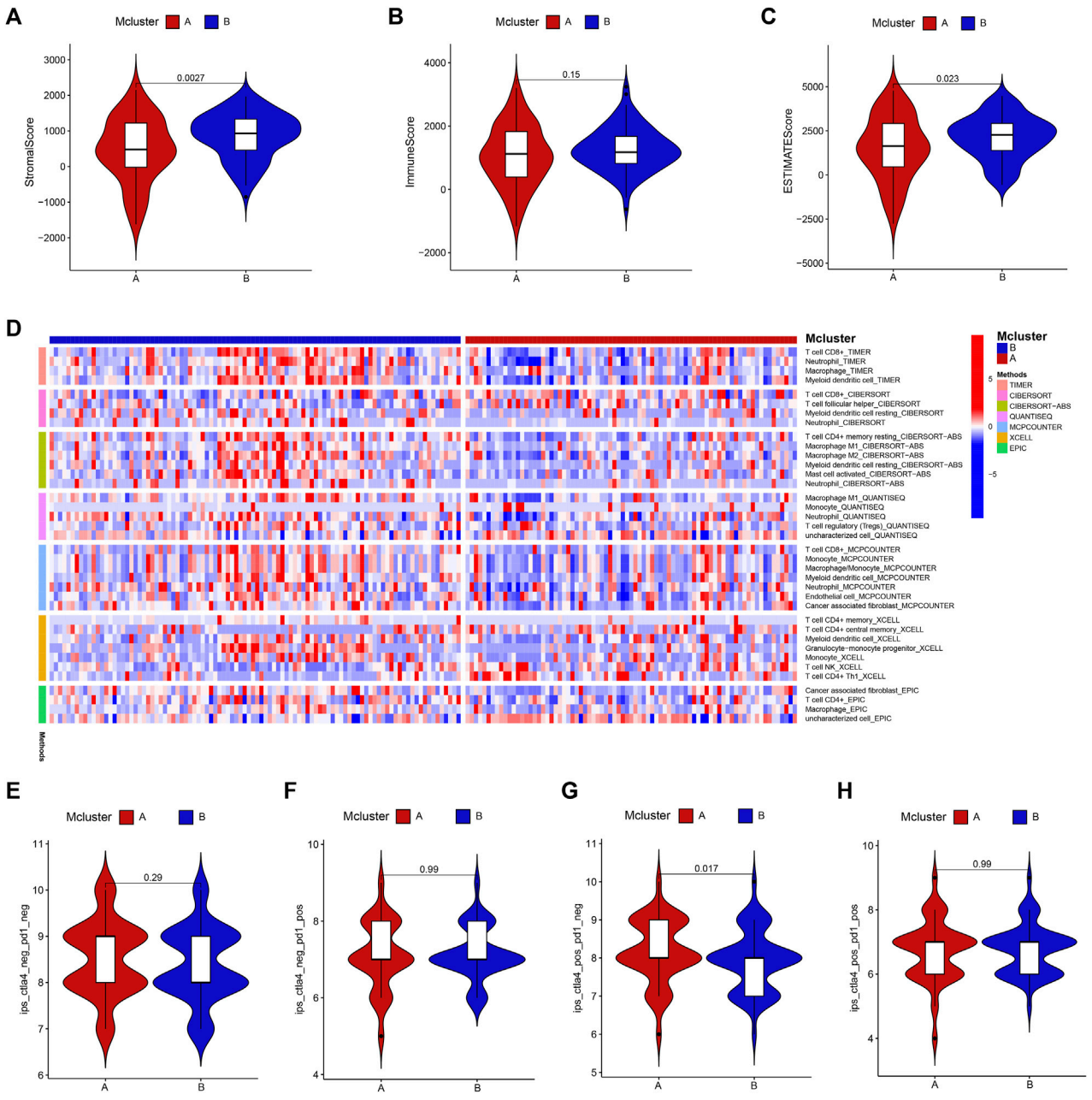


FIGURE 4
Immunoassay. The stroma **(A)**, immune **(B)** as well as ESTIMATE **(C)** scores between Mcluster A and B were evaluated using "ESTIMATE" algorithm. **(D)** Variations in infiltrated levels of different immune cell subpopulations in Mcluster A and B. The overall immunophenoscore **(E)**, immunophenoscore for PD-1 blocker **(F)**, CTLA-4 blocker **(G)**, and immunophenoscore for CTLA-4 and PD-1 blocker **(H)** between Mcluster A and B.

analysis identified the enriched pathways such as cell-matrix adhesion, endoderm development, endoderm formation, endodermal cell differentiation, extracellular matrix organization, formation of the primary germ layer, gastrulation, and integrin-mediated signaling pathway (Figure 6D). KEGG analysis revealed the enriched pathways containing amoebiasis, dilated cardiomyopathy, ECM-receptor interaction, focal adhesion, human papillomavirus infection, hypertrophic cardiomyopathy, PI3K-Akt signaling pathway, and small cell lung cancer (Figure 6E).

We further carried out consensus clustering analysis to investigate the prognostic usefulness of these 29 microbiome-related genes in PC. Five subtypes were identified for all PC patients: MRDEGclusters A through E (Figure 6F). PCA as well as t-SNE analysis clearly distinguished the five subtypes (Figures 6G, H). KM curves indicated that the five subtypes' prognoses varied significantly, with MRDEGcluster B having the best prognosis and MRDEGcluster D having the worst prognosis (Figure 6I).

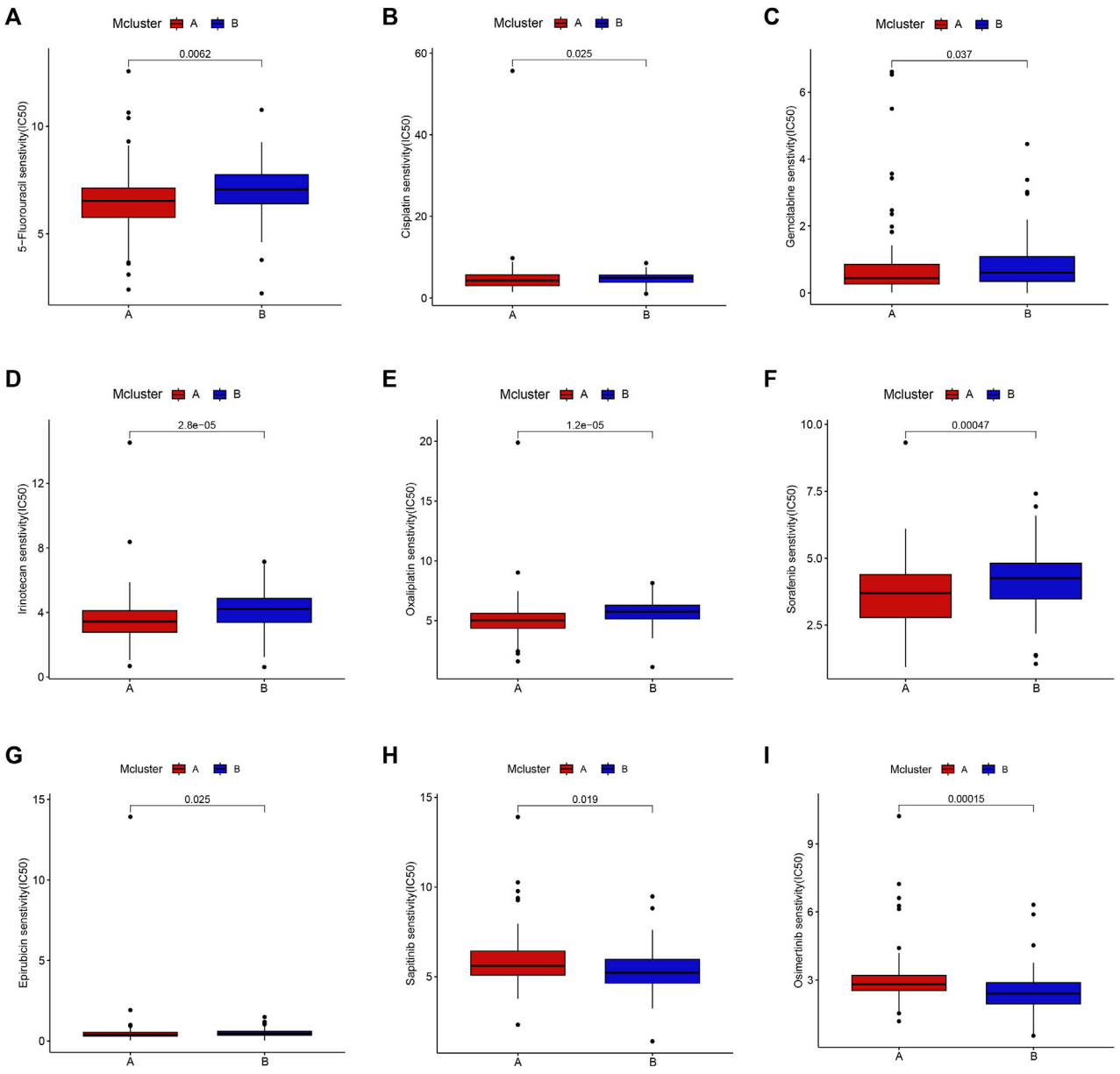


FIGURE 5
Drug sensitivity between Mcluster A and B. Patients with Mcluster A had high sensitivity to 5-Fluorouracil **(A)**, Cisplatin **(B)**, Gemcitabine **(C)**, Irinotecan **(D)**, Oxaliplatin **(E)**, Sorafenib **(F)**, and Epirubicin **(G)**. Patients with Mcluster B were more sensitive to Sapitinib **(H)** and Osimertinib **(I)**.

3.6 Building and testing a prognostic signature

For predicting each PC patient's prognosis more accurately, we constructed a prognostic signature utilizing microbiome-derived DEGs. Firstly, 21 genes linked to PC patients' prognosis were found using univariable Cox regression analysis Supplementary Figure S4. Then, utilizing LASSO regression to eliminate genes overfitting (Figures 7A, B). Finally, utilizing multivariable Cox regression to construct the prognostic model, which included two genes: LIPH and LAMA3 (Figure 7C). PC patients with low-risk scores had considerably longer overall survival times than those with high-risk scores (Figure 7D). The prognostic model's dependability was further attested to by the internal as well as external validation sets (Figures 7E, F). The AUC values of 1, 3, and 5-year survival rates were 0.726, 0.743, and 0.832 in the training set (Figure 7G), 0.713, 0.670, and 0.669 in the internal validation set (Figure 7H), and 0.568, 0.642, and 0.847 in the external validation set (Figure 7I), indicating good predictive value. PCA and t-SNE analyses clearly distinguished patients between high- and low-risk categories in the training (Figure 7J), internal validation (Figure 7K), and external validation sets (Figure 7L).

We used a Sankey diagram to illustrate the association between the prognostic model and intratumor microbiome-derived subtypes

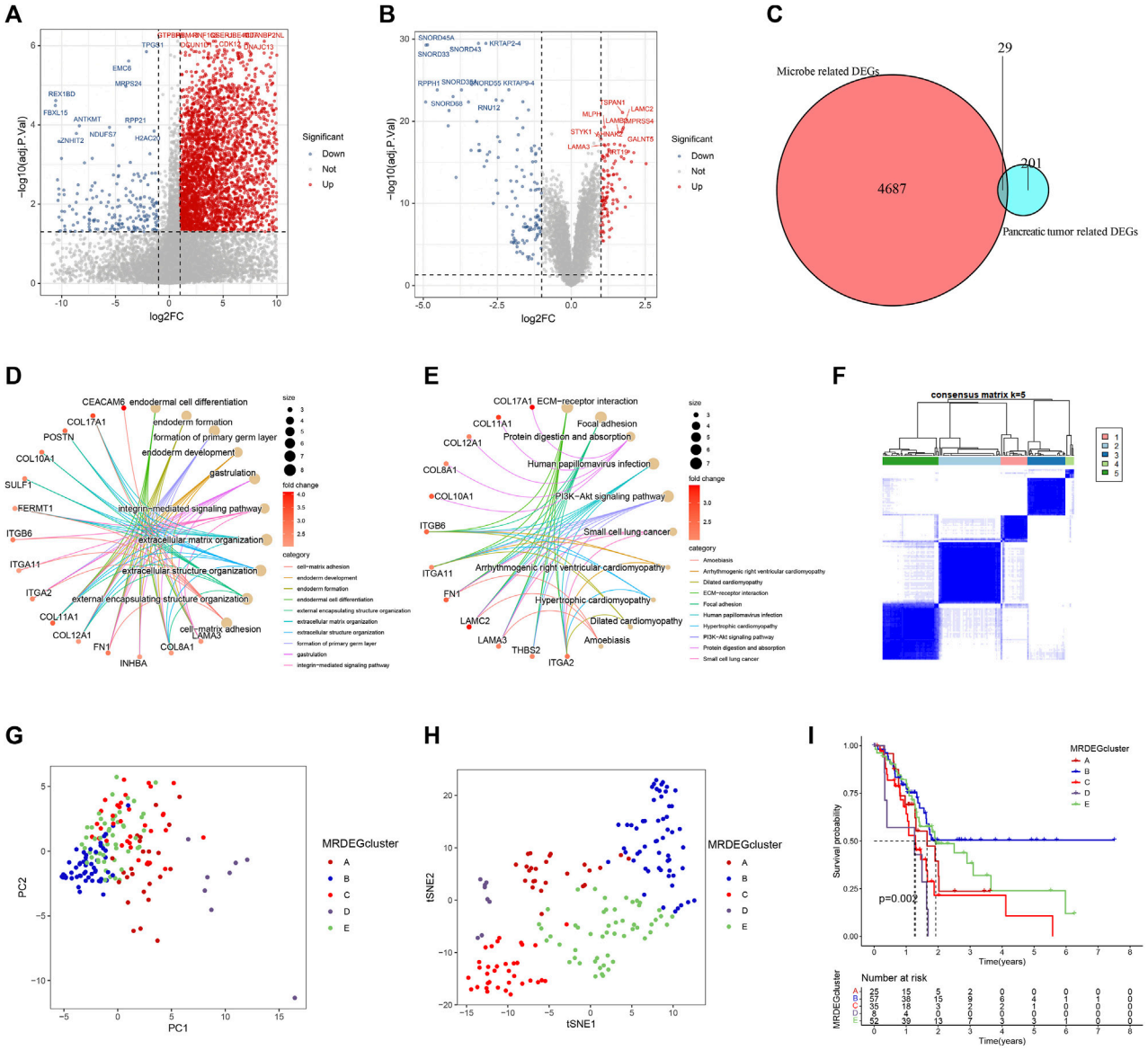


FIGURE 6
Identification and analysis of differentially expressed genes (DEGs). **(A)** Volcanic map of DEGs between Mcluster A and B. **(B)** Volcanic map of DEGs between PC tissues and normal tissues. **(C)** Intersection of DEGs. **(D)** GO enrichment analysis. **(E)** KEGG enrichment analysis. **(F)** Heatmap of consensus matrix when the cluster number was 5. PCA **(G)** and t-SNE **(H)** can clearly distinguish samples among different MRDEGclusters. **(I)** Survival curves among different MRDEGclusters.

(Figure 8A). PC patients in Mcluster A had considerably lower risk scores than those in Mcluster B (Figure 8B). Besides, it was found that Riemerella had remarkably higher abundance in patients with high scores, while Azohydromonas, Derxia, Hylemonella, Paucibacter, and Silanimonas had remarkably higher abundance in patients with low scores (Figure 8C). Subsequently, we further investigated the association between gene expression and microbial abundance. Results showed that the abundance of Riemerella was remarkably related positively to the expression of LIPH and LAMA3, while the abundance of Silanimonas and Hylemonella was remarkably related negatively to the expression of LIPH and LAMA3 (Figure 8D).

3.7 Correlation of clinicopathological features, independent prognostic analysis, and construction of nomogram prediction model

Risk scores did not differ remarkably between groups generated by age, gender, and M stage (Figures 9A, B, E). Patients with N1 stage had higher risk scores compared to those with N0 stage, which were approaching statistical significance (Figure 9D). Patients' risk scores were noticeably greater in those with higher T stage, TNM stage, and pathological grade (Figures 9C, F, G). Additionally, age and risk score were found to be independent

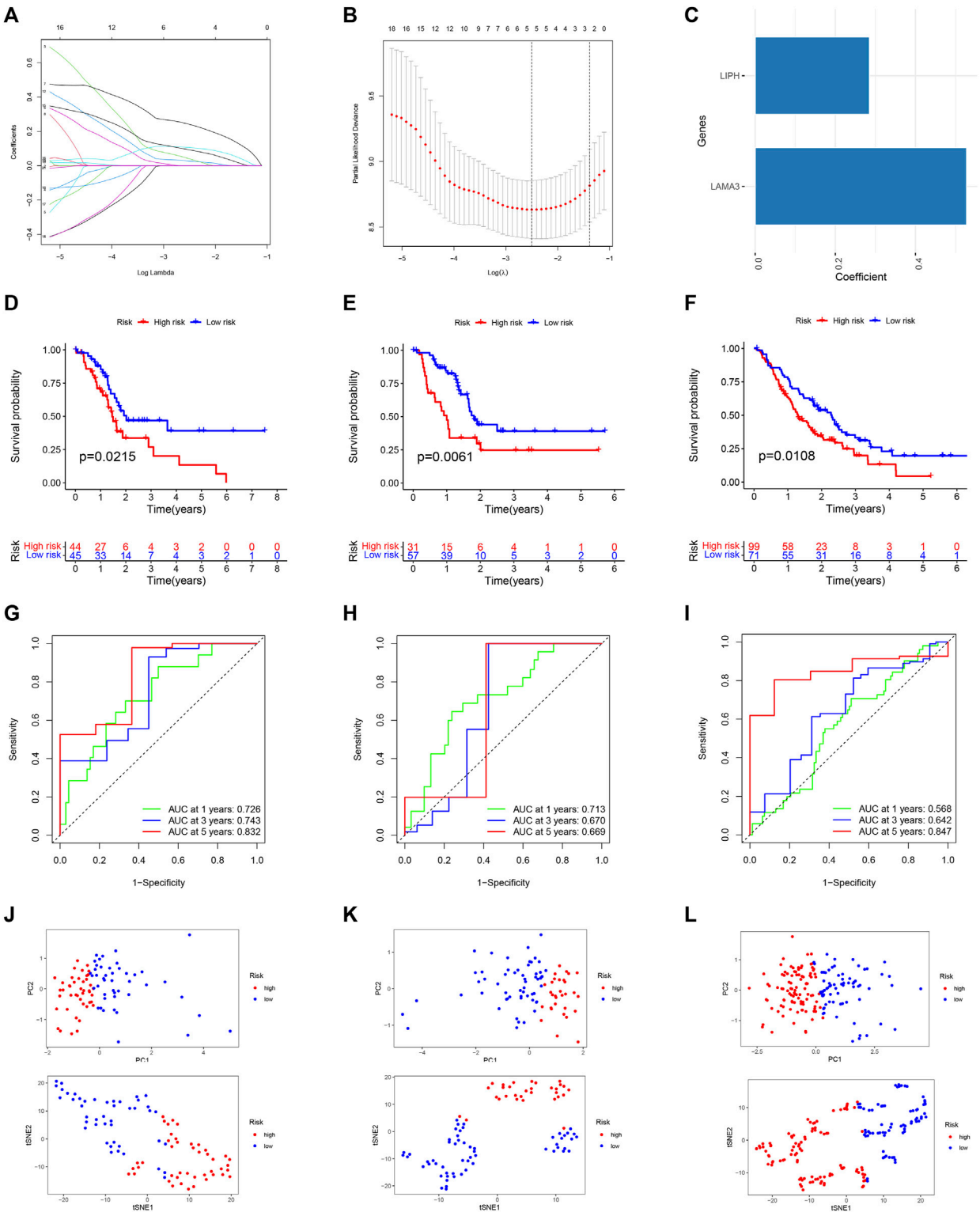


FIGURE 7
Building and testing a prognostic signature. **(A)** The coefficient path graph. **(B)** The cross validation curves. **(C)** Coefficient of LAMA3 and LIPH. Survival curve of the training **(D)**, internal validation **(E)**, and external GEO **(F)** dataset. The time-dependent ROC curve of the training **(G)**, internal validation **(H)**, and external GEO **(I)** dataset. PCA and t-SNE for the training **(J)**, internal validation **(K)**, and external GEO **(L)** dataset.

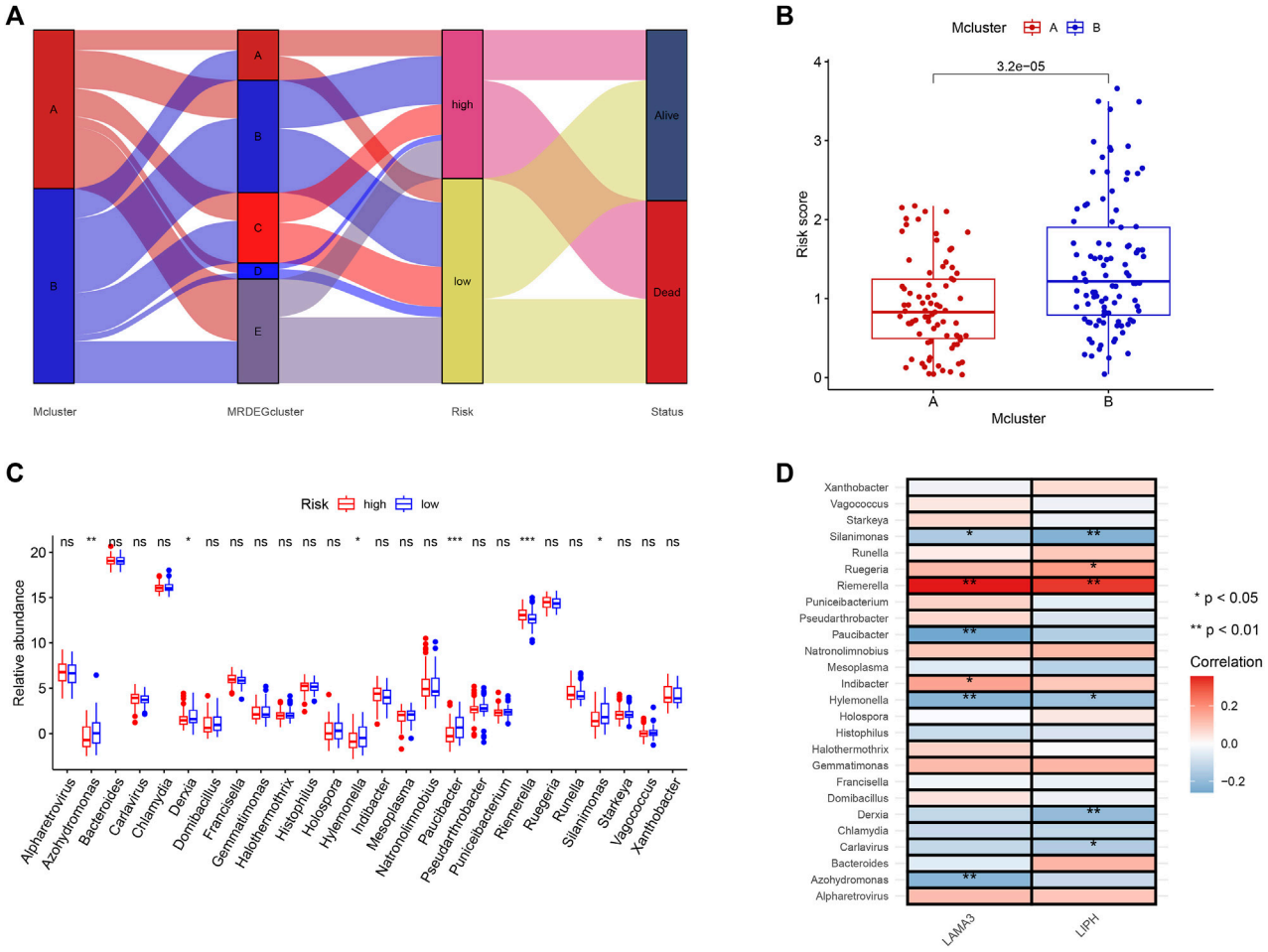


FIGURE 8
The correlation of microbiome-derived subtypes and signature. **(A)** Alluvial diagram for the microbiome-derived subtypes and signature. **(B)** Differences in risk score of Mcluster A and Mcluster B. **(C)** Variations in genera abundance between the high- and low-risk group. **(D)** Correlation of LAMA3 and LIPH and different genera. (ns, no significant; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

poor prognostic variables of PC in both the univariable and multivariable Cox regression analyses (Figures 9H, I). Then, for further evaluating the prognosis of PC patients, we created a nomogram prediction model utilizing clinicopathological variables and risk scores (Figure 9J). The calibration curve demonstrated that our nomogram model had strong predictive value because the predictive 1, 3, and 5-year survival rates were relatively close to the actual 1, 3, and 5-year survival rates (Figure 9K).

3.8 Drug sensitivity analysis in different risk groups

For exploring the potential relationship of risk score and drug sensitivity, the "oncoPredict" package was utilized to predict the drug sensitivity of PC patients. The results indicated that PC with low scores had higher sensitivity to cisplatin, epirubicin, fludarabine, irinotecan, KRAS (G12C) Inhibitor-12, oxaliplatin, and sorafenib (Figures 10A–G). On the other hand, PC with high scores exhibited higher sensitivity to trametinib and sapitinib (Figures 10H, I).

3.9 Experimental validation and single-cell analysis

Utilizing GEPIA platform, the mRNA expression of LAMA3 and LIPH between pancreatic tumour and normal tissues was investigated. The expression of LAMA3 was markedly increased within pancreatic tumour tissues (Figure 11A) and was linked to the advanced stage and poor prognosis (Figures 11B, C). Similarly, LIPH also exhibited higher expression in pancreatic tumour tissues (Figure 11D) and was linked to the advanced stage and poor prognosis (Figures 11E, F). Furthermore, the immunohistochemistry images from HPA database showed that compared to normal pancreatic tissues, the protein expression of LAMA3 and LIPH was higher within PC tissues (Figures 11G, H). To further validate the reliability of our study, we performed qRT-PCR to confirm the expression of LAMA3 and LIPH. Similarly, LAMA3 and LIPH had higher expressed levels in PC cell lines than normal pancreatic cell lines (Figures 12A, B). Lastly, we performed single-cell analysis using the single-cell dataset GSE111672 (containing 3 samples and 6122 cells) to further uncover the cell subpopulations within the tumour microenvironment of PC. In

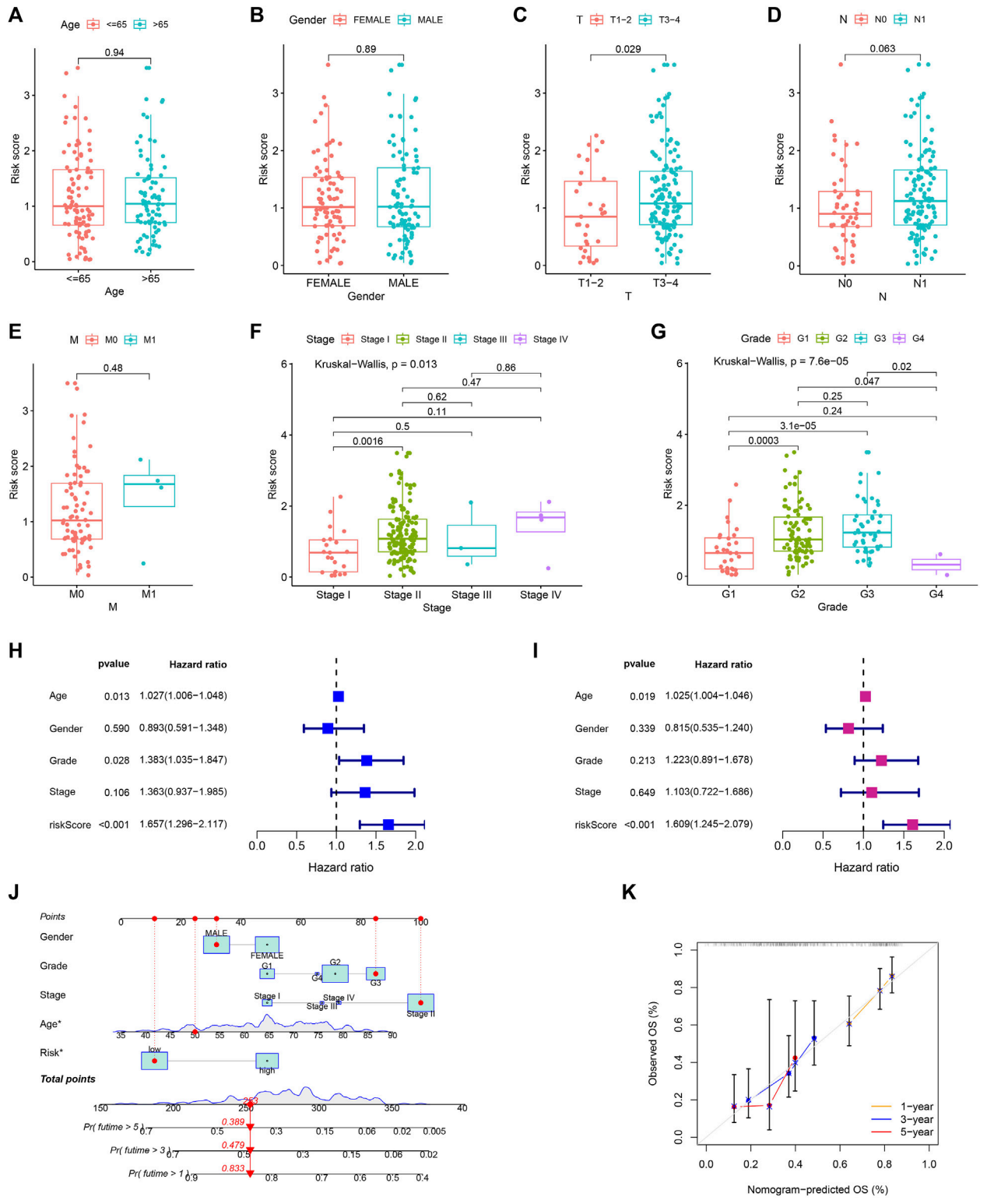


FIGURE 9
Independent prognostic analysis and constructing nomogram prediction model. The comparison of risk score in different age **(A)**, gender **(B)**, T **(C)**, N **(D)**, M **(E)**, TNM **(F)** stage, and pathological grade **(G)**. Forest map for univariate **(H)** and multivariate **(I)** Cox regression. **(J)** Nomogram prediction model. **(K)** Calibration curve of nomogram model.

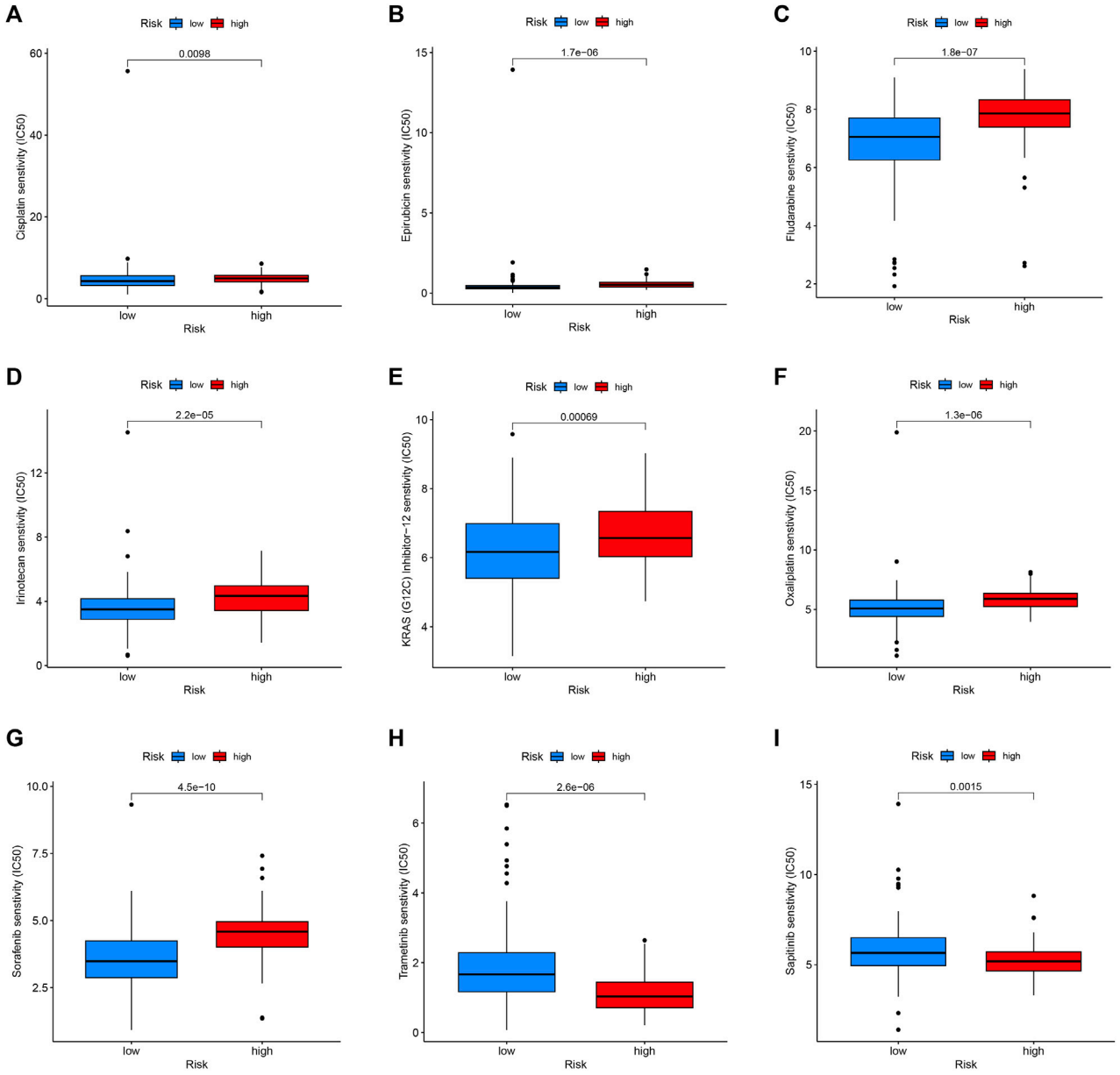


FIGURE 10
Drug sensitivity in the high- and low-risk group. Patients with low-risk scores were more sensitive to Cisplatin **(A)**, Epirubicin **(B)**, Fludarabine **(C)**, Irinotecan **(D)**, KRAS (G12C) Inhibitor-12 **(E)**, Oxaliplatin **(F)**, and Sorafenib **(G)**. Patients with high-risk scores had higher sensitivity to Trametinib **(H)** and Sapitinib **(I)**.

addition to cancer cells, the tumour microenvironment of PC is also rich in monocytes/macrophages, endothelial cells, and fibroblasts (Figures 13A–C). LIPH and LAMA3 exhibited relatively higher expression in cancer cells and neutrophils (Figures 13D–G).

4 Discussion

PC, a highly lethal malignancy characterized by early metastasis and resistance to anticancer treatments, has become the seventh most common cause of cancer-related death globally. Despite the rapid development of diagnostic and therapeutic strategies for malignancies, patients with PC are frequently discovered at a late stage, and current treatments have little effect (Zhu et al., 2018). Therefore, there is significant clinical value in developing biomarkers for PC early diagnosis and risk assessment. Recently, the regulatory function of microbiome in cancer occurrence and development has been intensively studied, which can influence the occurrence, progression, metastasis as well as therapy response in various tumours. In this study, the important role of microbiome in the outcome, tumour microenvironment heterogeneity, and treatment response for PC patients was investigated by integrating microbiome and transcriptome data, and first constructed the microbiome-related subtypes and signature in PC.

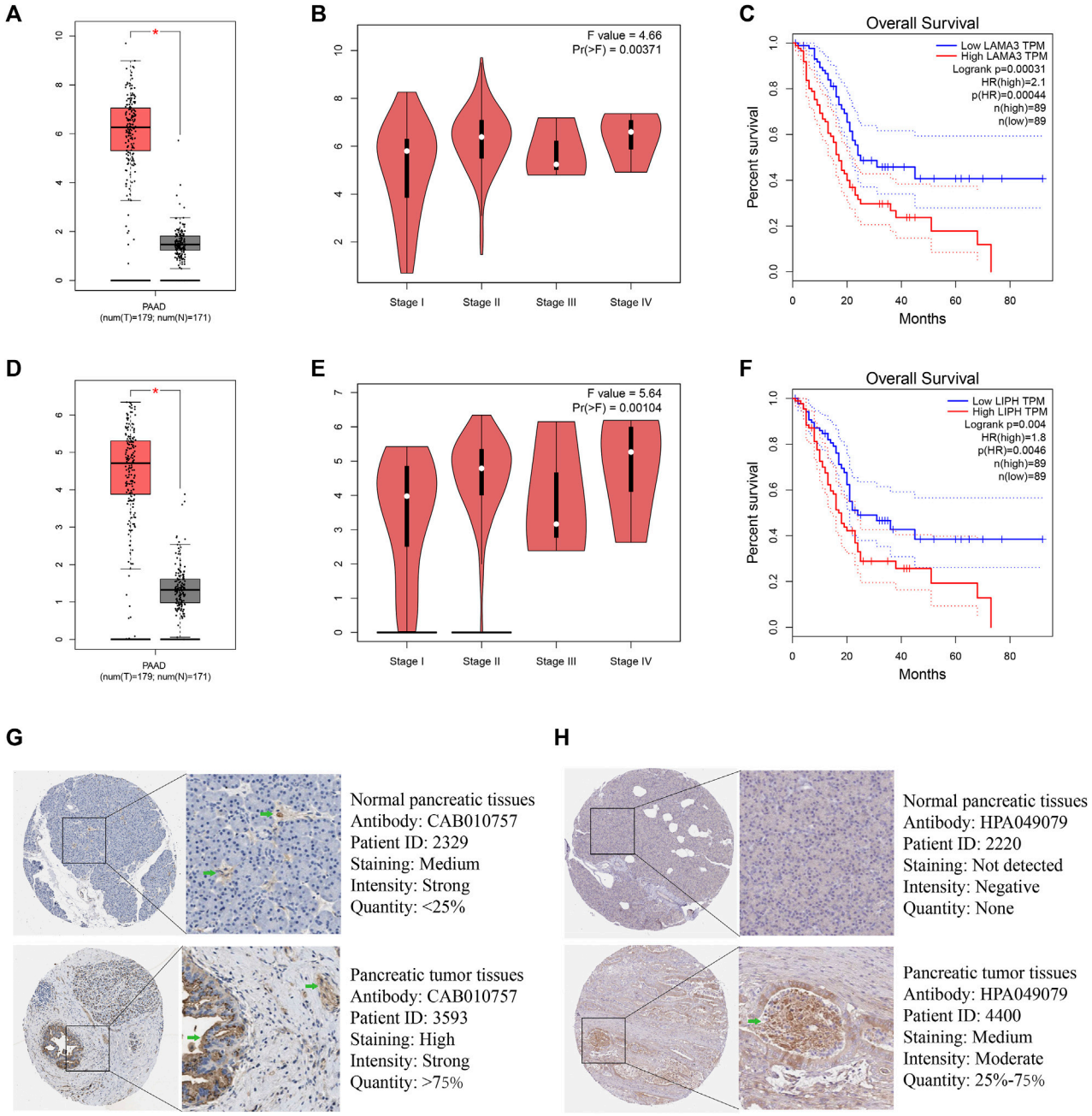


FIGURE 11
LAMA3 and LIPH. The expressed levels of LAMA3 in PC tumour tissue was remarkably higher within mRNA level **(A)**, and it was linked to TNM stage **(B)** and poor prognosis **(C)**. The expression of LIPH in PC tumour tissue was significantly higher within mRNA level **(D)**, and it was linked to TNM stage **(E)** and poor prognosis **(F)**. Immunohistochemical images indicated that the expression of LAMA3 **(G)** and LIPH **(H)** was significantly higher in pancreatic tumour tissues. (*$p < 0.05$).

We constructed the intratumor microbiome-derived subtypes by consensus cluster analysis. Survival analysis results suggested that Mcluster A had a remarkably better outcome compared with Mcluster B. What's more, compared with Mcluster B, the proportion of T1-2 stages, N0 stages, M0 stages, Stage I, and pathological grade G1 in Mcluster A was higher. Next, we analyzed the abundance differences of 26 genera in the two subtypes, and found that Azohydromonas, Derxia, Holospora, Hylemonella, Paucibacter, Silanimonas, Starkeya, and Xanthobacter have significantly higher abundance in Mcluster A, and Alpharetrovirus, Indibacter, Riemerella, and Ruegeria have significantly higher abundance in Mcluster B. It has been reported that Alpharetrovirus can achieve almost complete elimination of leukemia cells by enhancing the toxicity of NK cells to leukemia cells (Suerth et al., 2016). We speculated that the tumour cells of PC patients with Mcluster B were more malignant. In order to enhance the killing ability of NK cells

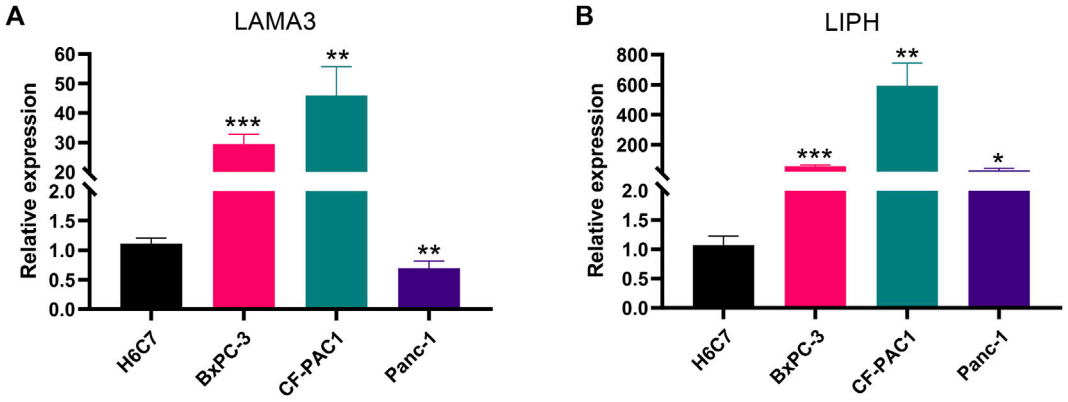


FIGURE 12
The validation of gene expression. LAMA3 **(A)** and LIPH **(B)** have higher expressed levels in PC cell lines than normal pancreatic cell lines. (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

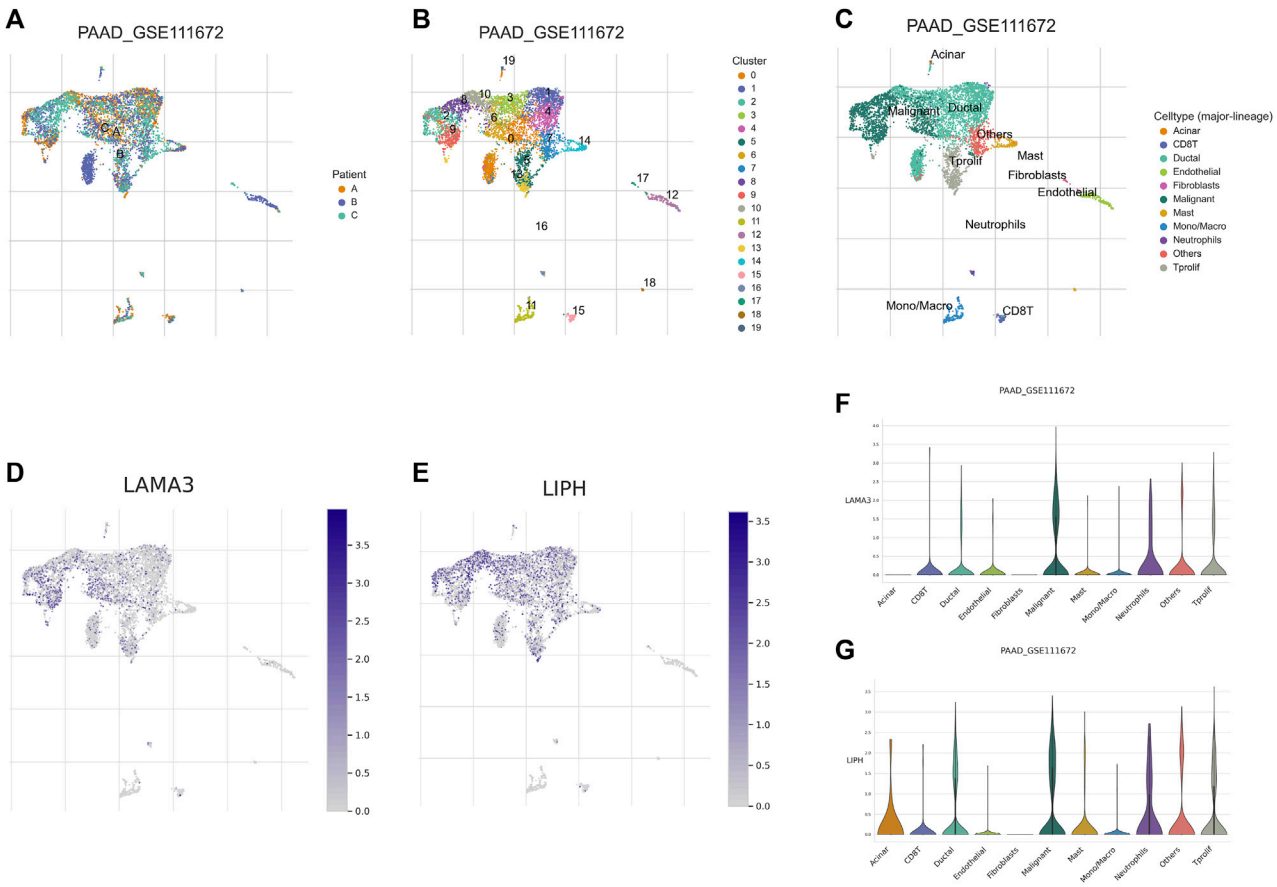


FIGURE 13
Single-cell analysis. **(A)** Annotation based on sample source. **(B)** Annotation based on cluster results. **(C)** Annotation based on the various cell subsets. **(D)** The distribution of LAMA3 expression within various cell subsets. **(E)** The distribution of LIPH expression within different cell subsets. **(F)** The relative expressed level of LAMA3 within various cell subsets. **(G)** The relative expressed level of LIPH within various cell subsets.

to fight against tumour cells, the body upregulated the level of Alpharetrovirus. Based on the above findings, microbiome was tightly connected with PC patients' outcomes.

Then, we delved into the molecular mechanisms underlying the differences in prognosis of patients with different subtypes of PC. GSEA revealed that signaling pathways associated with ion-gated

channels in tumour cells in patients with Mcluster A were remarkably enriched, while the activity of immune response signaling pathways in patients with Mcluster B were significantly enhanced. Ion-gated channels are responsible for tumour cell proliferation and are key factors in PC progression and invasion (Yee et al., 2012; Liu et al., 2018), and are also a key therapeutic target for PC (Yee, 2016). The microbiome with high abundance in Mcluster A patients may influence the prognosis of PC by regulating ion-gated channels. What's more, previous studies have revealed that microbiome can participate in the immune response, which resulted in the prognostic change in patients with PC. Therefore, microbiome can influence the progression as well as outcome of PC through regulating ion-gated channels or immune response pathways. We further analyzed the relationship between intratumor microbiome-derived subtypes and tumour microenvironment heterogeneity, and found that Mcluster B had higher stromal and ESTIMATE scores. Neutrophil, Treg, $CD8^+$ T cell, macrophages M1 and M2, cancer associated fibroblasts, myeloid dendritic cell, as well as activated mast cell had remarkably higher infiltrated levels within Mcluster B. $CD8^+$ T cells can recognize and eliminate tumor cells through multiple mechanisms (Borst et al., 2018; Terrén et al., 2019; Philip and Schietinger, 2022). Studies indicated that the higher infiltrated levels of $CD8^+$ T cell were linked to significantly longer survival time of PC patients (Carstens et al., 2017). Therefore, intratumor microbiome are likely to participate in shaping the tumour immune microenvironment, thereby affecting the immune response of tumor cells. The response of different intratumor microbiome-derived subtypes to immunotherapy was evaluated in this study. We found that the use of a CTLA-4 blocker was found to be more likely to be beneficial for PC patients in Mcluster B. Therefore, the study of microbiome in the tumour microenvironment of PC can help provide new strategies for the selection of immunotherapy for patients with PC.

For accurately predicting the prognosis for every PC patient, we utilized microbiome-related DEGs to construct and validate a prognostic signature. KM curves indicated that the survival time of PC patients with high-risk scores was remarkably lower, and ROC curves also indicated that the signature had a good predictive performance. What's more, we explored the differences in the abundance of microbiome between the high and low-risk categories, and found that Riemerella had a significantly higher abundance in PC with high-risk scores and was linked to the poor outcome of PC. Riemerella is a Gram-negative rod-shaped bacterium that can cause acute infectious disease as well as an inflammatory response (Afrin et al., 2018; Li et al., 2023). However, Riemerella has not been reported in PC. Our study provides a new therapeutic target for PC. Correlation analysis indicated that the abundance of Riemerella was significantly linked positively to the expression of LIPH and LAMA3. LIPH is a new member of the triglyceride lipase family located on human chromosomes. The protein encoded by LIPH can hydrolyze triglycerides and phospholipids to produce fatty acids, which can then promote intestinal absorption or serve as an energy source or energy reserve (Jin et al., 2002). Studies indicated that LIPH had a higher expressed level in breast tumor tissue, and it affected the distant metastasis of breast cancer by regulating CAPN2 and paxillin (Seki et al., 2014; Zhang et al., 2020). According to the findings of our investigation, LIPH can be exploited as a possible therapeutic target for PC because it was found to be increased expression in PC tissues and to be related to the disease's progression and bad prognosis. The LAMA3 gene can encode the α3 chain of laminin-5, which is an important cell membrane component and regulates cell adhesion and migration (Zhang et al., 2018; Xu et al., 2019). Studies indicated that LAMA3 is a promising target for cancer therapy since it may accelerate the growth and invasion of tumour cells (Xu et al., 2019; Shu et al., 2023). This agrees with our research suggesting that LAMA3 might be a useful treatment target for PC.

Drug-assisted therapy is one of the main ways to treat PC and can help improve the prognosis of patients. For example, modified FOLFIRINOX (containing oxaliplatin, irinotecan, leucovorin, and fluorouracil) and gemcitabine, as first-line chemotherapy regimens for PC, could result in 5-year disease-free survival rates of 26% and 19%, respectively, for patients with PC after surgery (Conroy et al., 2022). But the intricate PC tumour microenvironment frequently promotes the development of treatment resistance, which ultimately results in the failure of medication therapy. To improve treatment efficacy and prognosis, it is crucial to determine the medications to which each patient is sensitive. Our research assessed the relationship between intratumor microbiome-related subtypes and drug sensitivity. PC patients in Mcluster B or high-risk group had higher sensitivity to sapitinib, but PC patients in Mcluster A or low-risk group had higher sensitivity to cisplatin, irinotecan, oxaliplatin, sorafenib, and epirubicin. These findings provide a basis for individualized treatment of PC patients and are of great significance for improving the efficiency of drug treatment. The anti-tumor medication sapitinib has dual anti-tumor actions and can act on tumour blood vessels and tumour cells simultaneously (Gao et al., 2020; Attwa et al., 2023). Whether its curative effect on PC will be affected by intratumor microbiome and the specific mechanism still needs further basic research to explore.

However, our research has a few limitations that should be acknowledged. Firstly, this study belonged to retrospective research and was performed mainly based on data from public databases. Therefore, the prediction capability of our prognostic model should be validated in the prospective clinical research with large samples. Secondly, further investigation of molecular mechanism is required for exploring the function of intratumor microbiome in the occurrence and development of PC.

5 Conclusion

In the present study, we first constructed intratumor microbiome-derived subtypes in PC, and clarified the crucial role of microbiome in the outcome, tumor microenvironment shaping, and immunotherapy response for PC through multi-omics analysis, providing the novel microbiome-related targets for the treatment of PC. Meanwhile, we also built a prognostic signature utilizing intratumor microbiome-related genes to predict PC patients' outcomes. In conclusion, this study can provide a novel insight for the prognosis prediction and treatment decision-making of PC.

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.

Author contributions

The study's concept and design were created by BiZ, JL, and ZW. The information was gathered from the web database by BiZ and JL. BiZ, JL, and HL conducted data analysis. BiZ, BH, HL, and BS wrote the article. BoZ conducted experiments. CB, YL, and ZW revised the manuscript. All authors contributed to the article and approved the submitted version.

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

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2023.1244752/full#supplementary-material

SUPPLEMENTARY FIGURE S1
The relationship among the 26 prognostic-related genera.

SUPPLEMENTARY FIGURE S2
The cumulative distribution function (CDF) curve.

SUPPLEMENTARY FIGURE S3
The alteration of the area under CDF curve.

SUPPLEMENTARY FIGURE S4
Univariable Cox regression found 21 genes linked to the prognosis.

SUPPLEMENTARY TABLE S1
The abundance of 1406 genera in each pancreatic cancer sample.

SUPPLEMENTARY TABLE S2
44 prognostic genera were identified using the Kaplan–Meier method.

References

Afrin, F., Fernandez, C. P., Flores, R. A., Kim, W. H., Jeong, J., Chang, H. H., et al. (2018). Downregulation of common cytokine receptor γ chain inhibits inflammatory responses in macrophages stimulated with Riemerella anatipestifer. *Dev. Comp. Immunol.* 81, 225–234. doi:10.1016/j.dci.2017.12.009

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: Tool for the unification of biology. The gene Ontology consortium. *Nat. Genet.* 25, 25–29. doi:10.1038/75556

Attwa, M. W., AlRabiah, H., Mostafa, G. A. E., and Kadi, A. A. (2023). Development of an LC-MS/MS method for quantification of sapitinib in human liver microsomes: *In silico* and *in vitro* metabolic stability evaluation. *Molecules* 28, 2322. doi:10.3390/molecules28052322

Borst, J., Ahrends, T., Bąbała, N., Melief, C. J. M., and Kastenmüller, W. (2018). CD4+ T cell help in cancer immunology and immunotherapy. *Nat. Rev. Immunol.* 18, 635–647. doi:10.1038/s41577-018-0044-0

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* 68, 394–424. doi:10.3322/caac.21492

Carstens, J. L., Correa De Sampaio, P., Yang, D., Barua, S., Wang, H., Rao, A., et al. (2017). Spatial computation of intratumoral T cells correlates with survival of patients with pancreatic cancer. *Nat. Commun.* 8, 15095. doi:10.1038/ncomms15095

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2, 401–404. doi:10.1158/2159-8290.CD-12-0095

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* 18, 248–262. doi:10.1016/j.celrep.2016.12.019

Chi, H., Peng, G., Wang, R., Yang, F., Xie, X., Zhang, J., et al. (2022a). Cuprotosis programmed-cell-death-related lncRNA signature predicts prognosis and immune landscape in PAAD patients. *Cells* 11, 3436. doi:10.3390/cells11213436

Chi, H., Xie, X., Yan, Y., Peng, G., Strohmer, D. F., Lai, G., et al. (2022b). Natural killer cell-related prognosis signature characterizes immune landscape and predicts prognosis of HNSCC. *Front. Immunol.* 13, 1018685. doi:10.3389/fimmu.2022.1018685

Chi, H., Yang, J., Peng, G., Zhang, J., Song, G., Xie, X., et al. (2023a). Circadian rhythm-related genes index: A predictor for HNSCC prognosis, immunotherapy efficacy, and chemosensitivity. *Front. Immunol.* 14, 1091218. doi:10.3389/fimmu.2023.1091218

Chi, H., Zhao, S., Yang, J., Gao, X., Peng, G., Zhang, J., et al. (2023b). T-cell exhaustion signatures characterize the immune landscape and predict HCC prognosis via integrating single-cell RNA-seq and bulk RNA-sequencing. *Front. Immunol.* 14, 1137025. doi:10.3389/fimmu.2023.1137025

Conroy, T., Castan, F., Lopez, A., Turpin, A., Ben Abdelghani, M., Wei, A. C., et al. (2022). Five-year outcomes of FOLFIRINOX vs gemcitabine as adjuvant therapy for pancreatic cancer: A randomized clinical trial. *JAMA Oncol.* 8, 1571–1578. doi:10.1001/jamaoncol.2022.3829

Gao, H.-L., Gupta, P., Cui, Q., Ashar, Y. V., Wu, Z.-X., Zeng, L., et al. (2020). Sapitinib reverses anticancer drug resistance in colon cancer cells overexpressing the ABCB1 transporter. *Front. Oncol.* 10, 574861. doi:10.3389/fonc.2020.574861

Jin, W., Broedl, U. C., Monajemi, H., Glick, J. M., and Rader, D. J. (2002). Lipase H, a new member of the triglyceride lipase family synthesized by the intestine. *Genomics* 80, 268–273. doi:10.1006/geno.2002.6837

Kanehisa, M., and Goto, S. (2000). Kegg: Kyoto Encyclopedia of genes and Genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27

Kartal, E., Schmidt, T. S. B., Molina-Montes, E., Rodríguez-Perales, S., Wirbel, J., Maistrenko, O. M., et al. (2022). A faecal microbiota signature with high specificity for pancreatic cancer. *Gut* 71, 1359–1372. doi:10.1136/gutjnl-2021-324755

Li, J., Zhang, Y., Wang, Y., Zhang, Y., Shi, B., Gan, L., et al. (2023). Immunogenicity of live phoP gene deletion strain of Riemerella anatipestifer serotype 1. *Poult. Sci.* 102, 102294. doi:10.1016/j.psj.2022.102294

Liu, J., Hu, G., Gong, Y., Yu, Q., He, B., Li, W., et al. (2018). Silencing of TRPM8 inhibits aggressive tumor phenotypes and enhances gemcitabine sensitivity in pancreatic cancer. *Pancreatology* 18, 935–944. doi:10.1016/j.pan.2018.08.011

Maeser, D., Gruener, R. F., and Huang, R. S. (2021). oncoPredict: an R package for predicting *in vivo* or cancer patient drug response and biomarkers from cell line screening data. *Briefings Bioinforma.* 22, bbab260. doi:10.1093/bib/bbab260

Mao, A. W., Barck, H., Young, J., Paley, A., Mao, J.-H., and Chang, H. (2022). Identification of a novel cancer microbiome signature for predicting prognosis of human breast cancer patients. *Clin. Transl. Oncol.* 24, 597–604. doi:10.1007/s12094-021-02725-3

Mizrahi, J. D., Surana, R., Valle, J. W., and Shroff, R. T. (2020). Pancreatic cancer. *Lancet* 395, 2008–2020. doi:10.1016/S0140-6736(20)30974-0

Park, S. Y. (2018). Nomogram: An analogue tool to deliver digital knowledge. *J. Thorac. Cardiovasc. Surg.* 155, 1793. doi:10.1016/j.jtcvs.2017.12.107

Philip, M., and Schietinger, A. (2022). CD8+ T cell differentiation and dysfunction in cancer. *Nat. Rev. Immunol.* 22, 209–223. doi:10.1038/s41577-021-00574-3

Pontén, F., Schwenk, J. M., Asplund, A., and Edqvist, P.-H. D. (2011). The human protein Atlas as a proteomic resource for biomarker discovery: Review: The human protein Atlas. *J. Intern. Med.* 270, 428–446. doi:10.1111/j.1365-2796.2011.02427.x

Pushalkar, S., Hundeyin, M., Daley, D., Zambirinis, C. P., Kurz, E., Mishra, A., et al. (2018). The pancreatic cancer microbiome promotes oncogenesis by induction of innate and adaptive immune suppression. *Cancer Discov.* 8, 403–416. doi:10.1158/2159-8290.CD-17-1134

Riquelme, E., Zhang, Y., Zhang, L., Montiel, M., Zoltan, M., Dong, W., et al. (2019). Tumor microbiome diversity and composition influence pancreatic cancer outcomes. *Cell* 178, 795–806. doi:10.1016/j.cell.2019.07.008

Romero, J. M., Grünwald, B., Jang, G.-H., Bavi, P. P., Jhaveri, A., Masoomian, M., et al. (2020). A four-chemokine signature is associated with a T-cell–inflamed phenotype in primary and metastatic pancreatic cancer. *Clin. Cancer Res.* 26, 1997–2010. doi:10.1158/1078-0432.CCR-19-2803

Ryan, D. P., Hong, T. S., and Bardeesy, N. (2014). Pancreatic adenocarcinoma. *N. Engl. J. Med.* 371, 1039–1049. doi:10.1056/NEJMra1404198

Seki, Y., Yoshida, Y., Ishimine, H., Shinozaki-Ushiku, A., Ito, Y., Sumitomo, K., et al. (2014). Lipase member H is a novel secreted protein selectively upregulated in human lung adenocarcinomas and bronchioloalveolar carcinomas. *Biochem. Biophysical Res. Commun.* 443, 1141–1147. doi:10.1016/j.bbrc.2013.12.106

Sepich-Poore, G. D., Zitvogel, L., Straussman, R., Hasty, J., Wargo, J. A., and Knight, R. (2021). The microbiome and human cancer. *Science* 371, eabc4552. doi:10.1126/science.abc4552

Shu, C., Wang, W., Wu, L., Qi, C., Yan, W., Lu, W., et al. (2023). LINC00936/microRNA-221-3p regulates tumor progression in OvarianCancer by interacting with LAMA3. *PRA* 18, 66–79. doi:10.2174/1574892817666220316152201

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *CA A Cancer J. Clin.* 70, 7–30. doi:10.3322/caac.21590

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi:10.1073/pnas.0506580102

Suerth, J. D., Morgan, M. A., Kloess, S., Heckl, D., Neudörfl, C., Falk, C. S., et al. (2016). Efficient generation of gene-modified human natural killer cells via alpharetroviral vectors. *J. Mol. Med.* 94, 83–93. doi:10.1007/s00109-015-1327-6

Sun, D., Wang, J., Han, Y., Dong, X., Zheng, R., Ge, J., et al. (2021). Tisch: A comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. *Nucleic Acids Res.* 49, D1420–D1430. doi:10.1093/nar/gkaa1020

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). Gepia: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 45, W98–W102. doi:10.1093/nar/gkx247

Terrén, I., Orrantia, A., Vitallé, J., Zenarruzabeitia, O., and Borrego, F. (2019). NK cell metabolism and tumor microenvironment. *Front. Immunol.* 10, 2278. doi:10.3389/fimmu.2019.02278

Wang, Y., Yang, G., You, L., Yang, J., Feng, M., Qiu, J., et al. (2019). Role of the microbiome in occurrence, development and treatment of pancreatic cancer. *Mol. Cancer* 18, 173. doi:10.1186/s12943-019-1103-2

Wei, M.-Y., Shi, S., Liang, C., Meng, Q.-C., Hua, J., Zhang, Y.-Y., et al. (2019). The microbiota and microbiome in pancreatic cancer: More influential than expected. *Mol. Cancer* 18, 97. doi:10.1186/s12943-019-1008-0

Xiao, M., Liang, X., Yan, Z., Chen, J., Zhu, Y., Xie, Y., et al. (2022). A DNA-methylation-driven genes based prognostic signature reveals immune microenvironment in pancreatic cancer. *Front. Immunol.* 13, 803962. doi:10.3389/fimmu.2022.803962

Xu, S.-F., Zheng, Y., Zhang, L., Wang, P., Niu, C.-M., Wu, T., et al. (2019). Long non-coding RNA LINC00628 interacts epigenetically with the LAMA3 promoter and contributes to lung adenocarcinoma. *Mol. Ther. - Nucleic Acids* 18, 166–182. doi:10.1016/j.omtn.2019.08.005

Yee, N. S. (2016). "TRPM8 ion channels as potential cancer biomarker and target in pancreatic cancer," in *Advances in protein chemistry and structural biology* (Elsevier), 127–155. doi:10.1016/bs.apcsb.2016.01.001

Yee, N. S., Brown, R. D., Lee, M. S., Zhou, W., Jensen, C., Gerke, H., et al. (2012). TRPM8 ion channel is aberrantly expressed and required for preventing replicative senescence in pancreatic adenocarcinoma: Potential role of TRPM8 as a biomarker and target. *Cancer Biol. Ther.* 13, 592–599. doi:10.4161/cbt.20079

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

Yuan, Q., Deng, D., Pan, C., Ren, J., Wei, T., Wu, Z., et al. (2022). Integration of transcriptomics, proteomics, and metabolomics data to reveal HER2-associated metabolic heterogeneity in gastric cancer with response to immunotherapy and neoadjuvant chemotherapy. *Front. Immunol.* 13, 951137. doi:10.3389/fimmu.2022.951137

Zhang, B., Yuan, Q., Zhang, B., Li, S., Wang, Z., Liu, H., et al. (2023a). Characterization of neuroendocrine regulation- and metabolism-associated molecular features and prognostic indicators with aid to clinical chemotherapy and immunotherapy of patients with pancreatic cancer. *Front. Endocrinol.* 13, 1078424. doi:10.3389/fendo.2022.1078424

Zhang, J., Wang, H., Wang, Y., Dong, W., Jiang, Z., and Yang, G. (2018). Substrate-mediated gene transduction of LAMA3 for promoting biological sealing between titanium surface and gingival epithelium. *Colloids Surfaces B Biointerfaces* 161, 314–323. doi:10.1016/j.colsurfb.2017.10.030

Zhang, X., Zhuge, J., Liu, J., Xia, Z., Wang, H., Gao, Q., et al. (2023b). Prognostic signatures of sphingolipids: Understanding the immune landscape and predictive role in immunotherapy response and outcomes of hepatocellular carcinoma. *Front. Immunol.* 14, 1153423. doi:10.3389/fimmu.2023.1153423

Zhang, Y., Zhu, X., Qiao, X., Gu, X., Xue, J., Han, Y., et al. (2020). LIPH promotes metastasis by enriching stem-like cells in triple-negative breast cancer. *J. Cell Mol. Med.* 24, 9125–9134. doi:10.1111/jcmm.15549

Zhu, H., Li, T., Du, Y., and Li, M. (2018). Pancreatic cancer: Challenges and opportunities. *BMC Med.* 16 (214), 214–018. doi:10.1186/s12916-018-1215-3

Zhuo, Z., Lin, H., Liang, J., Ma, P., Li, J., Huang, L., et al. (2022). Mitophagy-related gene signature for prediction prognosis, immune scenery, mutation, and chemotherapy response in pancreatic cancer. *Front. Cell Dev. Biol.* 9, 802528. doi:10.3389/fcell.2021.802528

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 07 September 2023
DOI 10.3389/fphar.2023.1252805

OPEN ACCESS

EDITED BY
Linhui Wang,
Second Military Medical University, China

REVIEWED BY
Aimin Jiang,
Fudan University, China
Yinghao Cao,
Huazhong University of Science and Technology, China

*CORRESPONDENCE
Xinyu Gu,
stargu6932@163.com
Penghui Li,
lph0819@163.com

RECEIVED 04 July 2023
ACCEPTED 22 August 2023
PUBLISHED 07 September 2023

CITATION
Gu X, Li S, Ma X, Huang D and Li P (2023), Heterogeneity characterization of hepatocellular carcinoma based on the sensitivity to 5-fluorouracil and development of a prognostic regression model.
Front. Pharmacol. 14:1252805.
doi: 10.3389/fphar.2023.1252805

Heterogeneity characterization of hepatocellular carcinoma based on the sensitivity to 5-fluorouracil and development of a prognostic regression model

Xinyu Gu[1]*, Shuang Li[2], Xiao Ma[3], Di Huang[4] and Penghui Li[5]*

[1]Department of Oncology, The First Affiliated Hospital of Henan University of Science and Technology, Luoyang, China, [2]Hematology Department, Traditional Chinese Hospital of Luan, Lu'an, China, [3]School of Medicine, Zhejiang University, Hangzhou, China, [4]Department of Child Health Care, The Third Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [5]The Department of General Surgery, The First Affiliated Hospital of Henan University of Science and Technology, Luoyang, China

Background: 5-Fluorouracil (5-FU) is a widely used chemotherapeutic drug in clinical cancer treatment, including hepatocellular carcinoma (HCC). A correct understanding of the mechanisms leading to a low or lack of sensitivity of HCC to 5-FU-based treatment is a key element in the current personalized medical treatment.

Methods: Weighted gene co-expression network analysis (WGCNA) was used to analyze the expression profiles of the cancer cell line from GDSC2 to identify 5-FU-related modules and hub genes. According to hub genes, HCC was classified and the machine learning model was developed by ConsensusClusterPlus and five different machine learning algorithms. Furthermore, we performed quantitative reverse transcription-polymerase chain reaction (qRT-PCR) analysis on the genes in our model.

Results: A total of 19 modules of the cancer cell line were divided by WGCNA, and the most negative correlation with 5-FU was the midnight blue module, from which 45 hub genes were identified. HCC was divided into three subgroups (C1, C2, and C3) with significant overall survival (OS) differences. OS of C1 was the shortest, which was characterized by a high clinical grade and later T stage and stage. OS of C3 was the longest. OS of C2 was between the two subtypes, and its immune infiltration was the lowest. Five out of 45 hub genes, namely, *TOMM40L*, *SNRPA*, *ILF3*, *CPSF6*, and *NUP205*, were filtered to develop a risk regression model as an independent prognostic indicator for HCC. The qRT-PCR results showed that *TOMM40L*, *SNRPA*, *ILF3*, *CPSF6*, and *NUP205* were remarkably highly expressed in hepatocellular carcinoma.

Abbreviations: 5-FU, 5-fluorouracil; HCC, hepatocellular carcinoma; WGCNA, weighted gene co-expression network analysis; OS, overall survival; TS, thymidylate synthase; TME, tumor microenvironment; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; GDSC, Genomics of Drug Sensitivity in Cancer; DEG, differentially expressed gene; CDF, cumulative distribution function; SNV, single-nucleotide variant; CNV, copy-number variant; GBM, gradient boosting machine; LASSO, least absolute shrinkage and selection operator; SVM, support vector machines; ROC, receiver operating characteristic; AUC, area under the curve; Treg, regulatory T cell; MDSC, myeloid-derived suppressor cell.

Conclusion: The HCC classification based on the sensitivity to 5-FU was in line with the prognostic differences observed in HCC and most of the genomic variation, immune infiltration, and heterogeneity of pathological pathways. The regression model related to 5-FU sensitivity may be of significance in individualized prognostic monitoring of HCC.

KEYWORDS

5-fluorouracil, hepatocellular carcinoma, machine learning, genomic variation, immune infiltration, prognosis

Introduction

Hepatocellular carcinoma (HCC) is a critical global healthcare issue with a mortality-to-morbidity ratio as high as 91.6% (Villarruel-Melquiades et al., 2023). Patients with HBV/HCV infection, cirrhosis of any cause, alcoholism, non-alcoholic steatohepatitis, or family history of HCC are considered high-risk groups for HCC, especially among men aged over 40 years old (Xie et al., 2023). Surgical treatment, including hepatectomy and orthotopic liver transplantation, is widely used for tumor eradication (Kawaguchi et al., 2016), but it is also restricted by the applicability of patients and the availability of organs (Koza et al., 2023). In a clinical practice, more than 60% of HCC cases are diagnosed in the late stage, suggesting possible missed diagnostic opportunity. At present, the best choice for advanced HCC is systemic treatment, with sorafenib and lenvatinib as the first choice (Yang et al., 2023). However, patients treated with drugs will have to face the problem of drug resistance after drug treatment. Improving the understanding of the mechanism of HCC resistance is expected to bring further benefits to patients.

As a synthetic fluorinated pyrimidine analog, 5-fluorouracil (5-FU) enters cells as an anti-metabolite, imitates molecules vital to cell growth, interferes with basic biosynthetic activity by inhibiting the effect of thymidylate synthase (TS), or mistakenly mixes its metabolites into DNA and RNA, thereby inducing cytotoxicity (Blondy et al., 2020; Mafi et al., 2023). Since its approval by FDA in 1962, 5-FU has been widely applied alone or together with other drugs in treating various cancers, such as advanced head and neck squamous cell carcinoma (Yamauchi et al., 2023), colorectal cancer (Wosiak et al., 2023), gastric cancer (Kang et al., 2014), and metastatic breast cancer (Karapetis et al., 1999; Holmes et al., 2018). 5-Fu is also a widely used chemotherapeutic drug for patients with HCC. There is an urgent need to better improve the sensitivity of HCC to chemotherapy (Hu et al., 2016), and an accurate understanding of mechanisms that contribute to a lack of or a low sensitivity of HCC to 5-FU-based treatment is a critical component of the current trend of individualized medical care. Identifying and confirming current 5-FU-based predictive biomarkers, as well as developing novel targeted medicines for HCC therapy, may enhance patients' prognoses in the future (Vodenkova et al., 2020).

In this study, genes related to 5-Fu sensitivity were screened from large data sets for identification of HCC, in order to characterize the heterogeneity of HCC from molecular aspects and tumor microenvironment (TME). Genes suitable for constructing a risk model were identified from those related to 5-Fu sensitivity, hoping to provide a promising target for understanding 5-FU resistance of HCC.

Materials and methods

Data source

Clinical data and RNA sequencing of HCC were downloaded from the LIHC project of The Cancer Genome Atlas (TCGA) database (https://cancergenome.nih.gov), and a total of 50 corresponding paracancerous tissues and 365 HCC tumor tissues were incorporated into the analysis. A set of HCC chip data numbered GSE14520 were collected from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). Another set of data were downloaded from the HCCDB database (http://lifeome.net/database/hccdb/download.html), which provides expression profiles of HCC samples. Meanwhile, pan-cancer cell line drug sensitivity and genomic data resources were acquired from the Genomics of Drug Sensitivity in Cancer (GDSC) database (https://www.cancerrxgene.org/) (Yang et al., 2013).

Weighted gene co-expression network analysis (WGCNA)

WGCNA was performed for analyzing the expression profiles of cancer cell lines downloaded from GDSC2. Samples were clustered and used to construct a gene co-expression network, from which modules were identified and then related to external data. Key drivers in the "WGCNA" package (Langfelder and Horvath, 2008) of R were analyzed based on the relationships among the module. Under the selected parameters, the sampleTree function provided by "WGCNA" was used to cluster the HCC cell lines and present the outliers. The gene expression matrices of the retained samples were extracted, and Pearson correlation was computed to calculate the correlation between twisted genes. Under different power values, we selected the optimal soft threshold β by analyzing the scale independence and average connectivity of modules using the "pickSoftThreshold" function provided by the "WGCNA" package. The "scaleFreePlot" function was adopted to evaluate whether the topology of the network was scale-free. The hierarchical clustering of genes was implemented by using the "hclust" function. The distance clustering threshold (height = 0.25, deepSplit = 3) was set by the cutreeDynamic function in the dynamicTreeCut package, and the minimum number of genes was 30 in each module. The automatic module merge step was performed using the mergeCloseModules function. The interested modules were the most relevant to 5-fluorouracil; therefore, the IC_{50} value of 5-fluorouracil for module–trait relationships was analyzed.

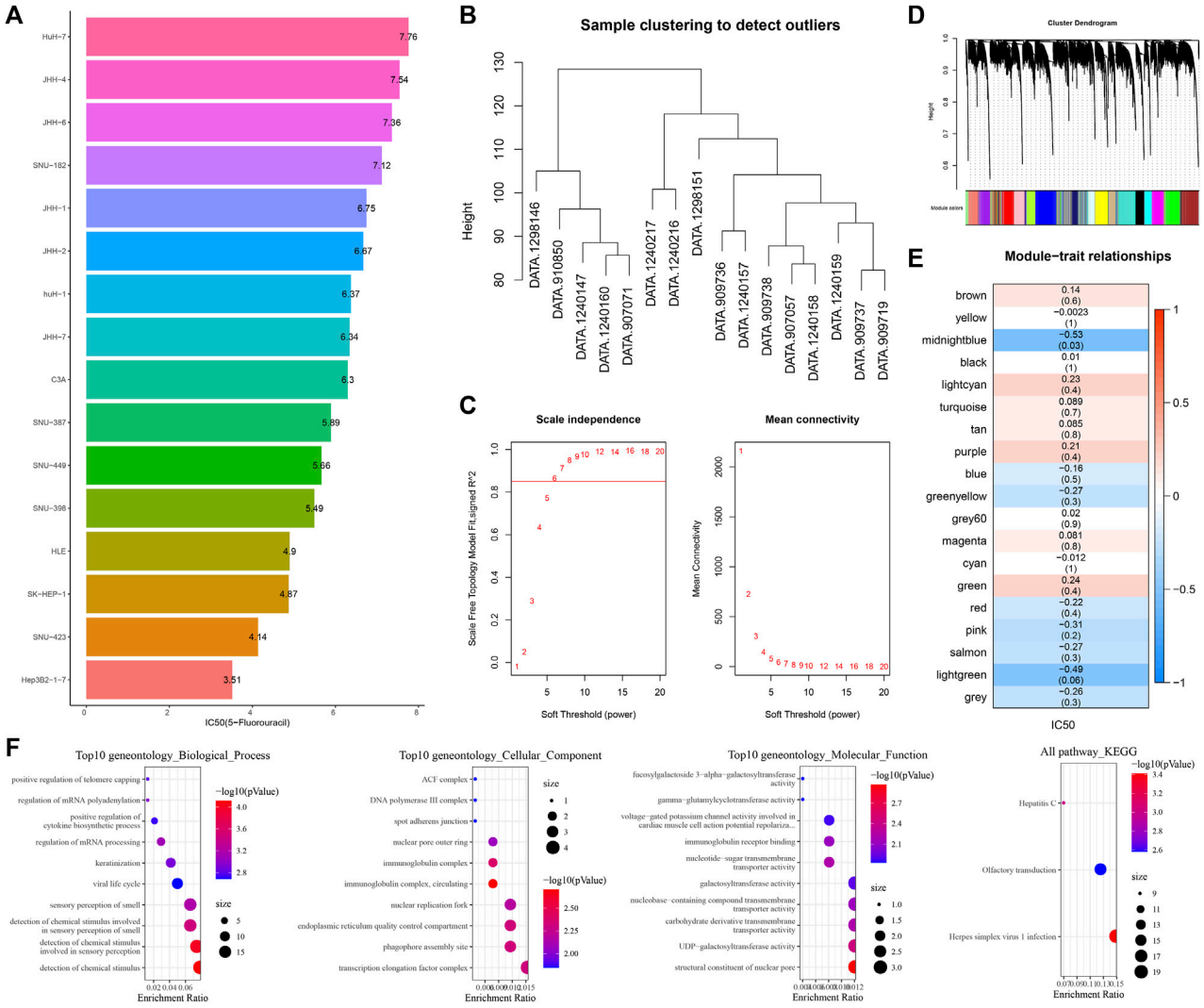


FIGURE 1
Identification of the gene module most related to 5-fluorouracil. **(A)** Bar chart shows the sensitivity of 5-fluorouracil in different hepatoma cell lines. **(B)** Clustering tree of cell samples from GDSC2. **(C)** Average connectivity corresponding to the scale-free fitting index and each soft threshold. **(D)** Cluster tree of genes in all hepatoma cell lines from GDSC2. **(E)** Correlation analysis of 19 clustered modules with 5-fluorouracil IC_{50}. The upper numbers represented correlation coefficients, and the lower numbers represented statistical *p*-values. **(F)** GO entries and KEGG pathways of gene enrichment in the midnight blue module.

Unsupervised clustering on HCC

The "limma" package (Ritchie et al., 2015) was employed to discriminate differentially expressed genes (DEGs) meeting FDR<0.05 and log2(Fold Change) > 1. DEGs were then screened by overlapping analysis with gene modules associated with 5-fluorouracil. The "ConsensusClusterPlus" package was applied to run the consensus clustering on the TCGA-LIHC data matrix (Wilkerson and Hayes, 2010). The initial step was to subsample 80% items and features. Each subsample was then partitioned into k groups. Afterward, consensus values were calculated and stored in a consensus matrix for each k-value. The output graphical plots included the consensus matrix plot and the empirical cumulative distribution function (CDF) plot.

Single-nucleotide variant (SNV) and copy-number variant (CNV) analyses

Genomic variation analysis

Genomic variation includes small insertions or deletions (indels), single-nucleotide variants, and CNVs. SNVs and CNVs, all belong to the category of genomic variation. After reading the MAF file from TCGA-LIHC, the generated MAF object was passed to the "maftools" package (Mayakonda et al., 2018) for SNV analysis and oncoplot drawing. GISTIC 2.0, which calculates a statistic involving the occurrence frequency and distortion amplitude, was employed to analyze CNV data. The characteristic of this method is to identify the regions of the genome, where anomalies occur more frequently than

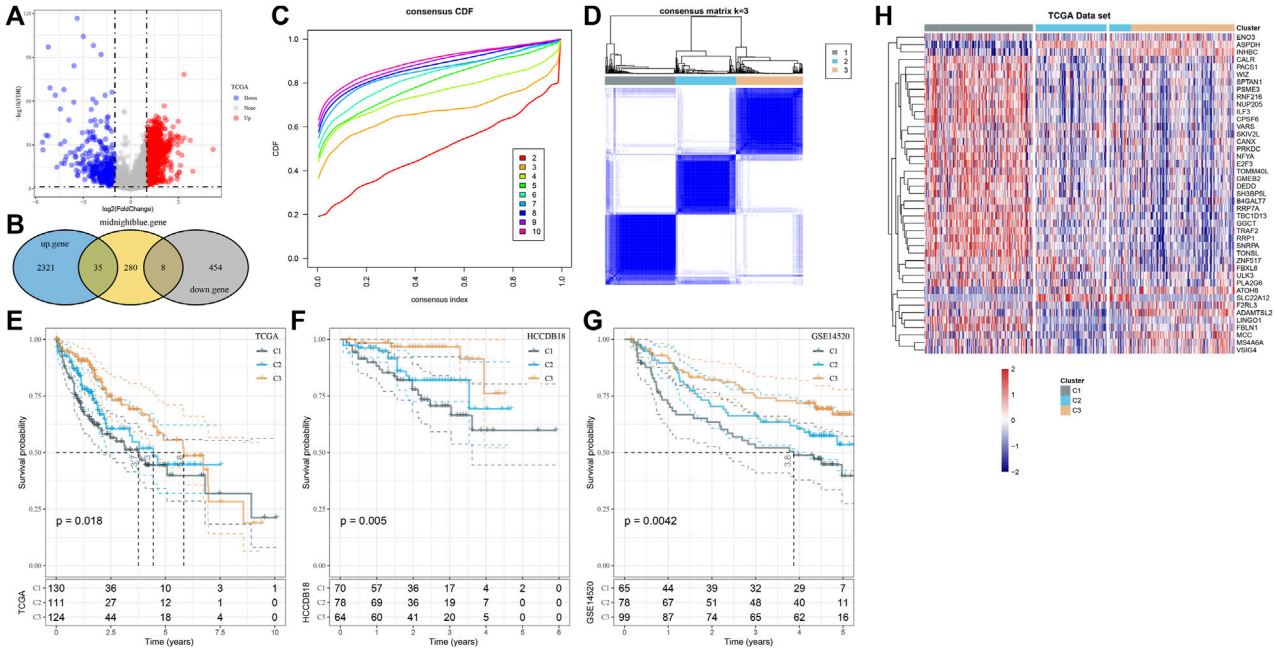


FIGURE 2
HCC was classified by identifying hub genes in the midnight blue module. **(A)** Difference analysis of log2 (TPM+1) of TCGA-LIHC between normal and HCC tissues. **(B)** The Hub gene of the midnight blue module was identified by overlap analysis of differential genes and the midnight blue module. **(C)** CDF plot displays consensus distributions for each k-value. **(D)** Consensus matrix shows the clustering partition of k = 3. The blue color represented the distance-based similarity between the samples. **(E–G)** Survival curves of three clusters in TCGA-LIHC, HCCB18, and GSE145203 cohorts. **(H)** Expression of 43 genes is shown in the form of a heatmap.

accidentally expected and gives more weight to high-amplitude events (high-level copy number gain or homozygous deletion) that are unlikely to represent random distortions. For each important region, the method defines a "peak region" with a maximum aberration frequency and amplitude (Beroukhim et al., 2007).

Immune cell infiltration analysis

The ESTIMATE algorithm, which leverages the properties of the TCGA-LIHC transcriptional profiles to infer the degrees of stromal and immune cell infiltration, was applied to determine the ESTIMATE score (Yoshihara et al., 2013). Different methods for assessing the level of immune infiltration, including CIBERSORT, ssGSEA, MCPcounter, and TIMER, were properly applied. CIBERSORT was used to measure the intra-sample (within-leukocyte) proportions of immune cell populations (Newman et al., 2015). Different from CIBERSORT, MCPcounter outputs the estimated abundance of each cell population, thereby enabling a comparison between samples to be expressed in arbitrary units (Becht et al., 2016). TIMER takes tissue specificity into account when estimating immune cell populations (Li et al., 2017), and this method helps identify associations between six types of immune cell infiltration and clustering in the TCGA-LIHC cohort.

Establishment of a risk stratification tool using multiple machine learning analysis

Univariate COX regression analysis identified prognostic genes from the intersection of DEGs and 5-fluorouracil-related gene modules, and introduced five different machine learning algorithms to complete the task of variable selection, including gradient boosting machine (GBM), least absolute shrinkage and selection operator (LASSO) regression, support vector machines (SVM), Decision Trees, and Random Forest. The intersection of genes selected by each machine learning algorithm was used for stepwise regression analysis in multiple linear regression to generate a fitting regression model to evaluate the risk of samples in different HCC cohorts.

Nomogram construction

This study integrated age, gender, T stage, stage, grade, and RiskScore information, and performed univariate COX and multivariate COX analyses to determine independent prognostic factors which influenced the prognosis of HCC. Based on these independent prognostic factors, we developed a nomogram for predicting HCC survival. Based on the actual and predicted survival outcomes, we developed calibration curves to validate the predictive power of the nomogram. In addition, we also graphed decision curves to determine the prognostic guidance value of the nomogram and RiskScore.

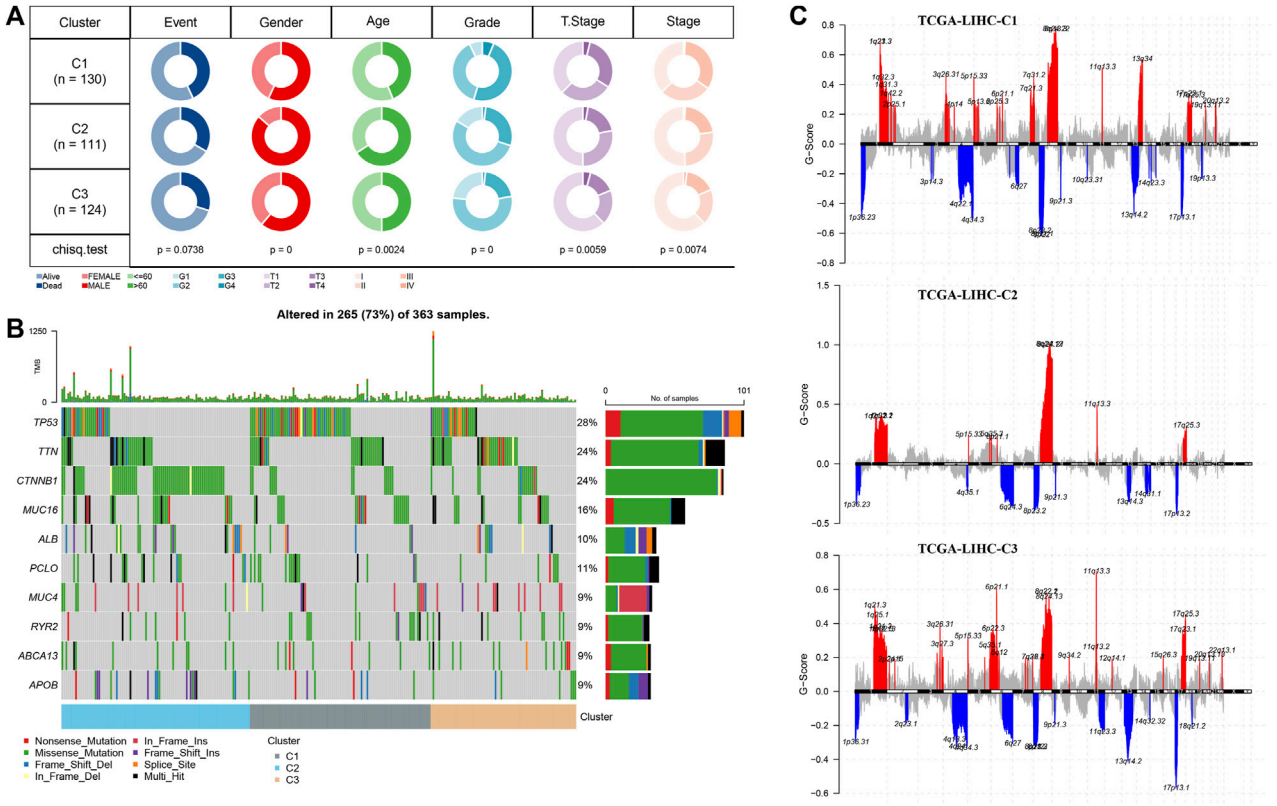


FIGURE 3
Clinical and genomic alteration features of molecular subtypes **(A)** Waterfall map of somatic mutation in three subtypes. **(B)** Manhattan plot shows the CNV situation of each subtype at the chromosomal level. **(C)** Analysis and comparison of clinical features of three subtypes. The red represents the q-arm of the chromosome, and the blue represents the p-arm of the chromosome.

Cell culture and transient transfection

HCC cell lines including Hep3B2.1-7 and Huh-7 were obtained from COBIOER (Nanjing, China). Hep3B2.1-7 and Huh-7 cells were cultured in DMEM F12 with 10% FBS (Gibco, Thermo Fisher, USA). Human liver epithelial cells (THLE-3) were obtained from ATCC (Manassas, VA, USA) and stored in BEGM (Lonza, Walkersville). Cells were grown at 37°C in a humidified environment containing 5% CO_2.

Quantitative reverse transcription polymerase chain reaction (qRT-PCR)

TRIzol reagent (Thermo Fisher, USA) was used to extract total RNA from the Hep3B2.1-7, Huh-7, and THLE-3 cell lines. Using FastStart Universal SYBR Green Master (Roche, Switzerland), quantitative reverse transcription polymerase chain reaction (qRT-PCR) was performed on the RNA extracted from each sample (2 μg) on a LightCycler 480 PCR System (Roche, USA). The cDNA was utilized as a template with a reaction volume of 20 μl (2 μl of cDNA template, 10 μl of PCR mixture, 0.5 μl of forward and reverse primers, and an appropriate water volume). The following procedures were utilized for the PCR reactions: cycling conditions started with an initial DNA denaturation phase at 95°C for 30 s, followed by 45 cycles at 94°C for 15 s, 56°C for 30 s, and 72°C for 20 s. Three separate analyses were performed on each sample. Based on the $2^{-\Delta\Delta CT}$ method, data from the threshold cycle (CT) were obtained and standardized to the levels of GAPDH in each sample. The expression levels of mRNA were compared to controls obtained from normal tissues. Sequence lists of primer pairs for the target genes are summarized in Supplementary Table S1.

Statistical analysis

All statistical analysis and verification were conducted in the R code. A chi-squared test was adopted to detect differences in clinical characteristics between subtypes. The survival of the samples was presented by the Kaplan–Meier curve. Statistical survival difference was analyzed using the log-rank test. The time-dependent receiver operating characteristic (ROC) curve and the area under the curve (AUC) of the risk layering tool were generated and calculated using the timeROC package. $p < 0.05$ meant that the difference was statistically significant. For the results of the statistical analyses, ns indicated no significance,

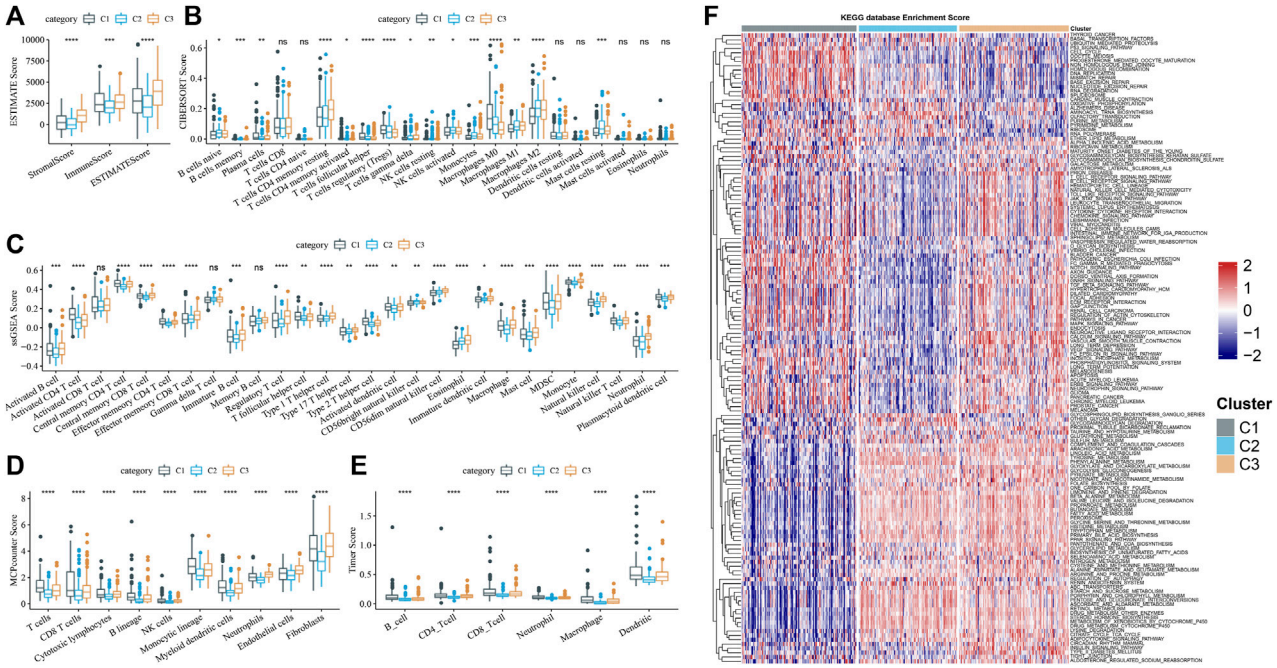


FIGURE 4
Immune filtration mode for three clusters **(A)** Subtype stromal score, immune score, and ESTIMATE score obtained from running ESTIMATE. **(B–E)** Enrichment differences among subtypes of immune cells judged by CIBERSORT, ssGSEA, MCPcounter, and TIMER. **(F)** Differences in enrichment of pathways in clusters were archived in a heatmap.

* indicated $p < 0.05$, ** indicated $p < 0.01$, *** indicated $p < 0.001$, and **** indicated $p < 0.0001$.

Results

Identification of the gene module most related to 5-fluorouracil

We initially examined the sensitivity of different HCC cell lines to 5-fluorouracil. The IC_{50} value of 5-fluorouracil was the lowest in Hep3B2-1-7 cells and the highest in HuH-7 cells, meaning that Hep3B2-1-7 cells were the most sensitive to 5-fluorouracil and HuH-7 cells had the strongest resistance to 5-fluorouracil (Figure 1A). All cell samples from GDSC2 were clustered (Figure 1B). The soft-threshold power satisfying the scale-free topology of the network was 6, the corresponding R2 value was 0.86, and the average connectivity was very close to 0 (Figure 1C). Next, all genes were clustered into 19 interacting modules (Figure 1D). Among the 19 clustering modules, their correlations with 5-fluorouracil resistance were analyzed. The result showed that midnight blue was the module with the highest significant negative correlation with 5-fluorouracil sensitivity (Figure 1E). We analyzed GO and KEGG annotation of genes within the midnight blue module. Biological process was annotated to regulation of mRNA processing, regulation of mRNA polyadenylation, and positive regulation of telomere capping in GO terms. The protein products of these genes might be the components of the transcription elongation factor complex, DNA polymerase III complex (Figure 1F).

HCC was classified by identifying hub genes in the midnight blue module

The differences between normal tissues and HCC tissues of log2 (TPM+1) of TCGA-LIHC were analyzed. A total of 2,356 genes with log2 (Fold Change) > 1 and FDR <0.05, as well as 462 genes with log2 (Fold Change) <-1 and FDR <0.05, were identified (Figure 2A), and 35 upregulated and 8 downregulated DEGs were also detected from the midnight blue module (Figure 2B). The samples of TCGA-LIHC were clustered according to the expression of the aforementioned 43 genes. The CDF plot helps in finding the k-value that reached the approximate maximum value of 3 (Figure 2C). The consensus matrix showed the clustering partition was k = 3 (Figure 2D). Complete separation of survival curves and overall survival (OS) of the three clusters in the detected TCGA-LIHC, HCCB18, and GSE145203 cohorts had significant differences among subtypes. Specifically, C3 had the most satisfactory survival outcome when compared with C1 and C2, while C1 had the shortest OS (Figures 2E–G). Next, the expression of 43 genes was also shown in a heatmap, which demonstrated that most genes were highly expressed in C1 than C3 and C2 (Figure 2H).

Clinical and genomic alteration features of molecular subtypes

Here, in the C1, C2, and C3 subtypes, we analyzed the status of genomic variations. The TP53 mutation rate of C1 was also significantly higher than that of C2 and C3. The mutation rate of CTNNB1 in C2 was the highest, which was significantly higher than that of C1 and C3. The mutation rate of TTN in C3 was the highest,

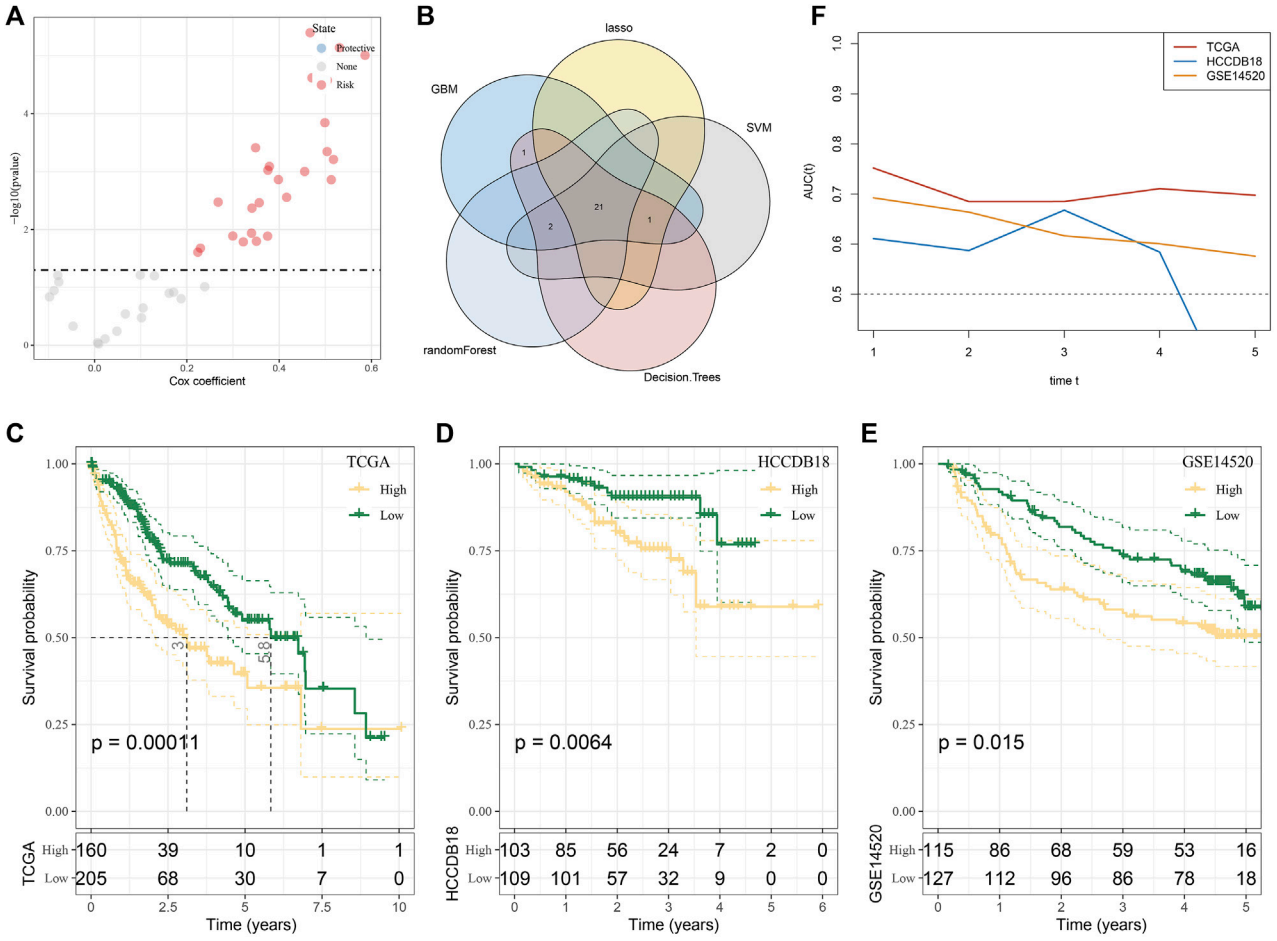


FIGURE 5
Selection and verification of genes suitable for constructing a risk model in the midnight blue module. **(A)** Volcanic map showing univariate Cox regression analysis for 43 genes. **(B)** Intersection of genes selected by LASSO regression, GBM, SVM, random forest, and decision tree. **(C–E)** Survival stratification curve of the regression model in the test set TCGA-LIHC and two independent verification sets, namely, HCCDB18 and GSE14520. **(F)** ROC curve of regression model in predicting 1–5 years survival of cases in TCGA-LIHC and GSE14520.

which was significantly higher than that of C1 and C2 (Figure 3A). From the Manhattan plot, we observed the CNV of each subtype at the chromosomal level, the number of high-level DNA copies amplified, and deleted in C2 was significantly less than that in C1 and C3 (Figure 3B). Comparison of the clinical characteristics of the three subtypes showed that there were more male subject samples than female subject samples in each subtype. Differences in sex, age distribution and grade, and T stage and stage characteristics were statistically significant among the three subtypes. Compared with the other two clusters, C2 had the highest proportion of male subjects and samples aged over 60 years old. C1 with the shortest OS was characterized by a high proportion of clinical grade and later T stage and stage (Figure 3C).

Immune filtration mode for three clusters

By running ESTIMATE, the ESTIMATE score, stromal score, and immune score of each cluster were calculated, which showed significant differences among the three clusters, and the level was the lowest in C2 (Figure 4A). Of the 22 immune cells provided by CIBERSORT, 15 showed significant differences in infiltration among three subtypes (Figure 4B). Memory cells, immunosuppressive cells (regulatory T cells (Treg) and myeloid-derived suppressor cells (MDSC)), and cytotoxic cells (CD8 T cells, natural killer (NK) cells, and NK T cells) identified in 28 TIL subpopulations showed differential infiltration among the three subtypes, and almost all of them had the least infiltration in C2 (Figure 4C). Combining the results of MCPcounter and TIMER analysis, the infiltration of CD4 T cell, T cells, B cells, macrophage, neutrophils, CD8 T cells, endothelial cells, and dendritic cells, and fibroblasts in C2 was significantly lower than that in C1 and C3 (Figures 4D, E). The differences in enrichment of pathways in clusters were archived in a heatmap, from which we could observe that the enrichment level of most pathways relevant to metabolism decreased in C1, such as linoleic acid metabolism, tyrosine metabolism, phenylalanine metabolism, and pyruvate metabolism (Figure 4F).

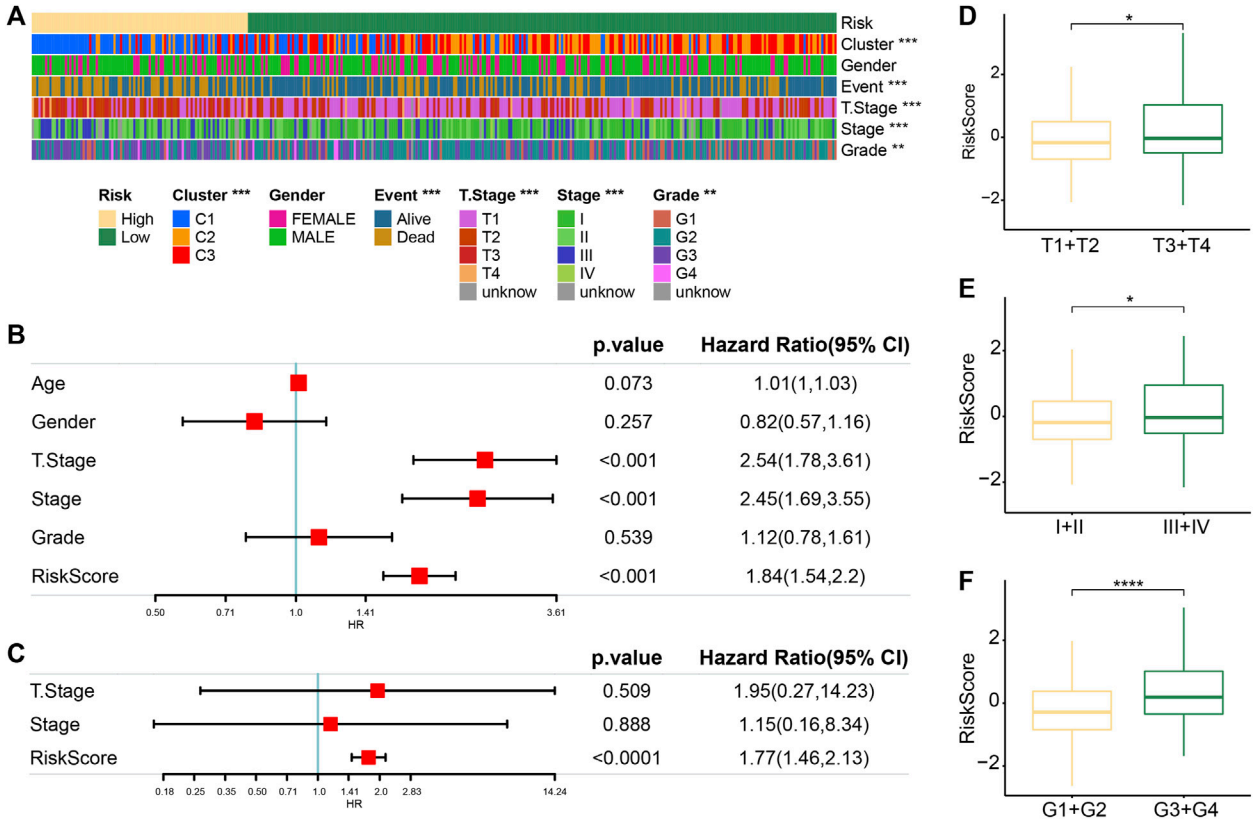


FIGURE 6
Regression model was an independent predictor of prognosis and clinical characteristics of HCC. **(A)** Heatmaps of clinical traits corresponding to risk score. **(B)** Univariate Cox regression analysis of the risk score and clinical traits. **(C)** Multivariate Cox regression analysis of prognostic traits of HCC based on univariate Cox regression analysis. **(D)** Difference in the risk score between samples stratified according to the T stage. **(E)** Differences in the risk score between samples stratified by stage. **(F)** Correlation between the grade and the risk score of the sample.

Selection and verification of genes suitable for constructing a risk model in the midnight blue module

In the midnight blue module, a total of 43 genes were identified as hub genes, and 25 HCC prognostic genes were screened from these hub genes by univariate Cox regression analysis (Figure 5A). Machine learning models of these genes were established based on machine learning methods, including LASSO regression, GBM, SVM, Random Forest, and Decision Tree. A total of 21 genes belonged to the intersection of five machine learning models (Figure 5B). The stepwise regression method screened five genes from the 21 genes suitable for constructing a risk model, including *TOMM40L*, *SNRPA*, *ILF3*, *CPSF6* and *NUP205*. The risk coefficient of each gene was obtained from multivariate Cox regression analysis, and a fitted regression model was generated: RiskScore = 0.293*TOMM40L+0.558*SNRPA-0.823*ILF3+0.493*CPSF6+0.464*NUP205. Regression models were used to calculate risk scores in the test set TCGA-LIHC and two independent verification sets HCCDB18 and GSE14520. A significant negative correlation with the sample OS in their cohorts was found, with the patients of a high-risk score showing a shorter survival time (Figures 5C–E). Meanwhile, the regression model showed stability and availability in predicting 1–5 year(s) OS of cases in TCGA-LIHC and GSE14520 (Figure 5F).

Regression model was an independent predictor of the prognosis and clinical characteristics of HCC

The heatmap of clinical characteristics corresponding to risk score was drawn. Differences in molecular subtypes, survival status, T stage and stage, and grade proportion were statistically significant between low-risk and high-risk groups. The high-risk group was characterized by a high C1 ratio, high mortality rate, later T stage and stage, and grade distribution, while these clinical traits in the low-risk group were significantly weaker (Figure 6A). The actual univariate Cox regression analysis showed that the risk score and T stage and stage were significantly correlated with HCC prognosis. The risk score was identified as an independent prognostic index of HCC by Multivariate Cox regression analysis (Figures 6B, C). Synthesizing information on T Stage, Stage, and RiskScore, we constructed the nomogram for assessing clinical outcomes of HCC patients at 1, 3, and 5 years (Supplementary Figure S1A). The calibration curves showed that the predicted clinical outcomes

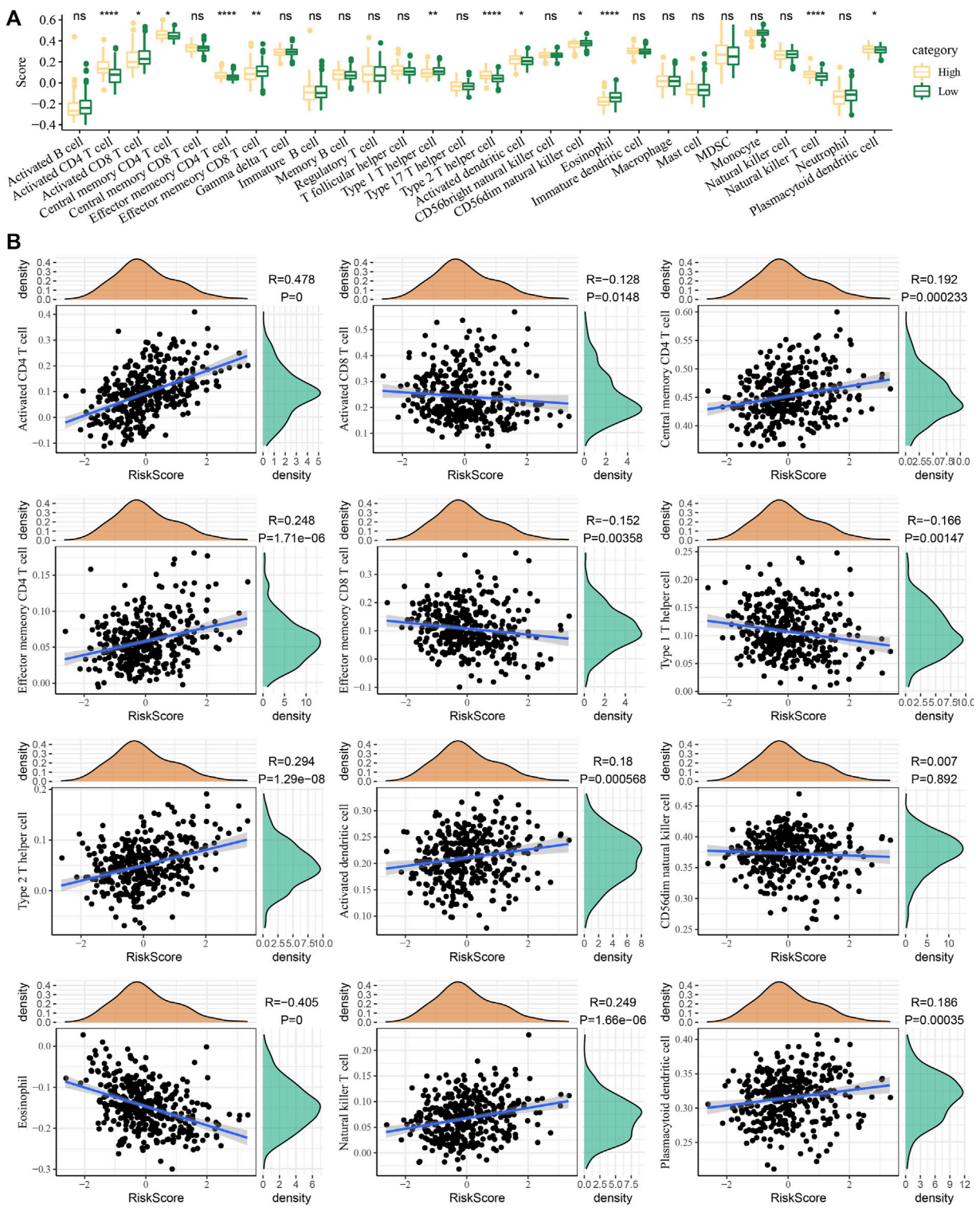


FIGURE 7
Relationship between the risk score and immune infiltration. **(A)** Immune cell infiltration was evaluated according to the risk score stratification. **(B)** Correlation of the ssGSEA score of the risk score and immune cells.

fit well with the actual observed clinical outcomes and that the nomogram had a good predictive value (Supplementary Figure S1B). In addition, the decision curve also showed that there is an excellent applicability of the nomogram and RiskScore in assessing clinical outcomes in HCC (Supplementary Figure S1C). The risk score of the T3–T4 stage, stage III–IV, and G3–G4 samples was significantly higher than that of the T1–T2 stage, stage I–II, and G1–G2 samples, respectively (Figures 6D–F).

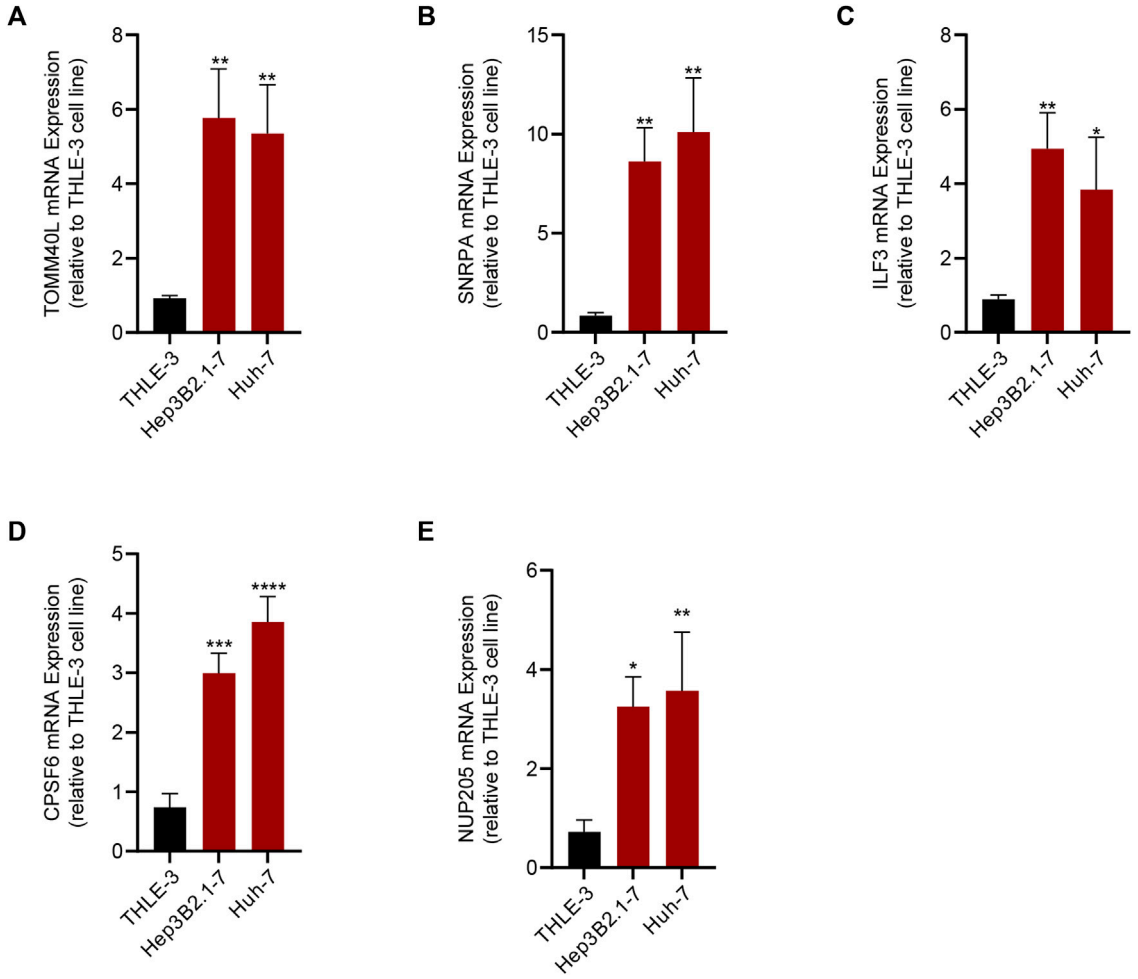


FIGURE 8
Results of qRT-PCR of the five genes that composed the RiskScore. **(A)** TOMM40L; **(B)** SNRPA; **(C)** ILF3; **(D)** CPSF6; **(E)** NUP205.

Relationship between the risk score and immune infiltration

The degree of immune cell infiltration was evaluated according to the risk score. We detected that the cells with the most different degrees of infiltration in the high-risk and low-risk groups were type 2 T helper cells, central memory CD4 T cells, type 1 T helper cells, plasmacytoid dendritic cells, effector memory CD4 T cells, activated CD4 T cells, activated CD8 T cells, eosinophils, natural killer T cells, CD56 dim natural killer cells, activated dendritic cells, and effector memory CD8 T cells (Figure 7A). The ssGSEA score correlation analysis of the risk score and immune cells showed that the degree of correlation between the 12 cells had infiltration and risk score differences between the high-risk and low-risk groups. The correlation between risk score and CD56 dim natural killer cells was almost negligible, and the other 11 cells showed a significant correlation with risk score. Among them, effector memory CD4 T cells, activated CD4 T cells, natural killer T cells, plasmacytoid dendritic cells, type 2 T helper cells, central memory CD4 T cells, and activated dendritic cells showed a significant positive correlation with the risk score, while activated CD8 T cells, effector memory CD8 T cells and type 1 T helper cells, and eosinophils were significantly negatively correlated with the risk sore (Figure 7B).

PCR validation of RiskScore

To verify the reliability of the RiskScore, we detected the expression of the five genes by PCR. The results of PCR corroborated the reliability of the RiskScore, and we found that *TOMM40L*, *SNRPA*, *ILF3*, *CPSF6*, and *NUP205* were significantly upregulated in the HCC cell lines Hep3B2.1-7 and Huh-7 compared to human normal liver epithelial cells THLE-3 (Figures 8A–E).

Discussion

The large-scale drug genome cell line database has in-depth multi-group characterization and extensive pharmacological characteristics of human cancer cell lines, and is an important tool to reveal the potential mechanism of inducing drug sensitivity of anticancer drug compounds (Kusch and Schuppert, 2020). This study explored how HCC

heterogeneity of HCC was affected by the molecules related to 5-FU sensitivity based on the expression profile of the cancer cell line in the largest public resource and the sensitivity data on 5-FU, a commonly used cancer chemotherapy drug. The WGCNA analysis identified the midnight blue gene module with the highest correlation with 5-FU sensitivity and 43 hub genes in the module. Three subgroups of HCC were defined according to their expression. This classification supported most of the genomic variation, TME, and pathological pathway heterogeneity observed in HCC.

Currently, 5-FU was the mainstream tumor chemotherapeutic agent (Blondy et al., 2020; Vodenkova et al., 2020). Accumulating evidence illustrated that 5-FU exhibited cytotoxicity by binding to DNA or RNA and modulating DNA synthesis-induced cell cycle disruption or apoptosis (Sethy and Kundu, 2021). Cell cycle abnormalities are typical in tumor cells, and inhibition of the tumor cell cycle is essential for suppressing cell proliferation and spreading, and even restoring immune cell surveillance functions (Liu et al., 2022). Based on the results of GO and KEGG annotations, we showed that midnight blue endogenous genes were closely associated not only with mRNA processing, regulation of mRNA polyadenylation, and positive regulation of the telomere capping pathway but also with the transcription elongation factor complex, DNA polymerase III complex synthesis. 5-FU disrupted the homologous recombination repair process in cells, leading to DNA damage and inhibition of proliferation in tumor cells (Srinivas et al., 2015). The mRNA processing, regulation of mRNA polyadenylation, and positive regulation of telomere capping were important regulators in the cell cycle in normal cells. The genes within the midnight blue module were recognized as gene modules sensitive to 5-FU treatment, suggesting the possibility that HCC might act through these biological processes when treated with 5-FU.

In terms of survival outcomes, C3 possessed the most satisfied survival outcome when compared with C1 and C2, and C1 demonstrated the shortest OS. Each subgroup also showed its own unique clinical characteristics, C2 had the highest proportion of male subjects and samples aged over 60 years old than C1 and C3. C1 with the shortest OS was characterized by a high proportion of clinical grade and later T stage and stage, and most metabolic pathways of this subtype were significantly inhibited. These bad characteristics were also reflected in OS, and C1 had the most unfavorable survival outcome. The gene with the highest mutation rate was TP53 in C1, CTNNB1 in C2, and TTN in C3. This indicated that C1 was a tumorigenesis subtype driven by TP53 mutation, C2 was a tumorigenesis subtype driven by CTNNB1 mutation, and C3 was a tumorigenesis subtype driven by TTN mutation. TP53 mutations and CTNNB1 mutations were most common in HCC. In a study by Gao et al. (2019), HCC patients characterized by TP53 mutations had a dysregulated cell cycle and DNA damage repair pathways, and TP53 was the gene with the highest mutation frequency. Low TP53 levels inhibited HCC development. Significant activation of metabolic reprogramming was demonstrated in patients enriched with CTNNB1 mutations. This phenomenon promoted glycolytic metabolic intensity and cell proliferation in HCC. It was also confirmed that the frequency of the R249S mutation in TP53 revealed the risk of HCC, and TP53 deletion increased the viability of hepatocellular carcinoma cells and the trend of poor prognosis (Lam et al., 2022). Interestingly, the C3 isoform might be a TTN mutation-driven molecular subtype that exhibited a high mutational profile. However, Kunadirek et al. (2021) noted that TTN mutations in blood predicted unfavorable prognostic status in HCC patients. The results of survival analysis in this study demonstrated that the C1 subtype predicted had an unfavorable prognosis, C2 subtype had a moderate prognosis, and C3 subtype had the best prognosis. However, we must point out that there were differences in sampling between them as tissue samples in our study were different from the blood samples in Kunadirek's. Second, HCC was a highly heterogeneous tumor both in terms of genomic composition and gene mutations (Jeng et al., 2015). The research challenges posed by the heterogeneity remain difficult to resolve. Overall, the C1 subtype in our study was enriched with TP53 mutations, and patients enriched with TP53 mutations tended to have unfavorable survival outcomes. Patients enriched with CTNNB1 mutations showed significant metabolic reprogramming activity, and inhibition of glycolytic signaling could be considered to target the C2 subtype to improve prognosis. Different treatment options could be considered for patients with C1 and C2 subtypes to achieve precision cancer treatment.

Personalized treatment for HCC patients has been increasingly recognized and applied in the clinical field. The development of risk models represents an important step toward personalized HCC monitoring. Although many risk models have been published, few are used in routine nursing to provide information for HCC monitoring decisions (Innes and Nahon, 2023). In this study, five of the 43 hub genes in the midnight blue module were used to develop a risk regression model, which was independent and had strong discriminating power for predicting HCC prognosis and indicating clinical traits. The pathological role and regulatory mechanism of some of them in cancerization have been recognized. SNRPA is a shear factor associated with microvascular invasion and promotes the metastasis of HCC by activating the NOTCH1/Snail pathway mediated by the circSEC62/miR-625–5p axis (Mo et al., 2023). ILF3 is overexpressed in patients with primary colorectal cancer and promotes tumor growth by directly regulating the mRNA stability of *SGOC* gene (Li et al., 2020). *CPSF6* is upregulated in HCC and induces metabolic changes in hepatocytes through the alternative polyadenylation of *NQO1* (Tan et al., 2021). Although the potential effects of these genes on cancer have been reported, the risk model integrating these genes was an innovative exploration and could be used as an indicator of the prognosis of HCC.

In summary, this study classified HCC subtypes based on the sensitivity to 5-FU. The results supported the prognostic differences observed in HCC and the heterogeneity of most genomic variations, TME, and pathological pathways. This study also provided an independent prognostic risk regression model integrated with five 5-FU-related genes, contributing to the study of individualized HCC monitoring.

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.

Author contributions

All authors contributed to this present work: XG and SL designed the study, and XM and DH acquired the data. PL drafted the manuscript, and XG revised the manuscript. All authors contributed to the article and approved the submitted version.

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/fphar.2023.1252805/full#supplementary-material

References

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17 (1), 218. doi:10.1186/s13059-016-1070-5

Beroukhim, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., et al. (2007). Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl. Acad. Sci. U. S. A.* 104 (50), 20007–20012. doi:10.1073/pnas.0710052104

Blondy, S., David, V., Verdier, M., Mathonnet, M., Perraud, A., and Christou, N. (2020). 5-Fluorouracil resistance mechanisms in colorectal cancer: from classical pathways to promising processes. *Cancer Sci.* 111 (9), 3142–3154. doi:10.1111/cas.14532

Gao, Q., Zhu, H., Dong, L., Shi, W., Chen, R., Song, Z., et al. (2019). Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. *Cell* 179 (5), 1240. doi:10.1016/j.cell.2019.10.038

Holmes, F. A., Hellerstedt, B. A., Pippen, J. E., Jr., Vukelja, S. J., Collea, R. P., Kocs, D. M., et al. (2018). Five-year results of a phase II trial of preoperative 5-fluorouracil, epirubicin, cyclophosphamide followed by docetaxel with capecitabine (wTX) (with trastuzumab in HER2-positive patients) for patients with stage II or III breast cancer. *Cancer Med.* 7 (6), 2288–2298. doi:10.1002/cam4.1472

Hu, Z., Lv, G., Li, Y., Li, E., Li, H., Zhou, Q., et al. (2016). Enhancement of anti-tumor effects of 5-fluorouracil on hepatocellular carcinoma by low-intensity ultrasound. *J. Exp. Clin. Cancer Res.* 35, 71. doi:10.1186/s13046-016-0349-4

Innes, H., and Nahon, P. (2023). Statistical perspectives on using hepatocellular carcinoma risk models to inform surveillance decisions. *J. Hepatol.* 2023, 5. doi:10.1016/j.jhep.2023.05.005

Jeng, K. S., Chang, C. F., Jeng, W. J., Sheen, I. S., and Jeng, C. J. (2015). Heterogeneity of hepatocellular carcinoma contributes to cancer progression. *Crit. Rev. Oncol. Hematol.* 94 (3), 337–347. doi:10.1016/j.critrevonc.2015.01.009

Kang, B. W., Kim, J. G., Kwon, O. K., Chung, H. Y., and Yu, W. (2014). Non-platinum-based chemotherapy for treatment of advanced gastric cancer: 5-fluorouracil, taxanes, and irinotecan. *World J. Gastroenterol.* 20 (18), 5396–5402. doi:10.3748/wjg.v20.i18.5396

Karapetis, C. S., Patterson, W. K., Pittman, K. B., Kotasek, D., and Sage, R. E. (1999). Treatment of metastatic breast cancer with continuous infusional 5 fluorouracil. *Aust. N. Z. J. Med.* 29 (4), 517–522. doi:10.1111/j.1445-5994.1999.tb00753.x

Kawaguchi, Y., Honda, G., Endo, I., Cherqui, D., and Kokudo, N. (2016). Current technical issues for surgery of primary liver cancer. *Liver Cancer* 6 (1), 51–58. doi:10.1159/000449345

Koza, A., Bhogal, R. H., Fotiadis, N., and Mavroeidis, V. K. (2023). The role of ablative techniques in the management of hepatocellular carcinoma: indications and outcomes. *Biomedicines* 11 (4), 1062. doi:10.3390/biomedicines11041062

Kunadirek, P., Chuaypen, N., Jenjaroenpun, P., Wongsurawat, T., Pinjaroen, N., Sirichindakul, P., et al. (2021). Cell-free DNA analysis by whole-exome sequencing for hepatocellular carcinoma: a pilot study in Thailand. *Cancers (Basel)* 13 (9), 2229. doi:10.3390/cancers13092229

Kusch, N., and Schuppert, A. (2020). Two-step multi-omics modelling of drug sensitivity in cancer cell lines to identify driving mechanisms. *PLoS One* 15 (11), e0238961. doi:10.1371/journal.pone.0238961

Lam, Y. K., Yu, J., Huang, H., Ding, X., Wong, A. M., Leung, H. H., et al. (2022). TP53 R249S mutation in hepatic organoids captures the predisposing cancer risk. *Hepatology* 78, 727–740. doi:10.1002/hep.32802

Langfelder, P., and Horvath, S. (2008). Wgcna: an R package for weighted correlation network analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559

Li, K., Wu, J. L., Qin, B., Fan, Z., Tang, Q., Lu, W., et al. (2020). ILF3 is a substrate of SPOP for regulating serine biosynthesis in colorectal cancer. *Cell Res.* 30 (2), 163–178. doi:10.1038/s41422-019-0257-1

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). Timer: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* 77 (21), e108–e110. doi:10.1158/0008-5472.CAN-17-0307

Liu, J., Peng, Y., and Wei, W. (2022). Cell cycle on the crossroad of tumorigenesis and cancer therapy. *Trends Cell Biol.* 32 (1), 30–44. doi:10.1016/j.tcb.2021.07.001

Mafi, A., Rezaee, M., Hedayati, N., Hogan, S. D., Reiter, R. J., Aarabi, M. H., et al. (2023). Melatonin and 5-fluorouracil combination chemotherapy: opportunities and efficacy in cancer therapy. *Cell Commun. Signal* 21 (1), 33. doi:10.1186/s12964-023-01047-x

Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28 (11), 1747–1756. doi:10.1101/gr.239244.118

Mo, Z., Li, R., Cao, C., Li, Y., Zheng, S., Wu, R., et al. (2023). Splicing factor SNRPA associated with microvascular invasion promotes hepatocellular carcinoma metastasis through activating NOTCH1/Snail pathway and is mediated by circSEC62/miR-625-5p axis. *Environ. Toxicol.* 38 (5), 1022–1037. doi:10.1002/tox.23745

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Sethy, C., and Kundu, C. N. (2021). 5-Fluorouracil (5-FU) resistance and the new strategy to enhance the sensitivity against cancer: implication of DNA repair inhibition. *Biomed. Pharmacother.* 137, 111285. doi:10.1016/j.biopha.2021.111285

Srinivas, U. S., Dyczkowski, J., Beissbarth, T., Gaedcke, J., Mansour, W. Y., Borgmann, K., et al. (2015). 5-Fluorouracil sensitizes colorectal tumor cells towards double stranded DNA breaks by interfering with homologous recombination repair. *Oncotarget* 6 (14), 12574–12586. doi:10.18632/oncotarget.3728

Tan, S., Zhang, M., Shi, X., Ding, K., Zhao, Q., Guo, Q., et al. (2021). CPSF6 links alternative polyadenylation to metabolism adaption in hepatocellular carcinoma progression. *J. Exp. Clin. Cancer Res.* 40 (1), 85. doi:10.1186/s13046-021-01884-z

Villarruel-Melquiades, F., Mendoza-Garrido, M. E., Garcia-Cuellar, C. M., Sanchez-Perez, Y., Perez-Carreon, J. I., and Camacho, J. (2023). Current and novel approaches in the pharmacological treatment of hepatocellular carcinoma. *World J. Gastroenterol.* 29 (17), 2571–2599. doi:10.3748/wjg.v29.i17.2571

Vodenkova, S., Buchler, T., Cervena, K., Veskrnova, V., Vodicka, P., and Vymetalkova, V. (2020). 5-fluorouracil and other fluoropyrimidines in colorectal cancer: past, present and future. *Pharmacol. Ther.* 206, 107447. doi:10.1016/j.pharmthera.2019.107447

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Wosiak, A., Szmajda-Krygier, D., Pietrzak, J., Boncela, J., and Balcerczak, E. (2023). Assessment of the influence of 5-fluorouracil on SMAD4 and TGFB1 gene expression, apoptosis induction and DNA damage in human cell lines. *Bioeng. (Basel).* 10 (5), 570. doi:10.3390/bioengineering10050570

Xie, D., Shi, J., Zhou, J., Fan, J., and Gao, Q. (2023). Clinical practice guidelines and real-life practice in hepatocellular carcinoma: a Chinese perspective. *Clin. Mol. Hepatol.* 29 (2), 206–216. doi:10.3350/cmh.2022.0402

Yamauchi, M., Minesaki, A., Ishida, T., Sato, Y., Okamura, S., Shuto, H., et al. (2023). Induction chemotherapy with 5-fluorouracil, cisplatin, and cetuximab in advanced head and neck squamous cell carcinoma. *Vivo* 37 (3), 1275–1280. doi:10.21873/invivo.13205

Yang, C., Zhang, H., Zhang, L., Zhu, A. X., Bernards, R., Qin, W., et al. (2023). Evolving therapeutic landscape of advanced hepatocellular carcinoma. *Nat. Rev. Gastroenterol. Hepatol.* 20 (4), 203–222. doi:10.1038/s41575-022-00704-9

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., et al. (2013). Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 41, D955–D961. doi:10.1093/nar/gks1111

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

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 14 September 2023
DOI 10.3389/fphar.2023.1249145

OPEN ACCESS

EDITED BY
Lin Qi,
Central South University, China

REVIEWED BY
Koulong Zheng,
Second Affiliated Hospital of Nantong University, China
Fangdie Ye,
Fudan University, China
Aierpati Maimaiti,
First Affiliated Hospital of Xinjiang Medical University, China, in collaboration with reviewer FY

*CORRESPONDENCE
Chang-Lin Zhai,
yesterdaygun@126.com
Bo Wang,
787553459@qq.com

†These authors have contributed equally to this work

RECEIVED 28 June 2023
ACCEPTED 04 September 2023
PUBLISHED 14 September 2023

CITATION
Yuan N, Pan H-H, Liang Y-S, Hu H-L, Zhai C-L and Wang B (2023), Identification of prognostic and diagnostic signatures for cancer and acute myocardial infarction: multi-omics approaches for deciphering heterogeneity to enhance patient management.
Front. Pharmacol. 14:1249145.
doi: 10.3389/fphar.2023.1249145

Identification of prognostic and diagnostic signatures for cancer and acute myocardial infarction: multi-omics approaches for deciphering heterogeneity to enhance patient management

Na Yuan[1†], Hai-Hua Pan[1†], Yan-Shan Liang[2], Hui-Lin Hu[1], Chang-Lin Zhai[1*] and Bo Wang[3*]

[1]The First Hospital of Jiaxing Affiliated Hospital of Jiaxing University, Jiaxing, Zhejiang, China, [2]Affiliated Dongguan Hospital, Southern Medical University, Dongguan, Guangdong, China, [3]The Second Affiliated Hospital of Jiaxing University, Jiaxing, Zhejiang, China

Patients diagnosed with cancer face an increased risk of cardiovascular events in the short term, while those experiencing acute myocardial infarction (AMI) have a higher incidence of cancer. Given limitations in clinical resources, identifying shared biomarkers offers a cost-effective approach to risk assessment by minimizing the need for multiple tests and screenings. Hence, it is crucial to identify common biomarkers for both cancer survival and AMI prediction. Our study suggests that monocyte-derived biomarkers, specifically WEE1, PYHIN1, SEC61A2, and HAL, hold potential as predictors for cancer prognosis and AMI. We employed a novel formula to analyze mRNA levels in clinical samples from patients with AMI and cancer, resulting in the development of a new risk score based on expression profiles. By categorizing patients into high-risk and low-risk groups based on the median risk score, we observed significantly poorer overall survival among high-risk patients in cancer cohorts using Kaplan-Meier analysis. Furthermore, calibration curves, decision curve analysis (DCA), and clinical impact curve analyses provided additional evidence supporting the robust diagnostic capacity of the risk score for AMI. Noteworthy is the shared activation of the Notch Signaling pathway, which may shed light on common high-risk factors underlying both AMI and cancer. Additionally, we validated the differential expression of these genes in cell lines and clinical samples, respectively, reinforcing their potential as meaningful biomarkers. In conclusion, our study demonstrates the promise of mRNA levels as biomarkers and emphasizes the significance of further research for validation and refinement.

KEYWORDS

signature, AMI, cancer, machine learning, prognosis

Introduction

Acute myocardial infarction (AMI) and cancer are significant contributors to morbidity and mortality globally (Psaty and Vasan, 2023). Despite limited references on the relationship between these two conditions, research indicates that patients diagnosed with cancer are at a higher short-term risk of experiencing cardiovascular events, while those with acute myocardial infarction have an increased incidence of cancer (Rinde et al., 2017; Leening et al., 2023). These potential links imply a latent connection exists between cancer survival and AMI incidence. Therefore, identifying common biomarkers for both cancer survival and AMI prediction is of utmost importance.

The observation that cancer patients face an elevated risk of cardiovascular events is a matter of concern. Studies have shown that cancer survivors experience cardiac complications, such as myocardial infarction, heart failure, and arrhythmias, at rates higher than the general population (Howard et al., 2019). It is hypothesized that this increased risk is multifactorial, involving both direct effects of cancer treatment (e.g., chemotherapy-induced cardiotoxicity) and shared risk factors between cancer and cardiovascular diseases (Shaikh and Shih, 2012). For instance, inflammation, oxidative stress, and endothelial dysfunction, which are common processes in both cancer and cardiovascular diseases, may contribute to the development of adverse cardiac events in cancer patients (Libby and Kobold, 2019). Therefore, identifying common biomarkers that can predict both cancer survival and AMI might help identify patients at higher risk for cardiovascular complications during cancer treatment.

On the other hand, the association between AMI and the subsequent occurrence of cancer has also been documented. Multiple studies have demonstrated an increased incidence of various types of cancer, including lung, colorectal, and hematological malignancies, among individuals with a history of AMI (Leening et al., 2023). Given the potential bidirectional relationship between AMI and cancer, it is vital to explore common biomarkers that might aid in early detection and improve patient outcomes for both conditions. In recent years, advances in molecular profiling technologies have paved the way for the discovery of potential shared biomarkers for cancer survival and AMI prediction (Zhao et al., 2022). Transcriptomics has emerged as a powerful tool for comprehensively analyzing gene expression patterns and identifying differentially expressed genes associated with both diseases (Wang et al., 2009; Ding et al., 2022). Integration of multi-omics data, combining transcriptomics with other molecular profiling techniques such as genomics and proteomics (Reel et al., 2021), holds even greater potential in unraveling the complex interplay between cancer and AMI. In our study, we used ovarian cancer (OC) samples to explore the common diagnostic and prognostic signature of AMI may seem unconventional at first. However, OC and AMI might share certain risk factors, such as obesity, diabetes, and smoking, which can predispose individuals to both conditions (Stewart et al., 2019; Kuhn et al., 2020). In this study, we used OC samples as an example. By studying OC samples, we can investigate if there are common molecular pathways or biomarkers (Aydin et al., 2019) associated with these shared pathways that might contribute to the development of both OC and AMI.

In our study, we applied a novel formula to analyze mRNA levels in clinical samples from AMI and OC. The calculation resulted in the generation of a new risk score based on the expression profiles. The utilization of risk scores demonstrates the ability to accurately predict the probability of AMI incidence and the prognosis of OC patients. Most important of all, these findings demonstrate the potential of utilizing shared biomarkers (WEE1, PYHIN1, SEC61A2, and HAL) to predict outcomes in both cancer and AMI patients.

Materials and methods

Pre-processing of bulk transcriptome data

For the AMI cohort, two independent datasets were analyzed on the GPL6244 platform, namely, GSE59867 (111 AMI patients and 46 stable CAD patients at admission) and GSE62646 (28 AMI patients and 14 stable CAD patients at admission) (Pan et al., 2023). The peripheral blood cohort of OC was obtained from GSE31682, which comprises 20 healthy controls and 48 OC patients. After excluding patients with incomplete follow-up information (Feng et al., 2022), we obtained RNA sequencing (RNA-seq) data from both the Cancer Genome Atlas (TCGA) database (Blum et al., 2018) and the International Cancer Genome Consortium (ICGC) database. Additionally, data from the Gene Expression Omnibus (GEO) database (Barrett et al., 2013) was obtained for the GPL570 platform ($n = 597$), which included GSE19829, GSE18520, GSE9891, GSE26193, GSE30161, and GSE63885. To integrate the ICGA and TCGA data and define the meta-RNA-seq dataset, the meta-microarray dataset was defined using the GPL570 platform. Finally, the "sva" package was utilized to effectively address and eliminate batch effects across the different datasets.

Pre-processing of single-cell RNA sequencing data

Considering the limited availability of human AMI single-cell RNA (scRNA) sequencing datasets, we employed a mouse single-cell sequencing dataset (GSE135310) as an alternative (Pan et al., 2023). This dataset includes single-cell RNA sequencing files for cardiac $CD45^+$ total leukocytes, isolated from mice subjected to AMI or sham surgery at various time points. As we lacked peripheral blood single-cell data from healthy individuals within the same batch, we focused on analyzing datasets obtained before and after chemotherapy from the same batch. In this regard, we obtained the GSE213243 dataset, comprising 2 peripheral blood samples from OC patients. In summary, we performed a series of data filtering steps to ensure the quality of our scRNA data. We retained cells with an expression of RNA features ranging from 200 to 2500 while keeping the percentage of mitochondrial RNA content below 10%. Additionally, we employed the Harmony algorithm to mitigate batch effects in our analysis. To annotate all clusters, we utilized the "SingleR" package for comprehensive annotation.

Clinical samples

As consistent with our previous publication (Pan et al., 2023), we performed mRNA validation of peripheral blood samples using the

same cohort of patients. In brief, ten early AMI patients and ten CAD patients were recruited from our hospital between January 2023 and March 2023. Blood samples were collected from the patients shortly after experiencing chest pain, before the administration of antiplatelet or anticoagulant drugs. Peripheral blood mononuclear cells (PBMCs) were isolated from the collected blood samples using established techniques (Boyum, , 1968).

Immunohistochemical techniques and RT-qPCR

Immunohistochemistry (IHC) staining involves the use of antibodies that are specifically designed to recognize and bind to target antigens of interest. The antibodies are labeled with a chromogenic or fluorescent dye, enabling the visualization and localization of the target molecules under a microscope. The IHC sections utilized in this study were obtained from the Human Protein Atlas (HPA) database. To validate the mRNA expression levels of AMI, PBMCs from patients were utilized. As for OC, cell lines were employed for the validation of mRNA expression levels. IOSE-80 (CP-H055), and SKOV3 (CL-0215) were purchased from Procell Life Science and Technology Co. Ltd. They were cultured in RPMI 1640 medium supplemented with 10% FBS and 1% penicillin/streptomycin at 37°C in a humidified incubator with 5% CO_2. In short, total RNA was extracted from samples using the FastPure Cell/Tissue Total RNA Isolation Kit V2 (Vazyme, Nanjing, China). Subsequently, RT-qPCR was performed on a LightCycler 480 II Real-time PCR instrument with the HiScript III All-in-one RT SuperMix Perfect for qPCR (Vazyme, Nanjing, China) and ChamQ universal SYBR qPCR Master Mix (Vazyme, Nanjing, China). We used the $2^{-\Delta\Delta Ct}$ method to calculate gene expression levels. The primer sequences used were designed based on previously published references and PrimerBank database (Wang et al., 2012), including WEE1 (Ma et al., 2022), PYHIN1 (Lee et al., 2019), SEC61A2, and HAL (Kozaczek et al., 2019).

CIBERSORT

CIBERSORT is a widely used computational tool for analyzing gene expression data and estimating the relative abundance of immune cell populations within complex tissue samples (Chen et al., 2018). It utilizes a deconvolution algorithm to infer cell type proportions from bulk RNA sequencing data. We used the mRNA expression profile data from GSE59867, GSE31682, and meta-RNA-seq datasets (TCGA-OV and ICGC-OV) as inputs. The algorithm utilizes a support vector regression model trained on the signature matrix to determine the relative abundance of each cell type.

Weighted gene co-expression network analysis

In our study, we utilized the Weighted Gene Co-expression Network Analysis (WGCNA) (Langfelder and Horvath, 2008) in order to investigate gene expression patterns and identify gene modules with similar expression profiles. WGCNA is a powerful tool that allows us to construct a co-expression network based on correlations between genes across different samples. By analyzing the module and module-trait (monocyte score from CIBERSORT) relationships, we can uncover biologically relevant modules associated with specific traits or phenotypes of interest. We established the optimal soft thresholding powers (β) for OC and AMI samples to be $\beta = 7$ and $\beta = 9$, respectively. Additionally, we ensured that each module consisted of no fewer than 50 genes.

Least absolute shrinkage and selection operator regression

We used cross-validation (10-flod) to determine the optimal value for λ. In short, our objective was to determine the optimal model λ value by constructing a penalty function with the occurrence of AMI as the endpoint event and the variation in gene expression as the variable for each sample. The Least absolute shrinkage and selection operator (LASSO) regression model is fit on the training set for each λ value (Cheng et al., 2022), and the performance is evaluated on the validation set using a chosen metric, such as mean squared error or area under the curve. The λ value that minimizes the error on the validation set is considered the optimal choice. Hence, we conducted a more comprehensive screening of potential diagnostic genes from the pool of monocyte-associated genes.

Establishment and validation of the logstic regression model

We utilized a nomogram based on Logistic regression to facilitate predictive modeling and risk assessment (Zhang et al., 2022). By fitting the logistic regression equation, we estimated the coefficients of the four variables (WEE1, PYHIN1, SEC61A2, and HAL) and captured their contributions to the probability of the outcome. Each predictor variable was assigned a corresponding point value based on its coefficient. The total points were summed up to determine the individual's predicted probability of the outcome. Moreover, we utilized calibration curves, decision curve analysis (DCA) curves, and clinical impact curves to validate the performance of the nomogram in GSE62646 and GSE59867.

Establishment and validation of the Cox regression model

The Cox regression model, also referred to as the proportional hazards model, is a widely employed statistical approach utilized in survival analysis to analyze time-to-event data. To construct the Cox regression model, we initially selected a group of potential predictor variables, namely, WEE1, PYHIN1, SEC61A2, and HAL. Subsequently, we performed model training by fitting the Cox regression equation to the meta-RNA-seq cohorts, thereby estimating the coefficients for each gene. The risk score was then calculated using the formula: risk score = Σ (Exp_i * $coef_i$), where coef and Exp represent the coefficient and expression of each gene, respectively. After establishing the Cox

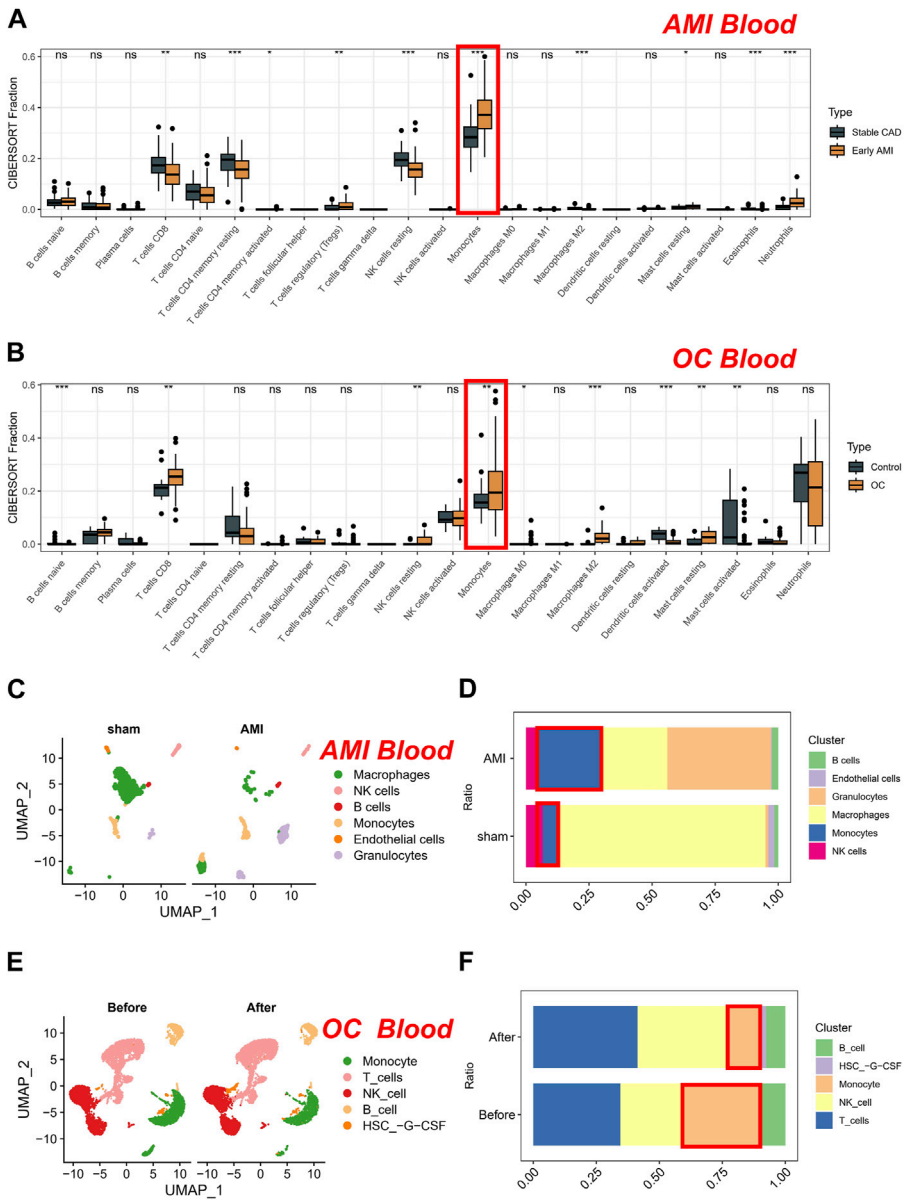


FIGURE 1
Analysis of cell proportions in AMI and CAD patients and OC patients. **(A)** Comparison of cell proportions using the CIBERSORT algorithm in peripheral blood microarray dataset of AMI and CAD patients. **(B)** Comparison of cell proportions using the CIBERSORT algorithm in peripheral blood microarray dataset of OC patients compared to normal individuals. **(C)** Dimensionality reduction analysis on single-cell data from sham and AMI. **(D)** The proportion of cells in sham and AMI groups. **(E)** Dimensionality reduction analysis on single-cell data from OC patients. **(F)** The proportion of cells in OC groups (Before and after chemotherapy).

regression model, we proceeded to validate its performance utilizing meta-microarray cohorts. We employed calibration plots, receiver operator characteristic (ROC) curves, and log-rank tests to assess the model's efficacy in distinguishing between high-risk and low-risk individuals.

Enrichment analysis

We employed hallmark gene sets (h.all.v7.5.1.symbols.gmt) (Liberzon et al., 2015), which are collections of genes representing key biological processes and signaling pathways. These gene sets cover a wide range of fundamental cellular activities, such as cell cycle regulation, DNA repair, immune response, and metabolism. By applying Gene Set Variation Analysis (GSVA) (Hanzelmann et al., 2013), we transformed our gene expression data into pathway enrichment scores for each sample. This was achieved by comparing the expression levels of genes within each hallmark gene set to the background distribution in our dataset. The resulting enrichment scores provided quantitative measures of the activity levels of these pathways in individual samples.

Statistical analysis

Statistical analysis was performed using R software (v4.1.2). To evaluate differences, the significance of most cases was assessed using the Wilcoxon rank-sum test. Statistical significance was defined as a p-value below 0.05 and indicated as $^{*}p < 0.05$, $^{**}p < 0.01$, or $^{***}p < 0.001$.

Results

Monocytes may serve as indicators for predicting AMI and the prognosis of cancer

Coronary artery disease (CAD) is a principal cause of morbidity and mortality worldwide. Patients with stable CAD are still at risk of AMI, which is a severe complication of CAD. Thus, in clinical settings, patients with CAD are commonly used as control groups to investigate changes in blood indicators among AMI patients. In our study, we first compared differences in cell proportions using the CIBERSORT algorithm on a peripheral blood microarray dataset consisting of AMI and CAD patients. AMI patients exhibit decreased levels of $CD8^{+}$ T cells, memory $CD4^{+}$ T cells, resting NK cells, and M2-type macrophages. Conversely, they have increased levels of Tregs, resting mast cells, neutrophils, and monocytes (Figure 1A). Subsequently, we further investigated the changes in cell content in the peripheral blood of OC patients compared to normal individuals. Interestingly, only the trend of monocytes cell changes was consistent with that of AMI patients. This suggests a potential common role of monocytes in both conditions (Figure 1B). To further validate the robustness of our results, we conducted dimensionality reduction analysis on single-cell data from both the sham and AMI patients, resulting in six distinct clusters that could be clearly distinguished: monocytes, endothelial cells, macrophages, granulocytes, NK cells, and B cells (Figure 1C). Of note, after AMI occurred, the proportion of monocytes increased, consistent with the results of bulk transcriptome analysis. Due to the lack of peripheral blood single-cell data from healthy individuals from the same batch, we analyzed the datasets obtained before and after chemotherapy from the same batch. Similarly, the single-cell data from patients were sorted into four clusters (Figure 1D), with a significant decrease in the proportion of monocytes observed after chemotherapy. Therefore, the proportion of monocytes in the peripheral blood of patients decreased when the tumor-load was reduced by chemotherapy. Conversely, monocytes proportion significantly increased when the tumor occurred (Figure 1E). Given that pathology is often considered the gold standard for cancer diagnosis, we re-assessed the prognostic value of monocytes (CD14 is a typical marker for monocytes) in the TCGA-OV cohort (cancer tissues). Our results also demonstrate that monocytes are a risk factor for OC prognosis. Specifically, as the expression of CD14 increases in bulk tissues, the prognosis of patients worsens (Supplementary Figure S1).

Biomarkers derived from monocytes using WGCNA in AMI and cancer

To enhance gene filtration from monocytes, we integrated the entire gene expression profile into WGCNA. Additionally, we utilized the score of monocyte expression in the CIBERSORT algorithm as a clinical feature to identify key modules most pertinent to monocyte expression. During the construction of the co-expression network, we observed optimal soft threshold powers of $\beta = 7$ for OC (Figure 2A) and $\beta = 9$ for AMI samples (Figure 2B). Through careful examination of correlation coefficients and p-values (Figures 2C, D), we determined that the purple and brown modules exhibited the strongest absolute correlation with the monocyte score in OC (Figure 2E), while in AMI, the purple and yellow modules displayed the strongest absolute correlation (Figure 2F). As a result, we designated these four modules as key modules and subsequently identified 23 overlapping genes within them (Figure 2G). Subsequently, we conducted further screening of the aforementioned 23 genes in the AMI dataset using the LASSO algorithm. Our objective was to determine the optimal model λ value by constructing a penalty function with the occurrence of AMI as the endpoint event and the variation in gene expression as the variable for each sample (Figures 2H, I). Consequently, we identified seven genes: HRH4, LTBR, WEE1, HAL, PYHIN1, S100A12, and SEC61A2. Of particular significance, based on these seven genes, we proceeded with prognostic validation utilizing the RNA-seq cohort of OC. Our findings indicated that WEE1, PYHIN1, and SEC61A2 served as risk factors impacting the prognosis of ovarian cancer, while HAL demonstrated a protective effect (Supplementary Figure S2).

In summary, our findings suggest that WEE1, PYHIN1, SEC61A2, and HAL derived from monocytes may serve as a potential predictor for AMI and cancer prognosis.

Establishment of the monocytes-related diagnostic signature in AMI

In the AMI dataset, we made noteworthy observations. Firstly, WEE1 ($p = 2.4e\text{-}11$) exhibited significant downregulation and displayed superior diagnostic potential, as evidenced by an AUC of 0.839 (95% CI: 0.768–0.899), as shown in Figure 3A. Similarly, PYHIN1 (Figure 3B) and SEC61A2 (Figure 3C) were markedly downregulated in AMI samples and demonstrated favorable diagnostic performance, with AUCs of 0.782 and 0.747, respectively. It is worth mentioning that HAL exhibited significant upregulation in AMI samples, yielding an AUC of 0.766 (Figure 3D). To further enhance the diagnostic accuracy of the model and facilitate its clinical application, we developed a nomogram utilizing the aforementioned four genes as predictors for AMI (Figure 3E). The calibration curve demonstrated the ability of the nomogram to accurately and reliably diagnose AMI among CAD patients (Figure 3F). Moreover, the DCA curve (Figure 3G) and decision curve analysis (Figure 3H) provided further evidence of its robust diagnostic capacity. Importantly, in the validation dataset (GSE62646), the calibration curve (Figure 3I), DCA curve (Figure 3J), and clinical impact curve (Figure 3K) consistently underscored the excellent external validation capability of the nomogram.

Establishment of the monocytes-related prognostic signature in cancer

Considering that cancer treatment primarily involves surgery and tissue samples are readily available, we conducted a comprehensive investigation into the expression of the

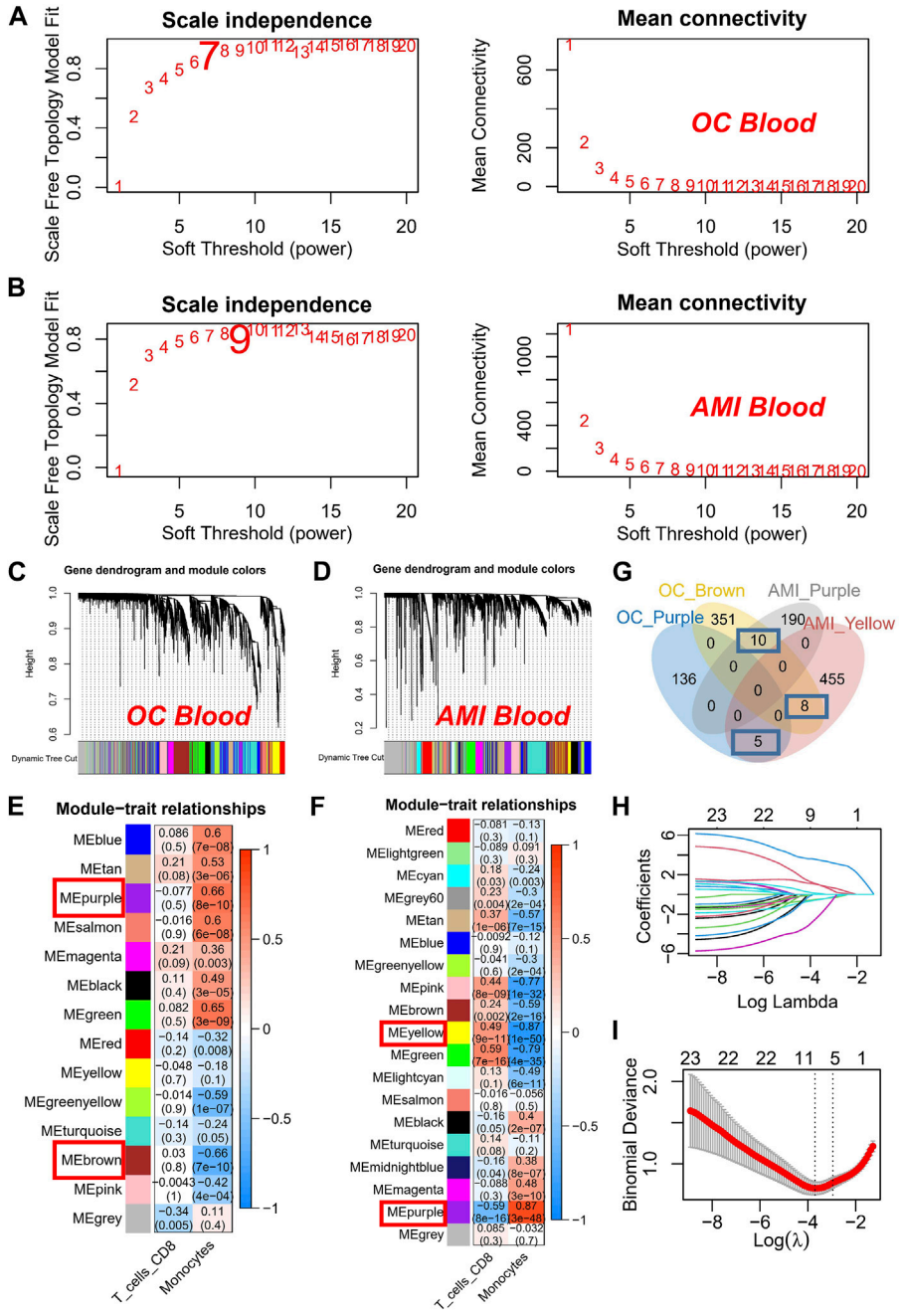


FIGURE 2
Construction and Analysis of Co-expression Networks for Monocyte-Related Genes. **(A)** Determination of soft threshold powers (β) for OC samples in the co-expression network construction. **(B)** Determination of soft threshold powers (β) for AMI samples in the co-expression network construction. **(C)** Clustering dendrograms, with dissimilarity based on topological overlap, together with assigned module colors in OC datasets. **(D)** Clustering dendrograms, with dissimilarity based on topological overlap, together with assigned module colors in AMI datasets. **(E)** Correlation coefficients and *p*-values used to identify key modules most correlated with the monocyte score in OC datasets. **(F)** Correlation coefficients and *p*-values used to identify key modules most correlated with the monocyte score in AMI datasets. **(G)** Venn plot of key modules. **(H)** The gene signature selection of optimal parameter (lambda). **(I)** LASSO coefficient profiles genes were selected by the optimal lambda.

aforementioned four genes across various tissue samples by combining the GTEx database and the HPA database. At both the protein and mRNA levels, WEE1 exhibited significantly higher expression in tumor samples (Figure 4A). Similarly, PYHIN1 displayed significantly higher mRNA expression in tumor samples, although no significant protein staining (Figure 4B). In contrast, HAL did not exhibit significant differences in mRNA expression between the two sample types, but protein upregulation was evident in tumor samples (Figure 4C). Unfortunately, an IHC antibody for SEC61A2 was unavailable. However, we characterized the protein's structure and confirmed its significant upregulation at the mRNA level in

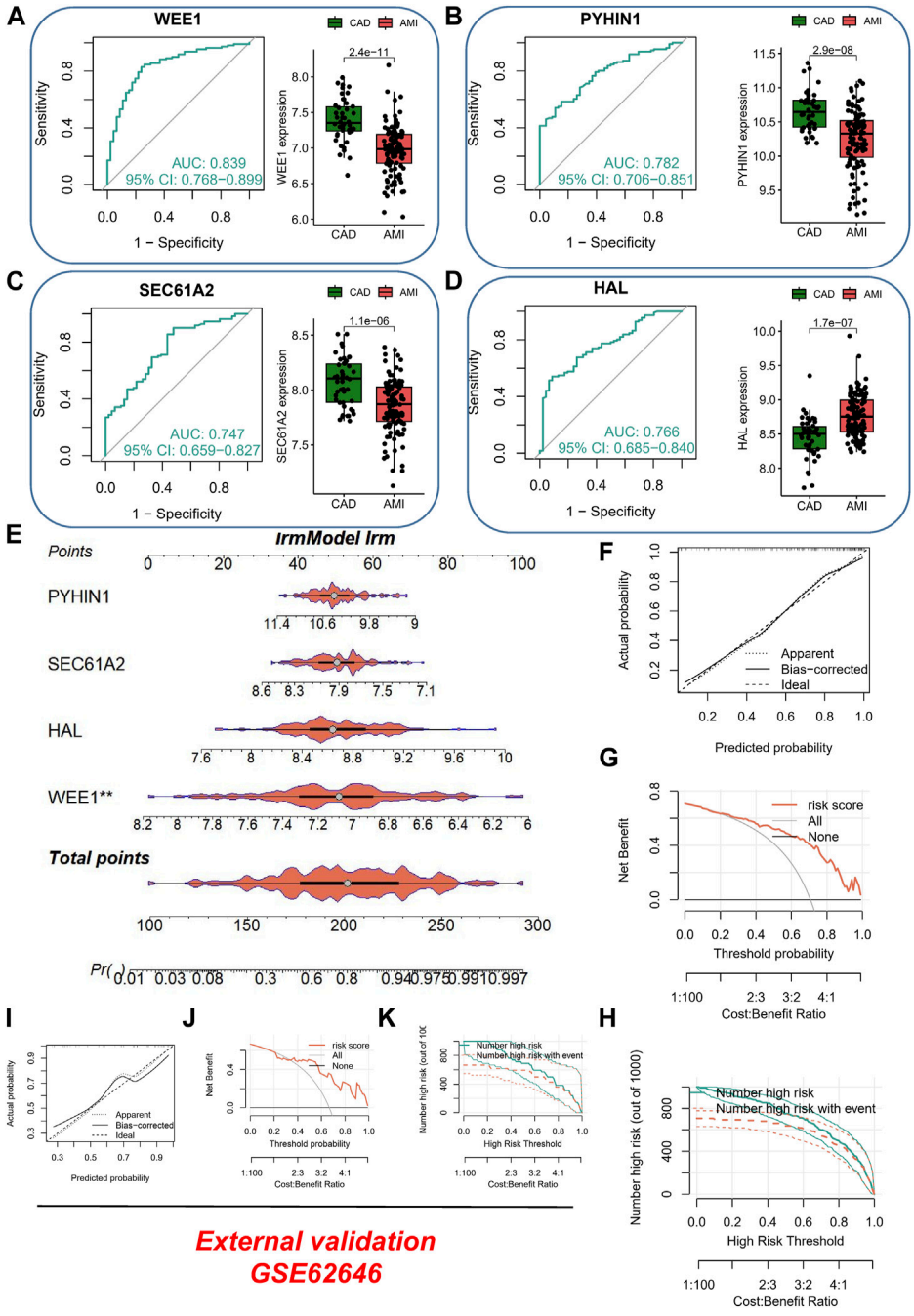


FIGURE 3
Evaluation of Diagnostic Performance and Clinical Application of the AMI Nomogram. **(A)** ROC curve and box plot of differential expression of WEE1 in different samples. **(B)** ROC curve and box plot of differential expression of PYHIN1 in different samples. **(C)** ROC curve and box plot of differential expression of SEC61A2 in different samples. **(D)** ROC curve and box plot of differential expression of HAL in different samples. **(E)** Development of a nomogram utilizing WEE1, PYHIN1, SEC61A2, and HAL as predictors for AMI diagnosis. **(F)** Calibration curve showing the accuracy and reliability of the nomogram in diagnosing AMI among CAD patients. **(G)** DCA curve demonstrating the diagnostic capacity of the nomogram. **(H)** Clinical impact curve illustrating the robustness of the nomogram for AMI diagnosis. **(I)** External validation of the nomogram using calibration curve. **(J)** External validation of the nomogram using DCA curve. **(K)** External validation of the nomogram using clinical impact curve.

normal samples (Figure 4D). Subsequently, to quantify the survival risk for each ovarian cancer patient, we developed a risk model utilizing a multi-factorial Cox formula (Figure 4D) based on the four aforementioned genes (WEE1, PYHIN1, SEC61A2, and HAL). The risk score for each OC patient was calculated using the equation: Risk score = (−0.195 × WEE1 expression) + (−0.335 × PYHIN1 expression) + (−0.191 × SEC61A2 expression) + (0.137 × HAL expression). Subsequently, survival curves were generated for each gene in the model, revealing intriguing findings. PYHIN1, SEC61A2, and WEE1 emerged as protective genes, suggesting that reduced expression of these genes may contribute to prolonged overall

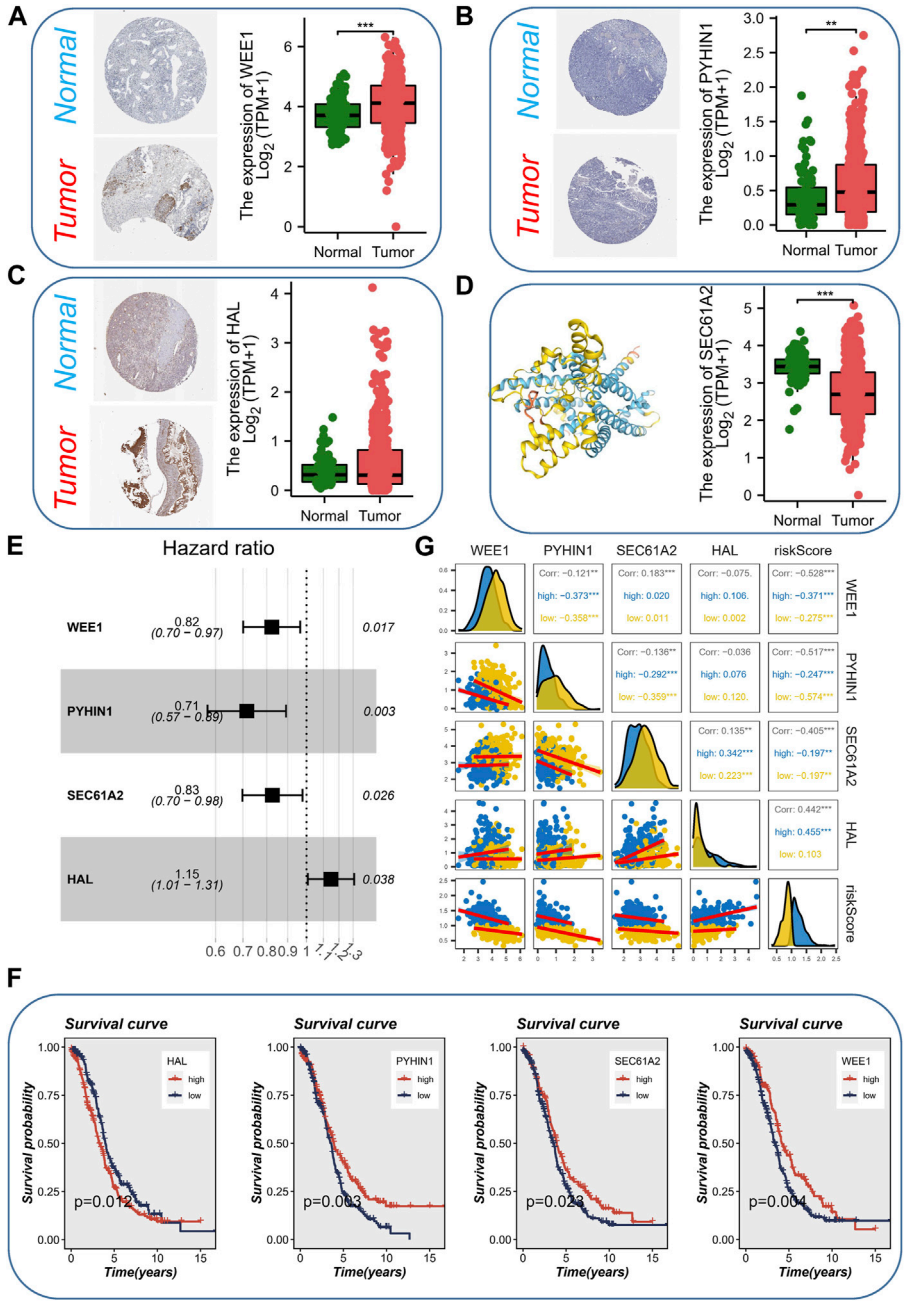


FIGURE 4
Expression Analysis and Prognostic Model Development. **(A)** The expression level (protein and mRNA) of WEE1 in different samples levels. **(B)** The expression level (protein and mRNA) of PYHIN1 in different samples levels. **(C)** The expression level (protein and mRNA) of HAL in different samples levels. **(D)** The expression level (mRNA) protein structure and of SEC61A2 in different samples levels. **(E)** Development of a risk model using a multi-factorial Cox formula based on the four genes (WEE1, PYHIN1, SEC61A2, and HAL). **(F)** Survival curves showing the impact of each gene in the model on overall survival. **(G)** Correlations among the genes involved in risk model.

survival in patients (Figure 4F). Conversely, HAL was identified as a high-risk gene, implying that increased expression of HAL may be associated with decreased patient survival. Notably, our expression level assessment demonstrated predominantly negative correlations among the genes (Figure 4G). Specifically, HAL displayed negative correlations with WEE1 and PYHIN1, while exhibiting a positive correlation with SEC61A2. Regarding risk scores, a noteworthy positive correlation was solely observed with HAL (R = 0.442, p < 0.001). The formula described above was applied to both the meta-RNA-seq cohort and the meta-microarray cohort, resulting in the generation of a new risk score based on the expression profile. Patients were then classified into high-risk and low-risk categories using the median risk score. Kaplan-Meier survival analysis conducted in the meta-RNA-seq cohort revealed that high-risk patients exhibited significantly worse overall survival

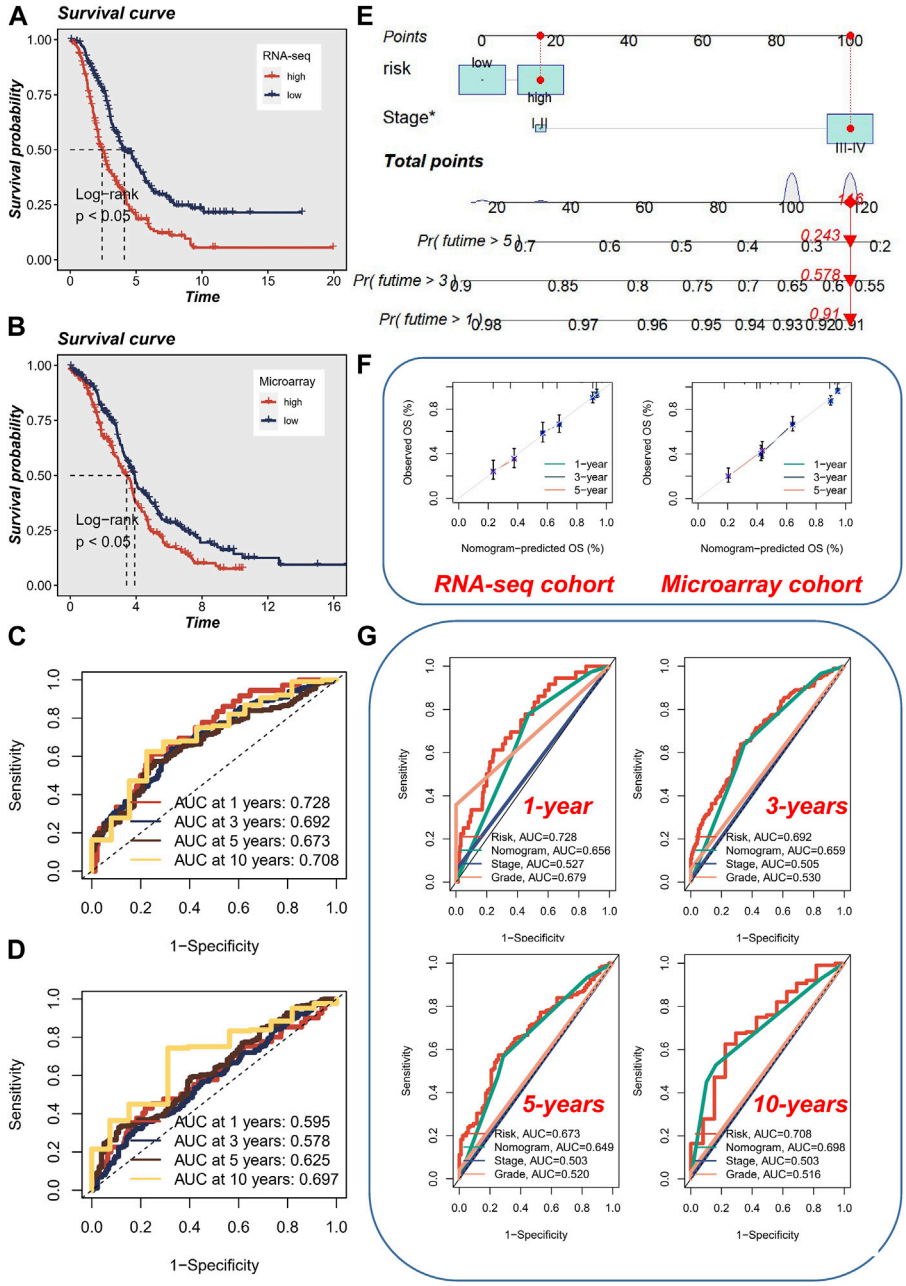


FIGURE 5
Prognostic Analysis and Nomogram Evaluation. **(A)** Kaplan-Meier survival analysis in the meta-RNA-seq cohort. **(B)** Kaplan-Meier survival analysis in the meta-microarray cohort. **(C)** Evaluation of prognostic accuracy using ROC curves for 1, 3, 5, and 10-year overall survival in the meta-RNA-seq cohort. **(D)** Evaluation of prognostic accuracy using ROC curves for 1, 3, 5, and 10-year overall survival in the meta-microarray cohort. **(E)** Utilization of a visual nomogram integrating FIGO staging and risk stratification to evaluate the risk of ovarian cancer patients. **(F)** Calibration curve of nomogram accuracy in both the meta-RNA-seq cohort and the meta-microarray cohort. **(G)** ROC analysis comparing the performance of the nomogram with other clinical models and risk scores.

compared to low-risk patients (Figure 5A). Similarly, in the meta-microarray cohort, the high-risk group had a lower likelihood of survival (Figure 5B). To assess the prognostic accuracy of the prognostic features, ROC curves for 1, 3, 5, and 10-year OS were analyzed. In the meta-RNA-seq training cohort, the AUC values for these time points were 0.728, 0.692, 0.673, and 0.708, respectively, (Figure 5C). Similarly, the meta-microarray validation set showed superior AUC values of 0.595, 0.578, 0.625, and 0.697 for the 1, 3, 5, and 10-year AUCs, respectively, (Figure 5D).

To ensure consistency with the AMI risk signature, we utilized a visual nomogram to evaluate the risk of OC patients by integrating FIGO staging and risk stratification (Figure 5E). The accuracy of nomogram was assessed and found to be superior in both the meta-RNA-seq cohort and the meta-microarray cohort (Figure 5F). Additionally, ROC

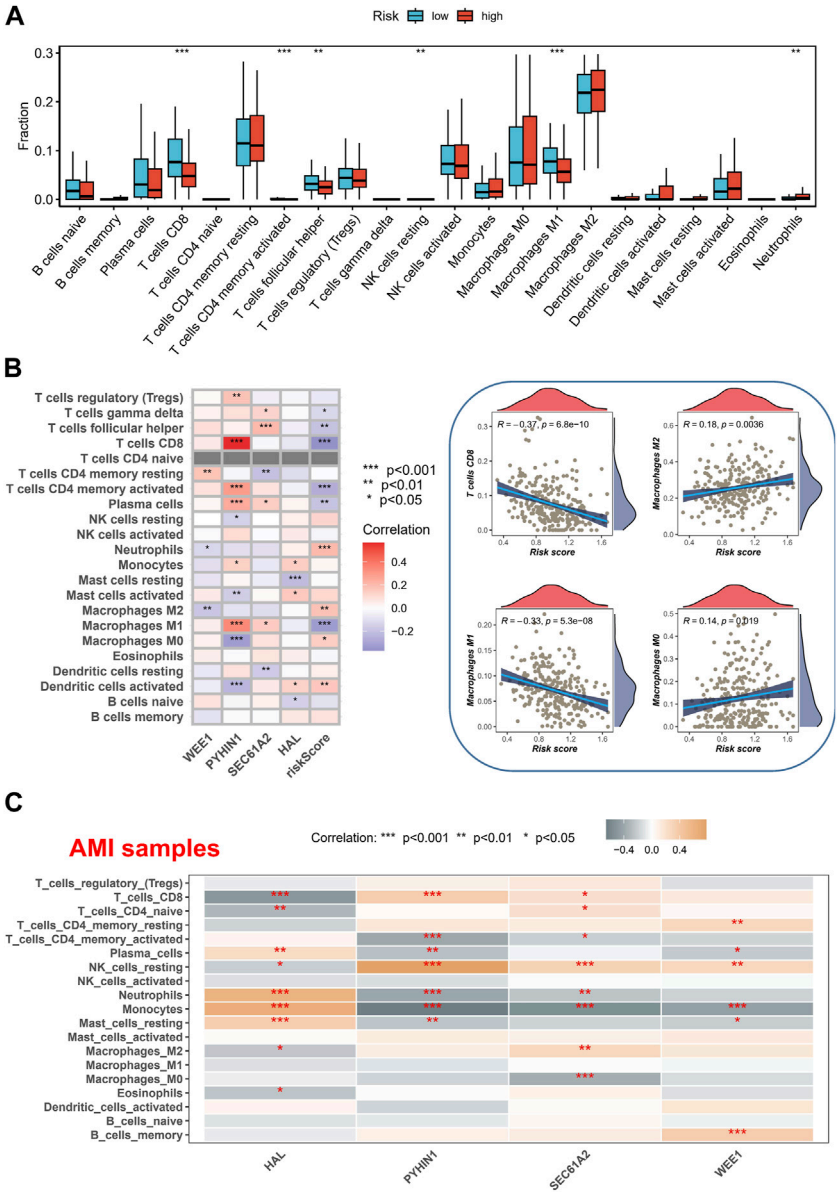


FIGURE 6
Immune Cell Composition Analysis and Model Gene Correlation. **(A)** Analysis of immune cell composition utilizing the CIBERSORT method in tissue samples of high-risk and low-risk groups. **(B)** Characterization of the correlation between gene expression and immune cell. **(C)** Characterization of the correlation between model genes and immune cells in AMI samples.

analysis was performed, indicating that early survival prediction using the nomogram outperformed other clinical models and risk scores, whereas for long-term survival prediction (>5 years), utilizing the risk score alone yielded better results (Figure 5G).

Immune cell infiltration in AMI and cancer

We utilized the CIBERSORT method to analyze the immune cell composition of tissue samples, comparing the high-risk and low-risk groups, and associating them with model genes. Interestingly, we discovered that the high-risk group exhibited lower levels of $CD8^+$ T cells and M1 macrophages (Figure 6A). $CD8^+$ T cells, typically referred to as cytotoxic T lymphocytes, secrete various cytokines involved in immune responses (Mittrucker et al., 2014), while M1-type macrophages are capable of producing pro-inflammatory cytokines (Mills et al., 2016). The reduction of $CD8^+$ T cells and M1 macrophages may indicate an "cold environment" in high-risk patients. Furthermore, we observed a significant positive correlation between the expression of the PYHIN1 gene and $CD8^+$ T cells, suggesting that the PYHIN1 gene primarily influences changes in the tumor microenvironment by regulating the proliferation of cytotoxic T lymphocytes (Figure 6B). We demonstrated the specific distribution of risk scores and different immune cell

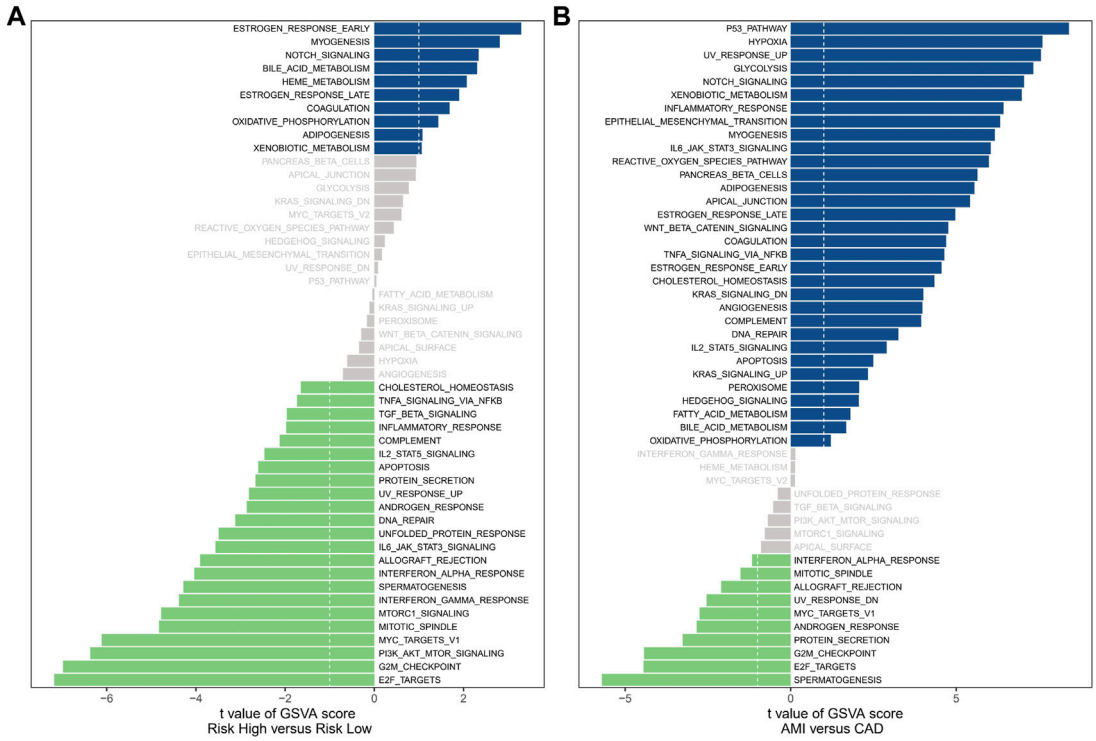


FIGURE 7
Enrichment analysis. **(A)** The results of Gene Set Variation Analysis in OC samples. **(B)** The results of Gene Set Variation Analysis in AMI samples.

types, revealing a negative correlation between $CD8^+$ T cells and risk scores. Moreover, it appears that changes in risk scores also influence the proportions of M0/M1/M2 macrophages. Furthermore, we characterized the correlation between model genes and immune cells in AMI samples, which similarly showed a strong significant positive correlation between PYHIN1 gene expression and $CD8^+$ T cells, potentially regulating the proliferation of resting NK cells (Figure 6C).

Enrichment analysis

The analysis of hallmark pathway gene features using the GSVA method reveals distinct differences between high-risk and low-risk groups in OC. A direct comparison between these groups demonstrates specific enrichment features. In the high-risk group, the top five enriched features include Estrogen Response Early, Myogenesis, Notch Signaling, Bile Acid Metabolism, and Heme Metabolism (Figure 7A). Conversely, the low-risk group exhibits the top five enriched features: E2F Targets, G2M Checkpoint, PI3K/AKT/MTOR Signaling, MYC Targets V1, and Mitotic Spindle. Additionally, we investigated the significantly different pathway signals between patients with AMI and coronary artery disease. The findings reveal substantial activation of the P53 Pathway, Hypoxia, and Notch Signaling in AMI patients (Figure 7B). Of particular interest is the shared activation of the Notch Signaling pathway, which may provide insight into the common high-risk factors underlying both AMI and cancer.

Validation of mRNA levels in clinical samples

To validate the reliability of the four prognostic genes identified, qRT-PCR testing was performed on both clinical samples and cell lines. In PBMC samples, we observed consistent differential expression patterns of the four genes with the results obtained from the microarray analysis (Figures 3A–D). Specifically, compared to CAD, WEE1, PYHIN1, and SEC61A2 were downregulated in AMI, while HAL exhibited upregulation (Supplementary Figure S3A). Furthermore, it is noteworthy that the validation performed at the cell lines also demonstrated consistent results with RNA-seq analysis. Specifically, when compared to IOSE-80 cells, SKOV3 cells exhibited upregulation in the expression of WEE1, PYHIN1, and HAL, while the expression of SEC61A2 was downregulated (Supplementary Figure S3B).

Discussion

Acute myocardial infarction (AMI) and cancer are two prevalent and devastating health conditions that contribute significantly to morbidity and mortality worldwide (Rinde et al., 2017). While extensive research exists on each of these diseases independently, the relationship between AMI and cancer remains relatively understudied. Common biomarkers could help identify patients who are at higher risk of developing either cancer or AMI, enabling tailored preventive strategies. By stratifying individuals based on their risk profiles, healthcare

providers can offer targeted interventions, such as lifestyle modifications, pharmacological interventions, and regular surveillance, with the goal of mitigating the risk of both diseases. Moreover, in a clinical setting where resources are often limited, utilizing shared biomarkers may offer a cost-effective approach to risk assessment by reducing the need for multiple tests and screenings.

In our study, our findings suggest that WEE1, PYHIN1, SEC61A2, and HAL derived from monocytes may serve as a potential predictor for AMI and cancer prognosis. WEE1, a protein kinase involved in cell cycle regulation, acts as a critical regulator of the DNA damage response pathway and plays a crucial role in maintaining genomic stability (Okabe et al., 2023; Su et al., 2023). Its overexpression has been observed in various tumor types, contributing to tumor growth, chemoresistance, and poor prognosis. Mohamed et al. (Mohamed et al., 2018) discovered that overexpression of cyclin-dependent kinase 1 (CDK1), CDK4, cyclin B1, and cyclin D1 in adult cardiomyocytes induces stable cell division, leading to significant cardiac regeneration after myocardial infarction. Importantly, they found that inhibition of Wee1, along with Tgf-β, made CDK1 and cyclin B dispensable, highlighting the role of WEE1 as a potential target for promoting cardiomyocyte proliferation. Chen and Gardner (Chen and Gardner, 2004) found that endothelin (ET) promotes proliferation of rat aortic smooth muscle cells by increasing CDK2 and CDC2 activity through the MEK/ERK/RSK signal transduction pathway. They observed that ET treatment led to phosphorylation and inactivation of the inhibitory kinase WEE1, along with upregulation of CDC25A phosphatase, highlighting the role of WEE1 in ET-dependent mitogenesis. PYHIN1, a member of the PYHIN (pyrin and HIN domain-containing) protein family, exerts complex functions ranging from tumor suppression to tumor promotion, depending on the specific tumor (Tong et al., 2019; Ding et al., 2022). Its involvement in DNA damage repair, cell cycle regulation, immune responses, and inflammation contributes to its multifaceted role in cancer progression. de Las Fuentes et al. (2013) investigated the role of SNP-loop diuretic interactions in hypertension across different ethnic groups. In their study on African Americans (AA) and European Americans (EA), they identified several promising loci, including genes such as NUDT12, CHL1, GRIA1, CACNB2, and PYHIN1 for systolic blood pressure (SBP) in AA, and ID3 for diastolic blood pressure (DBP) in AA. These findings suggest that PYHIN1 may play a role in the regulation of blood pressure and response to anti-hypertensive drugs, although no SNP reached genome-wide significance in this small study. Further research in more diverse populations is needed to identify additional variants. SEC61A2, a key component of the SEC61 protein complex, plays a crucial role in protein translocation across the endoplasmic reticulum (ER) membrane (Connerly et al., 2005; Vendrov et al., 2006) investigated NAD(P)H oxidase-mediated signaling in atherosclerosis and identified several genes regulated by thrombin-induced NAD(P)H oxidase, including SEC61A2, in vascular smooth muscle cells (VSMCs). They demonstrated that NAD(P)H oxidase plays a role in the regulation of CD44 and BMP4-Id signaling pathway, which are implicated in restenosis and atherosclerosis. These findings suggest that SEC61A2 and other genes controlled by NAD(P)H oxidase may have important implications for vascular lesion formation. *Homo sapiens* histidine ammonia-lyase (HAL) is an enzyme involved in the catabolism of histidine. It plays a crucial role in modulating histidine metabolism, which influences the immune response, and angiogenesis (Krzymuska, 1964; Blaeschke et al., 2019) conducted a study on pediatric medulloblastoma, a brain tumor with minimal mutational load and low immunogenicity. Despite this, they identified immunogenic tumor-specific peptides in each patient, including peptides derived from the HAL gene. These findings suggest that even in tumors with low mutational load, specific T-cell immunotherapy targeting neoantigens is feasible and may guide future therapeutic approaches. Yu et al. (2015) conducted a study on the association between genetic variants, histidine levels, and incident coronary heart disease (CHD). They identified three rare loss-of-function (LoF) variants in the HAL gene, which encodes histidine ammonia-lyase, and found that these variants had significant effects on blood histidine levels. Furthermore, high blood histidine levels were associated with a reduced risk of developing CHD, suggesting a potential protective role of histidine in both African Americans and European Americans. By identifying these genes as potential biomarkers for both cancer prognosis and AMI prediction, we aim to shed light on the shared molecular mechanisms underlying these diseases. However, further experimental studies are needed to elucidate the specific roles of WEE1, PYHIN1, SEC61A2, and HAL in disease occurrence, progression, and response to treatment in both cancer and AMI.

Moreover, we also applied a novel formula to analyze mRNA levels in clinical samples from a meta-RNA-seq cohort and a meta-microarray cohort. The calculation resulted in the generation of a new risk score based on the expression profiles. Patients were classified into high-risk and low-risk categories using the median risk score, and Kaplan-Meier survival analysis revealed that high-risk patients exhibited significantly poorer overall survival compared to low-risk patients in both cohorts. These findings demonstrate the potential of utilizing shared biomarkers to predict outcomes in cancer patients. Furthermore, our study assessed the prognostic accuracy of the obtained risk score using ROC curves for various time points of overall survival. In the meta-RNA-seq training cohort, the AUC values demonstrated moderate to good accuracy for predicting 1, 3, 5, and 10-year overall survival. Similarly, in the meta-microarray validation set, the AUC values indicated fair to good accuracy for the same time points. To ensure consistent presentation and enhance clinical utility, we integrated FIGO staging and risk stratification into a visual nomogram. Similarly, we created a nomogram for the diagnosis of AMI patients. The calibration curve demonstrated the ability of the nomogram to accurately and reliably diagnose AMI among CAD patients. Moreover, the DCA curve and decision curve analysis provided further evidence of its robust diagnostic capacity. Importantly, in the validation dataset (GSE62646), the calibration curve, DCA curve, and clinical impact curve

consistently underscored the excellent external validation capability of the nomogram.

In the context of cancer, the correlation between the risk score and immune signatures indicates that patients with a higher risk score may have a more dysregulated immune system, which could influence their response to immunotherapy. Immunotherapies, such as immune checkpoint inhibitors, have revolutionized cancer treatment by enhancing the immune system's ability to recognize and eliminate tumor cells. Therefore, our findings suggest that patients with a higher risk score may be more suitable candidates for immunotherapeutic approaches, as they may have a greater potential to respond to these treatments. Similarly, in the context of AMI, the correlation between the risk score and immune signatures implies that the immune response plays a crucial role in the pathogenesis and progression of cardiac injury. Targeting immune-related pathways involved in AMI may provide new avenues for therapeutic interventions. Modulating the immune response, reducing inflammation, and promoting tissue repair are potential strategies for improving outcomes in patients with AMI. Therefore, understanding the relationship between the risk score and immune signatures can guide the development of novel therapies targeted at modulating the immune response in the context of AMI.

Of particular interest is the shared activation of the Notch Signaling pathway, which may provide insight into the common high-risk factors underlying both AMI and cancer. The role of Notch signaling in cancer is complex and contributes to enhanced tumorigenesis through various mechanisms such as angiogenesis, drug resistance, and epithelial to mesenchymal transition. Inhibiting the Notch pathway has emerged as a promising therapeutic strategy, and studies have shown promising results with Notch inhibitory agents in reducing tumorigenic aggressiveness (Sen and Ghosh, 2023; Yu et al., 2023) investigated the role of Notch signaling in innate lymphoid cells (ILCs) in acute coronary syndrome. The study found that activation of the Notch signaling pathway was associated with a shift from ILC1 to ILC2 subsets in peripheral blood of AMI patients, and inhibiting Notch signaling increased ILC1 frequency and interferon-γ secretion while reducing ILC2 frequency and interleukin-5/interleukin-13 production. These findings suggest that Notch signaling may play a role in regulating ILC subsets in AMI patients. Liu et al. (2019) investigated the role of miR-29b and its effect on myocardial infarction (MI) in rats through the Notch signaling pathway. The study demonstrated that downregulation of miR-29b in the MI group was associated with increased expression of Notch1, DII4, Hesl, and NICD1, suggesting that miR-29b inhibits myocardial fibrosis and cardiac hypertrophy by activating the Notch signaling pathway, providing protection against MI. Matsuda et al. (2014) investigated the impact of Notch signaling on human cardiac stem cells (CSCs) and their therapeutic potential in an AMI rat model. They found that reducing Notch signaling by culturing CSCs at low plating density enhanced their proliferation, multi-differentiation potential, and therapeutic efficacy, highlighting the importance of optimizing culture conditions for CSCs in clinical applications.

One practical application of the common signature is its potential in guiding treatment decisions. By profiling the common signature in individual patients, clinicians could better stratify patients and predict their response to specific therapies. For example, if a patient with cancer or AMI has a dysregulated immune-related common signature, it suggests that they may be more likely to benefit from immunotherapeutic or immune-modulatory interventions. This information can help inform treatment selection and improve personalized medicine approaches. Furthermore, the common signature can also guide the development of novel therapeutic strategies. By targeting the shared dysregulated pathways identified in the common signature, researchers and pharmaceutical companies can develop new drugs or repurpose existing ones to effectively treat both cancer and AMI. This approach could lead to the development of combination therapies that simultaneously target the common dysregulated pathways, potentially improving treatment efficacy and patient outcomes. Additionally, the common signature can have implications for prognosis and risk stratification. By assessing the expression levels or activation states of the common signature genes, clinicians may be able to predict disease progression, recurrence, or complications in both cancer and AMI patients. This information can aid in tailoring surveillance strategies and determining appropriate follow-up care for patients at higher risk.

There are several limitations to consider in this study. Firstly, the sample size of the clinical samples used for mRNA analysis was not clearly stated, which may affect the generalizability and statistical power of the findings. Additionally, the study focused on a specific set of biomarkers derived from monocytes, and it is possible that other relevant biomarkers were not considered. The study also primarily relied on retrospective data analysis, which may introduce biases and limit causal interpretations. Further prospective studies are needed to validate the predictive value of these biomarkers in larger, diverse patient populations. Finally, although the shared activation of the Notch signaling pathway is mentioned as a potential underlying factor, the specific mechanistic links between the identified biomarkers, AMI, and cancer prognosis are not fully explored or elucidated. Future research should aim to investigate these mechanisms in order to better understand the biological significance of these biomarkers.

Conclusion

In conclusion, the identification of shared biomarkers for cancer survival and AMI prediction represents a critical step toward improving patient care for individuals affected by these conditions. By understanding the underlying pathophysiological mechanisms and implementing personalized preventive strategies, healthcare providers can potentially reduce the burden of both diseases and improve patient outcomes. Our study demonstrates the potential of utilizing mRNA levels as biomarkers and highlights the importance of further research in this area to validate and refine these findings.

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 the Ethics Committee of the First Hospital of Jiaxing Affiliated Hospital of Jiaxing University. 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

NY and H-HP conceived and designed the study. Y-SL was responsible for materials. H-LH and C-LZ drafted the article. BW revised the article critically. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by Zhejiang Provincial Natural Science Foundation of China (No. LGF21H020006).

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/fphar.2023.1249145/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Prognostic value of monocytes assessed in the TCGA-OV cohort.

SUPPLEMENTARY FIGURE S2
Forest plot of univariate cox regression analysis.

SUPPLEMENTARY FIGURE S3
Validation of mRNA levels in clinical samples. **(A)** mRNA expression of WEE1, PYHIN1, SEC61A2, and HAL in clinical samples. **(B)** mRNA expression of WEE1, PYHIN1, SEC61A2, and HAL in cell lines.

References

Aydin, S., Ugur, K., Aydin, S., Sahin, I., and Yardim, M. (2019). Biomarkers in acute myocardial infarction: Current perspectives. *Vasc. Health Risk Manag.* 15, 1–10. doi:10.2147/VHRM.S166157

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: Archive for functional genomics data sets--update. *Nucleic Acids Res.* 41, D991–D995. doi:10.1093/nar/gks1193

Blaeschke, F., Paul, M. C., Schuhmann, M. U., Rabsteyn, A., Schroeder, C., Casadei, N., et al. (2019). Low mutational load in pediatric medulloblastoma still translates into neoantigens as targets for specific T-cell immunotherapy. *Cytotherapy* 21, 973–986. doi:10.1016/j.jcyt.2019.06.009

Blum, A., Wang, P., and Zenklusen, J. C. (2018). SnapShot: TCGA-analyzed tumors. *Cell* 173, 530. doi:10.1016/j.cell.2018.03.059

Boyum, A. (1968). Separation of leukocytes from blood and bone marrow. Introduction. *Introd. Scand. J. Clin. Lab. Invest. Suppl.* 97, 7.

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12

Chen, S., and Gardner, D. G. (2004). Suppression of WEE1 and stimulation of CDC25A correlates with endothelin-dependent proliferation of rat aortic smooth muscle cells. *J. Biol. Chem.* 279, 13755–13763. doi:10.1074/jbc.M310064200

Cheng, X., Li, J., Feng, L., Feng, S., Wu, X., and Li, Y. (2022). The role of hypoxia-related genes in TACE-refractory hepatocellular carcinoma: Exploration of prognosis, immunological characteristics and drug resistance based on onco-multi-OMICS approach. *Front. Pharmacol.* 13, 1011033. doi:10.3389/fphar.2022.1011033

Connerly, P. L., Esaki, M., Montegna, E. A., Strongin, D. E., Levi, S., Soderholm, J., et al. (2005). Sec16 is a determinant of transitional ER organization. *Curr. Biol.* 15, 1439–1447. doi:10.1016/j.cub.2005.06.065

de Las Fuentes, L., Sung, Y. J., Schwander, K. L., Kalathiveetil, S., Hunt, S. C., Arnett, D. K., et al. (2013). The role of SNP-loop diuretic interactions in hypertension across ethnic groups in HyperGEN. *Front. Genet.* 4, 304. doi:10.3389/fgene.2013.00304

Ding, J. M., Lin, W. R., Fei, Z. D., and Chen, C. B. (2022b). PYHIN1 correlates with CD8+ T cells infiltration and confers good patient survival in oral cancer. *J. Dent. Sci.* 17, 551–559. doi:10.1016/j.jds.2021.06.014

Ding, J., Sharon, N., Bar-Joseph, Z., Wu, Z., Zhang, X., Li, B., et al. (2022a). Recent advances in quantum dots-based biosensors for antibiotics detection. *Nat. Rev. Genet.* 23, 355–364. doi:10.1016/j.jpha.2021.08.002

Feng, S., Xu, Y., Dai, Z., Yin, H., Zhang, K., and Shen, Y. (2022). Integrative analysis from multicenter studies identifies a WGCNA-derived cancer-associated fibroblast signature for ovarian cancer. *Front. Immunol.* 13, 951582. doi:10.3389/fimmu.2022.951582

Hanzelmann, S., Castelo, R., and Guinney, J. (2013). Gsva: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Howard, E., Steingart, R. M., Armstrong, G. T., Lyon, A. R., Armenian, S. H., Teresa Voso, M., et al. (2019). Cardiovascular events in cancer survivors. *Semin. Oncol.* 46, 426–432. doi:10.1053/j.seminoncol.2019.01.007

Kozaczek, M., Bottje, W., Greene, E., Lassiter, K., Kong, B., Dridi, S., et al. (2019). Comparison of liver gene expression by RNAseq and PCR analysis after 8 weeks of feeding soy protein isolate- or casein-based diets in an obese liver steatosis rat model. *Food Funct.* 10, 8218–8229. doi:10.1039/c9fo01387c

Krzymuska, A. (1964). Enzymes of histidine metabolism in normal and tumor tissues; histidase and urocanase activity. *Arch. Immunol. Ther. Exp. Warsz.* 12, 724–729.

Kuhn, F., Schiergens, T. S., and Klar, E. (2020). Acute mesenteric ischemia. *Visc. Med.* 36, 256–262. doi:10.1159/000508739

Langfelder, P., and Horvath, S. (2008). Wgcna: an R package for weighted correlation network analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559

Lee, Y. C., Chao, Y. L., Chang, C. E., Hsieh, M. H., Liu, K. T., Chen, H. C., et al. (2019). Transcriptome changes in relation to manic episode. *Front. Psychiatry* 10, 280. doi:10.3389/fpsyt.2019.00280

Leening, M. J. G., Bouwer, N. I., Ikram, M. A., Kavousi, M., Ruiter, R., Boersma, E., et al. (2023). Risk of cancer after ST-segment-elevation myocardial infarction. *Eur. J. Epidemiol.* 38, 853–858. doi:10.1007/s10654-023-00984-8

Libby, P., and Kobold, S. (2019). Inflammation: A common contributor to cancer, aging, and cardiovascular diseases-expanding the concept of cardio-oncology. *Cardiovasc Res.* 115, 824–829. doi:10.1093/cvr/cvz058

Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1, 417–425. doi:10.1016/j.cels.2015.12.004

Liu, Y., Wang, H., Wang, X., and Xie, G. (2019). MiR-29b inhibits ventricular remodeling by activating notch signaling pathway in the rat myocardial infarction model. *Heart Surg. Forum* 22, E019–E023. doi:10.1532/hsf.2079

Ma, L., Lin, Y., Sun, S. W., Xu, J., Yu, T., Chen, W. L., et al. (2022). KIAA1429 is a potential prognostic marker in colorectal cancer by promoting the proliferation via downregulating WEE1 expression in an m6A-independent manner. *Oncogene* 41, 692–703. doi:10.1038/s41388-021-02066-z

Matsuda, T., Miyagawa, S., Fukushima, S., Kitagawa-Sakakida, S., Akimaru, H., Horii-Komatsu, M., et al. (2014). Human cardiac stem cells with reduced notch signaling show enhanced therapeutic potential in a rat acute infarction model. *Circ. J.* 78, 222–231. doi:10.1253/circj.cj-13-0534

Mills, C. D., Lenz, L. L., and Harris, R. A. (2016). A breakthrough: Macrophage-directed cancer immunotherapy. *Cancer Res.* 76, 513–516. doi:10.1158/0008-5472.CAN-15-1737

Mittrucker, H. W., Visekruna, A., and Huber, M. (2014). Heterogeneity in the differentiation and function of CD8$^+$ T cells. *Arch. Immunol. Ther. Exp. Warsz.* 62, 449–458. doi:10.1007/s00005-014-0293-y

Mohamed, T. M. A., Ang, Y. S., Radzinsky, E., Zhou, P., Huang, Y., Elfenbein, A., et al. (2018). Regulation of cell cycle to stimulate adult cardiomyocyte proliferation and cardiac regeneration. *Cell* 173, 104–116. doi:10.1016/j.cell.2018.02.014

Okabe, S., Tanaka, Y., Moriyama, M., and Gotoh, A. (2023). WEE1 and PARP-1 play critical roles in myelodysplastic syndrome and acute myeloid leukemia treatment. *Cancer Cell Int.* 23, 128. doi:10.1186/s12935-023-02961-3

Pan, H. H., Yuan, N., He, L. Y., Sheng, J. L., Hu, H. L., and Zhai, C. L. (2023). Machine learning-based mRNA signature in early acute myocardial infarction patients: The perspective toward immunological, predictive, and personalized. *Funct. Integr. Genomics* 23, 160. doi:10.1007/s10142-023-01081-5

Psaty, B. M., and Vasan, R. S. (2023). The association of myocardial infarction with cancer incidence. *Eur. J. Epidemiol.* 38, 851–852. doi:10.1007/s10654-023-01019-y

Reel, P. S., Reel, S., Pearson, E., Trucco, E., and Jefferson, E. (2021). Using machine learning approaches for multi-omics data analysis: A review. *Biotechnol. Adv.* 49, 107739. doi:10.1016/j.biotechadv.2021.107739

Rinde, L. B., Smabrekke, B., Hald, E. M., Brodin, E. E., Njolstad, I., Mathiesen, E. B., et al. (2017). Myocardial infarction and future risk of cancer in the general population-the Tromso Study. *Eur. J. Epidemiol.* 32, 193–201. doi:10.1007/s10654-017-0231-5

Sen, P., and Ghosh, S. S. (2023). The intricate notch signaling dynamics in therapeutic realms of cancer. *ACS Pharmacol. Transl. Sci.* 6, 651–670. doi:10.1021/acsptsci.2c00239

Shaikh, A. Y., and Shih, J. A. (2012). Chemotherapy-induced cardiotoxicity. *Curr. Heart Fail Rep.* 9, 117–127. doi:10.1007/s11897-012-0083-y

Stewart, C., Ralyea, C., and Lockwood, S. (2019). Ovarian cancer: An integrated review. *Semin. Oncol. Nurs.* 35, 151–156. doi:10.1016/j.soncn.2019.02.001

Su, Y. L., Xiao, L. Y., Huang, S. Y., Wu, C. C., Chang, L. C., Chen, Y. H., et al. (2023). Inhibiting WEE1 augments the antitumor efficacy of cisplatin in urothelial carcinoma by enhancing the DNA damage process. *Cells* 12, 1471. doi:10.3390/cells12111471

Tong, Y., Song, Y., and Deng, S. (2019). Combined analysis and validation for DNA methylation and gene expression profiles associated with prostate cancer. *Cancer Cell Int.* 19, 50. doi:10.1186/s12935-019-0753-x

Vendrov, A. E., Madamanchi, N. R., Hakim, Z. S., Rojas, M., and Runge, M. S. (2006). Thrombin and NAD(P)H oxidase-mediated regulation of CD44 and BMP4-Id pathway in VSMC, restenosis, and atherosclerosis. *Circ. Res.* 98, 1254–1263. doi:10.1161/01.RES.0000221214.37803.79

Wang, X., Spandidos, A., Wang, H., and Seed, B. (2012). PrimerBank: A PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res.* 40, D1144–D1149. doi:10.1093/nar/gkr1013

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi:10.1038/nrg2484

Yu, B., Li, A. H., Muzny, D., Veeraraghavan, N., de Vries, P. S., Bis, J. C., et al. (2015). Association of rare loss-of-function alleles in HAL, serum histidine: Levels and incident coronary heart disease. *Circ. Cardiovasc Genet.* 8, 351–355. doi:10.1161/CIRCGENETICS.114.000697

Yu, H., Wei, Y., Dong, Y., and Chen, P. (2023). Regulation of notch signaling pathway to innate lymphoid cells in patients with acute myocardial infarction. *Immunol. Invest.* 52, 241–255. doi:10.1080/08820139.2022.2158856

Zhang, K., Feng, S., Ge, Y., Ding, B., and Shen, Y. (2022). A nomogram based on seer database for predicting prognosis in patients with mucinous ovarian cancer: A real-world study. *Int. J. Womens Health* 14, 931–943. doi:10.2147/IJWH.S372328

Zhao, S., Wu, Y., Wei, Y., Xu, X., and Zheng, J. (2022). Identification of biomarkers associated with CD8+ T cells in coronary artery disease and their pan-cancer analysis. *Front. Immunol.* 13, 876616. doi:10.3389/fimmu.2022.876616

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 15 September 2023
DOI 10.3389/fphar.2023.1240736

OPEN ACCESS

EDITED BY
Lin Qi,
Central South University, China

REVIEWED BY
Hongyu Zhao,
Hackensack University Medical Center, United States
Liu Xiufeng,
Sun Yat-sen University Cancer Center, China

*CORRESPONDENCE
Dehua Wu,
18602062748@163.com
Xie Weng,
wengxie.100@163.com
Yuming Chen,
chenyumingdg@163.com

†These authors have contributed equally to this work

RECEIVED 15 June 2023
ACCEPTED 04 September 2023
PUBLISHED 15 September 2023

CITATION
Kuang J, Zheng Z, Ma W, Zeng S, Wu D, Weng X and Chen Y (2023), Comprehensive analysis of Cuproplasia and immune microenvironment in lung adenocarcinoma.
Front. Pharmacol. 14:1240736.
doi: 10.3389/fphar.2023.1240736

Comprehensive analysis of Cuproplasia and immune microenvironment in lung adenocarcinoma

Junjie Kuang[1†], Zemao Zheng[2†], Wen Ma[3], Shaohui Zeng[4], Dehua Wu[5*], Xie Weng[6*] and Yuming Chen[1*]

[1]Dongguan Institute of Clinical Cancer Research, Dongguan Key Laboratory of Precision Diagnosis and Treatment for Tumors, Affiliated Dongguan Hospital, Southern Medical University, Dongguan, Guangdong, China, [2]Department of Respiratory and Critical Care Medicine, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, China, [3]Department of Medical Oncology, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, Guangzhou, Guangdong, China, [4]Shenzhen Hospital (Futian) of Guangzhou University of Chinese Medicine, Shenzhen, Guangdong, China, [5]Department of Radiation Oncology, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, China, [6]Cancer Center, Integrated Hospital of Traditional Chinese Medicine, Southern Medicine University, Guangzhou, Guangdong, China

Background: Trace elements such as copper are essential for human health. Recently the journal Nat Rev Cancer has put forward the concept of Cuproplasia, a way of promoting tumor growth through reliance on copper. We attempted to conduct a comprehensive analysis of Cuproplasia-related genes in lung adenocarcinoma (LUAD) to explore the mechanism of action of Cuproplasia-related genes in LUAD.

Method: Transcriptome data and clinical information of LUAD were obtained from TCGA-LUAD and GSE31210, and prognostic models of Cuproplasia-related genes were constructed and verified by regression analysis of GSVA, WGCNA, univariate COX and lasso. The signal pathways affected by Cuproplasia-related genes were analyzed by GO, KEGG and hallmarK pathway enrichment methods. Five immunocell infiltration algorithms and IMVIGOR210 data were used to analyze immune cell content and immunotherapy outcomes in the high-low risk group.

Results: In the results of WGCNA, BROWN and TURQUOISE were identified as modules closely related to Cuproplasia score. In the end, lasso regression analysis established a Cuproplasia-related signature (CRS) based on 24 genes, and the prognosis of high-risk populations was worse in TCGA-LUAD and GSE31210 datasets. The enrichment analysis showed that copper proliferation was mainly through chromosome, cell cycle, dna replication, g2m checkpoint and other pathways. Immunoinfiltration analysis showed that there were differences in the content of macrophages among the four algorithms. And IMVIGOR210 found that the lower the score, the more effective the immunotherapy was.

Conclusion: The Cuproplasia related gene can be used to predict the prognosis and immunotherapy outcome of LUAD patients, and may exert its effect by affecting chromosome-related pathways and macrophages.

KEYWORDS

Cuproplasia, lung adenocarcinoma (LUAD), prognostic models, gene signatures, immunotherapy

Introduction

There are currently more deaths from lung cancer than any other type of cancer worldwide (Bray et al., 2018). Lung adenocarcinoma (LUAD) is the most common pathological type of lung cancer, and the incidence of LUAD is increasing year by year. They tend to be younger, with fewer initial symptoms and more rapid onset, The fatality rate is high and the prognosis is poor. Most patients are diagnosed late period (Hong et al., 2020).

The human body contains trace elements, which constitute a minute fraction, 0.005%–0.01%, of its total composition, including iron, zinc, nickel, copper, selenium, iodine, manganese, cobalt, chromium, vanadium, fluorine, silicon, molybdenum, and tin (Costa et al., 2023; Frydrych et al., 2023; Xie et al., 2023; Yang et al., 2023). Despite their minimal content, trace elements are crucial for human health as they perform essential physiological and biochemical functions. Recent studies have revealed a significant association between trace elements and tumor progression and mortality (Aishajiang et al., 2023). One element of particular interest is copper, the level of which is critical for tumor development (van Renterghem et al., 2023). Both tumor tissue and serum copper levels in cancer patients are significantly elevated, with increased copper concentrations observed in a variety of tumors, including breast, prostate, lymphoma, cervical, lung, and stomach cancers (Denoyer et al., 2015).

Clinical evidence further underscores the importance of copper in cancer development. Specifically, abnormal copper accumulation in Wilson disease patients fosters malignant transformation of liver cells (Jopowicz and Tarnacka, 2023). Moreover, the Mitogen-activated protein kinases (MAPK) signaling pathways, which are intimately involved in the development of various cancers (Turski et al., 2012), are influenced by copper levels. Approximately 40%–50% of melanomas and other common tumors, such as lung, thyroid, and colorectal cancers, exhibit BRAF gene mutations, resulting in the constitutive activation of MAPK channels (Dankner et al., 2018). Dankner et al. (2018) discovered that intracellular copper directly binds and activates MAPK kinase (MEK), promoting MAPK signaling in *Drosophila melanogaster* and consequently leading to tumor development. Additionally, Brady DC et al. (Brady et al., 2014) demonstrated that copper ions enhance MEK phosphorylation of downstream extracellular signal-regulated kinase (ERK) through MEK interaction. Decreasing intracellular Ctr1 expression inhibits BRAF-mediated downstream ERK signaling pathway activation. Similarly, mutating the MEK1 copper binding site achieves the same inhibitory effect. Interestingly, Tsang T et al. (Tsang et al., 2020) identified MeK1-like copper binding sequences in unc-51-like kinase 1 (ULK1) and unc-51-like kinase 2 (ULK2). Further experiments confirmed that copper ions binding to ULK1 or ULK2 activates them, stimulating the autophagy pathway and resulting in mouse lung cancer proliferation.

The influence of copper ions extends to other signaling pathways as well. Some studies have demonstrated that copper ions bind to 3-phosphoinositide-dependent protein kinase-1 (PDK1), enhancing its interaction with serine/threonine protein kinase AKT (also known as protein kinase B) and activating AKT's oncogenic signaling in a phosphatidylinositol-3-kinase (PI3K)-dependent manner. Inhibiting the copper axis diminishes AKT signaling and suppresses tumor development, indicating a close relationship between the PI3K-PDK1-AKT axis and tumor proliferation (Guo et al., 2021).

Recently, the concept of Cuproplasia was first proposed in the journal nature reviews cancer to try to explain the cell proliferation mode dependent on copper ions. In this study, we carried out landscape description of the Cuproplasia genes summarized in this review in LUAD. The Cuproplasia risk score (CRS) was established by Least absolute shrinkage and selection operator (LASSO) regression analysis to predict the overall survival (OS) of LUAD patients, and multiple omics analysis was used to try to explain the specific role of Cuproplasia.

Materials and methods

Data collection and collation

In The Cancer Genome Atlas-lung adenocarcinoma (TCGA-LUAD) database, data categories were selected as transcrip-tome profiling and raw counts, including 535 primary lung adenocarcinoma samples and 59 normal samples. Clinical information including complete gender, age, survival time, survival status, and pathological stage of 201 cases was downloaded from UCSC for subsequent analysis. Transcriptome data and clinical information of 210 patients in GSE31210 were downloaded from the (Gene Expression Omnibus) GEO database. R software is used to process the data sets of TCGA and GEO. Gene expression levels are determined by the average value of expression if multiple probes correspond to the same gene. The Benjamini–Hochberg method is adopted to adjust the p value to control the false discovery rate (FDR). After removing duplicate genes and their expression in the original data downloaded from the TCGA database, the CPM function of the R software edgeR package was used to correct and standardize the data, and the mean CPM (count value of each transcript per million bases) ≤1 was deleted, and then we take the logarithm base 2. The copper proliferation gene was obtained from the literature (Ge et al., 2022).

Gene set variation analysis (GSVA) scores of Cuproplasia genes

GSVA is a method that can be applied to Microarray and RNA-sequence data sets of enrichment under the conditions of no parameters and without supervision by a trained scientist (Hänzelmann et al., 2013). GSVA can convert a gene-sample data matrix into gene-set - sample matrix. Based on this matrix, the enrichment of gene sets (such as KEGG pathway) in each sample can be further analyzed. Since GSVA is a gene-sample Enrichment matrix, downstream Analysis would be allowed more freedom than other Gene Set enrichment methods, such as Gene Set Enrichment Analysis (GSEA) (Yi et al., 2020). In this study, we scored the copper proliferation level of patients in TCGA-LUAD based on the Cuproplasia-related genes.

Weighted gene co-expression network analysis (WGCNA)

From whole-genome expression analysis, WGCNA provides valuable information about gene function and gene association, and can be used to detect module-membership (MM) of highly correlated genes and a module related to gene-significance (GS) that provides insight into the function of co-expressed genes, it can also assist in the identification of genes that play a critical role in the development of human diseases (Zhang and Horvath, 2005; Saris et al., 2009; Li et al., 2022; Cheng et al., 2023). The co-expression network was constructed by using the WGCNA package of R software to construct the TCGA data set. The Pearson coefficient between each gene was firstly calculated to convert it into a similar matrix, and the soft threshold β was automatically selected for network topology analysis through the pick soft threshold function of the WGCNA package. β can emphasize the strong and weak correlation between genes (Langfelder and Horvath, 2008) and β is set to 4, scale-free = 0.9. After β was determined, the similar matrix was transformed into an adjacency matrix, and then the adjacency matrix was transformed into a topological overlap matrix (TOM). The minimum number of genes in the module was set as 50, and the shear height was set as 0.25. The genes with similar expressions were placed in the same gene module through hierarchical clustering, and the threshold was set as 20,000 to eliminate outliers. Genes that express similar patterns are grouped into different modules. Gene modules closely related to tumorigenesis were selected based on correlation coefficients between genes and phenotypes (cancer tissue and normal samples). If a gene in the module has both large MM and GS, it is considered to be the core gene in the module, and the MM > 0.7 and GS > 0.35 was defined as the candidate core gene. Then, intersection of central genes selected by Cytohubba and core genes selected by modules was selected, and genes in the intersection were defined as the final key genes.

Acquisition of differential genes in cancer tissue and normal tissue

A powerful transcriptomics technique is differential gene expression analysis, which demonstrates quantitative changes in gene expression between normal and cancer cells based on molecular mechanisms. Such differences in gene expression can reveal potential biomarkers for specific diseases. We use R software packages for key genes Limma for differences in gene screening (Ritchie et al., 2015), selection criteria for: | log2 (fold—change) | ≥1 and the corrected *p* value (false discovery rate, FDR) ≤ 0.05.

Screening of prognostic related copper proliferating genes

Univariate Cox regression analysis was performed for differential gene expression and overall survival between tumor and normal tissues to screen out prognostic copper proliferating genes. The screening criteria were *p* value less than 0.05 (Lunn and McNeil, 1995).

LASSO regression analysis

The Least Absolute Shrinkage and Selection Operator (LASSO) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces (Li et al., 2023; Tu et al., 2023). By using the glmnet package, survival status was used as the dependent variable, and the expression value of the selected differential prognostic genes was used as the response variable. 1,000 Lasso regression analyses were carried out to reduce the number of genes, so as to reduce the error of the model and obtain a generalized linear model (Tibshirani, 1997). The Prognostic genes were identified using multivariate Cox proportional risk regression analysis. Predicting prognosis status was based on CRS. Molecular expressions of individual genes in the sample were taken into account in determining the multivariate Cox proportional risk regression scores for each patient. Following is a detailed description of the calculation formula:

$$CRS = \sum_{i=1}^{n} Exp_i * Coef_i$$

Receiver operating curve (ROC curve) of the relationship between CRS and prognosis

We calculated the CRS for the training dataset (TCGA-LUAD) and validation (GSE31210) using the prognosis model, and we divided the validation and training sets by the median CRS. The survival curve was drawn based on the survival information to obtain the survival status of high and low risk expression, and the prediction effect of the model was evaluated (≤0.05), the statistical method used in this process is the log-rank test. To evaluate regression models' predictive ability in 1-year, 3-year, and 5-year survival, the time-dependent ROC curve was calculated using R software's "survival ROC" package. When the AUC is larger than 0.5 and closer to 1, the prediction effect is better.

Functional enrichment analysis

To explore the pathogenesis and development mechanism of lung adenocarcinoma, gene function analysis (gene ontology, GO) and pathway analysis (Kyoto encyclopedia of genes and genomes, Genomes, KEGG) for detailed biological annotation and description of the function of gene products. GO covers molecular function (MF), cellular components (CC) and biological processes (BP). The functional information of a given gene was comprehensively summarized through enrichment analysis (Harris et al., 2004). KEGG incorporates information about genomics, chemical processes, and systematic functions. The method analyzes gene function from all angles of gene and molecular networks, which are thought to be responsible for identifying metabolic and functional pathways (Kanehisa and Goto, 2000). The path enrichment analysis was also carried out by hallmark.

Relationship between GSVA score and riskscore

Spearman method was used to analyze the correlation between GSVA score and riskScore of Cuproplasia gene set.

Immune checkpoint analysis

The expression levels of 10 common immune checkpoints (CTLA4, PDCD1, CD274, ICOS, LAG3, BTLA, TNFRSF14, NRP1, CD28, and CD44) were analyzed among high and low risk groups in the TCGA-LUAD cohort to determine whether the CRS could be used in the treatment of immune checkpoints.

Relationship between CRS and immunotherapy

Scoring and immunotherapy were performed using the IMVIGOR210 CoreBiologies. We downloaded the data of TCGA-LUAD patients from this dataset for analysis. The specific method of analysis was to compare immunotherapy outcomes in low-risk patients, including complete response (CR), partial response (PR), stable disease (SD) and progressive disease (PD).

Hierarchical analysis

To observe the effectiveness of riskScore in different clinical stages, LUAD patients were divided into four stages according to the American Joint Committee on Cancer staging (AJCC) staging method, and K-M survival curves of high and low risk groups were drawn in each stage according to the median CRS.

Analysis of immune cell infiltration

Immune cell infiltration and risk scoreUtilizing the xCell method (Aran et al., 2017), we assessed the enrichment levels of 64 immunological markers to evaluate the immune cell invading the microenvironment. Using techniques including CIBERSORTx (Steen et al., 2020), ssGSEA (Yi et al., 2020), quanTIseq (Finotello et al., 2019), TIMER, and MCPcell from the R package immunedeconv version 2.0.4, we conducted more exhaustive analyzes.

Cell culture

The human LUAD cell line A549 was purchased from the American Type Culture Collection (ATCC) and cultured in RPMI 1640 medium supplemented with 10% fetal bovine serum (FBS) and 105 IU/L penicillin and 0.1 g/L streptomycin, at 37°C in a humidified atmosphere containing 5% CO2.

siRNA transfection

Lipofectamine 2000 was used for transient transfection of siRNA. Cells were plated in advance and A549 cells in logarithmic growth phase were adjusted to a density of 1 × 10^5 cells/mL. When the confluence reached 60%, the siRNA complexes were added to the plated cells along with the transfection reagent in serum-free medium, and after 6 h, the medium was replaced with serum-containing medium. 48 h post-transfection, RNA or protein was extracted to assess transfection efficiency.

RT-qPCR

Cells were seeded in 6-well plates and transfected with si-LAG3 for 48 h, then the medium was discarded and cells were collected. Total RNA was extracted using the Trizol one-step method. The reaction system was prepared according to the instructions of the kit, and mRNA was reverse transcribed to cDNA at 37°C for 60 min. The PCR reaction system was prepared with 2.5×RealMaster Mix/20×SYBR solution 4.5 μL, 1 μL of each primer, 2 μL cDNA, and 1.5 μL of triple-distilled water, making up a total of 10 μL. The primer sequences are shown in Supplementary Table S1. The reaction conditions were 94°C for 15 min, followed by 40 cycles of 94°C for 20 s, 56°C for 30 s, and 68°C for 30 s. GAPDH was used as an internal control to calculate the expression of the target RNA. The primer sequence for si-LAG3 is as follows: GGAGACAAU GGCGACUUUA (5'-3').

Western blot

Cells were seeded in 6-well plates and transfected with si-LAG3 for 48 h. The medium was then discarded, and the cells were collected and lysed on ice for 30 min. After centrifugation, the supernatant was collected, and protein concentration was determined using a BCA protein assay kit. The proteins were then separated by SDS/PAGE electrophoresis and transferred onto PVDF membranes. The membranes were blocked, incubated with primary antibodies overnight at 4°C, then with secondary antibodies for 1 h at room temperature, and finally, the bands were visualized using an ECL detection kit.

CCK8 assay

Cells in logarithmic growth phase were prepared into a cell suspension at a density of 5 × 10^3 cells/mL, and 100 μL of the cell suspension was seeded into each well of a 96-well plate and incubated. At 1d, 2d, 3d, and 4d post-seeding, 10 μL of CCK-8 solution was added to each well and incubated for 1–2 h. The absorbance at 450.0 nm was measured using a microplate reader.

Transwell cell migration assay

Adherent cells were digested and resuspended in serum-free medium. The cell suspension was counted and diluted to a density of 2 × 10^6 cells/mL with serum-free medium. A pipette was used to mix the cell suspension thoroughly. Transwell chambers with 8.0 μm pores were placed into a 24-well plate, and 400 μL of the cell suspension was added to the upper chamber, while the lower chamber was filled with medium containing 20% FBS. The plate was incubated at 37°C. After 36 h, the chambers were removed, and the cells were fixed with 4% paraformaldehyde for 20 min. The paraformaldehyde was then washed off with ultrapure water, and the cells were stained with crystal violet for 5 min. Excess stain was washed off with water, and the cells in the upper chamber were gently removed with a cotton swab. The chambers were observed under a microscope, and images were captured from five different fields of view. The cells were counted using ImageJ software.

Wound healing assay

Logarithmically growing cells were resuspended in culture medium and seeded into 6-well plates. When the cell density reached 90%, a straight line was scratched in the center of each well using a 200 μL plastic pipette tip. The scratched cells and debris were washed off with PBS, and serum-free medium was added to each well. The plate was incubated at 37°C, and images of the wounds were captured at 0 h and 24 h. The wound area was measured using ImageJ software.

EDU/DAPI staining for cell proliferation

After seeding and transfecting cells with siRNA, 2xEDU solution was added to an equal volume of culture medium containing the experimental cells and co-incubated for 12 h. The medium was then discarded, and the cells were fixed with 2.5 mL of PBS containing paraformaldehyde for 15 min at room temperature. The fixative was removed, and the cells were washed three times with PBS. Then, 2.5 mL of permeabilization solution was added, and the cells were incubated for 20 min at room temperature. The permeabilization buffer was removed, and the cells were washed twice with PBS. The staining reaction mixture was prepared according to the instructions of the EDU staining kit, and 100 mL of the mixture was added to each well. The cells were incubated in the dark for 30 min at room temperature, then washed once with PBS. Under a fluorescence microscope, the proliferating cells stained with EDU appeared red, while the cell nuclei stained with DAPI appeared blue.

Statistical analysis

In comparing two categories with normal distributions and non-normally distributed data, statistical significance was determined using independent t-tests and Mann–Whitney U tests. Multiple category differences were compared using one-way analysis of variance (ANOVA) and Kruskal–Wallis analyses. The R package Hmisc 4.4.1 was employed to conduct Spearman and distance correlational analysis. Objects having a correlation coefficient greater than 0.5 were considered highly related. To determine the prognostic variables, Cox regression analysis was carried out. Prior to creating the survivorship curves using the R package survminer, the overall survival (OS) and CRS were calculated with the R package survival, and cutoff values were set with the R program survminer. Utilizing the R package Complex Heatmap 2.4.3, every heatmap was generated. Using the ggplot2 R software, data comparisons were illustrated. All statistical analyzes were performed on both sides of the data using R software. Statistical significance was determined by a p-value of 0.05.

Results

WGCNA result

Cuproplasia scores were performed for each patient in TCGA-LUAD through GSVA function in GSEA, and the scores of patients were shown in Supplementary Table S2 list.

The WGCNA process is as follows: The gene co-expression network was established (Figures 1B,C), and the process of module identification was shown in Figures 1A,D,E. The results show that the correlation coefficients between BROWN and TURQUOISE and copper proliferation mode are 0.41 and 0.38 respectively (Figure 1F). Therefore, BROWN and TURQUOISE are considered as modules closely related to copper proliferation. There were 1,593 genes in BROWN and 3,194 genes in TURQUOISE, and the causes of 4,787 of these modules were included in subsequent analyses ().

Construction and validation of CRS

A total of 9,064 differentially expressed genes were obtained between tumor tissues and normal tissues in the TCGA-LUAD dataset, and the results were presented in. We combined 9,064 differentially expressed genes with overall survival and performed univariate cox analysis to obtain 1,439 prognostic related genes. The intersection of 1,439 prognostic genes with the aforementioned BROWN and TURQUOISE genes obtained 214 prognostic genes related to copper proliferation.The formula of Lasso regression after dimensionality reduction is as follows: risk socres = −0.1004*PLA2G4F + 0.155*NIM1K + 0.0277*PLEK2 + 0.0402*SIRPA + 0.1713*FABP5 + 0.2555*PTX3 + 0.1306*CCT6A + 0.2177*STARD4 + 0.1917*SEC61G + −0.0174*BEX4 + 0.2549*LINC02535 + 0.2983*SHC1 + −0.269*MBOAT1 + 0.043*KRT81 + 0.0063*MT1X + −0.1954*BCAS4 + 0.0302*KRT6A + 0.0432*CD109 + −0.1846*HS3ST2 + 0.111*TINAG + 0.0055*LHX2 + 0.069*CNTNAP2 + 0.0063*KLK8 + 0.1446*CALML5。

According to K-M survival curves, the high risk group had a significantly higher survival rate than the low risk group ($p \leq 0.001$) (Figure 2A). ROC curve showed that the value of AUC in the first year, the third year and the fifth year were 0.785, 0.714, and 0.701 respectively, demonstrating fair prediction ability (Figure 2B).

Similar results were observed in the validation set GSE31210, with worse survival in the high-risk group ($p = 0.01$) (Figure 2C). These results suggest that the risk score constructed by Cuproplasia-related genes can predict the OS of LUAD patients and can be used as a prognostic indicator for patients.

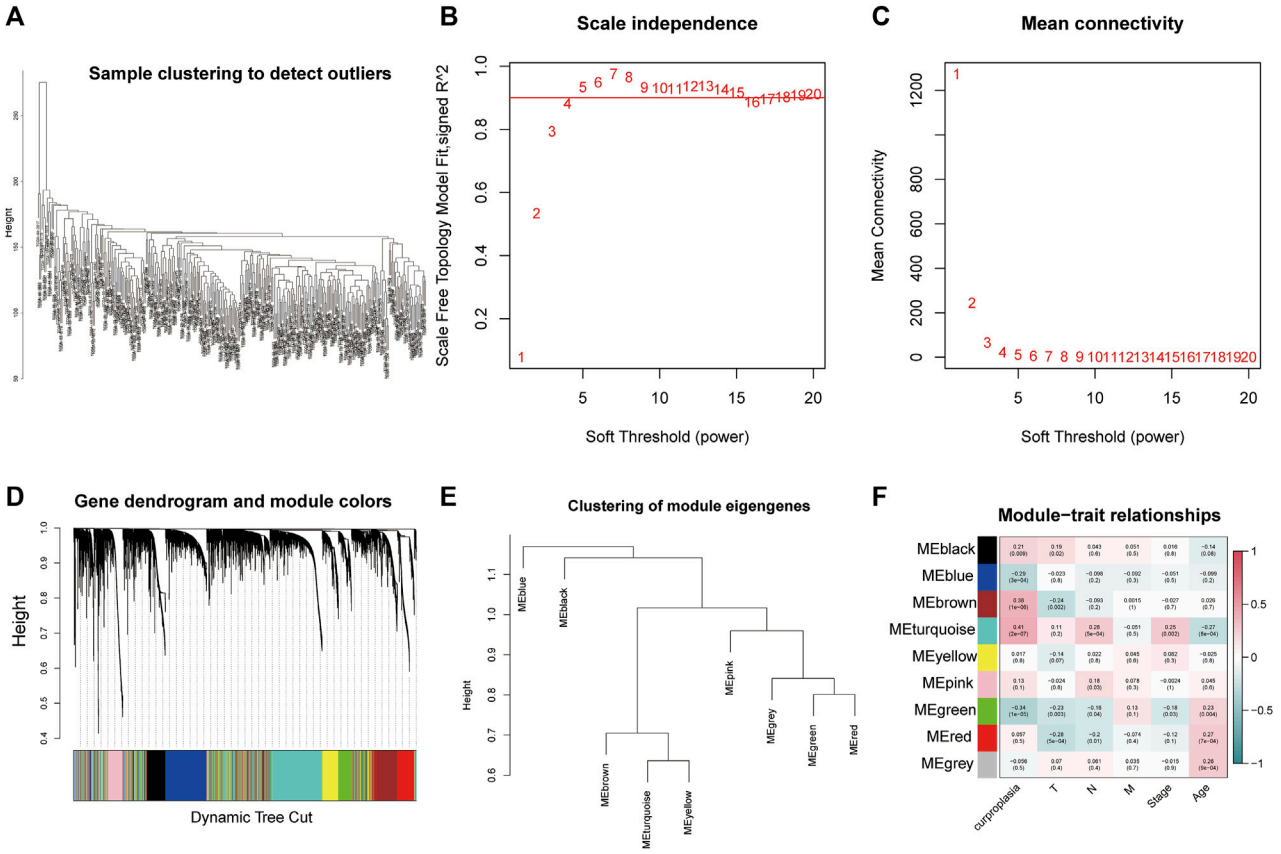


FIGURE 1
Weighted correlation network analysis results. **(A)** Sample clustering to detect outliers. **(B)** Scale independence results. **(C)** Mean connectivity. **(D)** Gene dendrogram and module colors. **(E)** Clustering of module eigengenes. **(F)** Module–trait relationships.

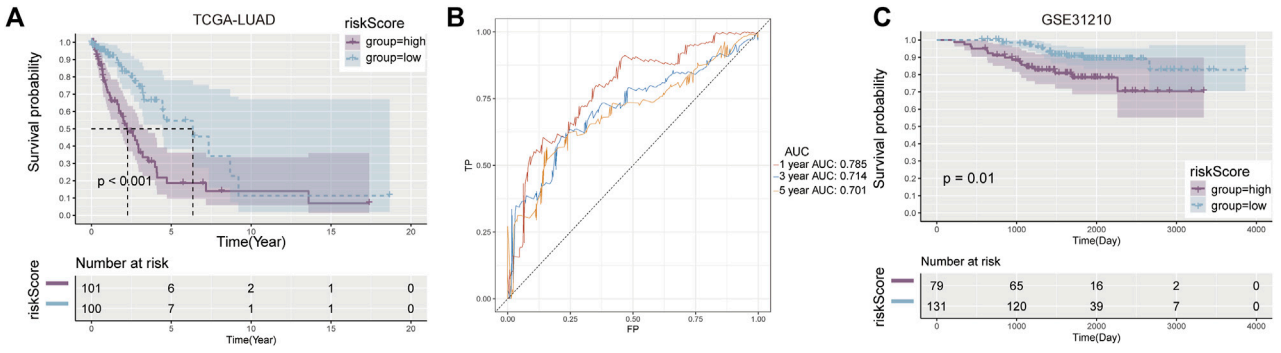


FIGURE 2
Development and validation of prognostic model. **(A)** Kaplan-Meier curves between high and low risk groups in the training cohort. **(B)** AUC values in the first, third and fifth year prognostic model. **(C)** Verify the K-M curve between high—and low-risk groups in the cohort.

Pathway of enrichment of Cuproplasia-related genes

GO, KEGG and Hallmark pathway enrichment methods were used to analyze the possible influence of copper proliferation-related genes on the differential genes between high and low risk groups. We showed the top ten pathways with the most significant *p* value of GO enrichment analysis, and the results were respectively enriched in mitotic sister chromatid segregation, nuclear chromosome segregation, sister chromatid segregation, chromosome segregation, chromosome centromeric region, condensed chromosome centromeric region, mitotic nuclear division, condensed chromosome, organelle fission, regulation of mitotic nuclear division. These results mainly suggest that risk scores are related to biological

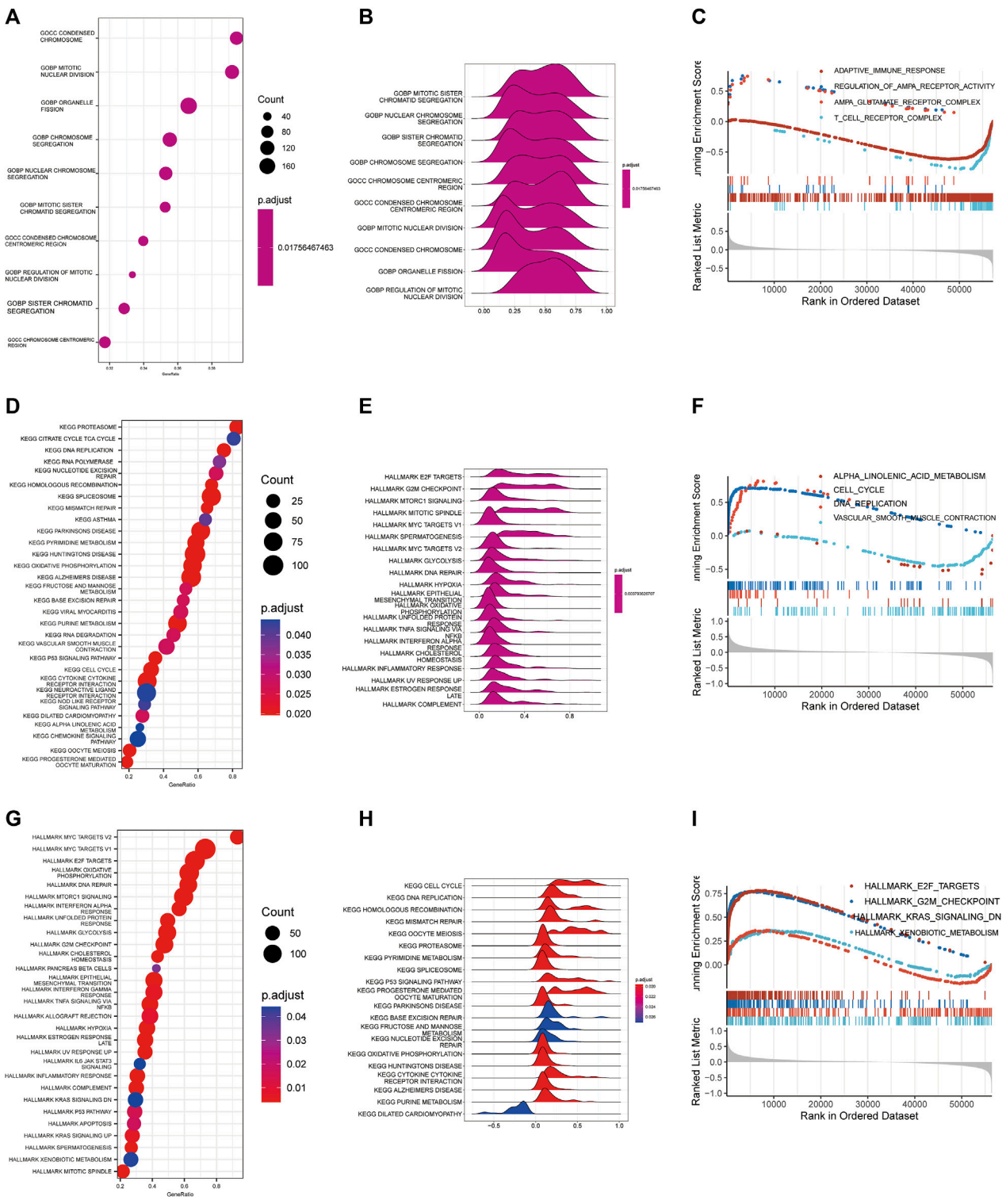


FIGURE 3
Functional enrichment analysis. Bubble diagram **(A)**, Ridge map **(B)**, GSEA diagram **(C)** of GO functional enrichment analysis. Bubble diagram **(D)**, Ridge map **(E)**, GSEA diagram **(F)** of GO functional enrichment analysis. Bubble diagram **(G)**, Ridge map **(H)**, GSEA **(I)** diagram of GO functional enrichment analysis.

processes such as chromosomes. The bubble diagram is used to sort enrichment pathways by GeneRatio, and is shown in Figure 3B according to *p*-value. Figure 3C shows the four pathways with the highest NES value in the GO analysis. KEGG analysis revealed that risk scores were closely related to the following 20 pathways, and the results are shown in Figures 3 D,E. The top four pathways with the most NES

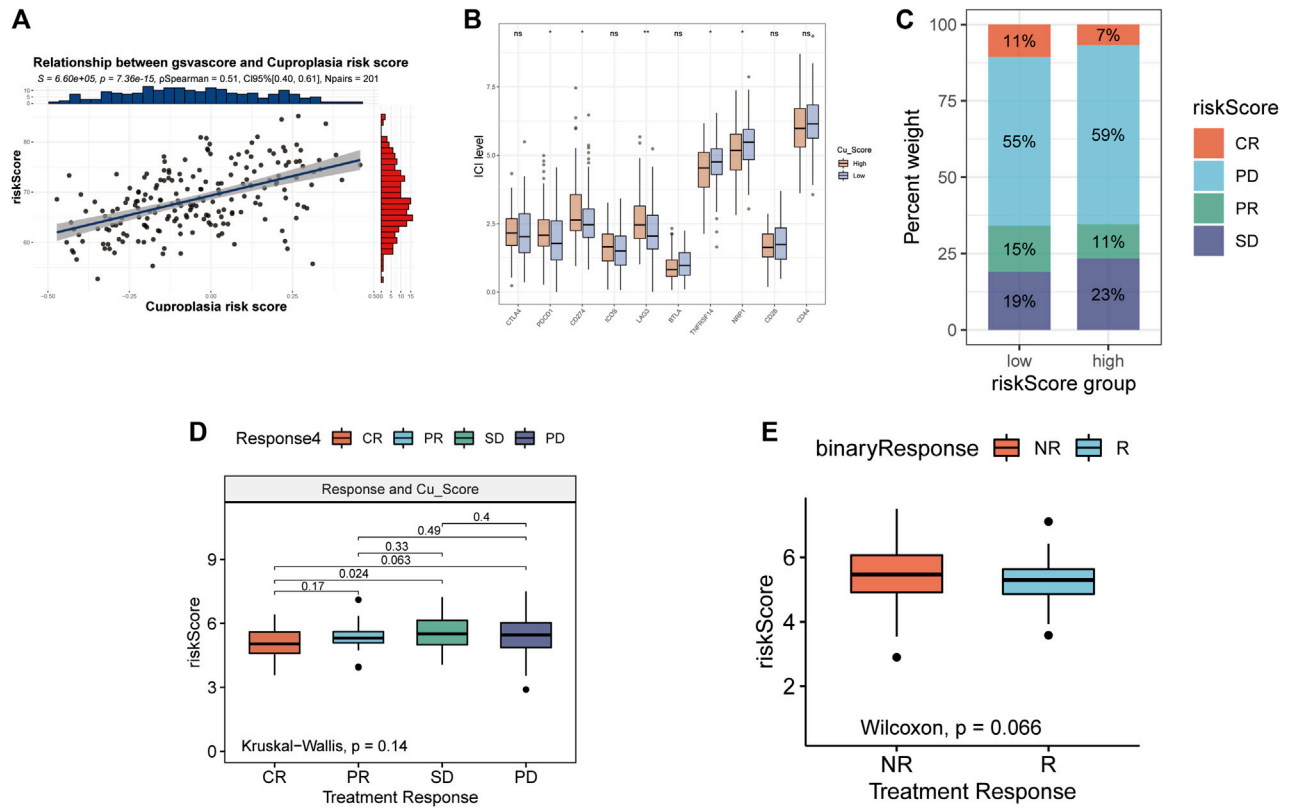


FIGURE 4
Relationship between risk scores and immunotherapy. **(A)** Relationship between gsvascore and riskScore. **(B)** The percentage of different treatment outcomes in the high-low risk group. **(C)** Expression levels of 10 immune checkpoints in different risk groups. **(D)** Risk scores in epidemic groups. **(E)** Risk scores for responders and non-responders.

score are alpha linolenic acid metabolism, cell cycle, DNA replication, vascular smooth muscle contraction (Figures 3F, G). It is suggested that Cuproplasia-related genes may play a role through these pathways and affect the OS of LUAD.HALLMARK enrichment analysis found that risk scores were closely related to the following pathways: E2F targets, G2M checkpoint, KRAS signaling down, xenobiotic meta (Figures 3H, I). LUAD's occurrence and development may be influenced by CRS through these pathways, which affects patients' survival chances as well.

Risksores is closely related to immunotherapy

The correlation between GSVA score of copper proliferation gene set and riskScore was observed, and a strong positive correlation was found (Figure 4A). Analysis of the expression levels of 10 immune checkpoints in the high-low risk group showed that PDCD1, CD274, LAG3 were highly expressed in the high-risk group, and TNFRSF14 and NRP1 were highly expressed in the low-risk group. These results suggest that CRS is a predictor of LUAD immune checkpoint therapy (Figure 4B). stack plots showed that there were more patients in the low-risk group with CR and that the difference in survival was not significant (Figure 4C). The IMVIGOR210 analysis showed that the lower the risk score, the better the treatment outcome (Figure 4D). However, binary comparisons did not show any significant differences between reactive and non-reactive groups (Figure 4E).

Risk scores are also prognostic indicators in different clinical stages

According to the AJCC, patients at different stages were divided into high-low risk group according to the median risk score, and it was found that patients in the high-risk group of stage III and stage IV patients also had worse prognosis (Figures 5C,D). However, the high-risk group in stage I and stage II also had a worse trend (Figures 5A,B). This suggests that the higher the stage, the greater the clinical value of the risk score. Figure 5E shows the relationship between risk scores and stages, and the results show that the risk scores of stage I, stage II and stage III are different. These results indicate that risk score is closely related to clinical stage and has better clinical predictive value in stage III and IV patients.

The relationship between risk scores and immune cells

The results of six immune cell analysis algorithms between high and low risk groups are shown in Figure 6, in which immune cells

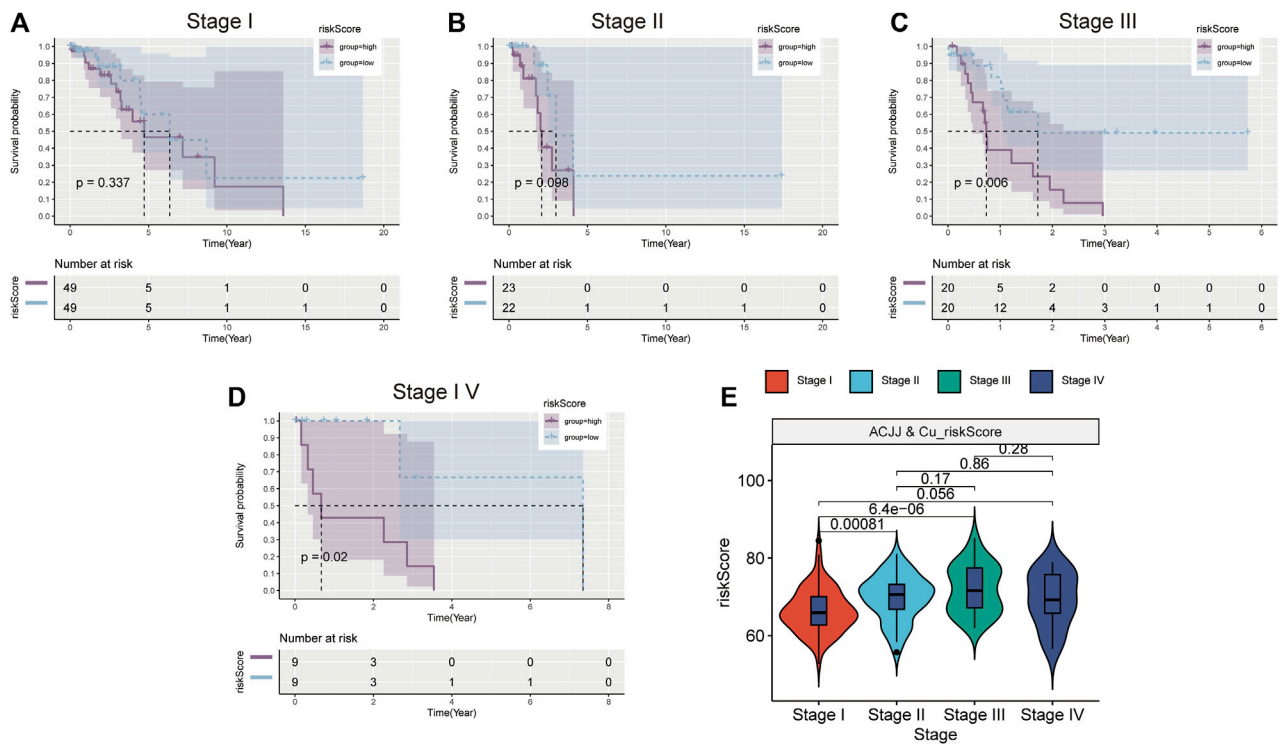


FIGURE 5
Stratified analysis based on clinical stage. K-M survival curve of high and low risk groups in Stage1 **(A)**, Stage II **(B)**, Stage III **(C)**, Stage IV **(D)**. **(E)** Comparison of risk scores in different stages.

with significant differences are specially labeled, and multiple types of immune cells are decreased. The analysis results of each algorithm are shown separately in Figure 7. CIBERSORT algorithm in Figure 7A shows that there are NK cells resting, T cells CD4 memory resting, T cells CD4 memory activated, Monocytes and Macrophages M0 The infiltrate contents of Dendritic cells resting and Mast cells resting were different between high and low risk groups. MCPcells showed that NK cella, Myeloid dendritic cells, and Endothelial cells were differentially expressed in high and low risk groups. quanTIseq analysis showed that the infiltrate content of acrophages.M2, Tregs and Dendritic.cell were significantly different among the high and low risk groups. CD4 T cell, Macrophage had significant difference between high and low risk groups. xcell analysis suggested significant differences in the content of a large number of immune cells between high and low risk groups. It should be noted that the content of macrophages was significantly different among the four algorithm analyses. These results suggest that copper proliferating genes may influence the development of LUAD patients through these immune cells.

Inhibition of cell proliferation and migration by LAG3 knockdown

To investigate the specific mechanism of LAG3 in LUAD, we initially constructed A549 cells with stable knockdown of LAG3. The results of qPCR and Western blot experiments indicated that the expression of LAG3 in A549 cells was significantly lower after si-LAG3 treatment compared to the control group of A549 cells (Figures 8A–C). The effect of LAG3 on cell proliferation was assessed using the CCK8 assay. The results revealed that knocking down LAG3 significantly inhibited the proliferation of A549 cells (Figure 8D). Results from the Transwell cell migration assay indicated that suppressing LAG3 expression significantly inhibited the migration of A549 cells (Figures 8E,F). The scratch wound healing assay also corroborated these results (Figures 8G,H). Furthermore, we evaluated cell proliferation activity by observing cell staining under a fluorescence microscope after EDU/DAPI staining. Our experimental results showed that the intensity of red fluorescence was significantly reduced in A549 cells after si-LAG3 treatment (Figures 8I,J), suggesting that downregulating LAG3 expression in LUAD A549 cells will inhibit tumor cell proliferation.

Discussion

Different trace elements serve distinct biological functions *in vivo*, contributing to the onset and progression of cancer to varying degrees (Cheng et al., 2019). Copper, a crucial cofactor of tyrosine kinase and ceruloplasmin, plays an indispensable role in organism growth and development. Research indicates that copper is vital for maintaining the body's normal biological functions due to its redox properties, which enable catalysis, oxidation, cell respiration, and various other life-sustaining activities (Li, 2020). Additionally, studies have highlighted that copper ions can modulate the

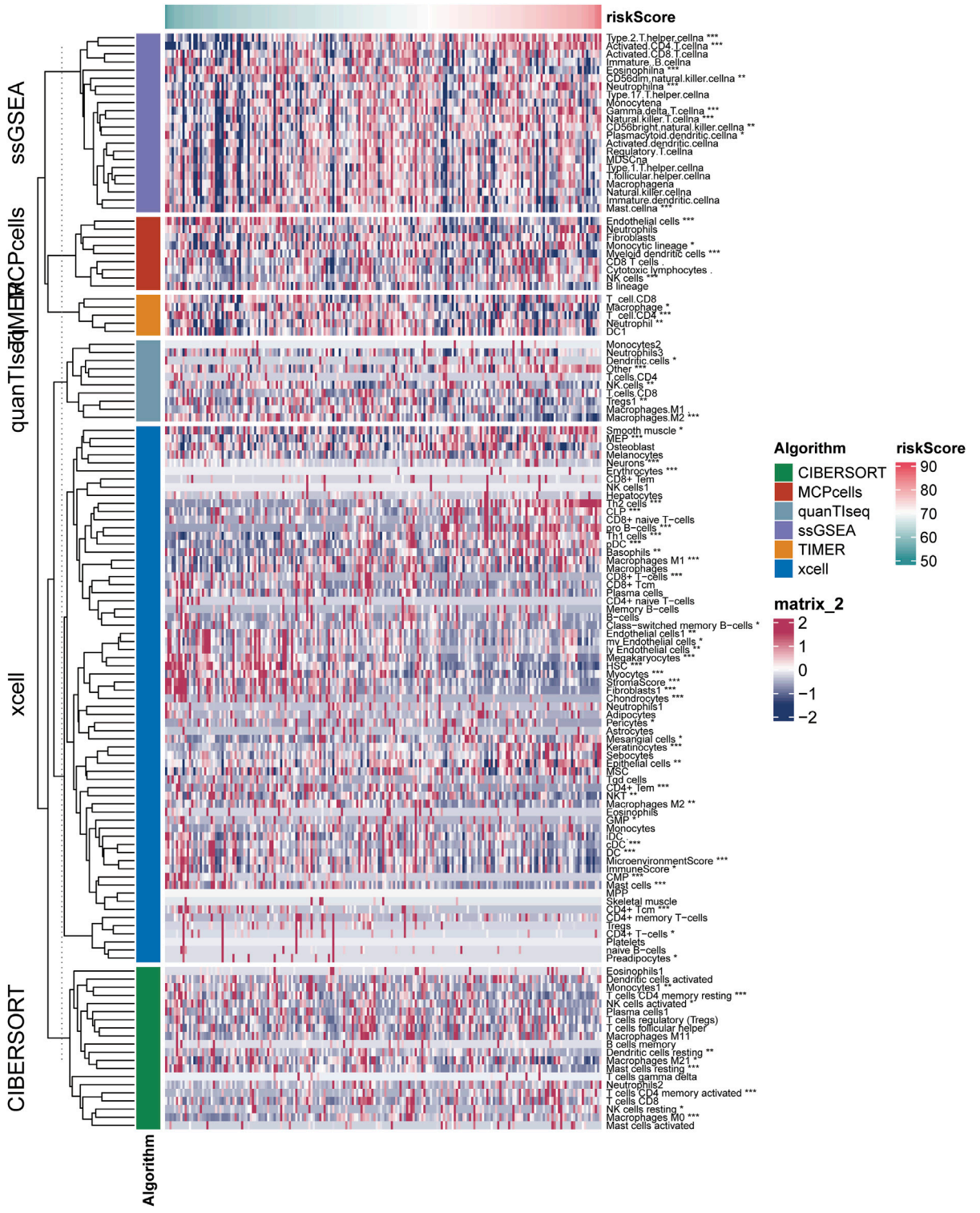


FIGURE 6
Results of six immune cell analysis algorithms in high—and low-risk groups.

oxidative phosphorylation and growth processes of tumors, participate in the carcinogenic signal transduction pathway via BRAF transduction and tumorigenesis, and exhibit potential toxicity to organisms (Ishida et al., 2013; O'Leary et al., 2019; Antsotegi-Uskola et al., 2020). Hence, it is imperative to maintain the copper ion content of an organism in a relatively stable state to prevent disrupting the homeostasis of the internal environment, inducing stress responses, and causing unnecessary harm. Recently, research on cuproptosis has garnered attention. The specific mechanism involves copper ions binding to fatty acylated

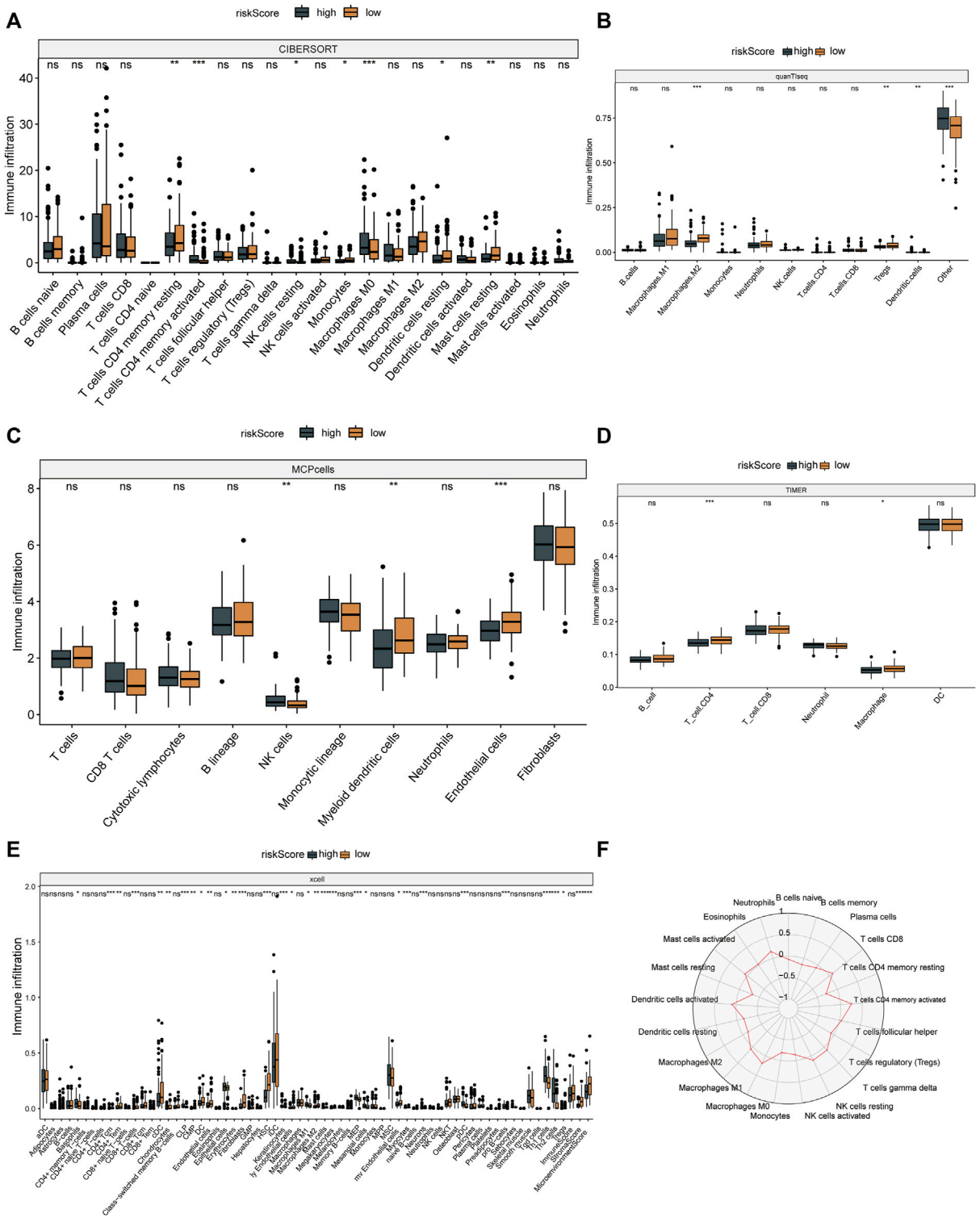


FIGURE 7
Immune cell analysis results. CIBERSORT **(A)**, quanTIseq **(B)**, mcp cells **(C)**, TIMER **(D)** and xcell **(E)** algorithms were used to analyze the immune cells in high and low risk groups. **(F)** Radar map of immune cells in relation to risk scores in CIBERSORT.

proteins in the mitochondrial respiration process of the TCA cycle, leading to protein aggregation. This, in turn, promotes the downregulation of Fe-S cluster proteins, inducing protein-toxic stress and ultimately resulting in cell death (Antsotegi-Uskola et al., 2020). The study of cuproptosis has piqued the interest of researchers, and some studies based on extensive transcriptome data have shed light on the mechanism of copper-induced cell death in lung adenocarcinoma (LUAD) (Wang X. et al., 2023; Wang T. et al.,

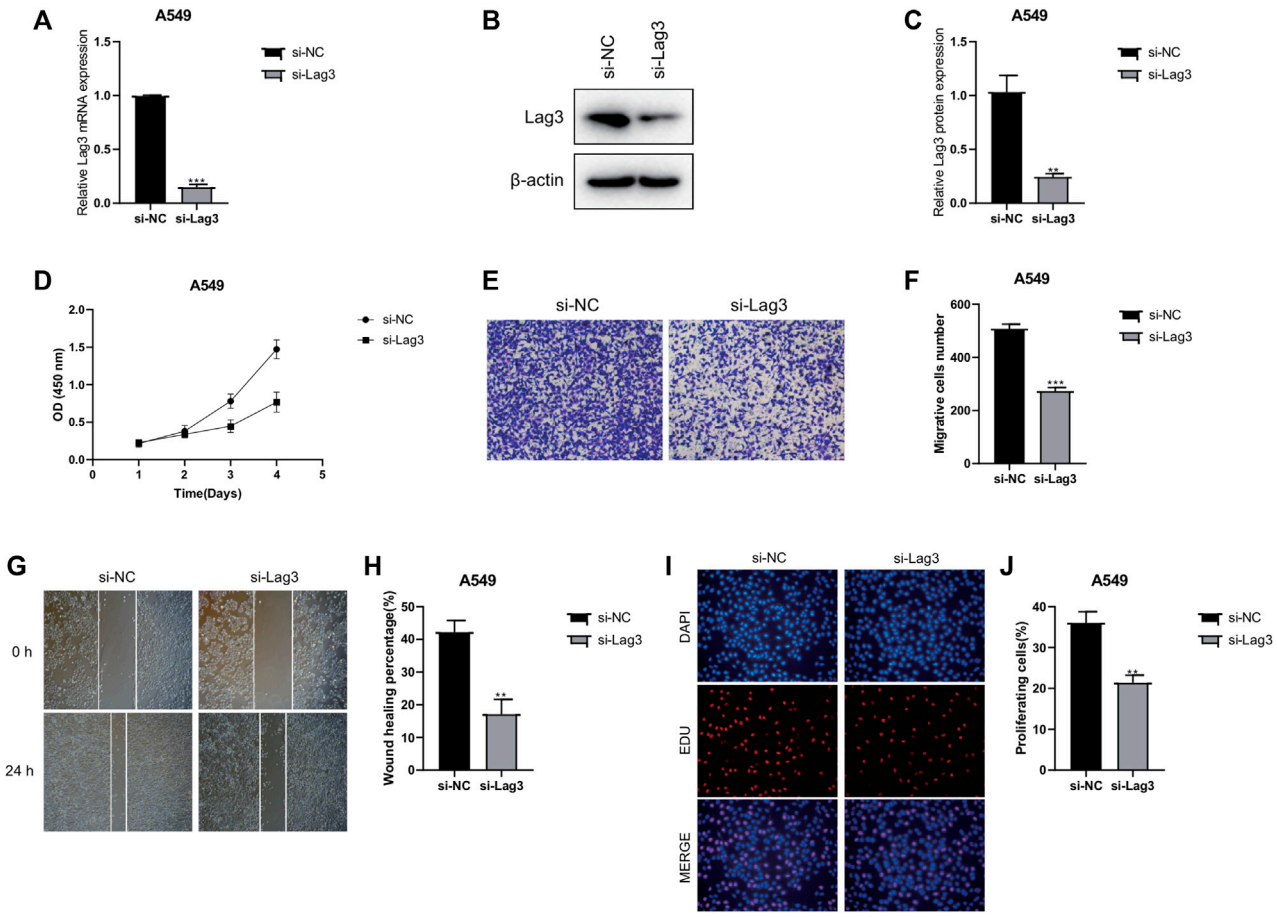


FIGURE 8
Effects of LAG3 Knockdown on A549 Cells. **(A–C)** Expression of LAG3 in A549 cells after si-LAG3 treatment assessed by qPCR and Western blot. **(D)** CCK8 assay showing the inhibition of A549 cell proliferation after LAG3 knockdown. **(E, F)** Transwell migration assay demonstrating the inhibitory effect of LAG3 suppression on A549 cell migration. **(G, H)** Scratch wound healing assay corroborating the inhibitory effect of LAG3 suppression on A549 cell migration. **(I, J)** Fluorescence microscopy images of EDU/DAPI stained A549 cells showing reduced red fluorescence intensity, indicative of inhibited cell proliferation, after si-LAG3 treatment. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

2023; Ma et al., 2023). The recent evolution of perspectives on Cuproplasia may pave the way for novel avenues of research into the role of copper in tumorigenesis.

In this study, we initially identified 40 Cuproplasia-related genes and scored 313 LUAD patients based on their transcriptome using the GSVA scoring method. Through WGCNA analysis, we screened 4,787 genes in the BROWN and TURQUOISE modules, which are closely related to the Cuproplasia score. Subsequently, 214 prognostic-related genes were identified through univariate Cox analysis, and a prognostic model was established based on 24 genes through lasso regression analysis. We found that risk scores could predict overall survival (OS) in both TCGA-LUAD and GSE31210 cohorts. These results suggest that genes related to copper proliferation play a significant role in the prognosis of LUAD patients. Among the genes included in the prognostic model, PLEK2 has been reported in five studies as a prognostic marker for lung adenocarcinoma (Cheng et al., 2019; Li, 2020; Ishida et al., 2013; O'Leary et al., 2019; Antsotegi-Uskola et al., 2020). Cell invasion, cell cycle, DNA damage, and DNA repair are positively correlated with PLEK2 expression in LUAD cells (Jiang et al., 2020; Zhang et al., 2020; Wu et al., 2021; Zhou et al., 2022). Promoter hypomethylation may underlie its upregulation (Zhang et al., 2020). *In vitro*, FABP5 regulates fat metabolism by diverting fat into complex lipid synthesis instead of catabolism. FABP5 is also essential for cell cycle progression, migration, and tumor growth *in vivo* (Garcia et al., 2022). EC61γ has been shown to promote LUAD proliferation, metastasis, and invasion through the EGFR signaling pathway, suggesting this gene as a potential therapeutic target. Previous literature has indicated that the BEX family has diagnostic and prognostic value in LUAD, and that BEX4 is associated with clinicopathologic features, especially in higher-grade LUAD. LINC02535 has been shown to promote LUAD development through the NF-κB signaling pathway, and further pan-carcinoma analysis has demonstrated extensive prognostic value for LINC02535 in pan-carcinoma. While the mechanism of action of these genes in LUAD has been partially explored, their correlation with copper proliferation has not been investigated.

In the subsequent functional enrichment analysis, GO was used to analyze the differential genes of patients in the high-low risk group, which were mainly concentrated in chromosome-related signaling pathways, and studies have shown that copper ions can promote the formation of reactive oxygen species, which can

damage DNA and chromatin (Garcia et al., 2022).The most significant pathways identified by KEGG analysis included cell cycle and DNA replication. The most significant pathways in HALLMARK enrichment analysis include G2M chekpoint. Studies have shown that copper oxide nanoparticles can reduce the activity of mouse embryonic fibroblasts and stop the cell cycle at G2M (Luo et al., 2014). The binding of copper and DNA bases has a concentration dependent tolerance relationship, which can reshape the integrity of DNA by affecting the structure of B-DNA, and affect the process of DNA replication and transcription (Govindaraju et al., 2013). These results further reveal the role of copper proliferation-related genes in LUAD and provide a new direction to explore the role of copper proliferation-related genes in LUAD in the future.

According to the immunocheckpoint analysis between high and low risk groups, the expression of PDCD1 CD274 LAG3 TNFRSF14 NRP1 was differentially expressed. Studies have shown that the variation of PDCD1 and CD274 genes regulates the risk and prognosis of LUAD and LUSC. The IMVIGOR210 analysis found that the lower the risk score, the better the treatment effect.

Subsequent stage-based stratified analysis found that patients in the high-risk group had worse prognosis, and the difference between the high-low risk groups was more significant in stage 3 and stage 4 patients. These results increased the application value of risk scores in LUAD. Subsequently, five algorithms were used to analyze the degree of immune cell infiltration between TCGA-LUAD high and low risk groups, and it was found that there were statistical differences in macrophages among the four algorithms, and the proportion of some types of macrophages was higher in high risk groups. Studies have shown that targeting macrophages in drug-resistant advanced LUAD is a therapeutic approach (Yin et al., 2023), and recent single-cell sequencing data further revealed that macrophages may play an important role in LUAD brain metastasis patients (Sun et al., 2022). These results suggest the research direction of copper proliferation-related genes in LUAD.

There are still some limitations in our study. First of all, we only described the genes related to copper proliferation through two publicly available data sets, which were not verified *in vitro* and *in vivo* experiments. Secondly, the accuracy of the prognostic model lacks the validation of multi-center data, which may limit the scope of the prognostic model. Subsequently, the predictive effects of immune checkpoints and immunotherapy still need to be further demonstrated in clinical trials. Finally, the effects of copper proliferation-related genes on chromatin and other signaling pathways still need further molecular biological experiments to prove.

Conclusion

In conclusion, we first conducted a comprehensive analysis of copper proliferating genes in the LUAD landscape and found that prognostic models constructed with copper proliferation-related genes could predict OS, immune checkpoint, and immunotherapy in LUAD. Further pathway enrichment analysis revealed that copper proliferation-related genes may affect chromatin structure and DNA replication and translation, thus influencing cell cycle. Meanwhile, immune cell infiltration analysis revealed that macrophages may be the key immune cells in this process.

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.

Author contributions

JK and ZZ performed the bioinformatic analysis and wrote the manuscript. YC, XW, DW, and JK conceive and design the experiments. WM and SZ revised the paper and contributed to the discussion, and All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Provincial subsidy funds for high-level hospital construction (301505).

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/fphar.2023.1240736/full#supplementary-material

References

Aishajiang, R., Liu, Z., Wang, T., Zhou, L., and Yu, D. (2023). Recent advances in cancer therapeutic copper-based nanomaterials for antitumor therapy. *Molecules* 28 (5), 2303. doi:10.3390/molecules28052303

Antsotegi-Uskola, M., Markina-Iñarrairaegui, A., and Ugalde, U. (2020). New insights into copper homeostasis in filamentous fungi. *Int. Microbiol.* 23 (1), 65–73. doi:10.1007/s10123-019-00081-5

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

Brady, D. C., Crowe, M. S., Turski, M. L., Hobbs, G. A., Yao, X., Chaikuad, A., et al. (2014). Copper is required for oncogenic BRAF signalling and tumorigenesis. *Nature* 509 (7501), 492–496. doi:10.1038/nature13180

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68 (6), 394–424. doi:10.3322/caac.21492

Cheng, L., Xiong, W., Li, S., Wang, G., Zhou, J., and Li, H. (2023). CRISPR-Cas9 screening identified lethal genes enriched in necroptosis pathway and of prognosis significance in osteosarcoma. *J. Gene Med.* 8, e3563. doi:10.1002/jgm.3563

Cheng, X., Zhou, Y. C., Zhou, B., Huang, Y. C., Wang, G. Z., and Zhou, G. B. (2019). Systematic analysis of concentrations of 52 elements in tumor and counterpart normal tissues of patients with non-small cell lung cancer. *Cancer Med.* 8 (18), 7720–7727. doi:10.1002/cam4.2629

Costa, M. I., Sarmento-Ribeiro, A. B., and Gonçalves, A. C. (2023). Zinc: From biological functions to therapeutic potential. *Int. J. Mol. Sci.* 24 (5), 4822. doi:10.3390/ijms24054822

Dankner, M., Rose, A. A. N., Rajkumar, S., Siegel, P. M., and Watson, I. R. (2018). Classifying BRAF alterations in cancer: new rational therapeutic strategies for actionable mutations. *Oncogene* 37 (24), 3183–3199. doi:10.1038/s41388-018-0171-x

Denoyer, D., Masaldan, S., La Fontaine, S., and Cater, M. A. (2015). Targeting copper in cancer therapy: 'Copper that cancer. *Metallomics* 7 (11), 1459–1476. doi:10.1039/c5mt00149h

Finotello, F., Mayer, C., Plattner, C., Laschober, G., Rieder, D., Hackl, H., et al. (2019). Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* 11 (1), 34. doi:10.1186/s13073-019-0638-6

Frydrych, A., Krośniak, M., and Jurowski, K. (2023). The role of chosen essential elements (Zn, Cu, Se, Fe, Mn) in food for special medical purposes (FSMPs) dedicated to oncology patients-critical review: State-of-the-Art. *Nutrients* 15 (4), 1012. doi:10.3390/nu15041012

Garcia, K. A., Costa, M. L., Lacunza, E., Martinez, M. E., Corsico, B., and Scaglia, N. (2022). Fatty acid binding protein 5 regulates lipogenesis and tumor growth in lung adenocarcinoma. *Life Sci.* 301, 120621. doi:10.1016/j.lfs.2022.120621

Ge, E. J., Bush, A. I., Casini, A., Cobine, P. A., Cross, J. R., DeNicola, G. M., et al. (2022). Connecting copper and cancer: From transition metal signalling to metalloplasia. *Nat. Rev. Cancer* 22 (2), 102–113. doi:10.1038/s41568-021-00417-2

Govindaraju, M., Shekar, H. S., Sateesha, S. B., Vasudeva Raju, P., Sambasiva Rao, K. R., Rao, K. S. J., et al. (2013). Copper interactions with DNA of chromatin and its role in neurodegenerative disorders. *J. Pharm. Anal.* 3 (5), 354–359. doi:10.1016/j.jpha.2013.03.003

Guo, J., Cheng, J., Zheng, N., Zhang, X., Dai, X., Zhang, L., et al. (2021). Copper promotes tumorigenesis by activating the PDK1-AKT oncogenic pathway in a copper transporter 1 dependent manner. *Adv. Sci. (Weinh).* 8 (18), e2004303. doi:10.1002/advs.202004303

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

Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., et al. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32, D258–D261. doi:10.1093/nar/gkh036

Hong, D., Xu, K., Zhang, L., Wan, X., and Guo, Y. (2020). Radiomics signature as a predictive factor for EGFR mutations in advanced lung adenocarcinoma. *Front. Oncol.* 10, 28. doi:10.3389/fonc.2020.00028

Ishida, S., Andreux, P., Poitry-Yamate, C., Auwerx, J., and Hanahan, D. (2013). Bioavailable copper modulates oxidative phosphorylation and growth of tumors. *Proc. Natl. Acad. Sci. U. S. A.* 110 (48), 19507–19512. doi:10.1073/pnas.1318431110

Jiang, H., Xu, S., and Chen, C. (2020). A ten-gene signature-based risk assessment model predicts the prognosis of lung adenocarcinoma. *BMC cancer* 20 (1), 782. doi:10.1186/s12885-020-07235-z

Jopowicz, A., and Tarnacka, B. (2023). Neurological wilson's disease signs-hepatic encephalopathy or copper toxicosis? *Diagn. (Basel)* 13 (5), 893. doi:10.3390/diagnostics13050893

Kanehisa, M., and Goto, S. (2000). Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi:10.1093/nar/28.1.27

Langfelder, P., and Horvath, S. (2008). Wgcna: an R package for weighted correlation network analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559

Li, H., Yu, L., Zhang, X., Shang, J., and Duan, X. (2022). Exploring the molecular mechanisms and shared gene signatures between rheumatoid arthritis and diffuse large B cell lymphoma. *Front. Immunol.* 13, 1036239. doi:10.3389/fimmu.2022.1036239

Li, H., Zhang, X., Shang, J., Feng, X., Yu, L., Fan, J., et al. (2023). Identification of NETs-related biomarkers and molecular clusters in systemic lupus erythematosus. *Front. Immunol.* 14, 1150828. doi:10.3389/fimmu.2023.1150828

Li, Y. (2020). Copper homeostasis: Emerging target for cancer treatment. *IUBMB Life* 72 (9), 1900–1908. doi:10.1002/iub.2341

Lunn, M., and McNeil, D. (1995). Applying Cox regression to competing risks. *Biometrics* 51 (2), 524–532. doi:10.2307/2532940

Luo, C., Li, Y., Yang, L., Zheng, Y., Long, J., Jia, J., et al. (2014). Activation of Erk and p53 regulates copper oxide nanoparticle-induced cytotoxicity in keratinocytes and fibroblasts. *Int. J. Nanomedicine* 9, 4763–4772. doi:10.2147/IJN.S67688

Ma, C., Li, F., He, Z., Zhao, S., Yang, Y., and Gu, Z. (2023). Prognosis and personalized treatment prediction in lung adenocarcinoma: An *in silico* and *in vitro* strategy adopting cuproptosis related lncRNA towards precision oncology. *Front. Pharmacol.* 14, 1113808. doi:10.3389/fphar.2023.1113808

O'Leary, C. G., Andelkovic, V., Ladwa, R., Pavlakis, N., Zhou, C., Hirsch, F., et al. (2019). Targeting BRAF mutations in non-small cell lung cancer. *Transl. Lung Cancer Res.* 8 (6), 1119–1124. doi:10.21037/tlcr.2019.10.22

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Saris, C. G., Horvath, S., van Vught, P. W., van Es, M. A., Blauw, H. M., Fuller, T. F., et al. (2009). Weighted gene co-expression network analysis of the peripheral blood from Amyotrophic Lateral Sclerosis patients. *BMC Genomics* 10, 405. doi:10.1186/1471-2164-10-405

Steen, C. B., Liu, C. L., Alizadeh, A. A., and Newman, A. M. (2020). Profiling cell type abundance and expression in bulk tissues with CIBERSORTx. *Methods Mol. Biol. Clift. NJ)* 2117, 135–157. doi:10.1007/978-1-0716-0301-7_7

Sun, H. F., Li, L. D., Lao, I. W., Li, X., Xu, B. J., Cao, Y. Q., et al. (2022). Single-cell RNA sequencing reveals cellular and molecular reprograming landscape of gliomas and lung cancer brain metastases. *Clin. Transl. Med.* 12 (11), e1101. doi:10.1002/ctm2.1101

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statistics Med.* 16 (4), 385–395. doi:10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

Tsang, T., Posimo, J. M., Gudiel, A. A., Cicchini, M., Feldser, D. M., and Brady, D. C. (2020). Copper is an essential regulator of the autophagic kinases ULK1/2 to drive lung adenocarcinoma. *Nat. Cell Biol.* 22 (4), 412–424. doi:10.1038/s41556-020-0481-4

Tu, W., Tu, Y., Tan, C., Zhong, H., Xu, S., Wang, J., et al. (2023). Elucidating the role of T-cell exhaustion-related genes in colorectal cancer: a single-cell bioinformatics perspective. *Funct. Integr. Genomics* 23 (3), 259. doi:10.1007/s10142-023-01188-9

Turski, M. L., Brady, D. C., Kim, H. J., Kim, B. E., Nose, Y., Counter, C. M., et al. (2012). A novel role for copper in Ras/mitogen-activated protein kinase signaling. *Mol. Cell Biol.* 32 (7), 1284–1295. doi:10.1128/MCB.05722-11

van Renterghem, A. W. J., van de Haar, J., and Voest, E. E. (2023). Functional precision oncology using patient-derived assays: bridging genotype and phenotype. *Nat. Rev. Clin. Oncol.* 20, 305–317. doi:10.1038/s41571-023-00745-2

Wang, T., Jiang, X., Lu, Y., Ruan, Y., and Wang, J. (2023b). Identification and integration analysis of a novel prognostic signature associated with cuproptosis-related ferroptosis genes and relevant lncRNA regulatory axis in lung adenocarcinoma. *Aging* 15, 1543–1563. doi:10.18632/aging.204561

Wang, X., Jing, H., and Li, H. (2023a). A novel cuproptosis-related lncRNA signature to predict prognosis and immune landscape of lung adenocarcinoma. *Transl. Lung Cancer Res.* 12 (2), 230–246. doi:10.21037/tlcr-22-500

Wu, Z. X., Huang, X., Cai, M. J., Huang, P. D., and Guan, Z. (2021). Development and validation of a prognostic index based on genes participating in autophagy in patients with lung adenocarcinoma. *Front. Oncol.* 11, 799759. doi:10.3389/fonc.2021.799759

Xie, J., Yang, Y., Gao, Y., and He, J. (2023). Cuproptosis: mechanisms and links with cancers. *Mol. Cancer* 22 (1), 46. doi:10.1186/s12943-023-01732-y

Yang, C., Zhu, B., Zhan, M., and Hua, Z. C. (2023). Lithium in cancer therapy: Friend or foe? *Cancers (Basel)* 15 (4), 1095. doi:10.3390/cancers15041095

Yi, M., Nissley, D. V., McCormick, F., and Stephens, R. M. (2020). ssGSEA score-based Ras dependency indexes derived from gene expression data reveal potential Ras addiction mechanisms with possible clinical implications. *Sci. Rep.* 10 (1), 10258. doi:10.1038/s41598-020-66986-8

Yin, J. Z., Shi, X. Q., Wang, M. D., Du, H., Zhao, X. W., Li, B., et al. (2023). Arsenic trioxide elicits anti-tumor activity by inhibiting polarization of M2-like tumor-associated macrophages via Notch signaling pathway in lung adenocarcinoma. *Int. Immunopharmacol.* 117, 109899. doi:10.1016/j.intimp.2023.109899

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4, Article17. doi:10.2202/1544-6115.1128

Zhang, W., Li, T., Hu, B., and Li, H. (2020). PLEK2 gene upregulation might independently predict shorter progression-free survival in lung adenocarcinoma. *Technol. Cancer Res. Treat.* 19, 1533033820957030. doi:10.1177/1533033820957030

Zhou, B., Zang, R., Zhang, M., Song, P., Liu, L., Bie, F., et al. (2022). Identifying novel tumor-related antigens and immune phenotypes for developing mRNA vaccines in lung adenocarcinoma. *Int. Immunopharmacol.* 109, 108816. doi:10.1016/j.intimp.2022.108816

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 25 September 2023
DOI 10.3389/fphar.2023.1280428

OPEN ACCESS

EDITED BY
Hongzhou Cai,
Nanjing Medical University, China

REVIEWED BY
Wanzun Lin,
Fudan University, China
Fahui Liu,
Xiamen University, China
Xu-Sheng Liu,
Hubei University of Medicine, China

*CORRESPONDENCE
Zicheng Xiao,
xzcheng1998@163.com

RECEIVED 20 August 2023
ACCEPTED 11 September 2023
PUBLISHED 25 September 2023

CITATION
Liu J, He Y, Zhou W, Tang Z and Xiao Z (2023), A glycosylation risk score comprehensively assists the treatment of bladder neoplasm in the real-world cohort, including the tumor microenvironment, molecular and clinical prognosis.
Front. Pharmacol. 14:1280428.
doi: 10.3389/fphar.2023.1280428

A glycosylation risk score comprehensively assists the treatment of bladder neoplasm in the real-world cohort, including the tumor microenvironment, molecular and clinical prognosis

Jinhui Liu[1,2], Yunbo He[1,2], Weimin Zhou[1,2], Zhuoming Tang[1,2] and Zicheng Xiao[1,2]*

[1]Department of Urology, Xiangya Hospital, Central South University, Changsha, China, [2]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

Background: Bladder cancer is a common urological cancer associated high significant morbidity and mortality rates. Immunotherapy has emerged as a promising treatment option, although response rates vary among patients. Glycosylation has been implicated in tumorigenesis and immune regulation. However, our current comprehensive understanding of the role of glycosylation in bladder cancer and its clinical implications is limited.

Methods: We constructed a training cohort based on the downloaded TCGA-BLCA dataset, while additional datasets (Xiangya cohort, GSE32894, GSE48075, GSE31684, GSE69795 and E-MTAB-1803) from Xiangya hospital, GEO and ArrayExpress database were obtained and used as validation cohorts. To identify glycosylation-related genes associated with prognosis, univariate Cox regression and LASSO regression were performed. A Cox proportional hazards regression model was then constructed to develop a risk score model. The performance of the risk score was assessed in the training cohort using Kaplan-Meier survival curves and ROC curves, and further validated in multiple validation cohorts.

Results: We classified patients in the training cohort into two groups based on glycosylation-related gene expression patterns: Cluster 1 and Cluster 2. Prognostic analysis revealed that Cluster 2 had poorer survival outcomes. Cluster 2 also showed higher levels of immune cell presence in the tumor microenvironment and increased activation in key steps of the cancer immune response cycle. We developed an independent prognostic risk score ($p < 0.001$) and used it to construct an accurate prognostic prediction nomogram. The high glycosylation risk score group exhibited higher tumor immune cell infiltration, enrichment scores in immune therapy-related pathways, and a tendency towards a basal subtype. Conversely, the low-risk score group had minimal immune cell infiltration and tended to have a luminal subtype. These findings were consistent in our real-world Xiangya cohort.

Conclusion: This multi-omics glycosylation score based on these genes reliably confirmed the heterogeneity of bladder cancer tumors, predicted the efficacy of

immunotherapy and molecular subtypes, optimizing individual treatment decisions.

KEYWORDS

glycosylation, multi-omics, tumor heterogeneity, immunotherapeutic efficacy, molecular subtype, bladder carcinoma

Introduction

Bladder cancer (BLCA) is a prevalent malignancy worldwide, characterized by high morbidity and mortality rates (Siegel et al., 2020). Non-muscle invasive bladder cancer (NMIBC) accounts for three-quarters of initial diagnoses, while the remaining cases are categorized as either muscle-invasive bladder cancer (MIBC) or BLCA with distant metastasis (Powles et al., 2022). Despite advancements in treatment options, such as surgery and chemotherapy, the prognosis for patients with advanced stage BLCA remains suboptimal (Antoni et al., 2017; Lenis et al., 2020). Recent studies have shown that monoclonal antibodies targeting PD-1 and its ligands have emerged as a therapeutic strategy with encouraging clinical benefits for metastatic BLCA (Rosenberg et al., 2016; Plimack et al., 2017; Rijnders et al., 2017; Hu et al., 2022). However, only a subset of patients can achieve beneficial (Jenkins et al., 2018; Schoenfeld and Hellmann, 2020). Although increased expression of PD-L1 on tumor cells and/or immune cells is currently used as a diagnostic method for immune therapies targeting PD-1, it only partially correlates with the clinical benefits of these drugs (Ma et al., 2016). Therefore, there is an urgent requirement to discover novel biomarkers that can assist in treatment decision-making and improve patient outcomes.

The tumor immune microenvironment (TIME) is an intricate milieu consisting of immune cells and immune-related molecules (Binnewies et al., 2018), and its importance in immunotherapy has gained widespread recognition. Chen DS et al. categorized TIME into three subtypes: “immune inflamed,” “immune excluded,” and “immune desert,” based on their distinct characteristics and potential responsiveness to immunotherapy (Chen and Mellman, 2017). Similarly, Duan Q et al. classified tumors as either “hot” or “cold” tumors depending on the level of immune infiltration (Duan et al., 2020). Cold and immune desert tumors are characterized by limited immune cell infiltration, resulting in a poor response to immunotherapy (Galon and Bruni, 2019). Emerging therapeutic strategies are focused on enhancing immune infiltration to transform the tumor microenvironment (TME), from a poorly infiltrated “cold” phenotype to an immune-rich “hot” phenotype (Bonaventura et al., 2019; Vonderheide, 2020). Therefore, analyzing the TIME is crucial for enhancing the efficacy of immunotherapy.

Glycosylation, a common post-translational protein modification process, occurs in all domains of life (Pinho and Reis, 2015). It involves the attachment of monosaccharides or polysaccharides (i.e., oligosaccharides or complex glycans) to specific residues of target proteins (Eichler, 2019). This modification has been reported to impact various biological processes, including protein secretion, degradation, transport to receptor interactions, and modulation of immune responses (Moremen et al., 2012; Varki, 2017). Glycosylation is associated with the pathogenesis of numerous prevalent diseases, including cancer (Eichler, 2019). Glycosylation modification affects tumorigenesis through its influence on growth, differentiation, metastasis, and immune surveillance. Altered glycosylation profiles have been detected in various types of cancer, including BLCA (Przybylo et al., 2002; RodrÍguez et al., 2018). For instance, the invasive capacity of BLCA cells has been linked to the N-glycosylation of cadherin (Przybylo et al., 2002). Furthermore, glycosylation has been implicated in the regulation of immune responses within the TIME (Badmann et al., 2020; Sun et al., 2021). However, our understanding of the glycosylation landscape in BLCA and its clinical implications is still limited. Given the critical role of glycosylation in both tumorigenesis and immune regulation, it is plausible that glycosylation patterns could serve as potential biomarkers for predicting the response to immunotherapy in BLCA patients.

In this study, our research objective is to develop a new glycosylation risk score based on a multi-omics study to evaluate the prognosis comprehensively and individually, immunophenotype, and tumor heterogeneity of BLCA patients. In addition, we aim to study the glycosylation risk score to provide valuable insights into the potential of BLCA patients to make treatment decisions such as personalized immunotherapy and improve their prognosis.

Materials and methods

Data collection

Training set

We established a dataset consisting of 408 patients with BLCA by selecting individuals from the Cancer Genome Atlas (TCGA) database. The mRNA expression matrix and clinical information corresponding to these patients were downloaded from the Genomic Data Commons (GDC, https://portal.gdc.cancer.gov/) (Colaprico et al., 2016). We converted the fragments per kilobase of exon model per million mapped fragments (FPKM) and count value in the original expression matrix to transcripts per kilobase of exon model per million mapped reads (TPM). Subsequently, we merged this data with clinical information to create a new dataset. After excluding 5 patients due to duplicated or missing follow-up data, a total of 403 patients formed the training cohort.

Validation cohorts

In our early-stage study (Li et al., 2021), we constructed a dataset called the Xiangya cohort and have uploaded it to the Gene Expression Omnibus (GEO) database. This dataset includes 56 patients with BLCA and encompasses complete survival information along with RNA-sequencing (RNA-seq) data (GSE188715). We also downloaded relevant data from the GEO

database (https://www.ncbi.nlm.nih.gov/geo/) to construct four additional external validation cohorts (GSE32894, GSE48075, GSE31684 and GSE69795). Duplicate patients or those with incomplete survival information were excluded during data preprocessing, resulting in a final inclusion of 224 (GSE32894), 73 (GSE48075), 93 (GSE31684) and 38 (GSE69795) individuals in the four cohorts, respectively. Download the dataset with accession number E-MTAB-1803 from the ArrayExpress database (https://www.ebi.ac.uk/arrayexpress/) as an additional external validation cohort.

Supplementary Table S1 displays the clinical information of patients in the training and six validation cohorts.

Consensus clustering

We obtained a list of 628 glycosylation-related genes from the gene set enrichment analysis (GSEA) (Supplementary Table S2). To analyze the expression pattern of these genes in training cohort, we utilized the consensus clustering function in the "ConsuClusterPlus" R package (Wilkerson and Hayes, 2010). The parameters were set as follows: distance = "manhattan", clusterAlg = "pam", maxK = 5, Reps = 1,500, pItem = 0.8, pFeature = 1. By applying this approach, we identified distinct glycosylation expression patterns.

Describing the TIME of BLCA

To characterize the TIME of BLCA, we utilized the tracking tumor immunophenotype (TIP) database (http://biocc.hrbmu.edu.cn/TIP/) (Xu et al., 2018) to obtained the activation levels of the 7-step Cancer Immunity Cycle (CIC) (Chen and Mellman, 2013). Furthermore, we compiled a summary of 22 immune checkpoint inhibitor (ICI) genes, 18 T cell-associated inflammatory signature (TIS) genes, and effector genes of various immune cells, including CD8 T cells, dendritic cells (DCs), macrophages, natural killer (NK) cells, and type 1 T helper (Th1) cells, based on our previous study (Hu et al., 2021a) (Supplementary Table S3).

Development of glycosylation risk score

To identify candidate genes associated with glycosylation patterns and clinical prognosis, we employed two methods: Univariate Cox analysis and the least absolute shrinkage and selection operator (LASSO) algorithm. The "glmnet" R package was utilized for the LASSO algorithm. Initially, Initially, we conducted univariate Cox analysis on a set of 628 genes and identified 30 genes that were strongly correlated with prognosis ($p < 0.005$). Subsequently, the LASSO algorithm was applied to further refine the prognostic genes. The "glmnet" R package facilitated this process. From the LASSO analysis, we identified 20 candidate genes. Finally, the glycosylation risk score was constructed using the Cox proportional hazard regression model with the "glmnet" R package, incorporating 20 genes.

$$Glycosylation\ Score = \sum \beta_i * RNA_i$$

Evaluation and verification of glycosylation risk score

In the training set, patients were divided into high-risk and low-risk groups based on their risk scores, using the median of the risk score as the threshold. Kaplan-Meier (K-M) survival curves were plotted and the log-rank test was performed using the "survminer" R package to assess the differences in survival between the two groups. The predictive accuracy of the risk score was evaluated using the time-dependent receiver operating characteristic (tROC) analysis, implemented with the "tROC" R package. Additionally, a nomogram was constructed incorporating clinical information related to prognosis and the glycosylation risk score. The predictive efficacy of the nomogram was verified using calibration curves.

For external validation, the same method used in the training set was applied to an independent cohort of BLCA patients. In brief, the risk scores were calculated using the glycosylation risk scoring formula, and patients were classified into high-risk and low-risk groups using the median risk score as the threshold. The survival outcomes between the two groups were compared using the K-M method and log-rank test. The predictive accuracy of the risk scores was evaluated using tROC analysis.

Identification of molecular subtypes of BLCA by glycosylation risk score

In our previous studies, our team conducted an extensive review and summary of the existing seven molecular typing criteria for BLCA, including the TCGA, UNC, and Consensus systems, et al. To achieve a unified classification approach, we utilized two R packages, namely, "BLCAsubtyping" and "ConsensusMIBC". Additionally, we incorporated BLCA-related pathways identified by Kamoun, A. et al. (Kamoun et al., 2020) (Supplementary Table S4). To enhance clinical applicability, we further reclassified the different molecular subtypes into "luminal" and "basal" subtypes, aiming to provide a more concise and efficient clinical guidance.

Statistical analysis

Correlations between variables were assessed using either Pearson or Spearman coefficients, depending on the nature of the data. Differences between binary groups in continuous variables were evaluated using the *t*-test or Mann-Whitney *U* test. To examine the survival prognosis, the K-M method was employed to generate survival curves, and statistical significance was determined using the log-rank test. The relationship between candidate genes and survival prognosis was determined through univariate Cox analysis, and the LASSO algorithm was used to select and refine the candidate genes for constructing the glycosylation risk score. The hazard ratio (HR) and independent prognostic values of the glycosylation risk score were calculated using univariate and multivariate Cox regression models. The glycosylation risk score was constructed using the Cox proportional hazard regression model, and its accuracy was assessed by drawing time-dependent receiver operating characteristic (ROC) curves and calculating the area under the curve (AUC). All statistical analyses were conducted using R software (version 4.22) with a significance level set at $p < 0.05$. The adjusted *p*-value was obtained using the false discovery rate (FDR) method, and all tests were two-sided.

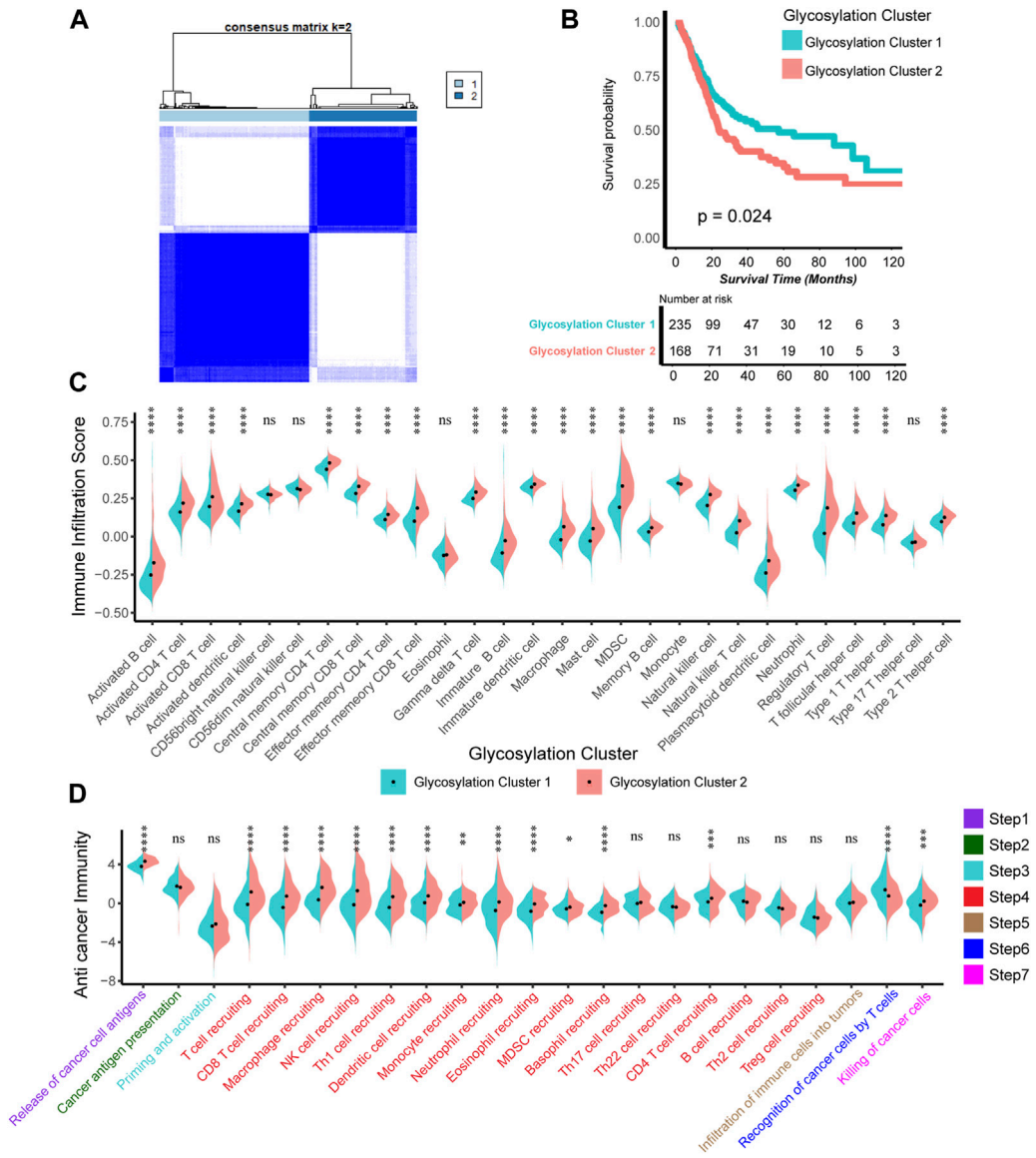


FIGURE 1
Construction of Glycosylation genes expression patterns related to prognosis and tumor immune microenvironment. **(A)** The unsupervised cluster analysis based on all the 628 Glycosylation-related genes; Light blue and dark blue lines represented Glycosylation cluster 1 and 2, separately. **(B)** Kaplan-Meier plot of OS between two Glycosylation-related patterns; Light green and red lines represented Glycosylation cluster 1 and 2, separately. **(C)** The different infiltration levels of 28 immune cells in the TME using ssGSEA algorithm between two Glycosylation-based patterns; Light green and red lines represent Glycosylation clusters 1 and 2, respectively. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$; ns, not statistically significant. **(D)** The different levels of anticancer immunity between two Glycosylation-based patterns; Light green and red lines represent Glycosylation cluster 1 and 2, respectively; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$; ns, not statistically significant.

Results

Construction of glycosylation genes expression patterns related to prognosis and tumor immune microenvironment

We constructed expression patterns based on glycosylation genes features using unsupervised clustering analysis in the TCGA-BLCA cohort, by "ConsenseClusterPlus" R package. And we found that dividing into two patterns was the most appropriate, named glycosylation cluster 1 and glycosylation cluster 2. (Figure 1A). Subsequently, we conducted a detailed analysis to investigate the disparities between the two mentioned glycosylation clusters. In terms of prognosis, compared to cluster 1, cluster 2 has a significantly poorer prognosis (p = 0.024, Figure 1B). As for the TIME, as depicted in Figure 1C (Supplementary Table S5), the infiltration level of most of immune cells including activated and immature B cell, activated and central memory CD4 T cell, activated and central memory CD8 T cell, natural killer cell and macrophage in TME was apparently higher in cluster 2 compared to cluster 1. In addition, in the 7-step CIC, cluster 2 exhibited a higher activation level in the

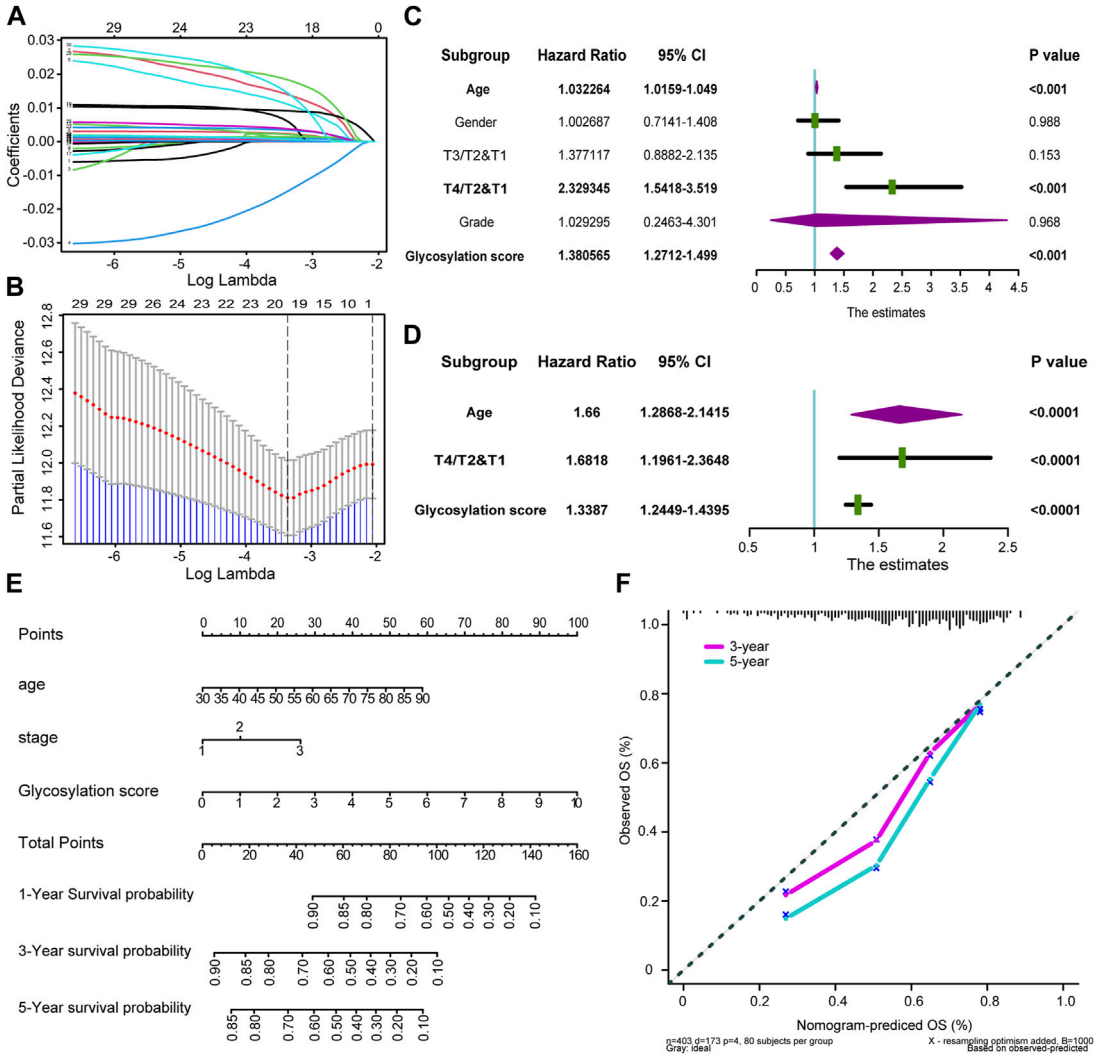


FIGURE 2
Developing Glycosylation-related risk scores and predicting clinical outcomes in multiple cohorts **(A)** Coefficients of Glycosylation-related prognosis genes value are shown by lambda parameter. **(B)** Partial likelihood deviance versus log (lambda) drawn by LASSO algorithm and 10-fold cross-validation. **(C,D)** Forest plots of univariate and multivariate Cox analysis of Glycosylation-based risk score combined with age, gender, tumor grade and stage of BLCA. **(E)** Nomogram developed by using age, tumor stage, and Glycosylation-based risk score. **(F)** Calibration curves of the nomogram.

main anti-tumor immune steps, including step 1 (release of cancer cell antigens), step 4 (recruitment of immune cells such as T cell, CD8 T, macrophage, NK cell, dendritic cell), step 6 (recognition of cancer cells by T cells) and step 7 (killing of cancer cells) (Figure 1D; Supplementary Table S6).

Developing glycosylation-related risk scores and predicting clinical outcomes in multiple cohorts

The completely different manifestations of these two glycosylation clusters mentioned above in the prognosis and TIME of BLCA aroused our interest. Therefore, we planned to develop a quantitative risk score utilizing the expression patterns of glycosylation genes. This risk score will be used to predict the clinical prognosis of each patient, aiming to achieve the accuracy treatment of BLCA.

Firstly, we selected 30 independent prognostic genes strongly associated with prognosis from glycosylation-related genes through univariate analysis ($p < 0.005$, Supplementary Table S7). Subsequently, LASSO regression helped us identify the 20 most suitable candidate genes in those 30 independent prognostic genes above for constructing glycosylation-related risk models (Figures 2A, B; Supplementary Table S8). And we choose the minimum lambda for the optimal cutoff value, and selected ten-fold cross validation method as correction. Finally, based on those 20 candidate glycosylation genes above, we employed the "glmnet" R software package to construct a Cox proportional risk regression model. This model allowed us to generate a risk score, known as the glycosylation-based risk score, in the TCGA-BLCA training cohort. The median glycosylation risk score would be used

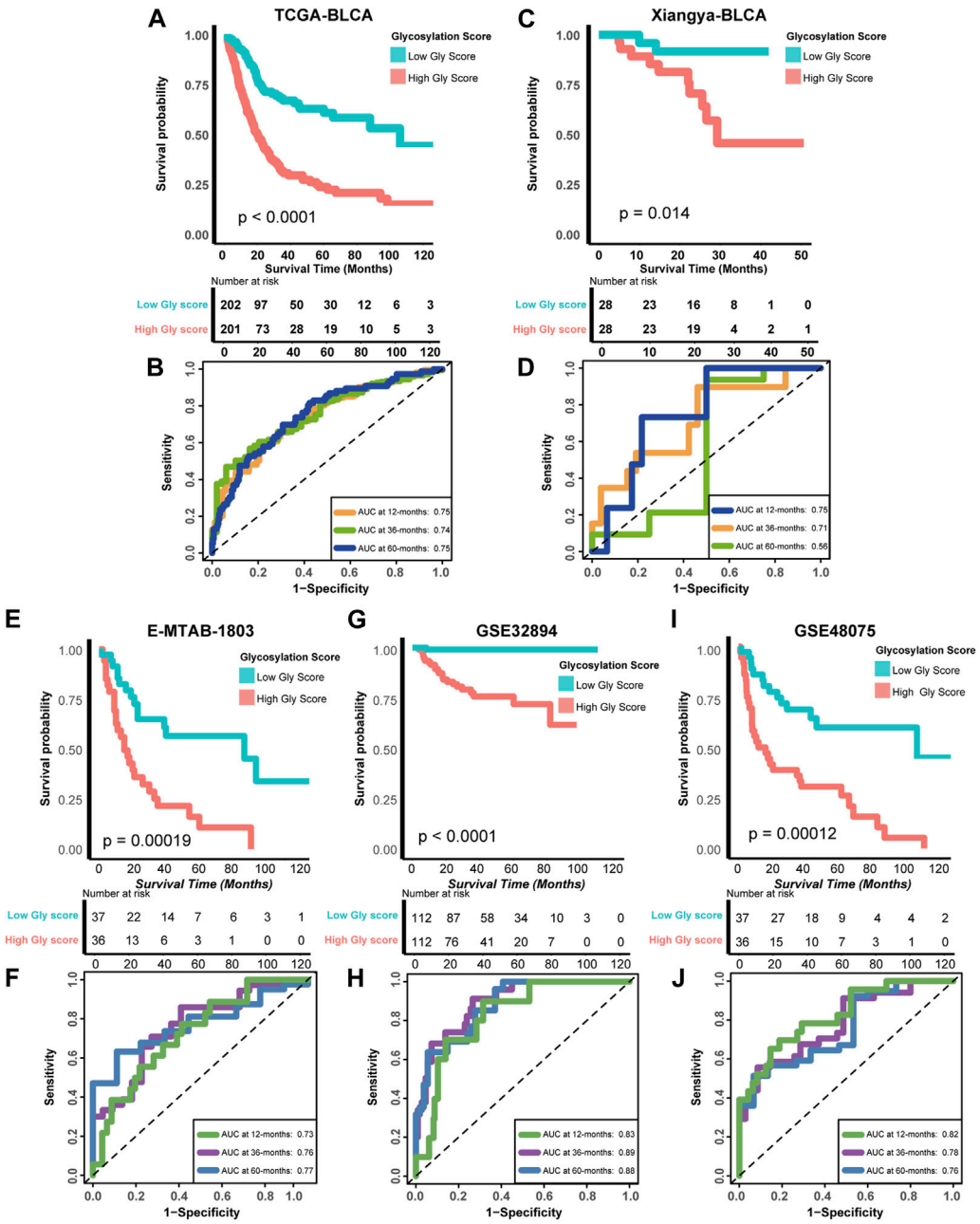


FIGURE 3
Verifying the accuracy of Glycosylation score in predicting prognosis in multiple cohorts **(A)** Kaplan-Meier (K–M) plot of OS between Glycosylation risk score groups in TCGA-BLCA cohort; Light red and green lines represented high and low Glycosylation risk score groups, separately. **(B)** The area under curves (AUCs) plot of Glycosylation risk score in TCGA-BLCA cohort. **(C,D)** K-M plot of OS between Glycosylation risk score groups and AUCs plot of the risk score the in Xiangya validation cohort, separately. **(E,F)** K-M plot of OS between Glycosylation risk score groups and AUCs plot of the risk score the in E-MTAB-1803 validation cohort, separately. **(G,H)** K-M plot of OS between Glycosylation risk score groups and AUCs plot of the risk score the in GSE32894 validation cohort, separately. **(I,J)** K-M plot of OS between Glycosylation risk score groups and AUCs plot of the risk score the in GSE48075 validation cohort, separately.

as a standard to classify the patients in cohorts into high and low score groups.

To investigate the potential clinical utility of glycosylation risk score, we first included it as an independent clinical indicator through univariate COX analysis. Our findings revealed that the glycosylation risk score, along with other clinical pathological factors such as tumor grade, stage of BLCA, age, and gender, significantly influenced prognosis ($p < 0.001$, Figure 2C). As shown in Figure 2D, subsequent multivariate Cox analysis demonstrated that the glycosylation risk score remains an independent prognostic indicator ($p < 0.001$). A glycosylation specific nomogram was developed using those independent prognostic factors identified by multivariate COX analysis (glycosylation risk score, age, and tumor stage) suggested that glycosylation risk score, like other

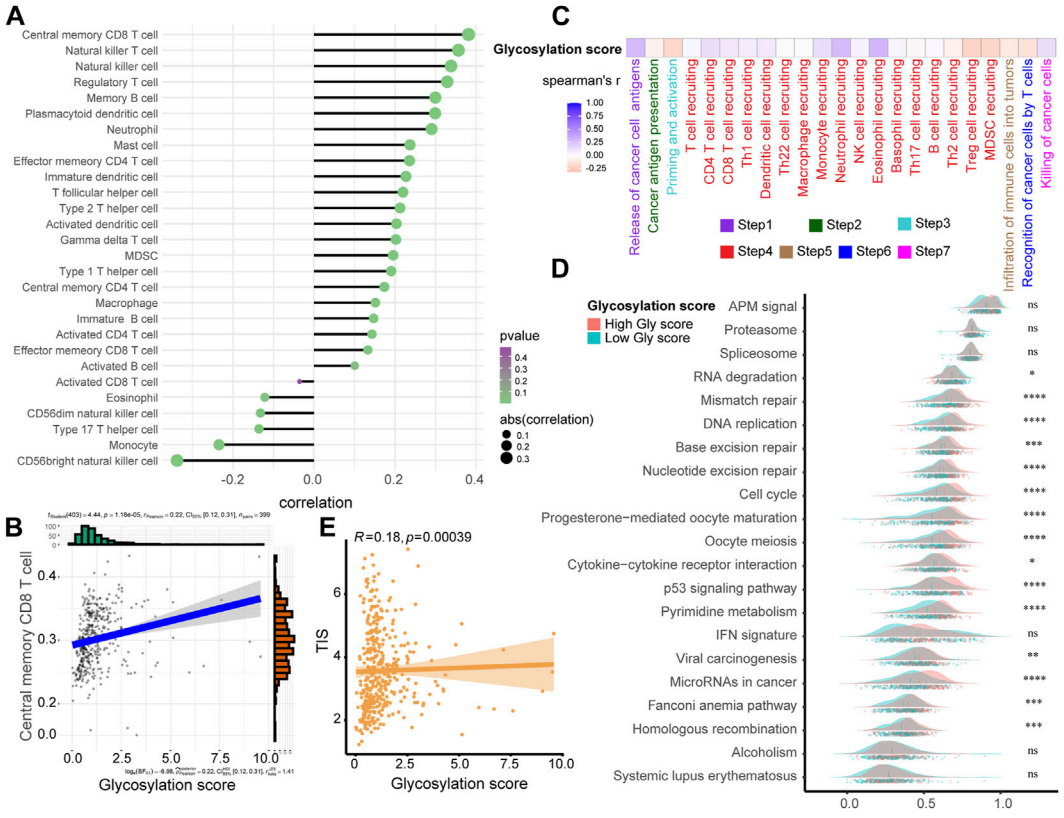


FIGURE 4
Exploring the relationship between Glycosylation score and TIME in the TCGA-BLCA cohort. **(A)** The association between Glycosylation risk score and immune cells in the Glycosylation in the TCGA-BLCA cohort, high Glycosylation score vs. low Glycosylation score. **(B)** The relationship between central memory CD8 T cells and Glycosylation score. **(C)** The association between Glycosylation risk score and cancer immunity cycles in the TCGA-BLCA cohort. **(D)** The different activated levels of gene signatures associated with ICB response between different Glycosylation-based risk score groups. Light red and green lines represent high and low Glycosylation-based risk score, respectively; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$; ns, not statistically significant. **(E)** The association between Glycosylation-based risk score and T cell-associated inflammatory signature (TIS) score.

clinical information, have important predictive value for prognosis (Figure 2E). The results indicated that glycosylation score accounted for a considerable proportion of the nomogram score, so its level could largely accurately predict the survival probability of a single patient at 1, 3, and 5 years. Furthermore, it was found that patients with higher glycosylation score had poorer prognosis, and there was an urgent need for new treatment methods to improve prognosis of these patients. In addition, the calibration curve indicated that the predicted OS of the glycosylation specific nomogram was very close to the real OS (Figure 2F), and the Q-Q plot had verified the normality of the above data (Supplementary Figure S1).

To further determine the prognostic significance of glycosylation risk score in BLCA, we validated its predictive value in multiple cohorts, including both public databases and our own real-world study. In the training cohort TCGA-BLCA, we observed that patients with a high glycosylation risk score had significantly worse prognosis compared to those with a low glycosylation risk score ($p < 0.0001$, Figure 3A). Additionally, the glycosylation risk score exhibited high accuracy in predicting 1-year, 3-year, and 5-year survival rates, with respective values of 0.75, 0.74, and 0.75 (Figure 3B). Meanwhile, in our real-world cohort (Xiangya BLCA cohort), the prognosis of the high glycosylation score group remained significantly poor ($p = 0.014$, Figure 3C) and its predictive accuracy was relatively high (1, 3, and 5 years accuracy: 0.75, 0.71 and 0.56 separately, Figure 3D). The above results remain robust: the prognosis of the high glycosylation score group presented obviously worse, in other public database cohorts, including E-MTAB-1803 ($p = 0.00019$, 1-year, 3-year, and 5-year accuracy: 0.73, 0.76 and 0.77 separately, Figures 3E, F), GSE32894 ($p < 0.0001$, 1-year, 3-year, and 5-year accuracy: 0.83, 0.89 and 0.88 separately, Figures 3G, H), GSE48075 ($p = 0.00012$, 1-year, 3-year, and 5-year accuracy: 0.82, 0.78 and 0.76 separately, Figures 3I, J), and two other GEO BLCA cohorts (Supplementary Figure S2).

The above results fully confirmed that glycosylation risk score can reliably predict the clinical outcomes of BLCA, and its predictive value had high accuracy and internal and external authenticity, which can be widely promoted to other cohorts.

Exploring the relationship between glycosylation score and TIME in the TCGA-BLCA cohort

The accurate prediction of glycosylation score for prognosis had sparked our interest in deeper research, therefore, we continued to investigate its association with the TIME in the TCGA-BLCA

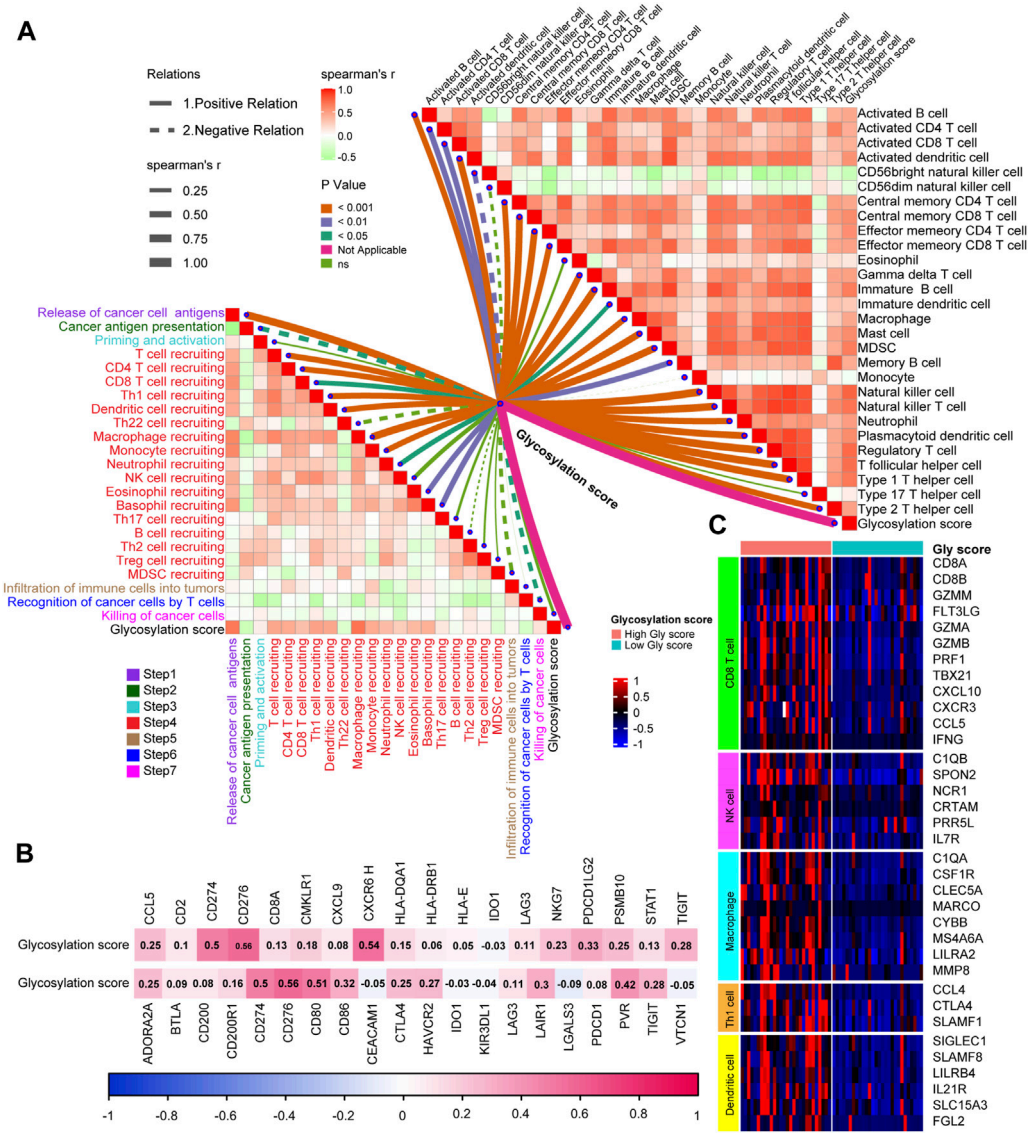


FIGURE 5
Verifying the relationship between Glycosylation score and TIME in a real-world BLCA cohort. **(A)** The association between Glycosylation-based risk score and cancer immunity cycles (left) and immune cells in the TME (right). The different types of lines represent the positive or negative relations. The different colors of the lines represent the *p* values of the relations, and the thickness of the lines represents the strength of the relations. **(B)** The association between Glycosylation risk score and T cell-associated inflammatory signature (TIS) genes (up) and immune checkpoint inhibitor (ICI) genes (down), separately. **(C)** The different expression patterns of effector genes of immune cells between different Glycosylation risk score groups, Light red and green lines represented high and low Glycosylation risk score groups, separately.

cohort. As shown in the single-sample gene set enrichment analysis (ssGSEA) analysis (Figure 4A; Supplementary Table S9), we found that compared to the low glycosylation score patients, the infiltration level of major tumor immune cells significantly increased in patients with high glycosylation score, such as central memory CD8/CD4 T cell, natural killer T cell, natural killer cell, regulatory T cell and memory B cell. The correlation between representative immune cells infiltration and glycosylation score were shown in Figure 4B and Supplementary Figure S3, Meanwhile, the activation of major steps of 7-step CIC, such as release of cancer cell antigens, immune cells recruiting and killing of cancer cells, was significantly higher in patients with high glycosylation score than those with low glycosylation score (Figure 4C; Supplementary Table S10). Furthermore, we examined the relationship between the glycosylation risk score and the enrichment score of gene features related to 21 immunotherapy-related pathways that summarized by Mariathasan et al. (2018). The findings revealed that the glycosylation score group exhibited higher pathway enrichment scores (Figure 4D). Finally, based on the TIS score of predicting immune checkpoint blocker (ICB) efficacy summarized by our team's previous research, we observed that the patients with higher glycosylation score also had higher TIS scores (Figure 4E). The consistency of the above results indicated that the high glycosylation score group was more inclined to express the "hot-immune" TME, and was predicted to be more sensitive to immunotherapy.

Verifying the relationship between glycosylation score and TIME in a real-world BLCA cohort

Based on glycosylation scores to predict the expression of TME immunophenotype in TCGA-BLCA, we verified how glycosylation affects the TIME in real BLCA patients in Xiangya Hospital. Like previous research ideas, in 7-step CIC (Figure 5A, left; Supplementary Table S11), we found that patients with high glycosylation scores were more activated in the main anti-tumor immune steps, including. Correspondingly, the level of immune cell infiltration in the TIME was significantly increased in patients with high scores (Figure 5A, right; Supplementary Table S12), including. Furthermore, as shown in Figure 5B (Supplementary Table S13), patients with high glycosylation scores also expressed more ICI (up) and TIS (down) related genes, further confirming the activity of immune cells in their TME. As for immune cell effector genes, the results strongly suggested that patients with higher glycosylation score express more effector genes for CD8+T cells, DC, macrophages, NK cells, and Th1 cells (Figure 5C). Based on the multiple verifications of glycosylation and immunity in TCGA-BLCA and Xiangya-BLCA, we conclude that patients with higher glycosylation scores often exhibited a "hot" TIME, which mad their efficacy in immunotherapy more ideal.

Glycosylation score guided precision medicine in BLCA by predicting molecular subtypes

The gene expression profiling of MIBC has revealed that it was a heterogeneous disease, that can be sub-grouped into a variety of molecular subtypes, and shared significantly different prognoses and responses to anti-tumor treatments (Sjödahl et al., 2012; Comprehensive molecular characterization of urothelial, 2014; Fu et al., 2018; McConkey and Choi, 2018). The most common and recognized molecular typing standards were as follows, consensus subtype (Kamoun et al., 2020), TCGA subtype (Robertson et al., 2017), Cartes d'Identité des Tumeurs-Curie (CIT) subtype (Rebouissou et al., 2014), Lund subtype (Marzouka et al., 2018), Baylor subtype (Mo et al., 2018), University of North Carolina (UNC) subtype (Damrauer et al., 2014), MDAnderson Cancer Center (MDA) subtype (Choi et al., 2014). Our previous research integrated and simplified the above 7 typing standards to promote the clinical implementation of BLCA molecular subtypes (Li et al., 2021).

We found high consistency in the results between the public training cohort TCGA-BLCA (Figure 6A, up) and our real-world research cohort Xiangya-BLCA (Figure 6A, down). Among all the classification criteria, the basal subtype was more inclined to obtain higher glycosylation score, while the luminal subtype was more inclined to obtain lower glycosylation score. And, patients with high glycosylation score tend to exhibit basal differentiation characteristics, such as EMT differentiation, Immune differentiation, basal differentiation, interferon response, and so on. Simultaneously, patients in the with low glycosylation score were more inclined to exhibit luminal differentiation, like luminal differentiation and urothelial differentiation. As for the accuracy of glycosylation score in predicting BLCA molecular subtype, in TCGA-BLCA (Figure 6B), most AUCs exceed 0.73, and in Xiangya BLCA (Figure 6C), most AUCs even exceed 0.87.

Previous studies (Seiler et al., 2017; Kamoun et al., 2020) have shown that the differentiation of the Basal subtype BLCA is lower than that of the Luminal subtype, resulting in poorer prognosis. However, it has a higher response rate to immunotherapy such as cisplatin and ICB. Our study showed that BLCA patients with high glycosylation score had poor prognosis but more immune cell infiltration due to a tendency towards lower differentiated Basal subtype.

Discussion

Neoadjuvant chemotherapy based on cisplatin, followed by radical cystectomy and urinary tract diversion, remains the standardized treatment plan for locally advanced MIBC since the early 21st century (Author Anonymous, 1999; Grossman et al., 2003). Glycosylation is involved in many fundamental cellular events, including cell migration, cell signaling, growth and intercellular adhesion, cell signaling, and growth, and is one of the most common post translational modifications of proteins (Fuster and Esko, 2005). And, abnormal glycosylation is also considered an indispensable part of the carcinogenesis process (Ni et al., 2014), including BLCA. The research on protein modification and tumor heterogeneity, as well as the prediction of immunotherapy efficacy such as ICB, remains a hot topic in many cancers (Liu et al., 2021; Liu et al., 2022). Therefore, our research was dedicated to deeply exploring the association between glycosylation and BLCA, with the goal of accurately predicting prognosis and individualized guidance for treatment of BLCA.

Firstly, based on the expression feature of 628 glycosylation genes in each TCGA-BLCA patient, we obtained the most appropriate two clusters through consensus clustering, named glycosylation cluster 1 and glycosylation cluster 2. The results indicated that the patients in glycosylation cluster 2 had poor prognosis but more immune cell infiltration into the TME. The poor prognosis and unsatisfactory treatment response of cancer are mostly related to complex TME (Siegel et al., 2021), with the role of immune cells and related pathways being important factors. Therefore, we hoped to further develop the quantitative value of glycosylation genes in predicting prognosis and immunophenotype of BLCA. In addition, in recent years, research on the mechanism of glycosylation in BLCA had made progress (Wu et al., 2021; Tan et al., 2022), but research on the development of risk score to evaluate the prognosis of BLCA was still lacking. Therefore, we constructed a model by screening candidate genes that were strongly correlated with prognosis and most representative of glycosylation gene expression characteristics, and for the first time developed a glycosylation risk score that can comprehensively predict the prognosis, immune phenotype, and molecular subtype of BLCA.

Tumor cells had a faster rate of protein glycosylation than normal cells (Beatson et al., 2016), and a prospective multi-omics study on ovarian cancer by Hu et al. (2020). Further demonstrated that there was a significant differential expression of glycosylation between cancer cells and normal cells, and the degree of glycosylation difference could be reflected by the expression of

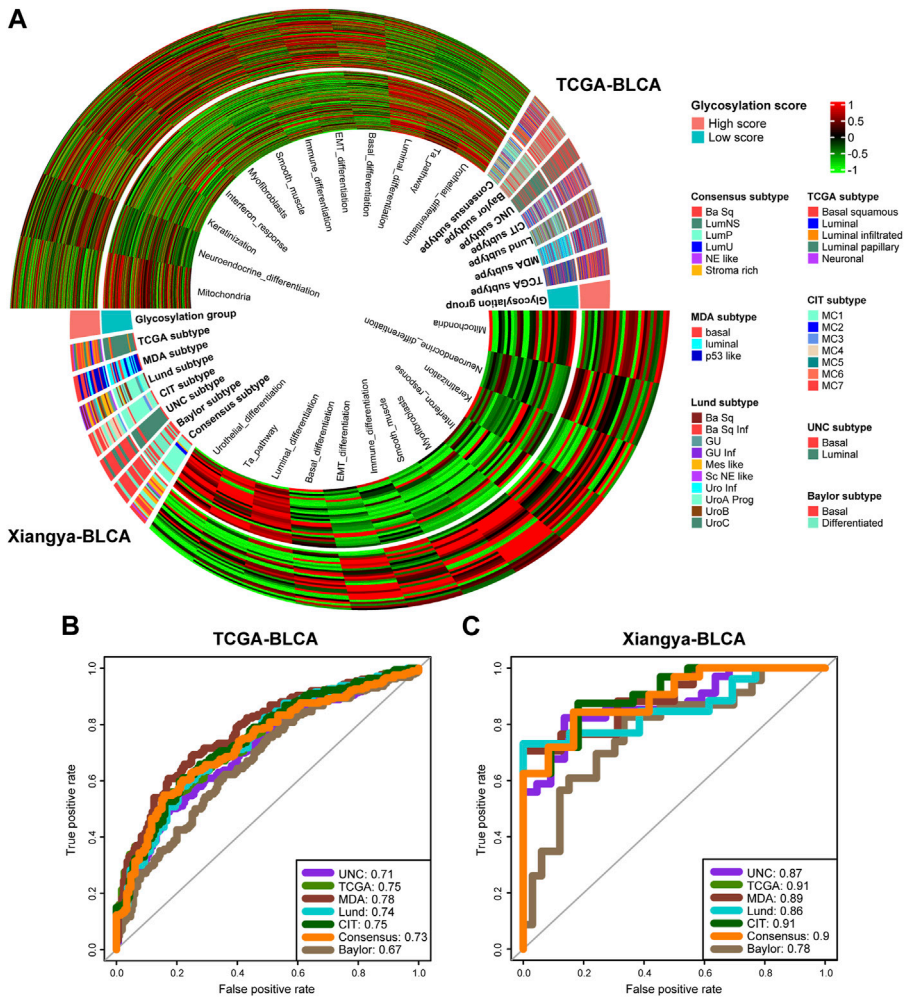


FIGURE 6
Glycosylation score guided precision medicine in BLCA by predicting molecular subtypes. **(A)** The heatmap of different Glycosylation risk score groups, seven molecular subtype classifications and bladder cancer associated signatures in the TCGA-BLCA (up) and Xiangya cohort (down). Activated or inhibited pathways are marked as red or green, separately. **(B,C)** ROC plot of the Glycosylation risk score for predicting seven molecular subtype classifications in BLCA in the TCGA-BLCA and Xiangya cohort.

glycoproteins in cancer cells. Some studies had shown that the abnormal mutation of the glycosylation related gene GALNT1 would lead to the occurrence and progression of a variety of cancers, including BLCA (Dyrskjøt et al., 2009). The activity of tumor infiltrating immune cells (TIICs), especially tumor infiltrating lymphocytes (TIL), in TME directly determines the survival outcome of tumor patients (Fridman et al., 2012), including early pT1 BLCA (Hülsen et al., 2020). In addition, the development of new targets, such as BCAT2, EMT-related signature and S100A5 (Xiao et al., 2022; Cai et al., 2023; Li et al., 2023), was playing an increasingly important role in immunotherapy for BLCA. A multicenter cohort study involving 709 patients (Bajorin et al., 2021) suggested that BLCA patients who still had a high risk of recurrence after surgery should be assisted with nivolumab. In this study, patients with high glycosylation scores had worse prognosis but presented a "hot" TIME (Duan et al., 2020) with high immune cell infiltration, and this result was highly consistent in the training set TCGA-BLCA and our own real-world cohort Xiangya-BLCA. Badmann S. et al.'s study (Badmann et al., 2020) provided a possible explanation for this phenomenon: in ovarian cancer, glycosylation could promote macrophage differentiation towards anti-inflammatory M2 type, leading to immune escape of cancer cells in immune activated TME.

Previous studies had shown that molecular typing can refine the prognosis and immune microenvironment of tumors. For example, in the study of breast cancer, it was found that the TILs infiltration level of different molecular subtypes of HR + breast cancer was quite different, in which TILs infiltration only prolongs OS, not disease-free survival (DFS) (Denkert et al.,2018). That was to say, different molecular subtypes exhibit different TIME (Goldberg et al., 2021). Rethinking the criteria for tumor molecular typing had become a hot topic, such as refining, updating, integrating, and simplifying. For example, the 5 mC regulator subtype system developed by our team in the previous study can accurately predict molecular typing in BLCA (Hu et al., 2021b). Moreover, the establishment of a consensus molecular subtype standard in gastric adenocarcinoma (GAC) to reclassify it and predict the response rate to ICB treatment (Wu et al., 2022), and a new standard developed by our team (Li et al.,

2021) that integrate multiple mainstream molecular subtypes of BLCA will bring molecular typing closer to tumor treatment practice. In addition, Miao et al. (2022) reported that a glycosylation related protein B3GNT5 was specifically overexpressed in basal-like breast cancer (BLBR), revealing the close relationship between glycosylation and cancer molecular subtype. In this study, BLCA patients with high glycosylation score tended to differentiate into basal subtype, and they had "hot" TIME characteristics, but had poor prognosis. However, patients with low glycosylation score exhibited opposite luminal subtype, as well as corresponding prognosis and immune phenotype. In summary, patients with high glycosylation score would have better expected efficacy in receiving immunotherapy such as ICB, so more efforts should be made to explore new immunotherapies to improve the prognosis after treatment. On the contrary, patients with low glycation score should focus more on the development of targeted therapies and other therapies. This result also confirmed the previous research on the impact of molecular subtype on tumor prognosis and immunity phenotype (Choi et al., 2014; Hodgson et al., 2018).

Finally, there are some limitations that need to be further explored and supplemented in future research in this study. First, the materials of this study were retrospective data, and the influence between glycosylation and prognosis, immunophenotype and molecular typing of BLCA mostly stops at the level of correlation. Therefore, we plan to take this study as a pre-study and carry out prospective research on glycosylation and immunotherapy and targeted therapy of BLCA in the follow-up series of studies. In addition, based on this study and more literature review, combined with experimental conditions, we will conduct research on the mechanism of glycosylation related genes affecting BLCA treatment, committed to developing new therapeutic targets to promote precise treatment of BLCA.

Conclusion

Our study constructed a glycosylation score related to BLCA through multi-omics data, and predicted the tumor heterogeneity, prognosis and immunophenotype of BLCA. Glycosylation score can reliably predict the efficacy of immunotherapy and molecular subtypes of BLCA, which is conducive to individualized treatment decisions of BLCA patients.

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 the Medical Ethics Committee of Xiangya Hospital Central South University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participant's legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

JL: Conceptualization, Methodology, Resources, Writing–original draft, Writing–review and editing, Formal Analysis, Investigation, Software, Visualization. YH: Conceptualization, Methodology, Resources, Visualization, Writing–original draft. WZ: Conceptualization, Methodology, Resources, Visualization, Writing–original draft. ZT: Conceptualization, Methodology, Resources, Visualization, Writing–original draft. ZX: Conceptualization, Methodology, Resources, Writing–original draft, Funding acquisition, Project administration, Supervision, Validation, Writing–review and editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China (81873626, 82070785, 82270816), Hunan Natural Science Foundation (2020JJ5884) and Hunan Province Young Talents Program (2021RC3027).

Acknowledgments

We sincerely thank all participants in the study.

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/fphar.2023.1280428/full#supplementary-material

References

Antoni, S., Ferlay, J., Soerjomataram, I., Znaor, A., Jemal, A., and Bray, F. (2017). Bladder cancer incidence and mortality: A global overview and recent trends. *Eur. Urol.* 71 (1), 96–108. doi:10.1016/j.eururo.2016.06.010

Author Anonymous, (1999). Neoadjuvant cisplatin, methotrexate, and vinblastine chemotherapy for muscle-invasive bladder cancer: A randomised controlled trial. International collaboration of trialists. *Lancet* 354 (9178), 533–540. doi:10.1016/S0140-6736(99)02292-8

Badmann, S., Heublein, S., Mayr, D., Reischer, A., Liao, Y., Kolben, T., et al. (2020). M2 macrophages infiltrating epithelial ovarian cancer express MDR1: A feature that may account for the poor prognosis. *Cells* 9 (5), 1224. doi:10.3390/cells9051224

Bajorin, D. F., Witjes, J. A., Gschwend, J. E., Schenker, M., Valderrama, B. P., Tomita, Y., et al. (2021). Adjuvant nivolumab versus placebo in muscle-invasive urothelial carcinoma. *N. Engl. J. Med.* 384 (22), 2102–2114. doi:10.1056/NEJMoa2034442

Beatson, R., Tajadura-Ortega, V., Achkova, D., Picco, G., Tsourouktsoglou, T.-D., Klausing, S., et al. (2016). The mucin MUC1 modulates the tumor immunological microenvironment through engagement of the lectin Siglec-9. *Nat. Immunol.* 17 (11), 1273–1281. doi:10.1038/ni.3552

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., et al. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* 24 (5), 541–550. doi:10.1038/s41591-018-0014-x

Bonaventura, P., Shekarian, T., Alcazer, V., Valladeau-Guilemond, J., Valsesia-Wittmann, S., Amigorena, S., et al. (2019). Cold tumors: A therapeutic challenge for immunotherapy. *Front. Immunol.* 10, 168. doi:10.3389/fimmu.2019.00168

Cai, Z., Chen, J., Yu, Z., Li, H., Liu, Z., Deng, D., et al. (2023). BCAT2 shapes a noninflamed tumor microenvironment and induces resistance to anti-PD-1/PD-L1 immunotherapy by negatively regulating proinflammatory chemokines and anticancer immunity. *Adv. Sci. (Weinheim, Baden-Wurttemberg, Ger.* 10 (8), e2207155. doi:10.1002/advs.202207155

Chen, D. S., and Mellman, I. (2017). Elements of cancer immunity and the cancer-immune set point. *Nature* 541 (7637), 321–330. doi:10.1038/nature21349

Chen, D. S., and Mellman, I. (2013). Oncology meets immunology: The cancer-immunity cycle. *Immunity* 39 (1), 1–10. doi:10.1016/j.immuni.2013.07.012

Choi, W., Porten, S., Kim, S., Willis, D., Plimack, E. R., Hoffman-Censits, J., et al. (2014). Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell* 25 (2), 152–165. doi:10.1016/j.ccr.2014.01.009

Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., et al. (2016). TCGAbiolinks: an R/bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 44 (8), e71. doi:10.1093/nar/gkv1507

Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507 (7492), (2014). 315–322. doi:10.1038/nature12965

Damrauer, J. S., Hoadley, K. A., Chism, D. D., Fan, C., Tiganelli, C. J., Wobker, S. E., et al. (2014). Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proc. Natl. Acad. Sci. U. S. A.* 111 (8), 3110–3115. doi:10.1073/pnas.1318376111

Denkert, C., von Minckwitz, G., Darb-Esfahani, S., Lederer, B., Heppner, B. I., Weber, K. E., et al. (2018). Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: A pooled analysis of 3771 patients treated with neoadjuvant therapy. *Lancet. Oncol.* 19 (1), 40–50. doi:10.1016/S1470-2045(17)30904-X

Duan, Q., Zhang, H., Zheng, J., and Zhang, L. (2020). Turning cold into hot: firing up the tumor microenvironment. *Trends Cancer* 6 (7), 605–618. doi:10.1016/j.trecan.2020.02.022

Dyrskjøt, L., Ostenfeld, M. S., Bramsen, J. B., Silahtaroglu, A. N., Lamy, P., Ramanathan, R., et al. (2009). Genomic profiling of microRNAs in bladder cancer: miR-129 is associated with poor outcome and promotes cell death *in vitro*. *Cancer Res.* 69 (11), 4851–4860. doi:10.1158/0008-5472.CAN-08-4043

Eichler, J. (2019). Protein glycosylation. *Curr. Biol. CB* 29 (7), R229–R231. doi:10.1016/j.cub.2019.01.003

Fridman, W. H., Pagès, F., Sautès-Fridman, C., and Galon, J. (2012). The immune contexture in human tumours: Impact on clinical outcome. *Nat. Rev. Cancer* 12 (4), 298–306. doi:10.1038/nrc3245

Fu, H., Zhu, Y., Wang, Y., Liu, Z., Zhang, J., Xie, H., et al. (2018). Identification and validation of stromal immunotype predict survival and benefit from adjuvant chemotherapy in patients with muscle-invasive bladder cancer. *Clin. Cancer Res. Official J. Am. Assoc. For Cancer Res.* 24 (13), 3069–3078. doi:10.1158/1078-0432.CCR-17-2687

Fuster, M. M., and Esko, J. D. (2005). The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat. Rev. Cancer* 5 (7), 526–542. doi:10.1038/nrc1649

Galon, J., and Bruni, D. (2019). Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nat. Rev. Drug Discov.* 18 (3), 197–218. doi:10.1038/s41573-018-0007-y

Goldberg, J., Pastorello, R. G., Vallius, T., Davis, J., Cui, Y. X., Agudo, J., et al. (2021). The immunology of hormone receptor positive breast cancer. *Front. Immunol.* 12, 674192. doi:10.3389/fimmu.2021.674192

Grossman, H. B., Natale, R. B., Tangen, C. M., Speights, V. O., Vogelzang, N. J., Trump, D. L., et al. (2003). Neoadjuvant chemotherapy plus cystectomy compared with cystectomy alone for locally advanced bladder cancer. *N. Engl. J. Med.* 349 (9), 859–866. doi:10.1056/NEJMoa022148

Hodgson, A., Liu, S. K., Vesprini, D., Xu, B., and Downes, M. R. (2018). Basal-subtype bladder tumours show a 'hot' immunophenotype. *Histopathology* 73 (5), 748–757. doi:10.1111/his.13696

Hu, J., Chen, J., Ou, Z., Chen, H., Liu, Z., Chen, M., et al. (2022). Neoadjuvant immunotherapy, chemotherapy, and combination therapy in muscle-invasive bladder cancer: A multi-center real-world retrospective study. *Cell Rep. Med.* 3 (11), 100785. doi:10.1016/j.xcrm.2022.100785

Hu, J., Othmane, B., Yu, A., Li, H., Cai, Z., Chen, X., et al. (2021b). 5mC regulator-mediated molecular subtypes depict the hallmarks of the tumor microenvironment and guide precision medicine in bladder cancer. *BMC Med.* 19 (1), 289. doi:10.1186/s12916-021-02163-6

Hu, J., Yu, A., Othmane, B., Qiu, D., Li, H., Li, C., et al. (2021a). Siglec15 shapes a non-inflamed tumor microenvironment and predicts the molecular subtype in bladder cancer. *Theranostics* 11 (7), 3089–3108. doi:10.7150/thno.53649

Hu, Y., Pan, J., Shah, P., Ao, M., Thomas, S. N., Liu, Y., et al. (2020). Integrated proteomic and glycoproteomic characterization of human high-grade serous ovarian carcinoma. *Cell Rep.* 33 (3), 108276. doi:10.1016/j.celrep.2020.108276

Hülsen, S., Lippolis, E., Ferrazzi, F., Otto, W., Distel, L., Fietkau, R., et al. (2020). High stroma T-cell infiltration is associated with better survival in stage pT1 bladder cancer. *Int. J. Mol. Sci.* 21 (21), 8407. doi:10.3390/ijms21218407

Jenkins, R. W., Barbie, D. A., and Flaherty, K. T. (2018). Mechanisms of resistance to immune checkpoint inhibitors. *Br. J. Cancer* 118 (1), 9–16. doi:10.1038/bjc.2017.434

Kamoun, A., de Reynies, A., Allory, Y., Sjodahl, G., Robertson, A. G., Seiler, R., et al. (2020). A consensus molecular classification of muscle-invasive bladder cancer. *Eur. Urol.* 77 (4), 420–433. doi:10.1016/j.eururo.2019.09.006

Lenis, A. T., Lec, P. M., Chamie, K., and Mshs, M. D. (2020). Bladder cancer: A review. *JAMA* 324 (19), 1980–1991. doi:10.1001/jama.2020.17598

Li, H., Chen, J., Li, Z., Chen, M., Ou, Z., Mo, M., et al. *S100A5 attenuates efficiency of anti-PD-L1/PD-1 immunotherapy by inhibiting CD8+ T cell-mediated anti-cancer immunity in bladder carcinoma*. Advanced Science (Weinheim, Baden-Wurttemberg, Germany), 2023: p. e2300110.

Li, H., Liu, S., Li, C., Xiao, Z., Hu, J., and Zhao, C. (2021). TNF family-based signature predicts prognosis, tumor microenvironment, and molecular subtypes in bladder carcinoma. *Front. Cell Dev. Biol.* 9, 800967. doi:10.3389/fcell.2021.800967

Liu, X. S., Liu, C., Zeng, J., Zeng, D. B., Chen, Y. J., Tan, F., et al. (2022). Nucleophosmin 1 is a prognostic marker of gastrointestinal cancer and is associated with m6A and cuproptosis. *Front. Pharmacol.* 13, 1010879. doi:10.3389/fphar.2022.1010879

Liu, X. S., Liu, J. M., Chen, Y. J., Li, F. Y., Wu, R. M., Tan, F., et al. (2021). Comprehensive analysis of hexokinase 2 immune infiltrates and m6A related genes in human esophageal carcinoma. *Front. Cell Dev. Biol.* 9, 715883. doi:10.3389/fcell.2021.715883

Ma, W., Gilligan, B. M., Yuan, J., and Li, T. (2016). Current status and perspectives in translational biomarker research for PD-1/PD-L1 immune checkpoint blockade therapy. *J. Hematol. Oncol.* 9 (1), 47. doi:10.1186/s13045-016-0277-y

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 554 (7693), 544–548. doi:10.1038/nature25501

Marzouka, N. A., Eriksson, P., Rovira, C., Liedberg, F., Sjodahl, G., and Hoglund, M. (2018). A validation and extended description of the Lund taxonomy for urothelial carcinoma using the TCGA cohort. *Sci. Rep.* 8 (1), 3737. doi:10.1038/s41598-018-22126-x

McConkey, D. J., and Choi, W. (2018). Molecular subtypes of bladder cancer. *Curr. Oncol. Rep.* 20 (10), 77. doi:10.1007/s11912-018-0727-5

Miao, Z., Cao, Q., Liao, R., Chen, X., Li, X., Bai, L., et al. (2022). Elevated transcription and glycosylation of B3GNT5 promotes breast cancer aggressiveness. *J. Exp. Clin. Cancer Res. CR* 41 (1), 169. doi:10.1186/s13046-022-02375-5

Mo, Q., Nikolos, F., Chen, F., Tramel, Z., Lee, Y. C., Hayashi, K., et al. (2018). Prognostic power of a tumor differentiation gene signature for bladder urothelial carcinomas. *J. Natl. Cancer Inst.* 110 (5), 448–459. doi:10.1093/jnci/djx243

Moremen, K. W., Tiemeyer, M., and Nairn, A. V. (2012). Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* 13 (7), 448–462. doi:10.1038/nrm3383

Ni, J., Jiang, Z., Shen, L., Gao, L., Yu, M., Xu, X., et al. (2014). β3GnT8 regulates the metastatic potential of colorectal carcinoma cells by altering the glycosylation of CD147. *Oncol. Rep.* 31 (4), 1795–1801. doi:10.3892/or.2014.3042

Pinho, S. S., and Reis, C. A. (2015). Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* 15 (9), 540–555. doi:10.1038/nrc3982

Plimack, E. R., Bellmunt, J., Gupta, S., Berger, R., Chow, L. Q. M., Juco, J., et al. (2017). Safety and activity of pembrolizumab in patients with locally advanced or metastatic urothelial cancer (KEYNOTE-012): A non-randomised, open-label, phase 1b study. *Lancet. Oncol.* 18 (2), 212–220. doi:10.1016/S1470-2045(17)30007-4

Powles, T., Bellmunt, J., Comperat, E., De Santis, M., Huddart, R., Loriot, Y., et al. (2022). Bladder cancer: ESMO clinical practice guideline for diagnosis, treatment and follow-up. *Ann. Oncol. Official J. Eur. Soc. For Med. Oncol.* 33 (3), 244–258. doi:10.1016/j.annonc.2021.11.012

Przybylo, M., Hoja-Lukowicz, D., Litynska, A., and Laidler, P. (2002). Different glycosylation of cadherins from human bladder non-malignant and cancer cell lines. *Cancer Cell Int.* 2, 6. doi:10.1186/1475-2867-2-6

Rebouissou, S., Bernard-Pierrot, I., de Reyniès, A., Lepage, M. L., Krucker, C., Chapeaublanc, E., et al. (2014). EGFR as a potential therapeutic target for a subset of muscle-invasive bladder cancers presenting a basal-like phenotype. *Sci. Transl. Med.* 6 (244), 244ra91. doi:10.1126/scitranslmed.3008970

Rijnders, M., de Wit, R., Boormans, J. L., Lolkema, M. P. J., and van der Veldt, A. A. M. (2017). Systematic review of immune checkpoint inhibition in urological cancers. *Eur. Urol.* 72 (3), 411–423. doi:10.1016/j.eururo.2017.06.012

Robertson, A. G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A. D., et al. (2017). Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* 171 (3), 540–556.e25. doi:10.1016/j.cell.2017.09.007

RodrÍguez, E., Schetters, S. T. T., and van Kooyk, Y. (2018). The tumour glyco-code as a novel immune checkpoint for immunotherapy. *Nat. Rev. Immunol.* 18 (3), 204–211. doi:10.1038/nri.2018.3

Rosenberg, J. E., Hoffman-Censits, J., Powles, T., van der Heijden, M. S., Balar, A. V., Necchi, A., et al. (2016). Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: A single-arm, multicentre, phase 2 trial. *Lancet (London, Engl.* 387 (10031), 1909–1920. doi:10.1016/S0140-6736(16)00561-4

Schoenfeld, A. J., and Hellmann, M. D. (2020). Acquired resistance to immune checkpoint inhibitors. *Cancer Cell* 37 (4), 443–455. doi:10.1016/j.ccell.2020.03.017

Seiler, R., Ashab, H. A. D., Erho, N., van Rhijn, B. W. G., Winters, B., Douglas, J., et al. (2017). Impact of molecular subtypes in muscle-invasive bladder cancer on predicting response and survival after neoadjuvant chemotherapy. *Eur. Urol.* 72 (4), 544–554. doi:10.1016/j.eururo.2017.03.030

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer statistics, 2021. *CA a Cancer J. For Clin.* 71 (1), 7–33. doi:10.3322/caac.21654

Siegel, R. L., Miller, K. D., Jemal, A., Fedewa, S. A., Butterly, L. F., Anderson, J. C., et al. (2020). Colorectal cancer statistics, 2020. *CA a Cancer J. For Clin.* 70 (1), 145–164. doi:10.3322/caac.21601

Sjödahl, G., Lauss, M., Lövgren, K., Chebil, G., Gudjonsson, S., Veerla, S., et al. (2012). A molecular taxonomy for urothelial carcinoma. *Clin. Cancer Res. Official J. Am. Assoc. For Cancer Res.* 18 (12), 3377–3386. doi:10.1158/1078-0432.CCR-12-0077-T

Sun, R., Kim, A. M. J., and Lim, S.-O. (2021). Glycosylation of immune receptors in cancer. *Cells* 10 (5), 1100. doi:10.3390/cells10051100

Tan, Z., Jiang, Y., Liang, L., Wu, J., Cao, L., Zhou, X., et al. (2022). Dysregulation and prometastatic function of glycosyltransferase C1GALT1 modulated by cHP1BP3/miR-1-3p axis in bladder cancer. *J. Exp. Clin. Cancer Res. CR* 41 (1), 228. doi:10.1186/s13046-022-02438-7

Varki, A. (2017). Biological roles of glycans. *Glycobiology* 27 (1), 3–49. doi:10.1093/glycob/cww086

Vonderheide, R. H. (2020). CD40 agonist antibodies in cancer immunotherapy. *Annu. Rev. Med.* 71, 47–58. doi:10.1146/annurev-med-062518-045435

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Wu, J., Tan, Z., Li, H., Lin, M., Jiang, Y., Liang, L., et al. (2021). Melatonin reduces proliferation and promotes apoptosis of bladder cancer cells by suppressing O-GlcNAcylation of cyclin-dependent-like kinase 5. *J. Pineal Res.* 71 (3), e12765. doi:10.1111/jpi.12765

Wu, X., Ye, Y., Vega, K. J., and Yao, J. (2022). Consensus molecular subtypes efficiently classify gastric adenocarcinomas and predict the response to anti-PD-1 immunotherapy. *Cancers* 14 (15), 3740. doi:10.3390/cancers14153740

Xiao, Z., Cai, Z., Deng, D., Tong, S., and Zu, X. (2022). An EMT-based risk score thoroughly predicts the clinical prognosis, tumor immune microenvironment and molecular subtypes of bladder cancer. *Front. Immunol.* 13, 1000321. doi:10.3389/fimmu.2022.1000321

Xu, L., Deng, C., Pang, B., Zhang, X., Liu, W., Liao, G., et al. (2018). Tip: A web server for resolving tumor immunophenotype profiling. *Cancer Res.* 78 (23), 6575–6580. doi:10.1158/0008-5472.CAN-18-0689

Glossary

BLCA	bladder cancer
NMIBC	non-muscle invasive bladder cancer
MIBC	muscle-invasive bladder cancer
TIME	tumor immune microenvironment
TME	tumor microenvironment
TCGA	the Cancer Genome Atlas
GDC	Genomic Data Commons
FPKM	fragments per kilobase of exon model per million mapped fragments
TPM	transcripts per kilobase of exon model per million mapped reads
GEO	Gene Expression Omnibus
RNA-seq	RNA-sequencing
GSEA	gene set enrichment analysis
TIP	tracking tumor immunophenotype
CIC	Cancer Immunity Cycle
ICI	immune checkpoint inhibitor
TIS	T cell-associated inflammatory signature
DCs	dendritic cells
NK	natural killer
Th1	type 1 T helper
LASSO	least absolute shrinkage and selection operator
K-M	Kaplan-Meier
tROC	time-dependent receiver operating characteristic
HR	hazard ratio
ROC	receiver operating characteristic
AUC	area under the curve
ssGSEA	single sample gene set enrichment analysis
ICB	immune checkpoint blocker
TIICs	tumor infiltrating immune cells
TIL	tumor infiltrating lymphocytes
DFS	disease-free survival
GAC	gastric adenocarcinoma
BLBR	basal-like breast cancer

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 26 September 2023
DOI 10.3389/fphar.2023.1264345

OPEN ACCESS

EDITED BY
Linhui Wang,
Second Military Medical University, China

REVIEWED BY
Zhengrui Li,
Shanghai Jiao Tong University, China
Haizhou Lou,
Zhejiang University, China
Zhaomin Yao,
Northeastern University, China

*CORRESPONDENCE
Qinhong Zhang,
zhangqh0451@163.com
Hao Chi,
Chihao7511@163.com
Guanhu Yang,
guanhuyang@gmail.com

†These authors have contributed equally to this work

RECEIVED 20 July 2023
ACCEPTED 15 September 2023
PUBLISHED 26 September 2023

CITATION
Li Y, Cai H, Yang J, Xie X, Pei S, Wu Y, Zhang J, Song G, Zhang J, Zhang Q, Chi H and Yang G (2023), Decoding tumor heterogeneity in uveal melanoma: basement membrane genes as novel biomarkers and therapeutic targets revealed by multi-omics approaches for cancer immunotherapy.
Front. Pharmacol. 14:1264345.
doi: 10.3389/fphar.2023.1264345

Decoding tumor heterogeneity in uveal melanoma: basement membrane genes as novel biomarkers and therapeutic targets revealed by multi-omics approaches for cancer immunotherapy

Yunyue Li[1†], Huabao Cai[2†], Jinyan Yang[3†], Xixi Xie[3†], Shengbin Pei[4], Yifan Wu[4], Jinhao Zhang[3], Guobin Song[3], Jieying Zhang[5], Qinhong Zhang[6*], Hao Chi[7*] and Guanhu Yang[8*]

[1]Queen Mary College, Medical School of Nanchang University, Nanchang, China, [2]Department of Neurosurgery, First Affiliated Hospital of Anhui Medical University, Hefei, China, [3]School of Stomatology, Southwest Medical University, Luzhou, China, [4]Department of Breast Surgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China, [5]First Teaching Hospital of Tianjin University of Traditional Chinese Medicine, Tianjin, China, [6]Heilongjiang University of Chinese Medicine, Harbin, China, [7]Clinical Medical College, Southwest Medical University, Luzhou, China, [8]Department of Specialty Medicine, Ohio University, Athens, OH, United States

Background: Uveal melanoma (UVM) is a primary intraocular malignancy that poses a significant threat to patients' visual function and life. The basement membrane (BM) is critical for establishing and maintaining cell polarity, adult function, embryonic and organ morphogenesis, and many other biological processes. Some basement membrane protein genes have been proven to be prognostic biomarkers for various cancers. This research aimed to develop a novel risk assessment system based on BMRGs that would serve as a theoretical foundation for tailored and accurate treatment.

Methods: We used gene expression profiles and clinical data from the TCGA-UVM cohort of 80 UVM patients as a training set. 56 UVM patients from the combined cohort of GSE84976 and GSE22138 were employed as an external validation dataset. Prognostic characteristics of basement membrane protein-related genes (BMRGs) were characterized by Lasso, stepwise multifactorial Cox. Multivariate analysis revealed BMRGs to be independent predictors of UVM. The TISCH database probes the crosstalk of BMEGs in the tumor microenvironment at the single-cell level. Finally, we investigated the function of ITGA5 in UVM using multiple experimental techniques, including CCK8, transwell, wound healing assay, and colony formation assay.

Results: There are three genes in the prognostic risk model (ADAMTS10, ADAMTS14, and ITGA5). After validation, we determined that the model is quite reliable and accurately forecasts the prognosis of UVM patients. Immunotherapy is more likely to be beneficial for UVM patients in the high-risk group, whereas the survival advantage may be greater for UVM patients in the low-risk

group. Knockdown of ITGA5 expression was shown to inhibit the proliferation, migration, and invasive ability of UVM cells in vitro experiments.

Conclusion: The 3-BMRGs feature model we constructed has excellent predictive performance which plays a key role in the prognosis, informing the individualized treatment of UVM patients. It also provides a new perspective for assessing pre-immune efficacy.

KEYWORDS

uveal melanoma, basement membrane genes, machine learning, multi-omics, tumor heterogeneity, cancer immunotherapy

1 Introduction

Uveal melanoma (UVM) is a rare yet aggressive primary intraocular malignancy arising from ocular melanocytes, constituting a small fraction of all melanomas (Singh et al., 2005; Singh et al., 2011; Chattopadhyay et al., 2016; Chi et al., 2022). It imposes significant threats to patients' visual function and overall survival, with a high mortality rate of up to 50% attributed to its severe clinical presentation, malignancy, and limited treatment options (Andreoli et al., 2015). Notably, UVM exhibits a propensity for metastasis, with approximately half of the patients developing distant organ metastases, most commonly involving the liver, lung, and bone (Rusňák et al., 2020). Unfortunately, current therapeutic modalities for UVM have shown limited efficacy in managing metastatic disease (Augsburger et al., 2009; Damato, 2018). As a result, recent research endeavors have focused on the development of targeted therapeutics and immunotherapeutic strategies, including immune checkpoint inhibitors, vaccines, and adoptive cell therapy, to address the unmet medical needs in UVM (Curran et al., 2010; Larkin et al., 2015; Bol et al., 2016). However, the underlying etiology and molecular mechanisms driving UVM remain largely elusive (Smit et al., 2020; Derrien et al., 2021; Katopodis et al., 2021). Consequently, there is a critical need to identify novel prognostic biomarkers and molecular targets that can accurately predict patient outcomes and facilitate personalized treatment approaches, ultimately improving the quality of life for individuals affected by UVM.

The basement membrane (BM) is a specialized extracellular matrix located at the basal aspect of epithelial tissues, primarily composed of collagen IV, laminin, heparan sulfate proteoglycans, BM-40, and nidogen (Timpl, 1989). Its crucial role in establishing and maintaining cellular polarity and providing mechanical support to tissues is well-recognized (Banerjee et al., 2022). Moreover, BMs play critical roles in various physiological processes, including embryonic development, organ morphogenesis, and adult tissue homeostasis (Li et al., 2003). Perturbations in BM protein expression and turnover have been implicated in tumorigenesis, and dysregulation of BM integrity has been associated with tumor metastasis (Valastyan and Weinberg, 2011; Naba et al., 2014). While BM-related genes have shown prognostic significance in several cancers, their role, and prognostic implications in uveal melanoma (UVM) remain poorly understood. To elucidate the immunological status of UVM patients and accurately predict prognosis, this study aimed to develop a novel risk-scoring system based on BM-related genes. The objective was to establish a theoretical foundation for personalized therapeutic interventions tailored to individual patients. By comprehensively characterizing the expression and functional relevance of BMRGs, this risk-scoring system would enable precise prognostic stratification and facilitate tailored treatment strategies in UVM.

Following the rapid advancement of bioinformatics (Song et al., 2022a; Zhao et al., 2022a; Jin et al., 2022), a considerable amount of research has been conducted to establish models for predicting the prognosis of UVM through machine learning. For example, Zheng et al. established an autophagy-related gene (ARG) risk model and validated it with TCGA and four external independent UVM cohorts, revealing that UVM patients with higher risk scores exhibited higher levels of immune cell infiltration and enrichment of tumor markers (Zheng et al., 2021); Lv et al. constructed a UVM prognostic model based on the Epithelial-mesenchymal transition (EMT) signature, which found that patients with high EMT scores potentially had higher response rates to immunotherapy (Lv et al., 2022); Yang et al. utilized immune markers systematically to develop a prognostic six-immune-gene signature via RNA sequencing data from TCGA and GEO databases for predicting the overall survival outcome of UVM patients (Yang et al., 2023). Meanwhile, several studies have reported that BMRG signatures could predict the prognosis of tumor survivors and provide a potential target for immunotherapy (Cai et al., 2022; Shen et al., 2023). However, BMRG-related models have not yet been established and validated for prognostic prediction in UVM patients (Song et al., 2022a).

In this study, we developed a prognostic model for UVM using the TCGA-UVM cohort. We carefully selected three reliable basement membrane-related genes (BMRGs) through a rigorous screening process and employed two machine learning techniques to construct the model. By integrating genetic information from UVM patients, we aimed to explore the prognostic value of these three BMRGs and develop novel tools to enhance therapeutic strategies. Our analysis involved assessing the interaction between BMRGs and the immune microenvironment, as well as evaluating the impact of BMRGs on immunotherapy and chemotherapy sensitivity. We eventually verified the functional role of the ITGA5, the gene with the highest absolute HR value, in UVM cells by an *in vitro* experiment. By leveraging advanced computational methods and integrating multi-dimensional data, we sought to gain insights into the role of BMRGs in determining the prognosis of UVM, identify potential avenues for improving treatment regimens, and offer possibilities for developing personalized therapeutic approaches. These findings have the potential to enhance patient outcomes and pave the way for further advancements in UVM research and clinical practice.

2 Materials and methods

2.1 Patient data sources

We utilized the TCGA-UVM cohort, obtained from the publicly available TCGA database, as our training set, consisting of gene expression profiles and clinical data from 80 tumor patients. To ensure accurate analysis, we performed preprocessing steps on the data. Initially, we converted the level 3 HTSeq-fragments per kilobase (FPKM) data into transcripts per million reads (TPM) to account for gene length and sequencing depth variations across samples. This conversion was done using a formula that normalized the TPM values. Subsequently, we applied a logarithmic transformation to the TPM values to normalize the data and enhance comparability between samples. It is important to note that due to significant variation in sample sizes among UVM patients at stages M and N, these stages were excluded from our analysis to ensure robustness and reliability. Furthermore, we incorporated two external validation datasets, GSE84976 and GSE22138, from the GEO database. It is worth noting that datasets GSE84976 and GSE22138 were merged together to act as a validation set, and in order to mitigate the effects of batch differences between the microarray expression data, we utilized the ComBat function in the R package "sva" to achieve correction for batch effects. For comprehensive details on these datasets, see Supplementary Material. These datasets included genetic profiles and clinical data from 56 UVM patients, and their inclusion aimed to enhance the validity and generalizability of our analyses. In the training cohort, we transcribed and analyzed tissue samples from eye cancer patients for comparative analyses to obtain genes that were aberrantly expressed in eye cancer patients. While in the external validation set, we only included samples from eye cancer patients analyzed. In addition, we required complete patient follow-up and clinical information in the cohort and complete micro-matrix data in the cohort to ensure data quality for subsequent bioinformatics analysis. By employing these rigorous preprocessing steps and integrating multiple datasets, we aimed to improve the accuracy and reliability of our findings, providing valuable insights into the molecular characteristics and clinical implications of UVM.

2.2 Consensus clustering analysis

To gain deeper insights into the mechanistic implications of BMRGs in UVM, we employed advanced analytical methodologies. The "Consensus Cluster Plus" R package (Zhao et al., 2022b; Wang et al., 2022) was leveraged to classify UVM patient samples into distinct clusters based on the expression patterns of BMRGs, thereby unveiling unique gene expression profiles associated with specific subtypes. Differential expression patterns of BMRGs across clusters, along with clinicopathological parameters, were visualized using the "pheatmap" R package (Bhattacherjee et al., 2019; Song et al., 2022b). To elucidate the distinct biological pathways and processes underlying these clusters, we retrieved the "c2.cp.kegg.v7.4.symbols.gmt" file from the MSigDB database (Liberzon et al., 2015) for genomic variation analysis via GSVA. Employing the "GSVA" R package (Hänzelmann et al., 2013), we systematically analyzed pathway differences between clusters, revealing noteworthy disparities in key pathways among diverse UVM subtypes. Furthermore, the Single Sample Genome Enrichment Analysis (ssGSEA) algorithm (Zhuang et al., 2021; Huang et al., 2023) was applied to assess the infiltration levels of immune cells and expression levels of immune checkpoints within the identified clusters. This integrative approach shed light on potential variations in the immune microenvironment across UVM subtypes, offering crucial insights into the prospective efficacy of immune checkpoint-based therapies in specific patient cohorts.

2.3 Model construction and validation

The dataset of basement membrane (BM) genes was obtained from the Basement Membrane BASE database (https://bmbase.manchester.ac.uk), comprising a comprehensive collection of 224 genes associated with the basement membrane protein. To explore the potential prognostic relevance of these genes, univariate Cox regression analysis was performed, resulting in the identification of 81 genes significantly associated with survival outcomes. To further refine the gene set and mitigate the risk of overfitting, we employed the LASSO (Least Absolute Shrinkage and Selection Operator) method, a powerful machine learning approach (Chi et al., 2023a; Chi et al., 2023b). The "glmnet" R package (Engebretsen and Bohlin, 2019; Ren et al., 2023) was utilized to implement LASSO, which involves adding a penalty term to the regression model. This penalty encourages the coefficients of less influential predictors to shrink toward zero, effectively selecting the most informative subset of predictors. By applying LASSO, we successfully narrowed down the candidate genes to eight. Subsequently, a stepwise multi-factor Cox regression model was employed to identify and estimate the coefficients of the core genes from the selected set. Through this iterative procedure, we ultimately derived a risk profile consisting of four BMRGs. For each patient, the risk score was calculated by combining the expression levels of these genes with their corresponding coefficients: Risk score = ExpressionmRNA1 × CoefmRNA1 + ExpressionmRNA2 × CoefmRNA2 + ExpressionmRNAn × CoefmRNAn. By leveraging these analytical approaches, we aimed to establish a robust and concise set of BMRGs with prognostic implications in order to facilitate risk stratification and inform personalized treatment strategies for patients.

2.4 Correlation between clinicopathological features and risk scores

Investigating the relationship between risk scores and relevant clinical features in patients with uveal melanoma (UVM) can provide valuable insights for clinical prognostic assessment. To visualize the associations between clinical features and the modeled genes, we employed the "pheatmap" R package (Lu et al., 2021) to generate heat maps displaying multiple groups of clinical features. To gain a deeper understanding of the differences in risk scores among various patient subgroups, we performed clinical analyses on the entire sample cohort. The patients were stratified

based on different clinical characteristics, including age (≤65 and >65 years), sex (male and female), pathological stage (II and III-IV), and T-stage (T2 and T3-4). Between-group differences were assessed using the "ggpubr" package (Whitehead et al., 2019). By examining the relationships between risk scores and clinical subgroups, we aimed to identify potential variations in risk profiles based on different demographic and pathological factors. These analyses would contribute to a more comprehensive understanding of the prognostic implications of risk scores in UVM patients and their clinical relevance.

2.5 Independent prognostic analysis and nomogram construction

To evaluate the independent prognostic value of the risk score in predicting uveal melanoma (UVM) outcomes, we conducted both univariate and multivariate Cox regression analyses. These analyses aimed to assess whether the risk score could serve as a reliable prognostic factor, independent of conventional clinicopathological characteristics. The "rms" R package (Zhang et al., 2022) was employed to construct a nomogram incorporating the risk score and clinicopathological features. This nomogram provided a visual tool for predicting the survival of patients in the TCGA-UVM cohort, enabling clinicians to estimate individual patient prognoses more accurately. To assess the predictive performance of the nomogram, we utilized the "ggDCA" R package (Mao et al., 2021) to develop decision curve analysis (DCA) and calibration curves. The DCA allowed us to evaluate the clinical benefits of using the nomogram compared to other predictive models or strategies. Calibration curves were generated to assess the calibration accuracy of the nomogram in predicting patient survival. These comprehensive analyses aimed to validate the prognostic value of the risk score and provide clinicians with a practical tool for prognostic assessment in UVM patients. By integrating the risk score with clinicopathological characteristics, the nomogram offered improved prognostic accuracy, ultimately enhancing patient management and treatment decision-making.

2.6 Establishing the equations for signatures

After scoring all UVM patients based on the risk model equation, we determined the median risk score using the 'survminer' R package. Subsequently, we categorized the patients into a low-risk group and a high-risk group. Survival curves were plotted for both groups to visually compare their survival outcomes. To evaluate the predictive performance of the risk model, we calculated the C-index using the 'pec' R package. The C-index provides a measure of concordance between predicted risk scores and actual survival outcomes. To further assess the predictive power of the genetic traits, we conducted an analysis of receiver operating characteristic (ROC) curves using the 'time-ROC' R package. ROC curves allow us to evaluate the sensitivity and specificity of the genetic traits in predicting survival outcomes. Additionally, decision curve analysis (DCA) was performed for the multi-factor Cox regression model using the 'ggDCA' R package. DCA provides insights into the clinical utility of the predictive model by assessing the net benefits of different strategies or models across a range of threshold probabilities. Through these analyses, we aimed to assess the predictive accuracy and clinical usefulness of the risk model in UVM patients. The survival curves, C-index, ROC curves, and DCA plots provide valuable information for understanding the prognostic value and potential application of genetic traits in UVM patient management and treatment decision-making.

2.7 Enrichment analysis

To analyze the Gene Ontology (GO) pathway, we utilized the "ClusterProfiler" R package (Song et al., 2023; Zhang et al., 2023). In the generated graphs, a *p*-value of less than 0.05 indicated a statistically significant difference, highlighting the enriched pathways and functional categories associated with the genes of interest. For further enrichment analysis, we conducted GSVA using the "GSVA" R package. The data from "c2. cp.kegg.v7.5.1. symbols.gmt" in the MSigDB database were utilized to explore the functional annotation and enrichment pathways. To visualize the results, heatmaps were generated using the 'heatmap' R package. Adjusted *p*-values of less than 0.05, obtained through the 'limma' R package, indicated the statistical significance of subgroup differences in the heatmap. Through functional enrichment analysis, we aimed to gain insights into the biological functions, pathways, and processes associated with differentially expressed genes related to BMRGs in UVM. These analyses contribute to a better understanding of the molecular mechanisms underlying UVM and provide valuable information on functional annotations and enriched pathways associated with BMRGs in the context of UVM.

2.8 Immuno-infiltration analysis

Multiple methods have been developed to quantify immune infiltration scores, including XCELL, TIMER, QUANTISEQ, MCPCOUNT, EPIC, CIBERSORT, and CIBERSORT-ABS. These methods offer diverse approaches for evaluating the presence and abundance of immune cells within the tumor microenvironment. To investigate the association between immune cells and risk scores, Spearman correlation analysis was employed, allowing for a comprehensive understanding of the immune landscape in UVM. Utilizing the immune cell profiles of UVM patients, we applied the ssGSEA method to stratify patients into distinct low- and high-risk groups based on their immune signatures. Furthermore, we examined the differential expression of 20 suppressive immune checkpoints between the identified high-risk and low-risk groups, shedding light on the potential influence of immune checkpoint blockade therapies. To assess and visualize the impact of immunotherapy in UVM patients, we utilized the widely adopted 'limma' and 'ggpubr' R packages. To expand our understanding of the genetic underpinnings related to cancer and immunity, we referred to the curated collection of genes provided by Xu et al., available on their website (Xu et al., 2018). Employing the R package "ggcor," we explored the correlation between risk scores and these two genetic traits, unraveling potential associations between genetic alterations and disease prognosis in UVM. Additionally, to predict immune infiltration estimates and immunotherapy response data,

we leveraged the computational tool ImmuCellAI (Miao et al., 2020). This powerful resource enables comprehensive analyses of the immune landscape and aids in guiding immunotherapeutic strategies for UVM patients.

2.9 TISCH analysis

The Tumor Immunological Single Cell Centre (TISCH) hosts a comprehensive single-cell RNA sequencing database that focuses on investigating the intricate tumor microenvironment (TME). This valuable resource facilitates detailed annotation of various single-cell types, enabling in-depth analysis of gene expression within distinct cellular populations. By examining gene expression patterns across different cell types, we can unravel the intricate variations within the tumor microenvironment of individual UVM patients, thus shedding light on the underlying heterogeneity of UVM. This comprehensive characterization of the TME aids in elucidating the complex dynamics and functional implications of different cell types within the UVM context.

2.10 Cell culture

The human uveal melanoma cells (MuM-2B, OCM-1) utilized in this investigation were generously provided by the Cell Resource Center at Shanghai Life Sciences Institute. These cells were cultivated under controlled conditions in Dulbecco's Modified Eagle's Medium (DMEM) (Gibco, United States), supplemented with 1% penicillin/streptomycin and 10% fetal bovine serum (FBS) (Gibco, United States), within a humidified incubator set at 37°C with a 5% CO2 atmosphere.

2.11 CCK-8 assay

To assess the impact of ITGA5 on the proliferative capacity of uveal melanoma (UVM) cells, the Cell Counting Kit-8 (CCK-8) assay was employed. UVM cells were cultured in 96-well microplates in triplicate, with each well initially seeded with 5,000 cells. Subsequent to transfection, the cells were subjected to treatment at 37°C for a duration of 2 h, utilizing 10 μL of CCK-8 solution (A311-01, vazyme, Nanjing, China) mixed with 90 μL of complete media in each well at specific time points (0, 24, 48, 72, or 96 h). Following the respective incubation periods, the absorbance of each well was quantified at 450 nm using a microplate reader.

2.12 Wound-healing assay

The wound healing assay was employed to evaluate the migratory behavior of MuM-2B and OCM-1 cells, providing valuable insights into their migratory patterns. The transfected cells were cultured in a six-well plate and incubated at 37°C until they reached approximately 80% confluence. To create a standardized wound, a sterile 200 μL pipette tip was carefully used to generate a linear scrape across the cell monolayers. Following this, the medium was replaced with serum-free medium after two washes with phosphate-buffered saline (PBS) to eliminate any cellular debris. The movement of cells into the wound area was monitored at 0 h and 48 h using an inverted microscope (Olympus, Japan), enabling the quantification of the distance traveled by the cells into the wound surface.

2.13 Transwell assay

Cell migration was assessed using the Transwell migration assay, which involved a 24-well Transwell plate equipped with 8 μm-pore membrane filters. Briefly, the bottom chamber of the Transwell plate was supplemented with media containing 10% fetal bovine serum (FBS), while the top chamber was coated with 2×10^5 cells suspended in serum-free medium. Following a 48-h incubation period, the cells that had migrated to the bottom chamber were fixed in 4% methanol for 10 min and subsequently stained with 0.1% crystal violet (Solarbio, Beijing, China) for 15 min.

2.14 Statistical analysis

The statistical analysis was conducted using R software version 4.1.3. To compare the overall survival (OS) between the high-risk and low-risk groups, Kaplan-Meier (KM) survival curves and log-rank tests were employed. In addition, Lasso regression analyses were performed to assess the potential relevance of BMRGs. A stepwise multivariate Cox regression analysis was then employed to construct a BMRG signature. The predictive performance of the model was evaluated using a time-dependent ROC curve. The relationship between the risk score and immune cell infiltration was assessed using Spearman correlation analysis. To compare the ratios of tumor immune infiltrating cells (TIIC), immunological checkpoints, and immune function between the two groups, the Wilcox test was applied. Statistical significance was determined by p-values <0.05, and a false discovery rate (FDR) < 0.05 was considered statistically significant. The CCK-8 data analysis was conducted using GraphPad Prism Software version 8.3.0. The mean values ±standard deviation (SD) were determined based on data obtained from three independent experiments. Statistical significance was assessed using analysis of variance (ANOVA), with a significance level set at $p < 0.05$.

3 Results

3.1 Consensus clustering determined the molecular subtypes of BMRGs

The primary study design is presented in Figure 1, illustrating the overall flow of the investigation. The cumulative distribution function (CDF) values demonstrated an increasing trend in relation to the consensus index, indicating successful classification. To assess cluster composition and quantity, the consensus matrix serves as an excellent visual tool. We generated a color-coded heat map based on the consensus matrix, which revealed higher intra-cluster correlations and lower inter-cluster correlations when considering k = 2. These findings strongly support the acceptance of two

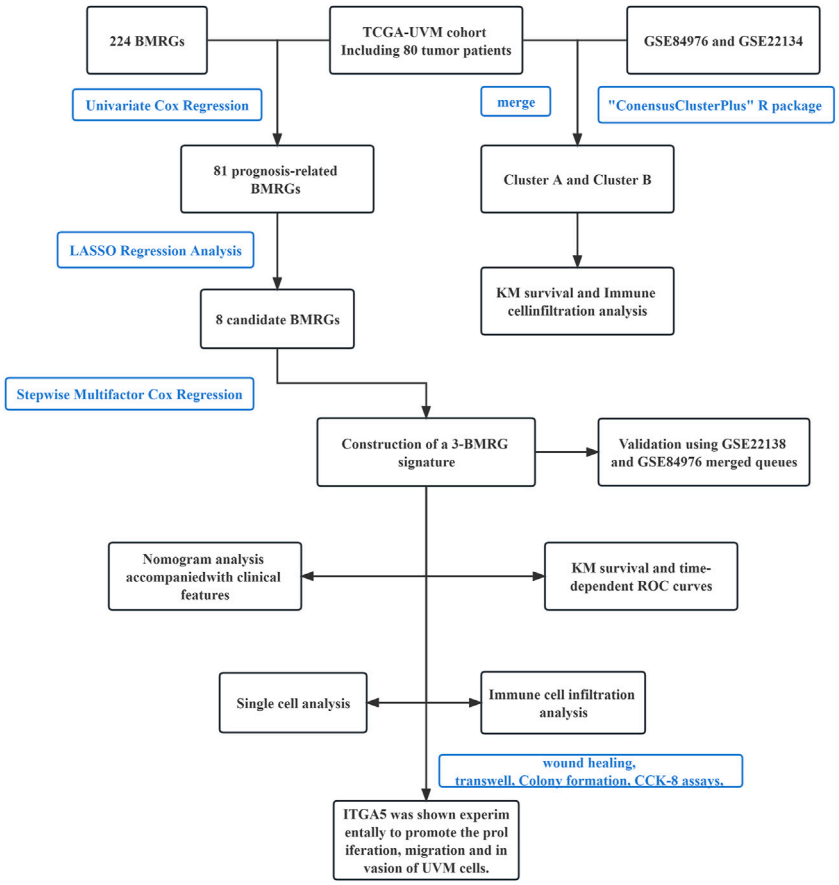


FIGURE 1
Flow chart of this study.

subtypes (Cluster A and Cluster B) for categorizing UVM patients. Based on the CDF curves and the Delta area, k = 2 represents the optimal point to achieve maximal inter-cluster differences as the clustering index "k" increases from 2 to 9. Consequently, we divided the UVM patients into two subgroups (Figures 2A–D).

Furthermore, we investigated the differential survival prognosis across clusters using the Cluster Survival R package. The results indicated that patients in cluster A exhibited significantly better survival prognoses than those in cluster B ($p < 0.001$) (Figure 2E).

Principal component analysis (PCA) was performed to visualize risk distribution among different patient groups. The PCA plot (Figure 2F) demonstrated distinct differences between Cluster A and Cluster B patients. Additionally, we conducted further analysis to explore metabolic variations in BMRGs between clusters A and B. The heat map revealed notable expression differences and clinical traits associated with BMRGs in cluster B (Figure 2G).

To investigate potential biological pathways, we performed an enrichment analysis using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database on the clustered samples. We explored correlations among various cancer-related pathways, such as apoptosis, transporters, and the MAPK signaling pathway (Figure 2H). Moreover, we employed the ssGSEA algorithm to assess the distribution and correlation of 23 tumor-infiltrating immune cells (TIICs) to guide immunotherapy. Notably, cluster B exhibited higher levels of immune cell infiltration compared to cluster A (Figure 2I). Considering the critical role of immune checkpoints in tumor immunotherapy effectiveness and their prominence within the tumor microenvironment (TME), we evaluated immune checkpoint expression between the two patient clusters. The analysis revealed significantly upregulated immune checkpoint expression in Cluster B patients, except for TMIGD2 and CD44. Based on these findings, we conclude that Cluster B demonstrates a more favorable response and effectiveness toward immunotherapy (Figure 2J).

3.2 Development and validation of the BMRGs signature

We developed a risk score model based on BMRGs to identify prognostic biomarkers in UVM patients. Differentially expressed BMRGs with prognostic value were selected using LASSO regression analysis, and the resulting LASSO regression curves and cross-validation plots are shown in Figures 3A,B, respectively. To address batch effects between GSE22138 and GSE84976 datasets, we employed the R package "Combat" for batch effect removal (Figures 3C,D). The prognostic index (PI) was calculated using the

10.3389/fphar.2023.1264345

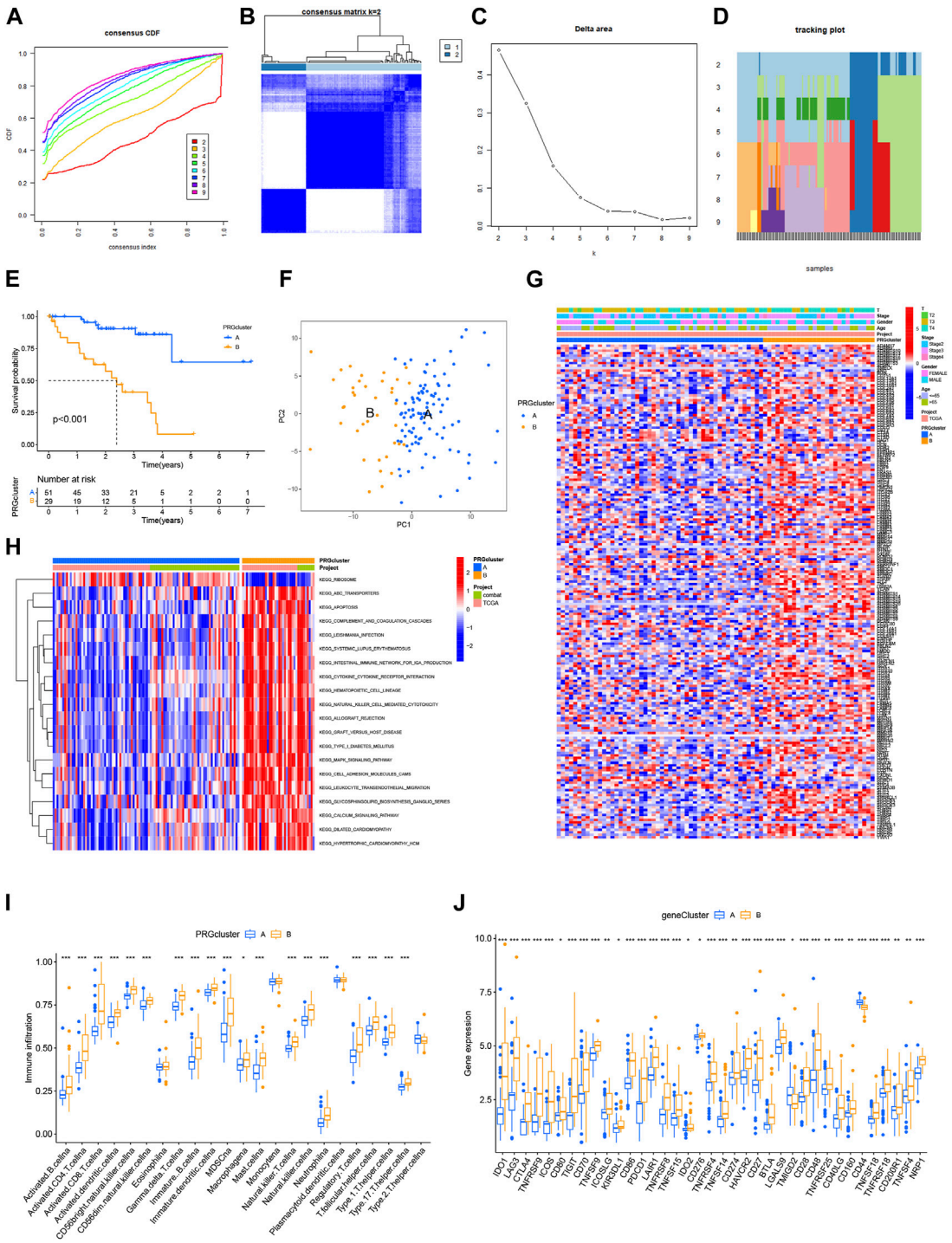


FIGURE 2
Consensus clustering determined the molecular subtypes of BMRGs. **(A)** Consensus clustering CDF with K = 2 to 9. **(B)** Consensus matrix heatmap for K = 2 clusters. **(C)** CDF plot illustrating the consensus clustering results for K = 2 to 9. **(D)** Tracking plot displaying the sample classification across K = 2 to 9 clusters. **(E)** Kaplan-Meier survival curves comparing the survival outcomes between Cluster A and Cluster B. **(F)** Principal component analysis (PCA) plot visualizing the distribution of samples. **(G)** Correlation analysis depicting the relationship between BMRGs expression and clinicopathological parameters. **(H)** Enrichment analysis of KEGG pathways in Cluster A and Cluster B. **(I)** Comparison of immune cell infiltration levels between clusters. **(J)** Differential expression of immune checkpoints between Cluster A and Cluster B. Statistical significance: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, ns > 0.05.

formula (−0.974 * ADAMTS10 exp.) + (1.015 * ADAMTS14 exp.) + (0.026 * CSF2 exp.) + (2.973 * ITGA5 exp.), and the risk score for each UVM patient was determined based on the median score using the equation. The optimal number of genes for cross-validation plots was 3, and the selected genes were ADAMTS10, ADAMTS14, and ITGA5.

10.3389/fphar.2023.1264345

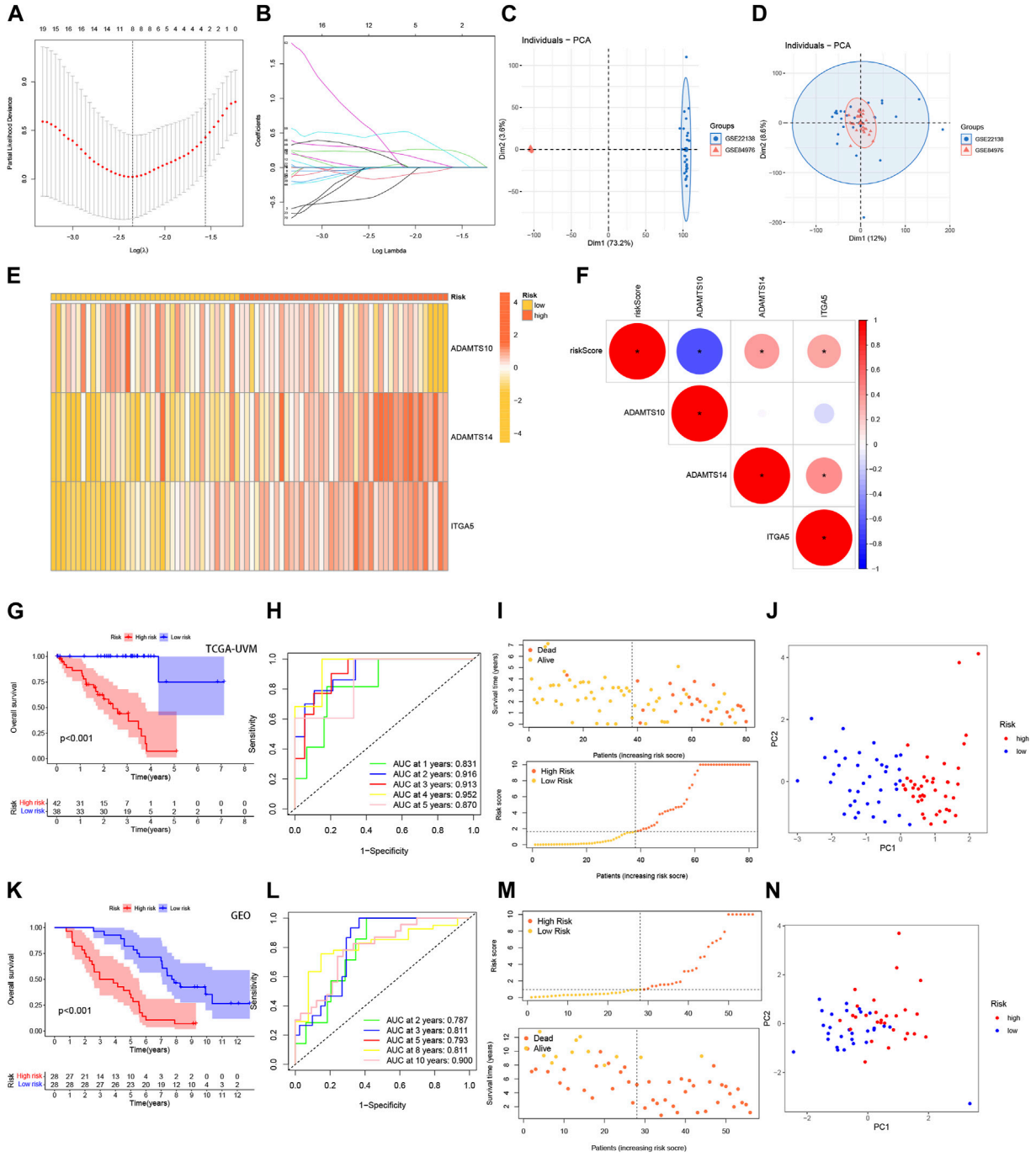


FIGURE 3
Development and validation of the BMRGs signature. **(A)** Ten-fold cross-validation for parameter selection using the LASSO model. **(B)** Profiles of LASSO coefficients. **(C)** Principal component analysis (PCA) plot of GSE22138. **(D)** PCA plot of GSE84976 after removing batch effects using Combat. **(E)** Heatmap illustrating the risk factors in high- and low-risk patients. **(F)** Correlation between three BMRGs and the risk score. **(G, K)** Kaplan-Meier curves comparing overall survival between low- and high-risk groups in the TCGA-UVM cohort and the GSE84976 cohort. **(H, L)** Time-dependent receiver operating characteristic (ROC) curves analysis of the TCGA-UVM cohort and the GSE84976 cohort. **(I, M)** Distribution of risk scores and survival status of UVM patients in the low- and high-risk groups in the TCGA-UVM cohort and the GSE84976 cohort. **(J, N)** PCA plot of the TCGA-UVM cohort and the GSE84976 cohort.

Further analysis revealed a strong correlation between the expression of the investigated BMRGs and the risk score. The risk score correlation heatmap (Figure 3E) and dot plot (Figure 3F) indicated that the expression levels of ADAMTS10 and ITGA5 were positively correlated with the risk score, while ADAMTS14 was negatively correlated. The TCGA-UVM cohort was used as the training set, and the de-batched GSE84976 dataset was used for validation. In the TCGA-UVM cohort, the low-risk group demonstrated significantly better prognostic outcomes ($p < 0.001$) (Figure 3G). The predictive

10.3389/fphar.2023.1264345

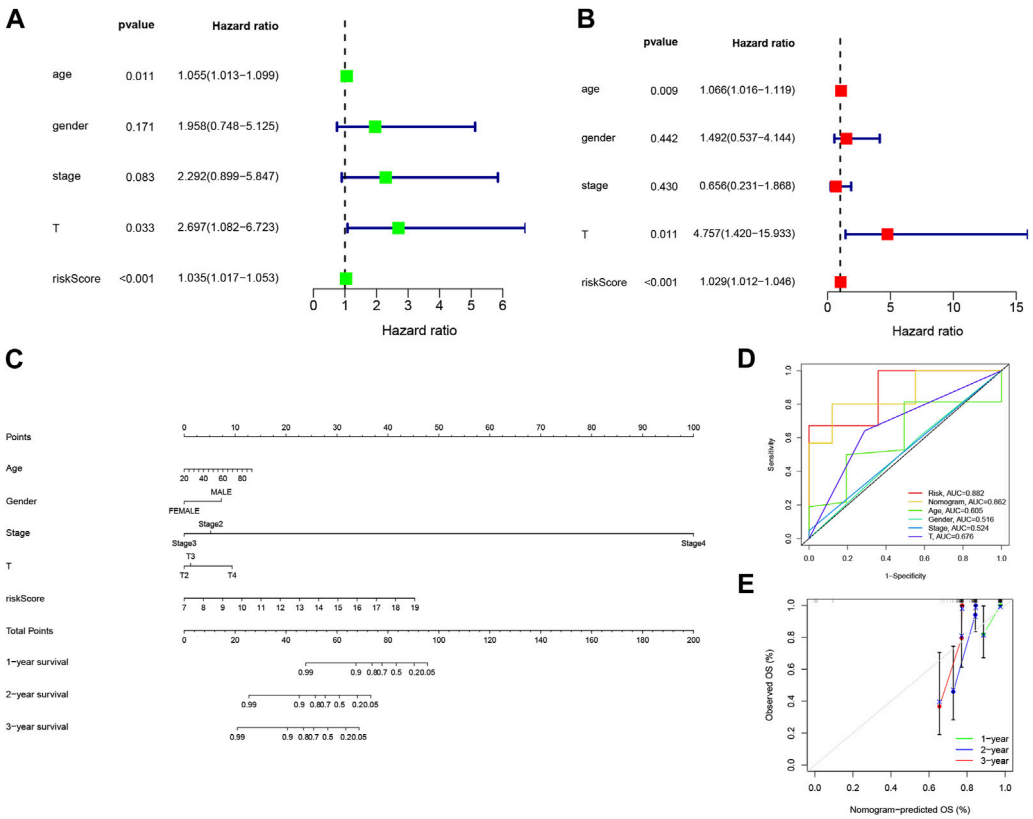


FIGURE 4
Building nomograms based on clinical characteristics. **(A)** Univariate Cox regression analysis of the signature and various clinical profiles. **(B)** Multivariate Cox regression analysis incorporating the signature and clinical characteristics. **(C)** Nomogram depicting age, gender, stage, T-stage, and risk score. **(D)** Calibration curves of the nomogram for 1-, 3-, and 5-year survival. **(E)** Time-dependent ROC curve.

model showed excellent performance as evidenced by the ROC curves, with high sensitivity and specificity reflected by the AUC values at 1, 2, 3, 4, and 5 years (0.831, 0.916, 0.913, 0.952, and 0.870) (Figure 3H). Moreover, there was an observed increase in mortality and a decrease in survival with higher risk scores (Figure 3I). Principal component analysis (PCA) clearly distinguished low-risk and high-risk patients from each other (Figure 3J). The results obtained in the GSE41613 cohort replicated those of the TCGA-UVM cohort (Figure 3K-N), indicating the reliability and consistency of our predictive model. In conclusion, our prediction model demonstrates high accuracy and reliability, providing valuable guidance for clinical management.

3.3 Construction of nomograms based on 3-BMRGs signatures with clinical features

Given the strong correlation between our constructed risk model and poor prognosis, we conducted univariate and multivariate Cox analyses to determine whether the prognostic characteristics based on the 3-BMRGs could serve as independent predictors of prognosis in UVM patients. In the univariate analysis, age ($p = 0.011$), T-stage ($p = 0.033$), and risk scores ($p < 0.001$) showed significant correlations with prognosis (Figure 4A). The subsequent multivariate analysis confirmed that age ($p = 0.009$), T-stage ($p = 0.011$), and risk scores ($p < 0.001$) remained accurate and independent predictors in this patient cohort (Figure 4B).

To enhance the clinical applicability and usability of the risk model, we developed Nomogram plots that incorporated age, gender, clinical stage, T-stage, and risk scores as predictors of survival probability at 1, 2, and 3 years for UVM patients. The risk score exhibited a substantial impact on predicting overall survival (OS), as demonstrated by the model analysis, indicating that the BMRGs-based risk model could provide more accurate prognostic predictions for UVM patients (Figure 4C). Additionally, we found that the risk score (AUC = 0.882) and Nomogram (AUC = 0.862) outperformed single independent clinical indicators in terms of predictive performance and discriminatory power (Figure 4D). Furthermore, the calibration analysis showed relatively consistent results between the predicted and observed 1-year, 3-year, and 5-year OS rates, as indicated by the calibration line closely aligning with the ideal 45-degree line (Figure 4E).

3.4 Clinical correlation and survival analysis of BMRGs in patients with UVM

A heat map was generated to visualize the correlation between the prognostic risk model identified using the 3-BMRGs and the clinical characteristics, risk scores, and expression levels of the 3-

10.3389/fphar.2023.1264345

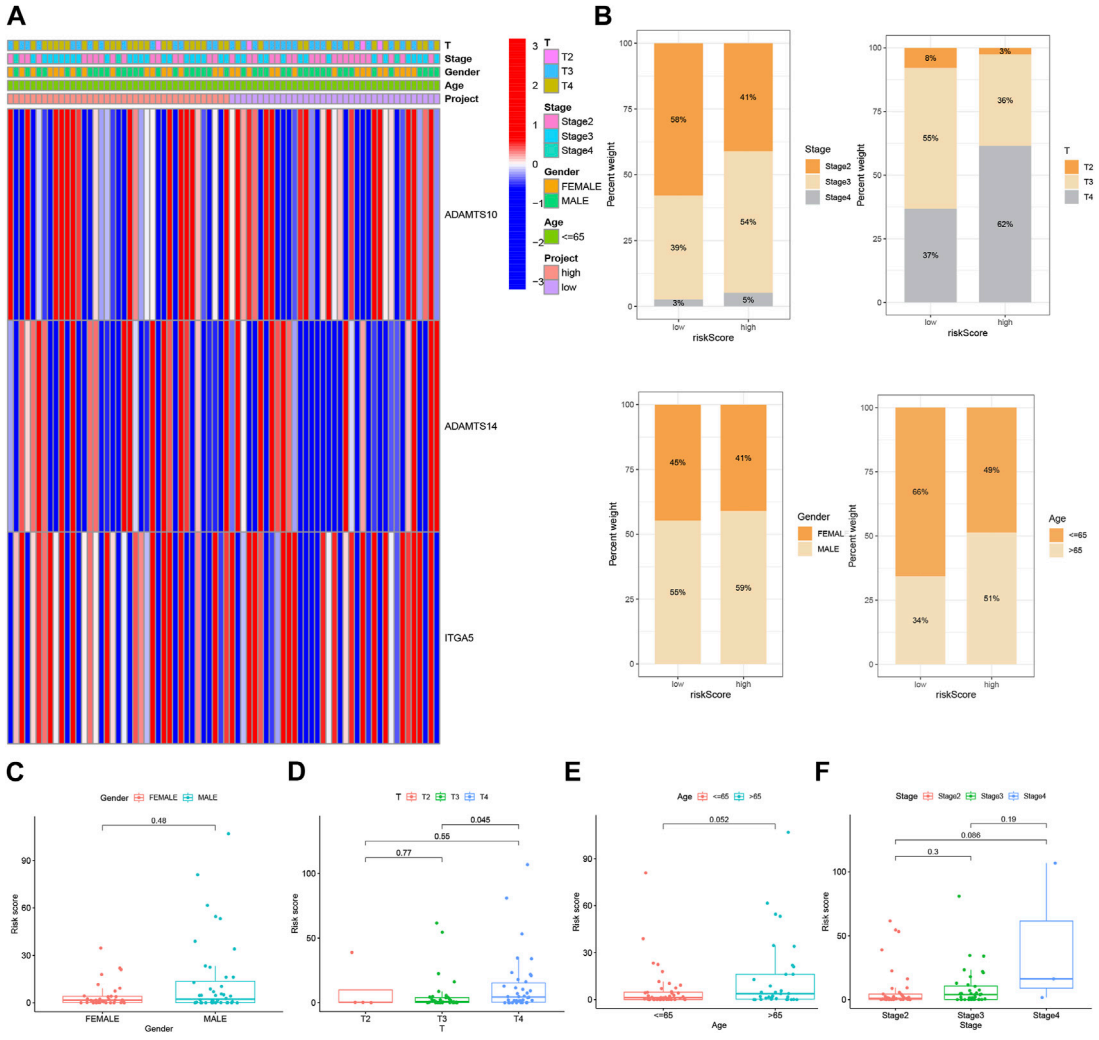


FIGURE 5
Clinical correlation and survival analysis of BMRGs in UVM patients. **(A)** Heatmap showing the relationship between clinical features and high-risk and low-risk scores in UVM patients. **(B)** Histogram presenting the distribution of clinical characteristics, including stage, T-stage, gender, and age percentages for each category. BMRGs can identify high-risk patients across different subgroups based on various clinicopathological traits. **(C)** Gender, **(D)** T-stage, **(E)** age, and **(F)** stage were analyzed for clinical correlation and survival in UVM patients.

BMRGs in all UVM patients from the TCGA dataset (Figure 5A). Additionally, we compared the distribution of patients with different clinicopathological features between the high-risk and low-risk groups (Figure 5B). To further examine the association between risk scores and clinicopathological characteristics, box plots were constructed for different subgroups based on gender (male and female), age (>65 and ≤65 years), clinical stage (II and III-IV), and T-stage (T2, T3, T4). Notably, the analysis revealed that patients with stage T4 had significantly higher risk scores compared to those with stage T3 (p = 0.045) (Figures 5C–F).

Considering the significant differences in individual clinical characteristics between the high-risk and low-risk groups for overall survival (OS), we further divided UVM patients into subgroups based on age (≤65 years, >65 years), gender (male and female), pathological stage (II and III-IV), and T-stage (T2 and T3-4). Remarkably, except for patients in stage T2, the low-risk subgroup exhibited a significant survival advantage with longer survival times compared to the high-risk subgroup (Figure 6). Based on these analyses, the 3-BMRGs risk model demonstrated its reliability as a clinical prediction tool for UVM patients.

3.5 3-BMRGs signatures exhibit superior performance compared to others in prognostic prediction

In order to assess the predictive performance of our BMRGs signature in UVM patients, we compared it with five previously published prognostic signatures, namely, the Xia signature, Xie signature, Zhang signature, Shi signature, and Hu signature. Using the same method, we calculated risk scores for each UVM sample in the entire TCGA cohort and found that our signature exhibited the highest correlation with survival outcomes (Figure 7AB, IJ). Despite successfully stratifying UVM patients

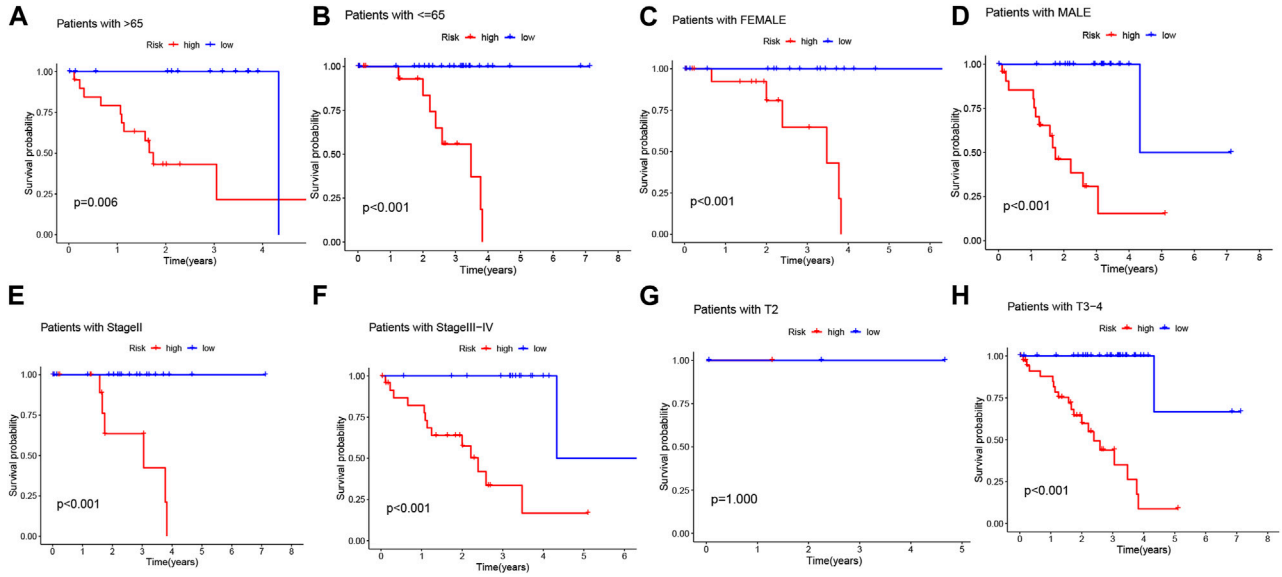


FIGURE 6
Clinical correlation and survival analysis of 3-BMRGs in UVM patients. **(A, B)** Age, **(C, D)** gender, **(E, F)** tumor grade, and **(G, H)** T-stage evaluated for clinical correlation and survival in UVM patients.

into two subgroups with significantly different prognoses, the AUC values of the five compared signatures at 1-, 3-, and 5-year survival were lower than those of our model (Figures 7E,F,K,L). Additionally, the C-index analysis demonstrated that our signature outperformed the other signatures (Figure 7M). Overall, our study indicates that our constructed BMRGs signature possesses excellent predictive ability in prognosticating UVM patients.

3.6 Functional enrichment analysis of DEGs in TCGA-UVM

In order to gain insights into the potential bioactivities and signaling pathways involved in UVM, and to understand the molecular mechanisms underlying UVM progression, we conducted Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis and gene ontology (GO) functional analysis. We applied stringent thresholds of FDR<0.05 and $p < 0.05$ to select significantly enriched items, as depicted in Figure 8A and Supplementary Table S1. The biological process (BP) analysis revealed enrichment in various processes such as rhythmic process, regulation of hormone levels, and circadian rhythm. Cellular component (CC) analysis highlighted correlations with neuronal cell body, presynaptic active zone, and terminal bouton, among others. Molecular function (MF) analysis indicated associations with functions such as G protein-coupled receptor binding, receptor-ligand activity, and signaling receptor activator. Furthermore, the KEGG enrichment analysis unveiled disease pathways including Circadian entrainment, Allograft rejection, and the Chemokine signaling pathway (Figure 8B). Moreover, the Gene Set Variation Analysis (GSVA) identified 50 significantly enriched pathways (Figure 8C). In-depth analysis revealed that in the low-risk population, pathways related to the Regulation of autophagy and RNA degradation were enriched. Conversely, in the high-risk population, pathway enrichment primarily involved immune and substance metabolism pathways such as leukocyte transendothelial migration and antigen processing and presentation. These findings contribute to our understanding of the molecular mechanisms underlying UVM and may provide valuable insights for the development of effective therapeutic strategies for UVM patients.

3.7 Risk score predicts TME and immune cell infiltration

Interactions between cancer cells and the tumor microenvironment (TME) play a crucial role in tumorigenesis, progression, and treatment outcomes. Tumor-infiltrating immune cells (TIICs) are integral components of the TME, and their distribution and alterations are closely associated with tumor progression. In this study, we investigated the relationship between risk scores and immune cell infiltration in the context of 3 BMRGs using seven algorithms: XCELL, TIMER, QUANTISEQ, MCPCOUNTER, CIBERSORT, CIBERSORT-ABS, and EPIC. Our results revealed a positive correlation between risk scores and the presence of T cell $CD8^+$ cells across multiple algorithms (Figure 9A). Furthermore, we analyzed the proportions of 22 immune cell infiltrates between the high-risk and low-risk groups of TCGA-UVM patients using the CIBERSORT algorithm. The results were visualized using stacked plots, demonstrating differences in immune cell composition between the two risk groups (Figure 9B). We also utilized the immune AI portal to assess immunotherapy response in UVM patients. Our analysis revealed that patients with higher risk

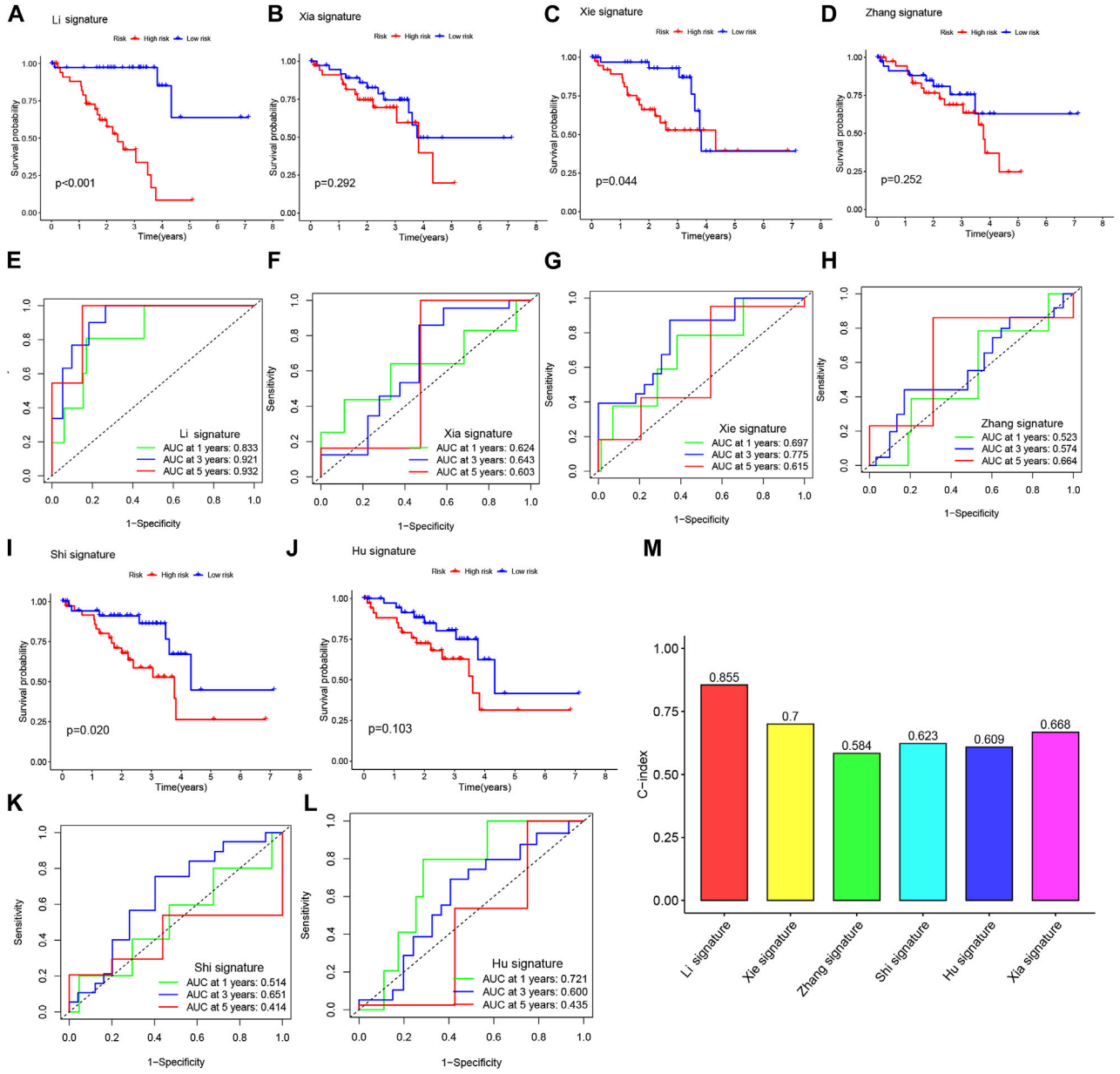


FIGURE 7
The BMRGs signature demonstrated superior prognostic prediction performance compared to other signatures. **(A, E)** Kaplan-Meier (KM) and receiver operating characteristic (ROC) curves of the BMRGs signature. **(B, F)** KM and ROC curves of the Xia signature. **(C, G)** KM and ROC curves of the Xie signature. **(D, H)** KM and ROC curves of the Zhang signature. **(I, K)** KM and ROC curves of the Shi signature. **(J, L)** KM and ROC curves of the Hu signature. **(M)** C-indexes of the six risk models.

scores were more likely to benefit from immunotherapy (Figures 9C,D), while those with lower risk scores exhibited a survival advantage. The ROC curves demonstrated the excellent performance of the 3-BMRGs biomarkers in predicting treatment outcomes for patients (Figure 9F). To further explore the immune profile of the tumor microenvironment, we plotted a correlation butterfly diagram to examine the relationship between risk scores and various steps of the tumor immune cycle. The analysis revealed a positive correlation between risk scores and most immune cycle steps, suggesting potential implications for immune modulation in UVM (Figure 9G).

3.8 Relationships between 3-BMRGs signatures and tumor microenvironment

In order to investigate the expression patterns of the 3-BMRGs in the tumor microenvironment, we utilized the cellular dataset UVM_GSE139829 obtained from the TISCH database. The distribution and numbers of 31 cell populations and 8 immune cell types in the UVM_GSE139829 dataset were analyzed and displayed (Figures 10A–D). Furthermore, we examined the expression of the 3-BMRGs in different immune cell populations. The expression of BMRGs was found to be lower in the

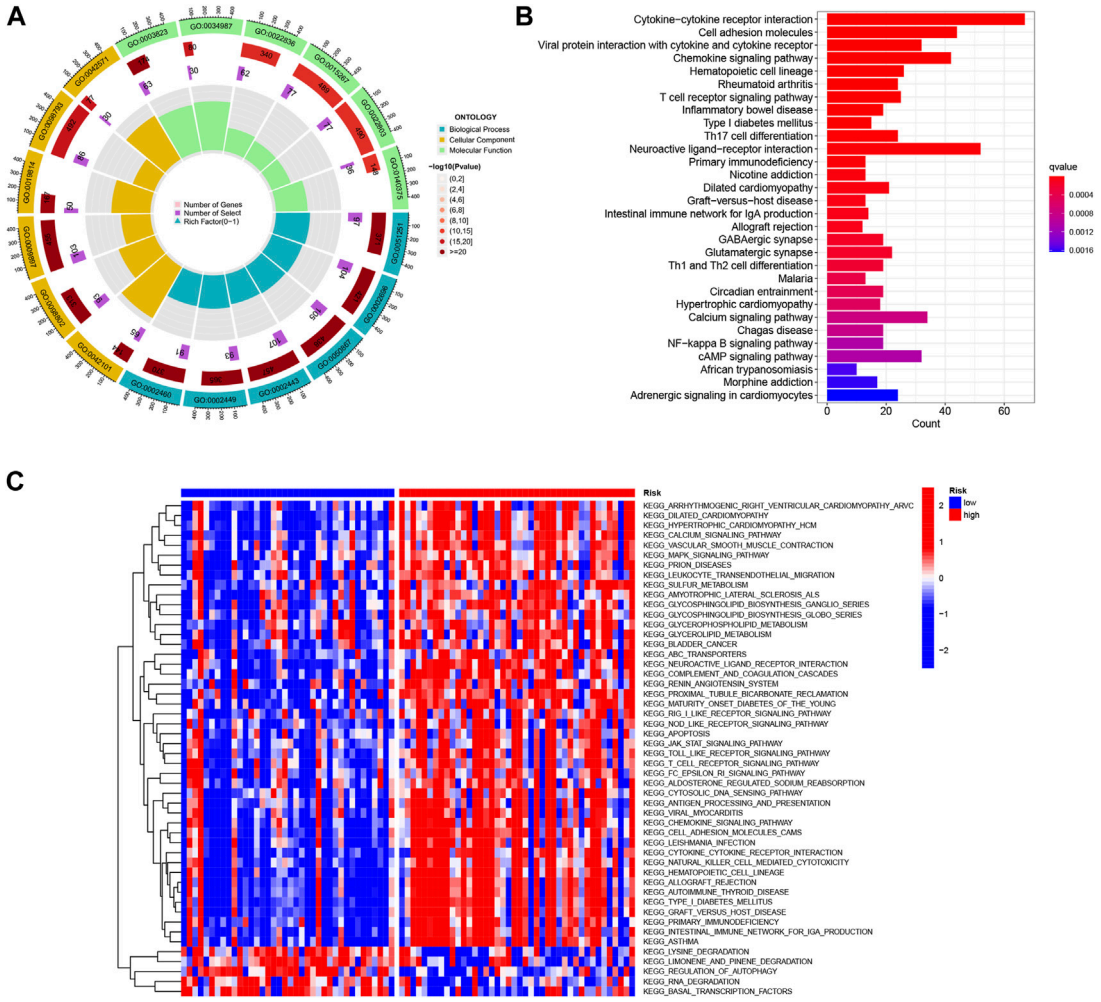


FIGURE 8
Functional enrichment analysis of Differentially Expressed Genes (DEGs) in TCGA-UVM was performed. **(A)** Gene Ontology (GO) enrichment analysis was conducted to investigate the differences in basement membrane x genes between UVM and normal samples. The analysis included biological processes (BP), cellular components (CC), and molecular functions (MF). **(B)** KEGG enrichment analysis was performed to identify enriched pathways associated with the DEGs. **(C)** Gene Set Variation Analysis (GSVA) was utilized to compare the enrichment scores between high-risk and low-risk cohorts, providing insights into the functional differences between these groups.

ADAMTS14 immune microenvironment (Figure 10E). On the other hand, ADAMTS10 and ITGA5 were expressed in various immune cell populations, as demonstrated in Figures 10F,G, respectively. Notably, ITGA5 showed predominant expression in CD8 Tex, Mono/Macro, and CD8T immune cell populations. These findings provide insights into the expression patterns of the 3-BMRGs within the immune cell landscape of UVM.

3.9 ITGA5 facilitates the proliferation, migration, and invasion of uveal melanoma (UVM) cells

Considering the identification of ITGA5 as a high-risk gene exhibiting a maximum absolute hazard ratio (HR) in uveal melanoma (UVM) patients, we conducted additional *in vitro* experiments to elucidate the specific role of ITGA5 in UVM. Knockdown systems targeting ITGA5 were established in OCM-1 and MUM-2C cell lines. The CCK-8 assay, and colony formation assay demonstrated a significant reduction in the proliferation rate of UVM cells following ITGA5 silencing (Figures 11A,B). Moreover, both the Transwell assay and the wound healing assay revealed diminished migration and invasiveness of UVM cells after ITGA5 knockdown, in comparison to cells transfected with si-NC (Figures 11C,D). Collectively, these findings provide evidence that ITGA5 functions as an oncogene, promoting the malignant characteristics of UVM cells, including proliferation, invasion, and migration.

4 Discussion

Although the prevalence of UVM is not extremely high, it accounts for 85% of all ocular melanomas, with up to 50% of patients of primary UVM developing distant metastases, 90% with liver damage, and a median survival of 4–5 months

10.3389/fphar.2023.1264345

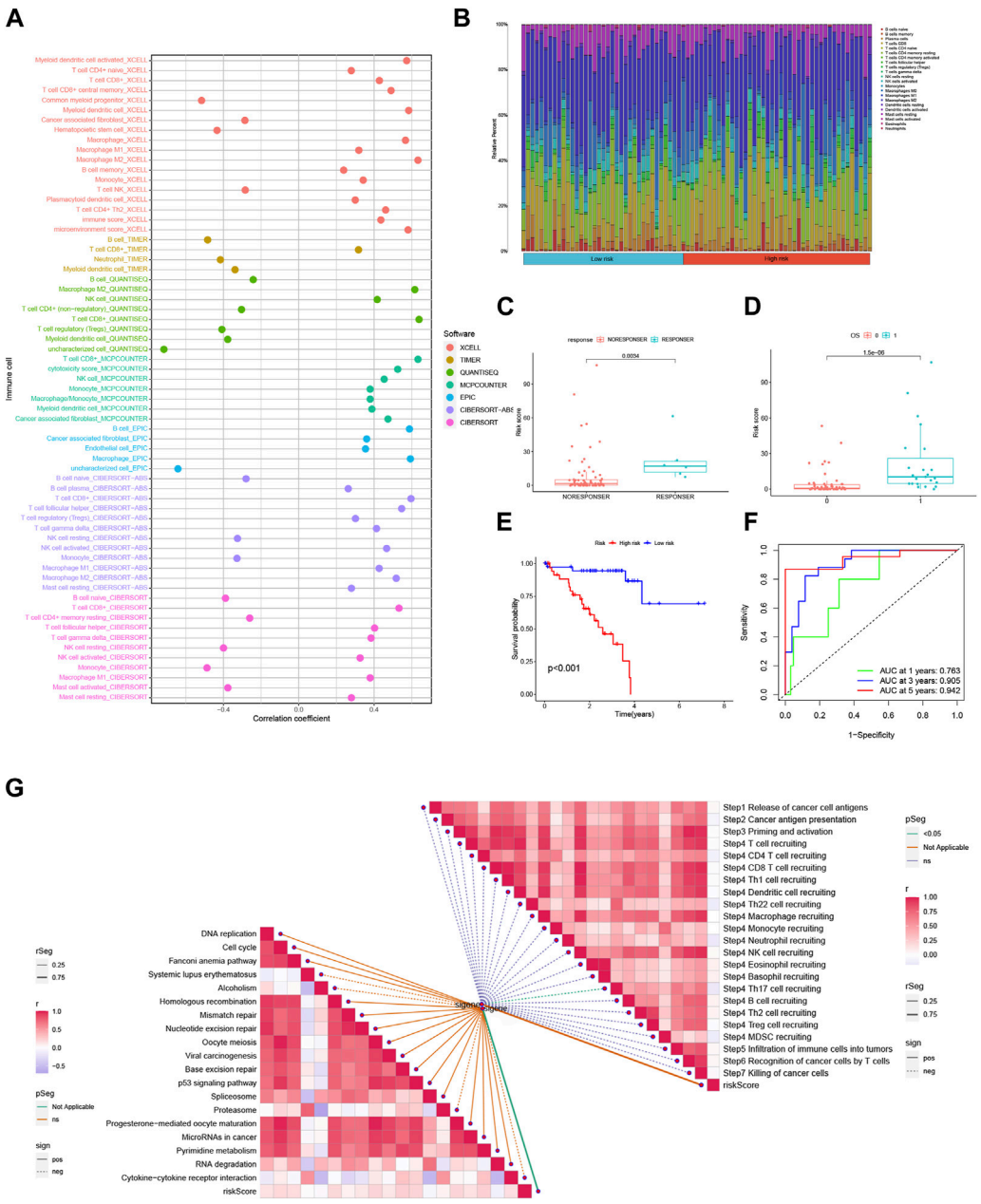


FIGURE 9
The risk score is predictive of the tumor microenvironment (TME) and immune cell infiltration. **(A)** An immune cell bubble plot was generated to visualize the composition of immune cell types. **(B)** A stacked plot illustrates the differences in immune cell infiltration between the high-risk and low-risk groups. **(C, D)** The expression levels of the 3-BMRGs were utilized to predict the response of patients to immune therapy. **(E)** Kaplan-Meier (KM) curves compare the survival outcomes between the high-risk and low-risk groups after receiving immunotherapy. **(F)** The receiver operating characteristic (ROC) curve analysis demonstrates the robust predictive performance of the marker model. (H) The correlation between risk scores and immune checkpoint blockade (ICB) response characteristics was examined. **(G)** The correlation between risk scores and each step of the tumor immunization cycle was investigated.

(Baggetto et al., 2005; Tabernero, 2007; Carvajal et al., 2017; Brouwer et al., 2019). Additionally, the metastatic rates for UVM throughout the course of 5 and 10 years are roughly 25% and 34%, respectively, and the mortality rate for UVM 1 year after metastasis is 80%. After diagnosis, the majority of patients with metastatic UVM have a survival span of 6–12 months for which metastatic UVM is virtually always challenging to treat. Due to the poor prognosis of UVM, few individuals can receive possibly curative surgery (Singh et al., 2005; Straatsma et al., 2018). Meanwhile, 5-year survival rates have remained essentially unchanged over the past 3 decades despite the development of efficient local therapy. There are currently no effective adjuvant systemic medications that have been proven to lower the risk of metastasis as well as actually extend survival, according to a recent review study (Triozzi and Singh, 2014).

Actually, early diagnosis and therapy are crucial to improving prognosis, while UVM diagnosis and prognosis prediction are

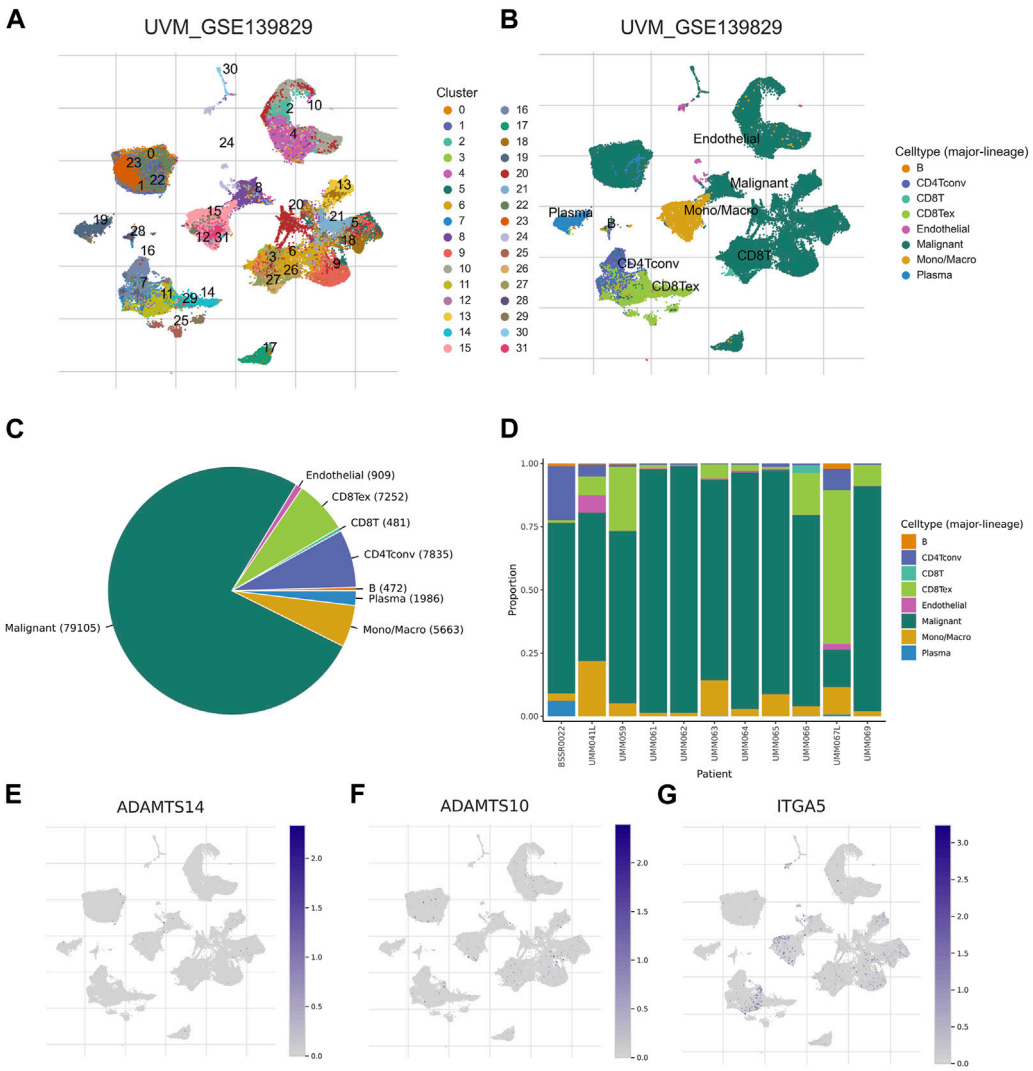


FIGURE 10
Association of BMRGs with the tumor microenvironment. Annotation of **(A)** 31 cell clusters and **(B)** 8 cell types in UVM_GSE139829. **(C, D)** The distribution and proportion of each cell type, including CD8T cells, endothelial cells, B cells, *etc.* Expression and percentage of **(E)** ADAMTS14, **(F)** ADAMTS10, **(G)** ITGA5.

currently reliant on clinical presentation and histological examination, which are insufficient to identify tumor heterogeneity and developmental patterns. Therefore, using the TCGA-UVM dataset, this study constructed a multigene prognostic model of genes related to basement membrane proteins from a molecular perspective in order to better predict the diagnosis and prognosis of UVM. It opens up new avenues for the investigation of individualized treatment strategies and prognosis prediction.

Defects in BM protein expression and turnover play a major role in the development of cancer, fibrosis, and diabetes (Tsilibary, 2003; Naba et al., 2014; Foster, 2017). Specifically, the overexpression of laminin, a component of BM protein, is closely associated with the overproliferation of certain tumor cells, such as those found in colon and breast cancer (Jayadev and Sherwood, 2017). Moreover, BM is significantly involved in the progression of tumors. During the early development of breast cancer, cancer cells invade through the BM foramen, which is a crucial step in metastasis (Sikic et al., 2022). In addition, the level of netrin-4 in BM is highly correlated with the prognosis of breast cancer, renal cancer, and uveal melanoma (Reuten et al., 2021). Several recent studies have already attempted to mine public databases to identify prognostic gene signatures related to BM proteins in tumors. For example, Cai et al. identified a 7-gene signature associated with basement membrane proteins that predicted the prognostic status of breast cancer patients and provided insights for immunotherapy (Cai et al., 2022); Zhou et al. developed a risk model using 8 BMRGs, which revealed that clear-cell renal cell carcinoma patients in the low-risk group had a better response to immunotherapy (Zhou et al., 2022); Lin et al. established a 7-BMRG signature and identified five small compounds that could potentially be used for the treatment of pancreatic cancer patients, providing new perspectives for a deeper understanding of this disease (Lin et al., 2023). Overall, these studies highlight the significance of basement membrane proteins in the precise treatment of tumors.

10.3389/fphar.2023.1264345

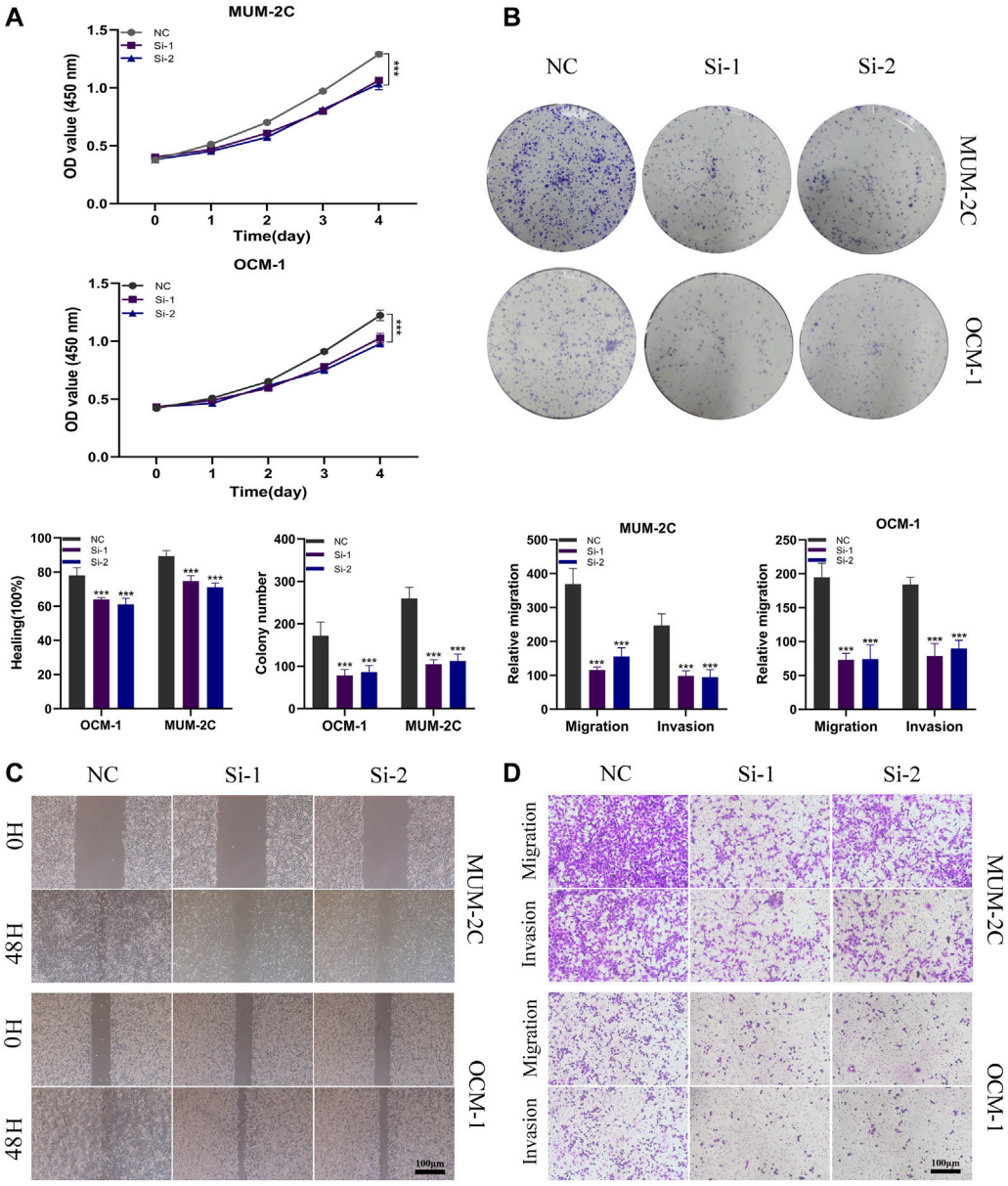


FIGURE 11
ITGA5 facilitates the proliferation, migration, and invasion of UVM cells. **(A)** CCK-8 assay showed that the proliferative capacity of UVM cells was significantly reduced after silencing of ITGA5. **(B)** Colony formation assays showed that the ability of UVM cells to form colonies was significantly reduced after ITGA5 silencing. **(C, D)** In wound healing and transwell assays, silencing of ITGA5 significantly reduced the migratory and invasive capacity of MuM-2B and OCM-1 cell lines.

In our study, we selected three basement membrane related-genes (ADAMTS10, ADAMTS14, and ITGA5) to create the novel prognostic model by utilizing Lasso regression analysis, SVM-RFE, and stepwise multiple COX regression analysis. Numerous studies have proven that the ADAMTS (a disintegrin and metalloproteinase with thrombospondin motif) family of proteins contributes to the development of malignant tumors, cell proliferation, apoptosis, migration, invasion, and angiogenesis (Held-Feindt et al., 2006; Rocks et al., 2006; Filou et al., 2015; Sun et al., 2015). ADAMTSs have a negative impact on the prognosis of patients with 24 tumors, in particular the patients with Adrenocortical carcinoma, Uveal Melanoma, Kidney renal clear cell carcinoma, Colon adenocarcinoma, Thyroid carcinoma, *etc.* (Wu et al., 2021). Although the precise mechanism by which ADAMTS play a role in tumor progression and metastasis remains uncertain, several research has explored these protein hydrolases and confirmed their relevance in various tumor types. ADAMTS10 expression is significantly downregulated in tumors (Sun et al., 2015). Furthermore, a variety of ADAMTSs, including ADAMTS20, ADAMTS10, and ADAMTS3, exhibit significant levels of methylation in a range of tumors. Analysis of the relationship between methylation and gene expression levels reveals a

negative relationship between the two, suggesting that the main function of ADAMTS methylation is to silence the ADAMTS gene, leading to a decrease in its expression (Wu et al., 2021). Besides, ADAMTS10 is frequently mutated in metastatic colorectal cancer, and mutated ADAMTS10 transcripts are actively expressed in the corresponding tumors implicating a possible role for ADAMTS10 in tumor metastasis (Oga et al., 2019). According to numerous research, the ADAMTS14 gene has been associated with an elevated likelihood of developing tumors. The expression of ADAMTS14 was identified to be considerably higher in human breast cancer tissues, according to Porter et al. (2004). As reported by Sheu et al., ADAMTS14 gene polymorphisms serve a part in the progression of hepatocellular carcinoma (Sheu et al., 2017). The expression of ADAMTS14 in oral squamous carcinoma (OSCC) is low. In OSCC patients, the downregulation of ADAMTS14 may be an effective independent prognostic marker for predicting overall survival because it is predictive of unfavorable clinicopathological characteristics (Lin et al., 2020). Furthermore, there is mounting evidence that Integrin A5 (ITGA5), which plays a major role in the adhesion, migration, and invasion of cancer cells, is highly expressed in several malignancies and contributes to tumor progression (Ohyagi-Hara et al., 2013). One of the markers of invasiveness in head and neck squamous cell carcinoma has been identified as ITGA5 (Yu et al., 2008). Additionally, a study demonstrated that pancreatic ductal malignant adenomas upregulate the ITGA5 gene. Silencing of ITGA5 inhibits the differentiation of human pancreatic stellate cells and reduces connective tissue formation (Kuninty et al., 2019). Furthermore, ITGA5 promotes the development, migration, and invasion of cells that undergo an epithelial-mesenchymal transition in oral cancer (Deng et al., 2019).

The 3-BMRGs we constructed proved to be an independent prognostic factor for UVM. Based on median risk ratings, patients with UVM were separated into high-risk and low-risk groups; there were notable prognostic differences between the two groups. The 3-BMRGs that we created turned out to be a reliable indicator of UVM's future. Based on median risk ratings, individuals with UVM were separated into high-risk and low-risk groups; there were notable prognostic differences between the two groups. Additionally, evaluations of the ROC and calibration curves revealed that the 3-BMRGs signature had excellent predictive power. To extend the predictive ability of the 3-BMRGs signature and to demonstrate its utility in the prognostic evaluation of UVM patients, we plotted a line graph based on clinical factors and risk scores. Meanwhile, we discovered that the 3-BMRGs signature has better predictive power than clinicopathological features, which could offer clinicians a basis for decision-making.

The tumor microenvironment (TME) is crucial to the metastasis and progression of cancer (Zhao et al., 2022c; Gong et al., 2022; Xiong et al., 2023). The TME comprises cancer cells, surrounding stromal cells, and tumor-infiltrating immune cells, with immune cells playing a dominant role in the TME (Hinshaw and Shevde, 2019; Shen et al., 2022). Through the strengthening of a weakened immune response to tumor cells and the resultant production of an immunological-mediated anti-tumor impact, immunotherapy has made significant strides in the treatment of cancers in recent years (Zhang et al., 2021). As a result, we assessed immune checkpoint expression and discovered that, with the exception of TMIGD2 and CD44, it was highly elevated in the high-risk group of UVM patients. Immune infiltration is closely related to immunotherapy's efficacy (Nishida and Kudo, 2020). The high-risk group exhibited higher levels of immune cell infiltration, which suggested that they responded more favorably to immunotherapy, according to the ssGSEA enrichment score. Meanwhile, we performed GO and KEGG enrichment analysis to provide more light on the biological pathways and putative molecular mechanisms associated with the BMRG signature. We noted that in the high-risk population, pathway enrichment mainly involved immune and substance metabolism pathways, such as leukocyte transendothelial migration, antigen processing, and presentation. In contrast, pathways related to the regulation of autophagy and RNA degradation were enriched in the low-risk population.

Immune checkpoint blockade has shown significant benefits in the treatment of malignant tumors. Nevertheless, its non-response rate and side effects have posed challenges in clinical practice (Hodi et al., 2010; Kennedy and Salama, 2020; Su et al., 2022). Thus, it is critical to identify individuals who are responsive to different immune checkpoint medicines based on the expression of immune checkpoint genes. Our model has demonstrated excellent results in this regard. In our study, we found that UVM patients with higher risk scores were more likely to benefit from immunotherapy, while patients with lower risk scores could have a higher survival advantage. Furthermore, the TME is closely correlated to the risk model that we built based on the BMRGs.

Our study suffers from the following limitations. Firstly, our study is retrospective, and based on data analysis in public databases with limited inclusion of UVM patients which still demands more clinical data and prospective studies to validate the model and improve the credibility of risk scores; Meanwhile, the extrapolation of our findings is limited due to the possible inherent bias and limitations of the TCGA-UVM and GEO cohorts themselves; In addition, the mechanisms by which BMRGs affect the prognosis of UVM patients are required to be further explored in more *in vivo* experiments.

5 Conclusion

To sum up, we have developed a model of the BMRG prognostic signatures including ADAMTS10, ADAMTS14, and ITGA5. Two external validation cohorts were employed to verify the reliability and applicability of the BMRGs scores. This constructed model exhibited robust predictive ability which could act as an independent prognostic factor for UVM, assisting clinicians to identify specific subgroups of patients who may benefit from immunotherapy and chemotherapy, and providing a novel strategy for individualized treatment of UVM patients.

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.

Author contributions

YL: Writing–original draft. HC: Validation, Writing–original draft. JY: Data curation, Visualization, Writing–original draft. XX: Data curation, Visualization, Writing–original draft. SP: Validation, Writing–original draft. YW: Writing–original draft. JZ: Writing–original draft. GS: Writing–original draft. JZ: Writing–original draft. QZ: Writing–review and editing. HC: Conceptualization, Data curation, Visualization, Writing–original draft, Writing–review and editing, Validation. GY: Conceptualization, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

All authors acknowledge the contributions from the TCGA and GEO project.

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

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2023.1264345/full#supplementary-material

References

Andreoli, M. T., Mieler, W. F., and Leiderman, Y. I. (2015). Epidemiological trends in uveal melanoma. *Br. J. Ophthalmol.* 99 (11), 1550–1553. doi:10.1136/bjophthalmol-2015-306810

Augsburger, J. J., Corrêa, Z. M., and Shaikh, A. H. (2009). Effectiveness of treatments for metastatic uveal melanoma. *Am. J. Ophthalmol.* 148 (1), 119–127. doi:10.1016/j.ajo.2009.01.023

Baggetto, L. G., Gambrelle, J., Dayan, G., Labialle, S., Barakat, S., Michaud, M., et al. (2005). Major cytogenetic aberrations and typical multidrug resistance phenotype of uveal melanoma: Current views and new therapeutic prospects. *Cancer Treat. Rev.* 31 (5), 361–379. doi:10.1016/j.ctrv.2005.05.001

Banerjee, S., Lo, W. C., Majumder, P., Roy, D., Ghorai, M., Shaikh, N. K., et al. (2022). Multiple roles for basement membrane proteins in cancer progression and EMT. *Eur. J. Cell. Biol.* 101 (2), 151220. doi:10.1016/j.ejcb.2022.151220

Bhattacherjee, A., Djekidel, M. N., Chen, R., Chen, W., Tuesta, L. M., and Zhang, Y. (2019). Cell type-specific transcriptional programs in mouse prefrontal cortex during adolescence and addiction. *Nat. Commun.* 10 (1), 4169. doi:10.1038/s41467-019-12054-3

Bol, K. F., van den Bosch, T., Schreibelt, G., Mensink, H. W., Keunen, J. E., Kiliç, E., et al. (2016). Adjuvant dendritic cell vaccination in high-risk uveal melanoma. *Ophthalmology* 123 (10), 2265–2267. doi:10.1016/j.ophtha.2016.06.027

Brouwer, N. J., Gezgin, G., Wierenga, A. P. A., Bronkhorst, I. H. G., Marinkovic, M., Luyten, G. P. M., et al. (2019). Tumour angiogenesis in uveal melanoma is related to genetic evolution. *Cancers (Basel)* 11 (7), 979. doi:10.3390/cancers11070979

Cai, J., Zhang, X., Xie, W., Li, Z., Liu, W., and Liu, A. (2022). Identification of a basement membrane-related gene signature for predicting prognosis and estimating the tumor immune microenvironment in breast cancer. *Front. Endocrinol. (Lausanne)* 13, 1065530. doi:10.3389/fendo.2022.1065530

Carvajal, R. D., Schwartz, G. K., Tezel, T., Marr, B., Francis, J. H., and Nathan, P. D. (2017). Metastatic disease from uveal melanoma: Treatment options and future prospects. *Br. J. Ophthalmol.* 101 (1), 38–44. doi:10.1136/bjophthalmol-2016-309034

Chattopadhyay, C., Kim, D. W., Gombos, D. S., Oba, J., Qin, Y., Williams, M. D., et al. (2016). Uveal melanoma: From diagnosis to treatment and the science in between. *Cancer* 122 (15), 2299–2312. doi:10.1002/cncr.29727

Chi, H., Peng, G., Yang, J., Zhang, J., Song, G., Xie, X., et al. (2022). Machine learning to construct sphingolipid metabolism genes signature to characterize the immune landscape and prognosis of patients with uveal melanoma. *Front. Endocrinol. (Lausanne).* 13, 1056310. doi:10.3389/fendo.2022.1056310

Chi, H., Gao, X., Xia, Z., Yu, W., Yin, X., Pan, Y., et al. (2023a). FAM family gene prediction model reveals heterogeneity, stemness and immune microenvironment of UCEC. *Front. Mol. Biosci.* 10, 1200335. doi:10.3389/fmolb.2023.1200335

Chi, H., Chen, H., Wang, R., Zhang, J., Jiang, L., Zhang, S., et al. (2023b). Proposing new early detection indicators for pancreatic cancer: Combining machine learning and neural networks for serum miRNA-based diagnostic model. *Front. Oncol.* 13, 1244578. doi:10.3389/fonc.2023.1244578

Curran, M. A., Montalvo, W., Yagita, H., and Allison, J. P. (2010). PD-1 and CTLA-4 combination blockade expands infiltrating T cells and reduces regulatory T and myeloid cells within B16 melanoma tumors. *Proc. Natl. Acad. Sci. U. S. A.* 107 (9), 4275–4280. doi:10.1073/pnas.0915174107

Damato, B. (2018). Ocular treatment of choroidal melanoma in relation to the prevention of metastatic death - a personal view. *Prog. Retin Eye Res.* 66, 187–199. doi:10.1016/j.preteyeres.2018.03.004

Deng, Y., Wan, Q., and Yan, W. (2019). Integrin α5/ITGA5 promotes the proliferation, migration, invasion and progression of oral squamous carcinoma by epithelial-mesenchymal transition. *Cancer Manag. Res.* 11, 9609–9620. doi:10.2147/CMAR.S223201

Derrien, A. C., Rodrigues, M., Eeckhoutte, A., Dayot, S., Houy, A., Mobuchon, L., et al. (2021). Germline MBD4 mutations and predisposition to uveal melanoma. *J. Natl. Cancer Inst.* 113 (1), 80–87. doi:10.1093/jnci/djaa047

Engebretsen, S., and Bohlin, J. (2019). Statistical predictions with glmnet. *Clin. Epigenetics* 11 (1), 123. doi:10.1186/s13148-019-0730-1

Filou, S., Korpetinou, A., Kyriakopoulou, D., Bounias, D., Stavropoulos, M., Ravazoula, P., et al. (2015). ADAMTS expression in colorectal cancer. *PLoS One* 10 (3), e0121209. doi:10.1371/journal.pone.0121209

Foster, M. H. (2017). Basement membranes and autoimmune diseases. *Matrix Biol.* 57-58, 149–168. doi:10.1016/j.matbio.2016.07.008

Gong, X., Chi, H., Strohmer, D. F., Teichmann, A. T., Xia, Z., and Wang, Q. (2022). Exosomes: A potential tool for immunotherapy of ovarian cancer. *Front. Immunol.* 13, 1089410. doi:10.3389/fimmu.2022.1089410

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

Held-Feindt, J., Paredes, E. B., Blömer, U., Seidenbecher, C., Stark, A. M., Mehdorn, H. M., et al. (2006). Matrix-degrading proteases ADAMTS4 and ADAMTS5 (disintegrins and metalloproteinases with thrombospondin motifs 4 and 5) are expressed in human glioblastomas. *Int. J. Cancer* 118 (1), 55–61. doi:10.1002/ijc.21258

Hinshaw, D. C., and Shevde, L. A. (2019). The tumor microenvironment innately modulates cancer progression. *Cancer Res.* 79 (18), 4557–4566. doi:10.1158/0008-5472.CAN-18-3962

Hodi, F. S., O'Day, S. J., McDermott, D. F., Weber, R. W., Sosman, J. A., Haanen, J. B., et al. (2010). Improved survival with ipilimumab in patients with metastatic melanoma. *N. Engl. J. Med.* 363 (8), 711–723. doi:10.1056/NEJMoa1003466

Huang, X., Chi, H., Gou, S., Guo, X., Li, L., Peng, G., et al. (2023). An aggrephagy-related LncRNA signature for the prognosis of pancreatic adenocarcinoma. *Genes. (Basel)* 14 (1), 124. doi:10.3390/genes14010124

Jayadev, R., and Sherwood, D. R. (2017). Basement membranes. *Curr. Biol.* 27 (6), R207–R211. doi:10.1016/j.cub.2017.02.006

Jin, W., Yang, Q., Chi, H., Wei, K., Zhang, P., Zhao, G., et al. (2022). Ensemble deep learning enhanced with self-attention for predicting immunotherapeutic responses to cancers. *Front. Immunol.* 13, 1025330. doi:10.3389/fimmu.2022.1025330

Katopodis, P., Khalifa, M. S., and Anikin, V. (2021). Molecular characteristics of uveal melanoma and intraocular tumors. *Oncol. Lett.* 21 (1), 9. doi:10.3892/ol.2020.12270

Kennedy, L. B., and Salama, A. K. S. (2020). A review of cancer immunotherapy toxicity. *CA Cancer J. Clin.* 70 (2), 86–104. doi:10.3322/caac.21596

Kuninty, P. R., Bansal, R., De Geus, S. W. L., Mardhian, D. F., Schnittert, J., van Baarlen, J., et al. (2019). ITGA5 inhibition in pancreatic stellate cells attenuates desmoplasia and potentiates efficacy of chemotherapy in pancreatic cancer. *Sci. Adv.* 5 (9), eaax2770. doi:10.1126/sciadv.aax2770

Larkin, J., Chiarion-Sileni, V., Gonzalez, R., Grob, J. J., Cowey, C. L., Lao, C. D., et al. (2015). Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N. Engl. J. Med.* 373 (1), 23–34. doi:10.1056/NEJMoa1504030

Li, S., Edgar, D., Fässler, R., Wadsworth, W., and Yurchenco, P. D. (2003). The role of laminin in embryonic cell polarization and tissue organization. *Dev. Cell.* 4 (5), 613–624. doi:10.1016/s1534-5807(03)00128-x

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell. Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Lin, Y. M., Lin, C. W., Lu, J. W., Yeh, K. T., Lin, S. H., and Yang, S. F. (2020). Decreased cytoplasmic expression of ADAMTS14 is correlated with reduced survival rates in oral squamous cell carcinoma patients. *Diagn. (Basel).* 10 (2), 122. doi:10.3390/diagnostics10020122

Lin, K., Xu, D., Wang, X., Shi, J., and Gao, W. (2023). Development of a basement membrane gene signature and identification of the potential candidate therapeutic targets for pancreatic cancer. *Gland. Surg.* 12 (2), 263–281. doi:10.21037/gs-23-24

Lu, Y., Liu, L., Pan, J., Luo, B., Zeng, H., Shao, Y., et al. (2021). MFG-E8 regulated by miR-99b-5p protects against osteoarthritis by targeting chondrocyte senescence and macrophage reprogramming via the NF-κB pathway. *Cell. Death Dis.* 12 (6), 533. doi:10.1038/s41419-021-03800-x

Lv, Y., He, L., Jin, M., Sun, W., Tan, G., and Liu, Z. (2022). EMT-related gene signature predicts the prognosis in uveal melanoma patients. *J. Oncol.* 2022, 5436988. doi:10.1155/2022/5436988

Mao, S., Yu, X., Shan, Y., Fan, R., Wu, S., and Lu, C. (2021). Albumin-bilirubin (ALBI) and monocyte to lymphocyte ratio (MLR)-Based nomogram model to predict tumor recurrence of AFP-negative hepatocellular carcinoma. *J. Hepatocell. Carcinoma* 8, 1355–1365. doi:10.2147/JHC.S339707

Miao, Y. R., Zhang, Q., Lei, Q., Luo, M., Xie, G. Y., Wang, H., et al. (2020). ImmuCellAI: A unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy. *Adv. Sci. (Weinh)* 7 (7), 1902880. doi:10.1002/advs.201902880

Naba, A., Clauser, K. R., Whittaker, C. A., Carr, S. A., Tanabe, K. K., and Hynes, R. O. (2014). Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver. *BMC Cancer* 14, 518. doi:10.1186/1471-2407-14-518

Nishida, N., and Kudo, M. (2020). Immune phenotype and immune checkpoint inhibitors for the treatment of human hepatocellular carcinoma. *Cancers (Basel)* 12 (5), 1274. doi:10.3390/cancers12051274

Oga, T., Yamashita, Y., Soda, M., Kojima, S., Ueno, T., Kawazu, M., et al. (2019). Genomic profiles of colorectal carcinoma with liver metastases and newly identified fusion genes. *Cancer Sci.* 110 (9), 2973–2981. doi:10.1111/cas.14127

Ohyagi-Hara, C., Sawada, K., Kamiura, S., Tomita, Y., Isobe, A., Hashimoto, K., et al. (2013). miR-92a inhibits peritoneal dissemination of ovarian cancer cells by inhibiting integrin α5 expression. *Am. J. Pathol.* 182 (5), 1876–1889. doi:10.1016/j.ajpath.2013.01.039

Porter, S., Scott, S. D., Sassoon, E. M., Williams, M. R., Jones, J. L., Girling, A. C., et al. (2004). Dysregulated expression of adamalysin-thrombospondin genes in human breast carcinoma. *Clin. Cancer Res.* 10 (7), 2429–2440. doi:10.1158/1078-0432.ccr-0398-3

Ren, Q., Zhang, P., Lin, H., Feng, Y., Chi, H., Zhang, X., et al. (2023). A novel signature predicts prognosis and immunotherapy in lung adenocarcinoma based on cancer-associated fibroblasts. *Front. Immunol.* 14, 1201573. doi:10.3389/fimmu.2023.1201573

Reuten, R., Zendehroud, S., Nicolau, M., Fleischhauer, L., Laitala, A., Kiderlen, S., et al. (2021). Basement membrane stiffness determines metastases formation. *Nat. Mater* 20 (6), 892–903. doi:10.1038/s41563-020-00894-0

Rocks, N., Paulissen, G., Quesada Calvo, F., Polette, M., Gueders, M., Munaut, C., et al. (2006). Expression of a disintegrin and metalloprotease (ADAM and ADAMTS) enzymes in human non-small-cell lung carcinomas (NSCLC). *Br. J. Cancer* 94 (5), 724–730. doi:10.1038/sj.bjc.6602990

Rusňák, Š., Hecová, L., Kasl, Z., Sobotová, M., and Hauer, L. (2020). Therapy of uveal melanoma A review. *Cesk Slov. Oftalmol.* 77 (1), 1–13. doi:10.31348/2020/10

Shen, Y., Chi, H., Xu, K., Li, Y., Yin, X., Chen, S., et al. (2022). A novel classification model for lower-grade glioma patients based on pyroptosis-related genes. *Brain Sci.* 12 (6), 700. doi:10.3390/brainsci12060700

Shen, J., Wei, Z., Lv, L., He, J., Du, S., Wang, F., et al. (2023). A model of basement membrane-associated gene signature predicts liver hepatocellular carcinoma response to immune checkpoint inhibitors. *Mediat. Inflamm.* 2023, 7992140. doi:10.1155/2023/7992140

Sheu, M. J., Hsieh, M. J., Chou, Y. E., Wang, P. H., Yeh, C. B., Yang, S. F., et al. (2017). Effects of ADAMTS14 genetic polymorphism and cigarette smoking on the clinicopathologic development of hepatocellular carcinoma. *PLoS One* 12 (2), e0172506. doi:10.1371/journal.pone.0172506

Sikic, L., Schulman, E., Kosklin, A., Saraswathibhatla, A., Chaudhuri, O., and Pokki, J. (2022). Nanoscale tracking combined with cell-scale microrheology reveals stepwise increases in force generated by cancer cell protrusions. *Nano Lett.* 22 (18), 7742–7750. doi:10.1021/acs.nanolett.2c01327

Singh, A. D., Bergman, L., and Seregard, S. (2005). Uveal melanoma: Epidemiologic aspects. *Ophthalmol. Clin. North Am.* 18 (1), 75–84. viii. doi:10.1016/j.ohc.2004.07.002

Singh, A. D., Turell, M. E., and Topham, A. K. (2011). Uveal melanoma: Trends in incidence, treatment, and survival. *Ophthalmology* 118 (9), 1881–1885. doi:10.1016/j.ophtha.2011.01.040

Smit, K. N., Jager, M. J., de Klein, A., and Kiliç, E. (2020). Uveal melanoma: Towards a molecular understanding. *Prog. Retin Eye Res.* 75, 100800. doi:10.1016/j.preteyeres.2019.100800

Song, B., Wu, P., Liang, Z., Wang, J., Zheng, Y., Wang, Y., et al. (2022a). A novel necroptosis-related gene signature in skin cutaneous melanoma prognosis and tumor microenvironment. *Front. Genet.* 13, 917007. doi:10.3389/fgene.2022.917007

Song, B., Chi, H., Peng, G., Song, Y., Cui, Z., Zhu, Y., et al. (2022b). Characterization of coagulation-related gene signature to predict prognosis and tumor immune microenvironment in skin cutaneous melanoma. *Front. Oncol.* 12, 975255. doi:10.3389/fonc.2022.975255

Song, G., Peng, G., Zhang, J., Song, B., Yang, J., Xie, X., et al. (2023). Uncovering the potential role of oxidative stress in the development of periodontitis and establishing a stable diagnostic model via combining single-cell and machine learning analysis. *Front. Immunol.* 14, 1181467. doi:10.3389/fimmu.2023.1181467

Straatsma, B. R., Diener-West, M., Caldwell, R., Engstrom, R. E., and Collaborative Ocular Melanoma Study Group* (2018). Mortality after deferral of treatment or no treatment for choroidal melanoma. *Indian J. Ophthalmol.* 66 (10), 1395–1400. doi:10.4103/ijo.IJO_1499_18

Su, K., Guo, L., Ma, W., Wang, J., Xie, Y., Rao, M., et al. (2022). PD-1 inhibitors plus anti-angiogenic therapy with or without intensity-modulated radiotherapy for advanced hepatocellular carcinoma: A propensity score matching study. *Front. Immunol.* 13, 972503. doi:10.3389/fimmu.2022.972503

Sun, Y., Huang, J., and Yang, Z. (2015). The roles of ADAMTS in angiogenesis and cancer. *Tumour Biol.* 36 (6), 4039–4051. doi:10.1007/s13277-015-3461-8

Tabernero, J. (2007). The role of VEGF and EGFR inhibition: Implications for combining anti-VEGF and anti-EGFR agents. *Mol. Cancer Res.* 5 (3), 203–220. doi:10.1158/1541-7786.MCR-06-0404

Timpl, R. (1989). Structure and biological activity of basement membrane proteins. *Eur. J. Biochem.* 180 (3), 487–502. doi:10.1111/j.1432-1033.1989.tb14673.x

Triozzi, P. L., and Singh, A. D. (2014). Adjuvant therapy of uveal melanoma: Current status. *Ocul. Oncol. Pathol.* 1 (1), 54–62. doi:10.1159/000367715

Tsilibary, E. C. (2003). Microvascular basement membranes in diabetes mellitus. *J. Pathol.* 200 (4), 537–546. doi:10.1002/path.1439

Valastyan, S., and Weinberg, R. A. (2011). Tumor metastasis: Molecular insights and evolving paradigms. *Cell.* 147 (2), 275–292. doi:10.1016/j.cell.2011.09.024

Wang, X., Ye, F., Xiong, M., Xiu, B., Chi, W., Zhang, Q., et al. (2022). Cross-talk of four types of RNA modification proteins with adenosine reveals the landscape of multivariate prognostic patterns in breast cancer. *Front. Genet.* 13, 943378. doi:10.3389/fgene.2022.943378

Whitehead, M. J., McCanney, G. A., Willison, H. J., and Barnett, S. C. (2019). MyelinJ: An ImageJ macro for high throughput analysis of myelinating cultures. *Bioinformatics* 35 (21), 4528–4530. doi:10.1093/bioinformatics/btz403

Wu, G., Li, J., Xu, Y., Che, X., Chen, F., and Wang, Q. (2021). A new survival model based on ADAMTSs for prognostic prediction in clear cell renal cell carcinoma. *J. Oncol.* 2021, 2606213. doi:10.1155/2021/2606213

Xiong, J., Chi, H., Yang, G., Zhao, S., Zhang, J., Tran, L. J., et al. (2023). Revolutionizing anti-tumor therapy: Unleashing the potential of B cell-derived exosomes. *Front. Immunol.* 14, 1188760. doi:10.3389/fimmu.2023.1188760

Xu, L., Deng, C., Pang, B., Zhang, X., Liu, W., Liao, G., et al. (2018). Tip: A web server for resolving tumor immunophenotype profiling. *Cancer Res.* 78 (23), 6575–6580. doi:10.1158/0008-5472.CAN-18-0689

Yang, B., Fan, Y., Liang, R., Wu, Y., and Gu, A. (2023). Identification of a prognostic six-immune-gene signature and a nomogram model for uveal melanoma. *BMC Ophthalmol.* 23 (1), 2. doi:10.1186/s12886-022-02723-1

Yu, Y. H., Kuo, H. K., and Chang, K. W. (2008). The evolving transcriptome of head and neck squamous cell carcinoma: A systematic review. *PLoS One* 3 (9), e3215. doi:10.1371/journal.pone.0003215

Zhang, Z., Hao, R., Guo, Q., Zhang, S., and Wang, X. (2021). TP53 mutation infers a poor prognosis and is correlated to immunocytes infiltration in breast cancer. *Front. Cell. Dev. Biol.* 9, 759154. doi:10.3389/fcell.2021.759154

Zhang, Z., Zhang, T., Zhou, L., and Guan, J. (2022). Identification of diagnostic genes and effective drugs associated with osteoporosis treatment by single-cell RNA-seq analysis and network Pharmacology. *Mediat. Inflamm.* 2022, 6830635. doi:10.1155/2022/6830635

Zhang, J., Peng, G., Chi, H., Yang, J., Xie, X., Song, G., et al. (2023). CD8 + T-cell marker genes reveal different immune subtypes of oral lichen planus by integrating single-cell RNA-seq and bulk RNA-sequencing. *BMC Oral Health* 23 (1), 464. doi:10.1186/s12903-023-03138-0

Zhao, S., Zhang, L., Ji, W., Shi, Y., Lai, G., Chi, H., et al. (2022a). Machine learning-based characterization of cuprotosis-related biomarkers and immune infiltration in Parkinson's disease. *Front. Genet.* 13, 1010361. doi:10.3389/fgene.2022.1010361

Zhao, S., Chi, H., Ji, W., He, Q., Lai, G., Peng, G., et al. (2022b). A bioinformatics-based analysis of an anoikis-related gene signature predicts the prognosis of patients with low-grade gliomas. *Brain Sci.* 12 (10), 1349. doi:10.3390/brainsci12101349

Zhao, Y., Wei, K., Chi, H., Xia, Z., and Li, X. (2022c). IL-7: A promising adjuvant ensuring effective T cell responses and memory in combination with cancer vaccines? *Front. Immunol.* 13, 1022808. doi:10.3389/fimmu.2022.1022808

Zheng, Z., Zhang, L., Tu, Z., Deng, Y., and Yin, X. (2021). An autophagy-related prognostic signature associated with immune microenvironment features of uveal melanoma. *Biosci. Rep.* 41 (3). doi:10.1042/BSR20203812

Zhou, T., Chen, W., Wu, Z., Cai, J., and Zhou, C. (2022). A newly defined basement membrane-related gene signature for the prognosis of clear-cell renal cell carcinoma. *Front. Genet.* 13, 994208. doi:10.3389/fgene.2022.994208

Zhuang, W., Sun, H., Zhang, S., Zhou, Y., Weng, W., Wu, B., et al. (2021). An immunogenomic signature for molecular classification in hepatocellular carcinoma. *Mol. Ther. Nucleic Acids* 25, 105–115. doi:10.1016/j.omtn.2021.06.024

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 13 October 2023
DOI 10.3389/fgene.2023.1230998

OPEN ACCESS

EDITED BY
Lin Qi,
Central South University, China

REVIEWED BY
Tian Tian,
Children's Hospital of Philadelphia, United States
Michael Anton Bauer,
University of Arkansas for Medical Sciences, United States

*CORRESPONDENCE
Mohammad Shahbaz Khan,
mskhan2@childrensnational.org

RECEIVED 29 May 2023
ACCEPTED 18 September 2023
PUBLISHED 13 October 2023

CITATION
Khan MS, Hanif W, Alsakhen N, Jabbar B, Shamkh IM, Alsaiari AA, Almehmadi M, Alghamdi S, Shakoori A, Al Farraj DA, Almutairi SM, Hussein Issa Mohammed Y, Abouzied AS, Rehman A-U and Huwaimel B (2023), Isoform switching leads to downregulation of cytokine producing genes in estrogen receptor positive breast cancer.
Front. Genet. 14:1230998.
doi: 10.3389/fgene.2023.1230998

Isoform switching leads to downregulation of cytokine producing genes in estrogen receptor positive breast cancer

Mohammad Shahbaz Khan[1]*, Waqar Hanif[2], Nada Alsakhen[3], Basit Jabbar[4], Israa M. Shamkh[5], Ahad Amer Alsaiari[6], Mazen Almehmadi[6], Saad Alghamdi[7], Afnan Shakoori[7], Dunia A. Al Farraj[8], Saeedah Musaed Almutairi[8], Yasser Hussein Issa Mohammed[9], Amr S. Abouzied[10,11], Aziz-Ur Rehman[12] and Bader Huwaimel[10,13]

[1]Children's National Hospital, Washington, DC, United States, [2]Department of Bioinformatics, Department of Sciences, School of Interdisciplinary Engineering & Science (SINES), National University of Sciences and Technology (NUST), Islamabad, Pakistan, [3]Department of Chemistry, Faculty of Science, The Hashemite University, Zarqa, Jordan, [4]Centre of Excellence in Molecular Biology, University of the Punjab, Lahore, Pakistan, [5]Chemo and Bioinformatics Lab, Bio Search Research Institution, Giza, Egypt, [6]Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Taif University, Taif, Saudi Arabia, [7]Laboratory Medicine Department, Faculty of Applied Medical Sciences, Umm Al-Qura University, Makkah, Saudi Arabia, [8]Department of Botany and Microbiology, College of Science, King Saud University, Riyadh, Saudi Arabia, [9]Department of Biochemistry, Faculty of Applied Science, University of Hajjah, Hajjah, Yemen, [10]Department of Pharmaceutical Chemistry, College of Pharmacy, University of Hail, Hail, Saudi Arabia, [11]Department of Pharmaceutical Chemistry, National Organization for Drug Control and Research (NOD CAR), Giza, Egypt, [12]Keystone Pharmacogenomics LLC, Bensalem, PA, United States, [13]Medical and Diagnostic Research Center, University of Hail, Hail, Saudi Arabia

Objective: Estrogen receptor breast cancer (BC) is characterized by the expression of estrogen receptors. It is the most common cancer among women, with an incidence rate of 2.26 million cases worldwide. The aim of this study was to identify differentially expressed genes and isoform switching between estrogen receptor positive and triple negative BC samples.

Methods: The data were collected from ArrayExpress, followed by preprocessing and subsequent mapping from HISAT2. Read quantification was performed by StringTie, and then R package ballgown was used to perform differential expression analysis. Functional enrichment analysis was conducted using Enrichr, and then immune genes were shortlisted based on the ScType marker database. Isoform switch analysis was also performed using the IsoformSwitchAnalyzeR package.

Results: A total of 9,771 differentially expressed genes were identified, of which 86 were upregulated and 117 were downregulated. Six genes were identified as mainly associated with estrogen receptor positive BC, while a novel set of ten genes were found which have not previously been reported in estrogen receptor positive BC. Furthermore, alternative splicing and subsequent isoform usage in the immune system related genes were determined.

Conclusion: This study identified the differential usage of isoforms in the immune system related genes in cancer cells that suggest immunosuppression due to the dysregulation of CXCR chemokine receptor binding, iron ion binding, and cytokine activity.

KEYWORDS

estrogen receptor, breast cancer, isoform switching, differential expression, functional enrichment, cytokine, immunosuppression, CXCR chemokine receptor

Introduction

Breast cancer (BC) is one of the most commonly diagnosed global malignancies and is a leading cause of mortality among women. BC is a heterogeneous disease involving multiple environmental and genetic factors such as age, hormones, unhygienic diet, or toxic environmental exposure. The BRCA1 and BRCA2 tumor suppressor genes play a significant role in BC development (Li et al., 2017). Despite advances in treatments like chemotherapy, endocrine therapy, and human epidermal growth factor receptor-2 (HER2)-targeted therapy, the chance of relapse and BC metastasis remains a great challenge (Zhu and Yu, 2022). BC is a global health challenge as the most commonly diagnosed cancer, with an estimated incidence of 2.26 million cases worldwide according to GLOBOCAN 2020 global cancer statistics. The reported BC incidence rate is higher in Asia at 45.4% (Sung et al., 2021). There are different types of BC, depending on which cells in the breast become cancerous. Estrogen receptor positive (ERP) and triple-negative BC (TNBC) are the most aggressive types of BC. ERP BC is characterized by the presence of estrogen receptors (ERs) on tumor cells that help them grow and proliferate rapidly based on estrogen fueling. It is the largest subtype of BC as it involves the expression and activity of the estrogen receptor. It is estimated that approximately 80% of BCs are ERP (Lamb et al., 2019). TNBC is defined as a type of BC with a negative expression of ER, progesterone receptor (PR), and HER2. The mortality rate of TNBC is higher because of its high invasiveness and because approximately 46% of TNBC patients are more likely to have distant metastasis (Yin et al., 2020).

The ERP BC microenvironment (BCM) consists of immune cells, fibroblasts, adipocytes, mesenchymal stem cells, extracellular matrix, and tumor-associated macrophages (TAMs) (Munir et al., 2021). During breast tumorigenesis, tumor cells escape the immune surveillance by modifying surface antigens and altering their surrounding environment (Segovia-Mendoza and Morales-Montor, 2019). Chemokine, a family of signaling proteins, functions to induce leukocyte migration. Chemokine CC receptor type 5 (CCR5) is a cell surface receptor that has a high affinity for chemotropic cytokines called chemokines. A 32-bp deletion in this receptor (CCR5Δ32) results in a non-functional and deformed receptor which, in turn, results in the activation and invasion of immune cells at the site of tumorigenesis and ultimately leads to its progression (Fatima et al., 2019). In mammalian cells, alternative splicing (AS) is a key mechanism of gene expression regulation. AS occurs when intron and exon elements become rearranged by splicing at different splice-sites, resulting in multiple RNA transcripts. AS regulation is influenced by multiple factors such as cancer or other diseases (Vitting-Seerup and Sandelin, 2017). It occurs when there is differential usage of gene transcripts between different conditions (Baralle and Giudice, 2017). Thus, gene expression should be analyzed at the isoform level because isoform switching (IS) with predicted functional consequences is more common and important in dysfunctional cells (Kahraman et al., 2020).

RNA sequencing (RNA-seq) is a proven quantitative tool for the expression estimation of cells and facilitates the detection and identification of novel transcripts generated by AS. This study identified differential isoform usage (DIU) across conditions (ERP vs. TNBC) in immune system-related genes that may assist targeted therapies for ERP BC. Identifying novel biomarkers and isoform switching may pave the way for the early detection and successful treatment of ERP BC (Chen et al., 2022).

Methodology

Overview of the protocol

The data were collected from ArrayExpress: E-MTAB-4993. Further processing and analysis were performed by RNA-seq analysis consisting of preprocessing, mapping, quantifying, and differential expression analysis (DEA) methods (Costa-Silva et al., 2017). Isoform switching and DIU were ultimately detected in immune system-related dysregulated genes in ERP vs. TNBC. This study considered two biological conditions of BC, ERP, and TNBC. The raw data comprised 63 samples (ERP = 51, TNBC = 12).

RNA-seq data preprocessing and mapping

The data obtained from ArrayExpress were in the form of raw reads and required preprocessing and quality control to reduce noise by trimming poor quality reads, adaptors, and primers. The first step in data preprocessing was quality assessment, which was performed by the FastQC tool to generate individual quality reports for each sample (v0.11.9) (Rostovskaya et al., 2022). The fastp tool (v0.20.0) was then used to trim poor quality reads to remove primer and adaptor content, resulting in filtered reads (Chen et al., 2018). These filtered reads then underwent quality check analysis by FastQC. Next, the filtered reads were used as input in the HISAT2 tool (v2.1.0) for mapping against the reference genome of *Homo sapiens* (GRCh38) (Kim et al., 2015). This generated files in SAM (sequence alignment map) format which contained aligned reads. Mapping rates indicative of the quality of RNA sequencing are presented in Supplementary File 6.

Read quantification and DEA

Before read quantification, the SAM files were first converted to BAM (binary alignment map) format, which is the compressed and binary format of aligned reads, using Samtools (v1.16) (Danecek et al., 2021). Next, BamUtil (v1.0.15) was used to remove duplicates (deduplication) from mapped reads (Jun et al., 2017). The quantification of deduplicated sorted reads was then performed using StringTie (v2.2.0) (Shumate et al., 2022) in three steps. In the first step, the StringTie assembler was employed to assemble the aligned reads of each sample into a transcriptome. In the second step, the full set of transcriptome assemblies was passed to the StringTie merge module to merge the genomic features among all samples to create a consistent set of transcripts across all samples. In the final step, this merged assembly was used to estimate the transcript abundances (Pertea et al., 2016). The identification of differentially expressed genes (DEGs) and their enrichment analysis offers biological insights into the processes that are affected by certain conditions (Frazee et al., 2015). R package ballgown (v3.15) was used to perform differential gene expression (DGE) analysis of all the transcripts and abundances in ERP vs. TNBC (Frazee et al., 2014). Criteria of a *p*-value less than 0.05 and log2 FoldChange value of <1.5 and >1.5 were used to identify biologically and statistically significant DEGs in ERP vs. TNBC. DEGs were graphically represented by the volcano plot (Nisar et al., 2021).

Functional enrichment analysis

For Gene Ontology (GO) and pathway enrichment analysis, the Enrichr package was used (Xie et al., 2021). The analysis of both up- and downregulated DEGs was performed separately. The plotEnrich () function was used to plot bar charts of biological processes (BP), molecular functions (MF), cellular components (CC), and KEGG pathways. The results were ordered according to *p*-value.

Identification of immune system-related genes

The ScType cell marker database was used to filter the genes involved in immune functions (regulation of immune cells through signaling pathway and immune response against tumors) from the DEGs identified in the previous step (Gonzalez et al., 2018; Ianevski et al., 2022). Genes common to the DEGs set and the ScType database (immune system) were selected for further analysis.

Identification of isoform switching in DEGs

Isoform switch analysis was performed to identify transcript-level expression profiles between ERP and TNBC to detect potential functional consequences resulting from isoform switch. The IsoformSwitchAnalyzeR package (v1.16.0) was used for this analysis (Vitting-Seerup and Sandelin, 2017). The package's input was the quantification files from StringTie, transcript files, a file containing merged annotations of all samples, and a design file containing sample IDs and relevant condition status. The IsoformSwitchTestDEXSeq () function was used to identify DIU based on differential isoform (dIF) cutoff. A dIF criteria of 0.1 was used to find the relative abundances of all isoforms of a gene between two sample groups, and gene ExpressionCutoff of 0.5 was applied. The open reading frames (ORF) were analyzed using the analyzeORF() function, where the longest orfMethod was selected in order to shortlist only long ORFs due to their functional importance. The longest ORFs were then extracted using the extractSequence () function that outputs two files—one containing nucleotide sequences and the other including protein sequences. The functional consequences of ORFs were identified in order to add functional knowledge to the transcripts. Four types of functional consequences were identified for ORFs: coding potential, protein domains, signal peptides, and intrinsically disordered regions (IDRs). The coding potential of the genes was identified through the CPC2 tool that takes nucleotide sequences as an input. The Pfam tool was used to predict protein domains. The signal peptides of ORFs were identified through SignalP, whereas intrinsically disordered regions (IDRs) were predicted by the IUPred3 tool. To assign the predicted functional consequences to the transcripts, R package was employed using functions such as analyzeCPC2 (), analyzeSignalP (), analyzePFAM (), and analyzeIUPred2A (). Moreover, the switchPlot () function was used to plot the shortlisted immune system-related genes that were dysregulated in ERP BC.

Results

Identification of upregulated and downregulated genes

The gene expression profiling of 63 samples (ERP = 51, TNBC = 12) by ballgown R/Bioconductor identified 15,947 DEGs between ERP and TNBC tissue samples. Genes with no annotation were filtered out, with 9,771 DEGs remaining. Using DEA, 86 genes were identified as upregulated (logFC >1.5 and *p*-value <0.05) and 117 were downregulated (logFC < −1.5 and *p*-value <0.05) (Figure 1). The top 10 upregulated DEGs were FOXA1, RHOB, AR, CMBL, AGR2, ESR1, TFF3, SYBU, CBLC, and DNALI1 (Table 1; Supplementary Information S1); the top 10 downregulated DEGs were CENPW, EN1, A2ML1, TMSB15A, FOXC1, KRT16, SLC7A5, CDK6, MELTF, and CA9 (Table 2; Supplementary Information S2).

Gene Ontology analysis

Both up- and downregulated DEGs were subjected to GO enrichment analysis. This analysis revealed BP, CC, and MF that were affected due to change in gene expression. Upregulated DEGs of biological processes were enriched in steroid hormone-mediated signaling pathway, intracellular steroid hormone receptor signaling pathway, regulation of smooth muscle cell proliferation, and response to estrogen, indicating that upregulated genes are involved in the regulation of breast stem cells, increased cell proliferation, increased estrogen hormone and cancerous T-cells, angiogenesis, and excessive mitochondrial and sodium ion transport (Figure 2A; Table 3).

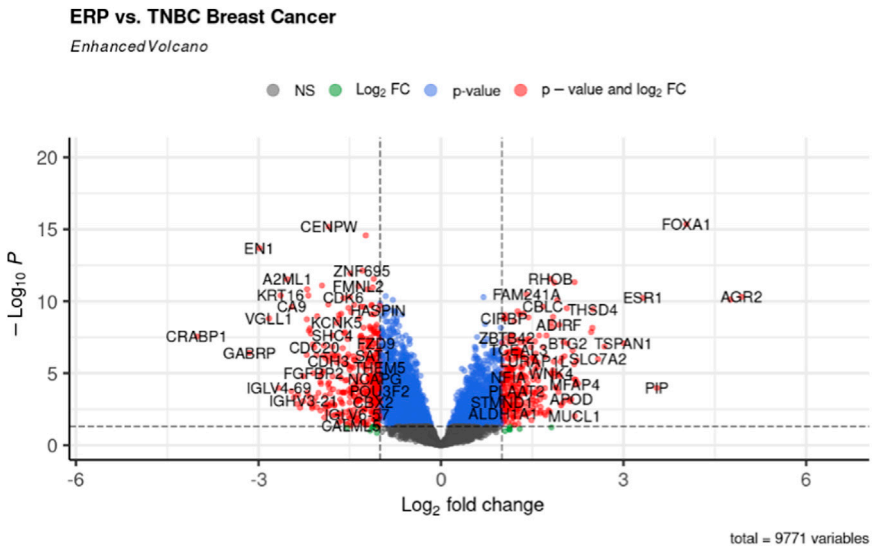


FIGURE 1
Volcano plot for DEGs. The red dots represent up- (right) and downregulated (left) DEGs. Upregulated genes having logFC >1.5 and *p*-value <0.05 can be seen on the right of the plot; downregulated genes having logFC < −1.5 and *p*-value <0.05 can be seen on the left of the plot.

TABLE 1 Top 10 differentially expressed upregulated genes.

Gene name	*p*-value	log2FoldChange	Expression
FOXA1	4.44E-16	4.03	Up
RHOB	2.75E-12	1.79	Up
AR	4.69E-12	2.19	Up
CMBL	5.16E-12	1.86	Up
AGR2	4.93E-11	4.93	Up
ESR1	5.81E-11	3.32	Up
TFF3	7.72E-11	4.75	Up
SYBU	8.24E-11	1.91	Up
CBLC	2.37E-10	1.67	Up
DNALI1	3.04E-10	2.06	Up

TABLE 2 Top 10 differentially expressed downregulated genes.

Gene name	*p*-value	log2FoldChange	Expression
CENPW	6.66E-16	−1.84	Down
EN1	2.06E-14	−2.98	Down
A2ML1	2.95E-12	−2.52	Down
TMSB15A	8.07E-12	−1.95	Down
FOXC1	1.47E-11	−2.20	Down
KRT16	3.93E-11	−2.63	Down
SLC7A5	3.98E-11	−2.18	Down
CDK6	5.90E-11	−1.60	Down
MELTF	1.66E-10	−1.85	Down
CA9	2.41E-10	−2.45	Down

The downregulated DEGs enriched in mitotic spindle organization, chemokine-mediated signaling pathway, cellular response to chemokine, microtubule cytoskeleton organization involved in mitosis, antimicrobial humoral immune response mediated by antimicrobial peptides, neutrophil chemotaxis, granulocyte chemotaxis, and the attachment of mitotic spindle microtubules to kinetochore and kinetochore organization indicated that they may have cancer development-related functions because of disrupted cell signaling and a dysregulated cell cycle due to incorrectly organized proteins and a suppressed immune system (Figure 2B; Table 4). Alternatively, the upregulated DEGs of molecular functioning show transcription coactivator binding, RNA polymerase II general transcription initiation factor binding, epidermal growth factor receptor binding, BMP receptor binding, SH3 domain binding, ATPase binding, general transcription initiation factor binding, metallocarboxypeptidase activity, and IgG binding (Figure 3A; Table 5). This denotes disrupted cell signaling, increased cell proliferation, growth, differentiation, and epithelial–mesenchymal transition (EMT) due to upregulated transcription. On the other hand, the downregulated DEGs were mainly enriched in CXCR3 and CXCR chemokine receptor binding, chemokine and cytokine activity, peptidase inhibitor activity, L-leucine transmembrane transporter activity, and chitinase activity, which indicate suppressed immune system response and increased abnormal proteins which may result in cancer progression and development (Figure 3B; Table 6). The cellular component enrichment of upregulated DEGs, collagen-containing extracellular matrix, elastic fiber, Golgi lumen, intracellular

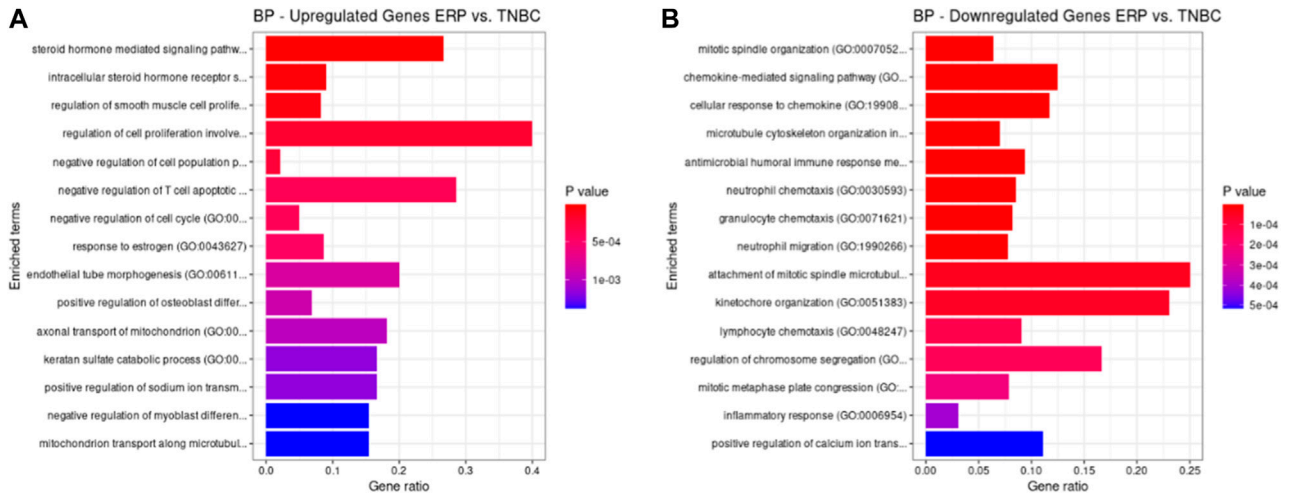


FIGURE 2
GO biological processes up- and downregulated by DEGs. Bar chart plots of top 15 BPs in ERP vs. TNBC. The *x*-axis is the gene ratio, while the color represents *p*-value. **(A)** Steroid hormone-mediated signaling pathway and response to estrogen and endothelial tube morphogenesis as significant upregulated BP. **(B)** Chemokine-mediated signaling pathway, cellular response to chemokine, granulocyte chemotaxis, and attachment of mitotic spindle microtubules to kinetochore as significant downregulated BP.

TABLE 3 GO analysis of BP of upregulated DEGs according to Enrichr (*p*-value<0.05).

Biological process	Gene ratio	*p*-value	Genes
Steroid hormone-mediated signaling pathway	4/15	4.19E-07	BMP4; AR; PGR; ESR1
Intracellular steroid hormone receptor signaling pathway	4/44	3.79E-05	AR; SCGB2A1; PGR; ESR1
Regulation of smooth muscle cell proliferation	4/49	5.82E-05	BMP4; ELN; OGN; APOD
Regulation of cell proliferation involved in heart morphogenesis	2/5	0.00018	BMP4; TBX3
Negative regulation of cell population proliferation	8/379	0.00022	BMP4; AR; BTG2; CAMK2N1; ERBB4; OGN; APOD; TBX3
Negative regulation of cell cycle	4/80	0.00039	BMP4; BTG2; CAMK2N1; RHOB
Response to estrogen	3/35	0.00045	GSTM3; AR; ESR1
Endothelial tube morphogenesis	2/10	0.00080	BMP4; RHOB
Positive regulation of osteoblast differentiation	3/44	0.00089	BMP4; NPNT; IL6ST
Axonal transport of mitochondrion	2/11	0.00098	MAPT; SYBU
Keratan sulfate catabolic process	2/12	0.0011	OMD; OGN
Positive regulation of sodium ion transmembrane transport	2/12	0.0011	FXYD1; WNK4
Negative regulation of myoblast differentiation	2/13	0.0013	BMP4; TBX3
Mitochondrion transport along microtubule	2/13	0.0013	MAPT; SYBU

organelle lumen, basement membrane, and sodium:potassium-exchanging ATPase complex indicate that these may have been involved in the initiation and progression of cancer due to changes in the cellular cytoskeleton, membrane remodeling, and alterations in protein secretions (Figure 4A; Table 7). Moreover, the downregulated DEGs of cellular components show intermediate filament, intermediate filament cytoskeleton, polymeric cytoskeletal fiber, spindle, cornified envelope, desmosome, and endoplasmic reticulum lumen, which indicate cancer development and progression due to disrupted cytoskeletal proteins, dysregulated cell division, misfolded proteins, and DNA damage (Figure 4B; Table 8).

KEGG pathway analysis

KEGG pathways were predicted using Enrichr for the DEGs to identify biological pathways that are disrupted due to the up- and downregulation of genes involved in those pathways. As indicated in Figure 5A, upregulated genes such as BMP4, GSTM3, FOS,

TABLE 4 GO analysis of BP of downregulated DEGs according to Enrichr (*p*-value<0.05).

Biological process	Gene ratio	*p*-value	Genes
Mitotic spindle organization	10/157	2.89E-08	CDC20; STMN1; NUF2; CDCA8; BIRC5; KIF23; KIF2C; BUB1; NDC80; AURKB
Chemokine-mediated signaling pathway	7/56	3.58E-08	CXCL10; CXCL9; FOXC1; CXCL11; CXCL8; CXCL13; CCL18
Cellular response to chemokine	7/60	5.85E-08	CXCL10; CXCL9; FOXC1; CXCL11; CXCL8; CXCL13; CCL18
Microtubule cytoskeleton organization involved in mitosis	9/128	6.26E-08	CDC20; STMN1; NUF2; CDCA8; BIRC5; KIF2C; BUB1; NDC80; AURKB
Antimicrobial humoral immune response mediated by antimicrobial peptide	6/64	2.00E-06	CXCL10; CXCL9; CXCL11; CXCL8; CXCL13; KRT6A
Neutrophil chemotaxis	6/70	3.40E-06	CXCL10; CXCL9; CXCL11; CXCL8; CXCL13; CCL18
Granulocyte chemotaxis	6/73	4.36E-06	CXCL10; CXCL9; CXCL11; CXCL8; CXCL13; CCL18
Neutrophil migration	6/77	5.95E-06	CXCL10; CXCL9; CXCL11; CXCL8; CXCL13; CCL18
Attachment of mitotic spindle microtubules to kinetochore	3/12	4.13E-05	NUF2; KIF2C; NDC80
Kinetochore organization	3/13	5.35E-05	CENPW; NUF2; NDC80
Lymphocyte chemotaxis	4/44	0.00012	CXCL10; CXCL11; CXCL13; CCL18
Regulation of chromosome segregation	3/18	0.00014	KIF2C; BUB1; AURKB
Mitotic metaphase plate congression	4/51	0.00022	NUF2; CDCA8; KIF2C; NDC80
Inflammatory response	7/230	0.00041	CXCL10; CXCL11; CXCL9; CXCL8; KRT16; CXCL13; CCL18
Positive regulation of calcium ion transmembrane transport	3/27	0.00051	CXCL10; CXCL9; CXCL11

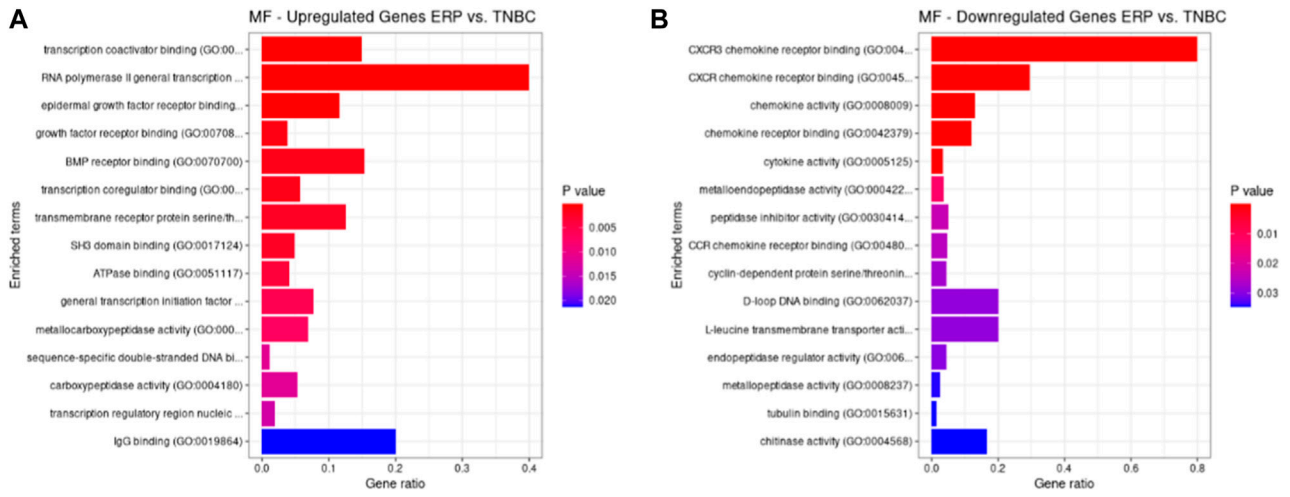


FIGURE 3
GO molecular functions up- and downregulated by DEGs. Bar chart plots of top 15 MF in ERP vs. TNBC. *X*-axis is the gene ratio, while the color represents *p*-value. **(A)** Transcription coactivator binding, RNA polymerase II general transcription initiation factor binding, epidermal growth factor receptor binding, and BMP receptor binding as significant upregulated MF. **(B)** CXCR3 chemokine receptor binding, CXCR chemokine receptor binding, chemokine activity, and L-leucine transmembrane transporter activity as significant downregulated MF.

HMGCS2, ADH1B, COL4A5, ESR1, NAT1, IL6ST, and PGR were enriched in pathways in cancer, chemical carcinogenesis, ECM–receptor interaction, tyrosine metabolism, valine, leucine and isoleucine degradation, the PI3K-Akt signaling pathway, estrogen signaling pathway, caffeine metabolism, signaling pathways that regulate the pluripotency of stem cells, and BC (Table 9). This indicates that the upregulation of genes promotes pathways that are mainly involved in DNA repair, cell motility and

TABLE 5 GO analysis of MF of upregulated DEGs according to Enrichr (*p*-value<0.05).

Molecular function	Gene ratio	*p*-value	Genes
Transcription coactivator binding	3/20	8.30E-05	AR; PGR; ESR1
RNA polymerase II general transcription initiation factor binding	2/5	0.00018	AR; ESR1
Epidermal growth factor receptor binding	3/26	0.00018	ERBB4; AGR2; CBLC
Growth factor receptor binding	4/105	0.0010	ERBB4; AGR2; CBLC; IL6ST
BMP receptor binding	2/13	0.0013	BMP4; GDF15
Transcription coregulator binding	3/53	0.0015	AR; PGR; ESR1
Transmembrane receptor protein serine/threonine kinase binding	2/16	0.0021	BMP4; GDF15
SH3 domain binding	3/62	0.0024	CBLC; EVL; MAPT
ATPase binding	3/73	0.0038	AR; PGR; ESR1
General transcription initiation factor binding	2/26	0.0055	AR; ESR1
Metallocarboxypeptidase activity	2/29	0.0068	CPA3; CPE
Sequence-specific double-stranded DNA binding	8/712	0.0114	FOXA1; AR; ERBB4; FOSB; FOS; LMX1B; ESR1; TBX3
Carboxypeptidase activity	2/38	0.0116	CPA3; CPE
Transcription regulatory region nucleic acid binding	4/212	0.0132	FOXA1; AR; ERBB4; FOS
IgG binding	1/5	0.0213	PIP

TABLE 6 GO analysis of MF of downregulated DEGs according to Enrichr (*p*-value<0.05).

Molecular function	Gene ratio	*p*-value	Genes
CXCR3 chemokine receptor binding	4/5	5.54E-09	CXCL10; CXCL11; CXCL9; CXCL13
CXCR chemokine receptor binding	5/17	3.68E-08	CXCL10; CXCL9; CXCL11; CXCL8; CXCL13
Chemokine activity	6/46	2.73E-07	CXCL10; CXCL9; CXCL11; CXCL8; CXCL13; CCL18
Chemokine receptor binding (GO:0042379)	6/50	4.54E-07	CXCL10; CXCL9; CXCL11; CXCL8; CXCL13; CCL18
Cytokine activity	6/173	0.00054	CXCL10; CXCL11; CXCL9; CXCL8; CXCL13; CCL18
Metalloendopeptidase activity	3/82	0.01235	ADAMDEC1; MMP7; MMP1
Peptidase inhibitor activity	2/40	0.02289	A2ML1; PI3
CCR chemokine receptor binding	2/42	0.02508	CXCL13; CCL18
Cyclin-dependent protein serine/threonine kinase regulator activity	2/44	0.02735	CCNB2; CDKN2A
D-loop DNA binding	1/5	0.02891	RAD51AP1
L-Leucine transmembrane transporter activity	1/5	0.02891	SLC7A5
Endopeptidase regulator activity	2/46	0.02970	A2ML1; PI3
Metallopeptidase activity	3/121	0.03415	ADAMDEC1; MMP7; MMP1
Tubulin binding	5/307	0.03443	STMN1; BIRC5; KIF23; KIF2C; FAM83D
Chitinase activity	1/6	0.03459	CHI3L2

proliferation, cell cycle regulation, the inhibition of apoptosis, and increased EMT, resulting in tumor development and prognosis. The downregulated genes such as CXCL8, CXCL10, CXCL11, CDC20, CDK6, CDKN2A, CXCL13, and SHC4 (Table 10) were enriched in chemokine signaling pathway, toll-like receptor signaling pathway, cell cycle, bladder cancer, IL-17 signaling pathway, cellular senescence, p53 signaling pathway, and microRNAs in cancer (Figure 5B). The downregulation of genes involved in these pathways plays a crucial role in the tumor microenvironment by disrupting immune response, cell cycle arrest in the G2/M phase, increased cell growth, metastasis, proliferation and invasiveness, and the angiogenic potential of cancer cells. The analysis revealed that cancer-related pathways that were dysregulated due to DEGs have also been reported in various other cancers.

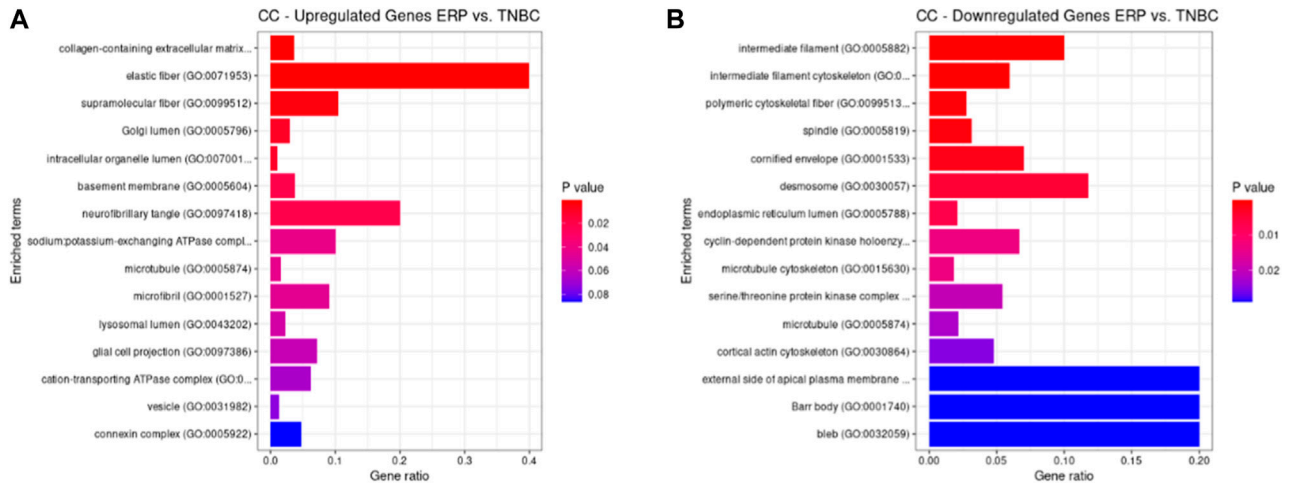


FIGURE 4
GO cellular components up- and downregulated by DEGs. Bar chart plots of top-15 CC in ERP vs. TNBC. *X*-axis is the gene ratio, while the color represents *p*-value. **(A)** Collagen-containing extracellular matrix, elastic fiber, Golgi lumen, and intracellular organelle lumen as significant upregulated CC. **(B)** Intermediate filament, intermediate filament cytoskeleton, polymeric cytoskeletal fiber, spindle, and desmosome as significant downregulated CC.

TABLE 7 GO analysis of CC of upregulated DEGs according to Enrichr (*p*-value<0.05).

Cellular component	Gene ratio	*p*-value	s
Collagen-containing extracellular matrix	14/380	8.27E-10	CPA3; TPSB2; COL14A1; GDF15; ELN; HTRA1; NPNT; ASPN; THSD4; THBS4; MFAP4; CILP; OGN; COL4A5
Elastic fiber	2/5	0.00018	MFAP4; ELN
Supramolecular fiber	2/19	0.00298	MFAP4; ELN
Golgi lumen	3/100	0.00919	MUCL1; OMD; OGN
Intracellular organelle lumen	9/848	0.01058	BMP4; MUCL1; COL14A1; ERBB4; OMD; OGN; COL4A5; ABAT; HMGCS2
Basement membrane	2/52	0.02108	COL4A5; THBS4
Neurofibrillary tangle	1/5	0.02131	MAPT
Sodium:potassium-exchanging ATPase complex	1/10	0.04218	FXYD1
Microtubule	3/182	0.04380	KIF12; MAPT; SYBU
Microfibril	1/11	0.04630	MFAP4
Lysosomal lumen	2/86	0.05294	OMD; OGN
Glial cell projection	1/14	0.05856	MAPT
Cation-transporting ATPase complex	1/16	0.06664	FXYD1
Vesicle	3/226	0.07374	OGN; TSPAN1; SYBU
Connexin complex	1/21	0.08656	GJC3

Selection of immune system genes

A total of 86 upregulated and 117 downregulated genes were used as a query against the marker genes database of the ScType R package, revealing that three upregulated and 10 downregulated genes were directly involved in immune system-related functions (Table 11, 12).

Isoform switching

Isoform switching analysis was performed on shortlisted immune system related genes, facilitating the identification of known and novel isoform switches from RNA-seq derived quantification data. Of 10 downregulated immune genes, three (STMN1, MELTF, and CXCL8) were found to have isoforms that were significantly used in

TABLE 8 GO analysis of CC of downregulated DEGs according to Enrichr (*p*-value<0.05).

Cellular component	Gene ratio	*p*-value	Genes
Intermediate filament	5/50	1.08E-05	SYNM; KRT16; PKP1; KRT75; KRT6C
Intermediate filament cytoskeleton	5/84	0.00013	SYNM; KRT16; PKP1; KRT75; KRT6C
Polymeric cytoskeletal fiber	7/256	0.00077	SYNM; KRT16; PKP1; KIF23; KIF2C; KRT75; KRT6C
Spindle	6/192	0.00093	CDC20; BIRC5; KIF23; KIF2C; FAM83D; AURKB
Cornified envelope	3/43	0.00203	PKP1; PI3; DSC3
Desmosome	2/17	0.00435	PKP1; DSC3
Endoplasmic reticulum lumen	6/285	0.00664	SPP1; COL9A3; MELTF; MSLN; MFGE8; CP
Cyclin-dependent protein kinase holoenzyme complex	2/30	0.01326	CCNB2; CDK6
Microtubule cytoskeleton	6/331	0.01326	CDC20; CCNB2; KIF23; KIF2C; FAM83D; AURKB
Serine/threonine protein kinase complex	2/37	0.01977	CCNB2; CDK6
Microtubule	4/182	0.02221	BIRC5; KIF23; KIF2C; AURKB
Cortical actin cytoskeleton	2/42	0.02508	GYS2; SLC2A1
External side of apical plasma membrane	1/5	0.02891	SLC7A5
Barr body	1/5	0.02891	MACROH2A2
Bleb	1/5	0.02891	ANLN

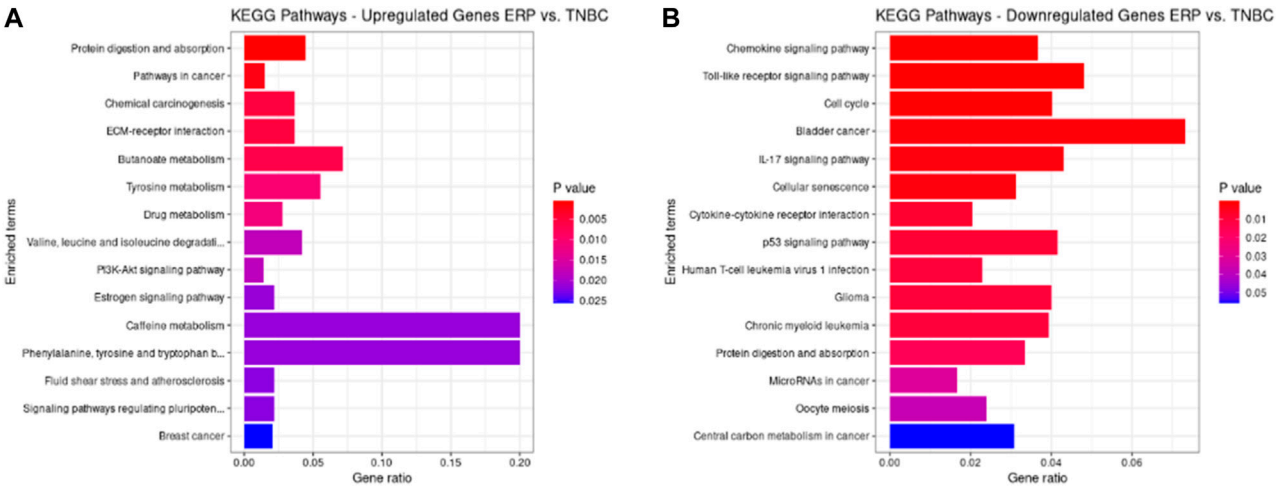


FIGURE 5
KEGG pathways analysis of upregulated DEGs. Bar chart plots of top 15 KEGG pathways in ERP vs. TNBC. *X*-axis is the gene ratio, while the color represents *p*-value. **(A)** Pathways in cancer, chemical carcinogenesis, ECM–receptor interaction, and tyrosine metabolism as significant upregulated pathways. **(B)** Chemokine signaling pathway, toll-like receptor signaling pathway, cell cycle, bladder cancer, and IL-17 signaling pathway as significant downregulated pathways.

ERP and have been validated through the Expression Atlas (Supplementary Information Table S3). Moreover, no significant isoform switch was observed in upregulated immune genes.

Isoform usage

STMN1, MELTF, and CXCL8 represent significant switches in isoform usage across ERP vs. TNBC, as shown in sashimi plots in Supplementary Information Figures S1–S3 respectively. By comparing the isoform usage across conditions, it was revealed that STMN1 has a single isoform (ENST00000485226) which was overexpressed in ERP. CXCL8 also has one isoform (ENST00000483500) which was significantly used in ERP. Furthermore, it was found that a novel isoform (MSTRG.29921.1) of MELTF was overexpressed in ERP (Figure 6).

Gene Ontology analysis of downregulated immune genes indicates that these genes may be involved in key immune system molecular functions such as CXCR chemokine receptor

TABLE 9 Pathway prediction for upregulated DEGs according to Enrichr (*p*-value<0.05).

KEGG pathway	Gene ratio	*p*-value	Genes
Protein digestion and absorption	4/90	0.00061	CPA3; COL14A1; ELN; COL4A5
Pathways in cancer	8/530	0.00200	BMP4; AR; GSTM3; IGF2; COL4A5; FOS; IL6ST; ESR1
Chemical carcinogenesis	3/82	0.00532	GSTM3; NAT1; ADH1B
ECM–receptor interaction	3/82	0.00532	CHAD; COL4A5; THBS4
Butanoate metabolism	2/28	0.00642	ABAT; HMGCS2
Tyrosine metabolism	2/36	0.01047	ADH1B; TAT
Drug metabolism	3/108	0.01133	GSTM3; NAT1; ADH1B
Valine, leucine, and isoleucine degradation	2/48	0.01813	ABAT; HMGCS2
PI3K–Akt signaling pathway	5/354	0.01831	ERBB4; CHAD; IGF2; COL4A5; THBS4
Estrogen signaling pathway	3/137	0.02129	PGR; FOS; ESR1
Caffeine metabolism	1/5	0.02131	NAT1
Phenylalanine, tyrosine, and tryptophan biosynthesis	1/5	0.02131	TAT
Fluid shear stress and atherosclerosis	3/139	0.02211	BMP4; GSTM3; FOS
Signaling pathways regulating pluripotency of stem cells	3/139	0.02211	BMP4; IL6ST; TBX3
BC	3/147	0.02555	PGR; FOS; ESR1

TABLE 10 Pathway prediction for downregulated DEGs according to Enrichr (*p*-value<0.05).

KEGG pathway	Gene ratio	*p*-value	Genes
Chemokine signaling pathway	7/190	0.00012	SHC4; CXCL10; CXCL11; CXCL9; CXCL8; CXCL13; CCL18
Toll-like receptor signaling pathway	5/104	0.00036	CXCL10; CXCL11; CXCL9; CXCL8; SPP1
Cell cycle	5/124	0.00081	CDC20; CCNB2; CDK6; CDKN2A; BUB1
Bladder cancer	3/41	0.00176	CXCL8; CDKN2A; MMP1
IL-17 signaling pathway	4/93	0.00217	CXCL10; CXCL8; MMP1; LCN2
Cellular senescence	5/160	0.00251	CCNB2; CDK6; CXCL8; CDKN2A; MYBL2
Cytokine–cytokine receptor interaction	6/294	0.00769	CXCL10; CXCL11; CXCL9; CXCL8; CXCL13; CCL18
p53 signaling pathway	3/72	0.00868	CCNB2; CDK6; CDKN2A
Human T-cell leukemia virus 1 infection	5/219	0.00936	CDC20; CCNB2; MMP7; CDKN2A; SLC2A1
Glioma	3/75	0.00970	SHC4; CDK6; CDKN2A
Chronic myeloid leukemia	3/76	0.01006	SHC4; CDK6; CDKN2A
Protein digestion and absorption	3/90	0.01585	KCNK5; COL9A3; KCNN4
MicroRNAs in cancer	5/299	0.03125	SHC4; CDK6; CDKN2A; STMN1; KIF23
Oocyte meiosis	3/125	0.03706	CDC20; CCNB2; BUB1
Central carbon metabolism in cancer	2/65	0.05559	SLC7A5; SLC2A1

binding, chemokine activity, iron ion binding, and cytokine activity (Figure 7).

Discussion

BC is the most prevalent type of cancer worldwide. It is thus essential to understand and explore ways to prevent its occurrence while identifying the genetic changes that are more susceptible to its incidence. The present study identified 9,771 DEGs, of which 86 genes were significantly upregulated and 117 were downregulated. The identified upregulated genes were FOXA1, RHOB, AR, CMBL, AGR2, ESR1, TFF3, SYBU, CBLC, and DNALI1; the downregulated genes were CENPW, EN1, A2ML1, TMSB15A, FOXC1, KRT16, SLC7A5, CDK6, MELTF, and CA9. Functional enrichment analysis of these DEGs revealed that the

TABLE 11 Shortlisted upregulated DEGs involved in the immune system (*p*-value<0.05).

Gene name	*p*-value	LogFC	Function	Expression
CPA3	1.02E-05	1.84048	Generates mature protease; released by mast cells	Up
THBS4	1.60E-05	1.95670	Adhesive glycoprotein	Up
CXCL14	5.20E-05	2.25235	Chemotactic factor for monocytes	Up

TABLE 12 Shortlisted downregulated DEGs involved in the immune system (*p*-value<0.05).

Gene name	*p*-value	LogFC	Function	Expression
MELTF	1.66E-10	−1.85184	Cell surface glycoprotein	Down
STMN1	4.25E-08	−1.53543	Integrate intracellular regulatory signals	Down
CXCL8	5.51E-05	−1.97224	Chemotactic factor	Down
CXCL11	0.00020	−1.67806	Regulate cell trafficking	Down
PI3	0.00020	−1.59618	Antimicrobial peptide	Down
CXCL10	0.00022	−2.01352	Stimulates monocytes, natural killer, and T-cells migration	Down
CD24	0.27436	−1.86584	Essential role in cell differentiation	Down
CD24P4	0.28050	−1.83388	Pseudogene	Down
CCL18	0.00771	−1.73928	Chemotactic factor, attracts only lymphocytes	Down
CXCL13	0.01855	−1.62905	Chemotactic factor for B-lymphocytes	Down

intracellular steroid hormone receptor signaling pathway, chemokine-mediated signaling pathway, kinetochore organization, pathways in cancer, BC, toll-like receptor signaling pathway, and cell cycle were the most dysregulated biological pathways and processes.

In the present study, FOXA1 was found to be upregulated and has been reported to inhibit STAT2, a transcription factor and its target IFN signaling pathway in BC; this may result in cancer progression due to suppressed immune response (He et al., 2021). Furthermore, upregulated RHOB results in ER-α (estrogen receptor alpha) overexpression that leads to increased estrogen uptake by BC cells which helps them grow and proliferate (Médale-Giamarchi et al., 2013). It has been reported that AR overexpression increases the transcription of genes involved in the cell cycle, resulting in increased proliferation of prostate cancer cells (Formaggio et al., 2021). This study found that CMBL (*p*-value: 0.00003) is suppressed in TNBC compared to non-TNBC types of BC, such as ERP. It encodes a cysteine hydrolase that cleaves cyclic esters which activate an angiotensin receptor blocker that helps lower blood pressure (Guo et al., 2017). Upregulated AGR2 is found in BC due to ER signaling and endoplasmic reticulum stress, and it results in increased cell proliferation, survival, and metastasis in BC (Ann et al., 2018). Moreover, ESR1 upregulation makes BC cells more prone to estrogen uptake which may lead to the increased growth and proliferation of cancer cells (Lei et al., 2019). According to the literature, TFF3 acts as an oncogene because it regulates other genes (FOXA1, HER2, and AR) involved in EMT, thus promoting invasiveness, survival, and increased proliferation in multiple carcinomas such as gastric cancer, mammary carcinoma, and prostate cancer (Yuan et al., 2017). It has been reported that SYBU, a microtubule-associated protein, is overexpressed in hepatocellular carcinoma (HCC), which results in disrupted cell cycle and increased proliferation (Zheng and Yu, 2021). Breast tumor formation is increased by CBLC overexpression, which suppresses TGF-β (transforming growth factor beta). This results in the deactivation of its target Smad3 pathway which is responsible for proliferation, differentiation, and apoptosis (Kang et al., 2012). This study found that DNALI1 (*p*-value: 0.0000148), a flagellar protein, is overexpressed in BC, which has not been reported previously for any other carcinoma.

According to the literature, CENPW was downregulated in BC and HCC. It is involved in kinetochore organization and centromere complex assembly. This downregulation results in subsequent function disruption, resulting in chromosomal instability due to mis-segregation of chromosomes (Liu and Liu, 2022). It has been reported that EN1 is downregulated in lung cancer due to altered DNA methylation which promotes cell proliferation and differentiation (Jiang et al., 2017). The downregulation of A2ML1, a protease inhibitor, results in MAPK pathway mutation, which leads to apoptotic resistance and uncontrolled cell division in BC (Li et al., 2016). FOXC1 suppression induces ER-α expression in BC cells, which helps in increased estrogen uptake, resulting in the growth and proliferation of tumor cells (Wang et al., 2017). It has been reported that KRT16 is overexpressed in basal-like TNBC, along with increased expression of EMT-associated proteins. In contrast, in luminal A and B subtypes of BC which include ER^+ and PR^+ tumors, KRT16 expression was suppressed, but E-cadherin (CDH1), an EMT protein was overexpressed,

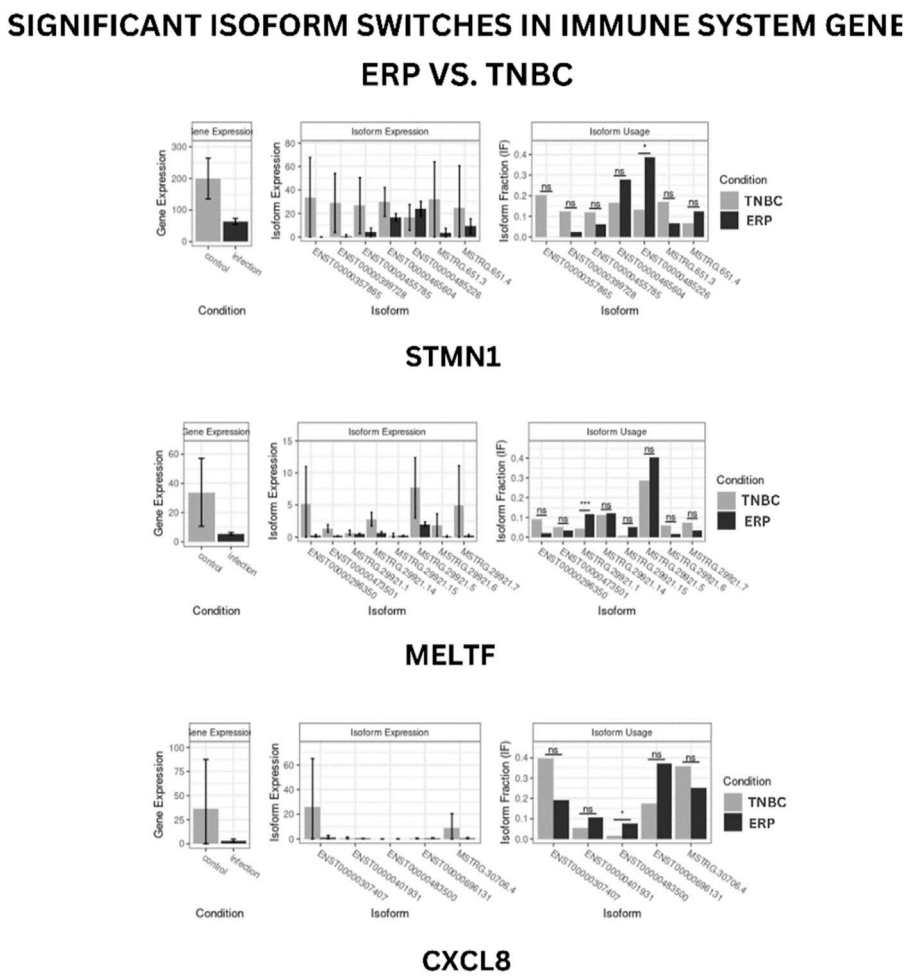


FIGURE 6
Isoform switch of differentially expressed immune genes. "*" represents significant isoform usage.

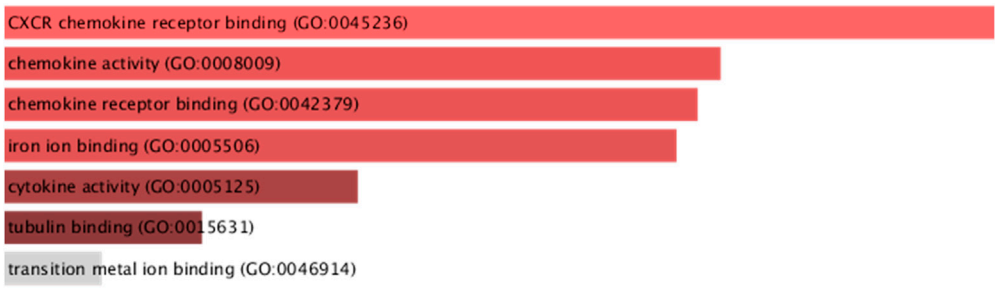


FIGURE 7
GO molecular functions of downregulated immune system genes (STMN1, MELTF, and CXCL8) having significant isoform switches (p-value <0.05).

leading to metastasis due to increased cellular motility (Elazezy et al., 2021). SLC7A5 has been reported to be overexpressed in TNBC due to its glutamine transporting activity to tumor cells for energy production—TNBC is thus glutamine dependent and requires glutaminase for its catabolism. In addition, cells with increased proliferation use transaminases to catabolize glutamate, in contrast to glutamate dehydrogenase (GLUD), to reduce ammonia production. However, ER^+ tumors are glutamine-independent and show increased GLUD expression (Wang et al., 2020). It has been reported that the downregulation of CDK6 also suppresses its interacting gene, RB1—a tumor suppressor gene. This results in dysregulated cell growth,

apoptosis, and increased proliferation in tumor cells (Knudsen et al., 2020). MELTF downregulation also dysregulates its interacting genes such as ACO2, a gene-encoding Krebs' cycle enzyme. The disruption of Krebs' cycle enzymes leads to the production of oncometabolites, which stabilize hypoxia-inducible factor 1 and activate cell growth signaling by regulating DNA methylation—crucial factors in cancer progression (Sajnani et al., 2017). CA9 suppression leads to the disruption of interacting genes such as HIF3A and EPAS1 which are involved in regulating hypoxic conditions. Such conditions are favorable for the increased proliferation of tumor cells (Jun et al., 2017). TMSB15A has been reported to be upregulated in TNBC; it plays a crucial role in the organization of the cytoskeleton, which is responsible for cancer cell motility and is involved in cancer metastasis (Darb-Esfahani et al., 2012).

KEGG pathway enrichment analysis revealed protein digestion and absorption, pathways in cancer, chemical carcinogenesis, and ECM–receptor interaction as upregulated pathways, while downregulated pathways include chemokine signaling, toll-like receptor signaling, cell cycle, bladder cancer, IL-17 signaling, cellular senescence, cytokine–cytokine receptor interaction, and p53 signaling. We found that the protein digestion and absorption pathway was upregulated, which demonstrates the use of proteins as an alternative fuel by tumor cells to fulfill their metabolic needs. This occurs due to the limited supply and metabolism of glucose by cancer cells (Lieu et al., 2020). Moreover, it was found that pathways in cancer were upregulated that involve the disruption of the ErbB, p-53-mediated apoptotic, and GSK3 signaling pathways, which are involved in DNA repair, cell growth, migration, differentiation, and metabolism (Yip and Papa, 2021). The upregulation of chemical carcinogenesis activates certain hormonal pathways that make mammary glands more susceptible to carcinogenesis due to altered DNA repair genes (Rodgers et al., 2018). Furthermore, upregulated ECM–receptor interaction results in interaction with HMMR and SDC1 genes, the dysregulation of which promotes BC cell motility and differentiation (Yeh et al., 2018). A downregulated chemokine signaling pathway and cytokine–cytokine receptor interaction cannot recruit immune cells (leukocytes) to the tumor microenvironment, thus resulting in tumor progression (Gil Del Alcazar et al., 2020). The downregulation of toll-like receptor signaling pathways results in non-recognition and the escape of cancer cells from the immune system, leading to the invasiveness, migration, and angiogenic potential of cancer cells (Javaid and Choi, 2020). The disruption of the cell cycle at the G2/M phase results in cells that contain damaged DNA and genomic instability, a hallmark of cancer (Thu et al., 2018). It has been reported that increased ER-α in BC cells suppresses bladder cancer cell growth by downregulating INPP4B which, in turn, suppresses the AKT signaling pathway (Hsu et al., 2014). Research has found that the IL-17 signaling pathway becomes downregulated due to increased estrogen receptor expression, resulting in dysregulated PD-1/PD-L1 and $CD8^+$ T cell expression—a suppressed immune response (Shuai et al., 2020). Furthermore, downregulated cellular senescence results in increased cell proliferation and tumor development (Milczarek, 2020). Moreover, a downregulated p53 signaling pathway cannot perform DNA damage repair and cell death, thereby facilitating the increased growth and metastasis of tumor cells (Marei et al., 2021).

According to the GTEx portal, cells express an average of 3.42 transcripts per gene (Tung et al., 2020). The expression dominance of major isoform transcripts compared to others from the same gene is crucial for normal cellular homeostasis (Hu et al., 2017). However, splicing regulation is often disrupted in cancer with a dominant expression of alternative transcripts in a tumor microenvironment (TME) which promotes switches that contribute to tumor progression and metastasis (Kahraman et al., 2020). The interaction pattern of the cancer-specific most dominant transcript (cMDT) differs from the generally expressed isoform of normal cells because of changes caused by alternative splicing; this could affect protein domains due to mutations caused by tumor and subsequent disruption in cancer-related pathways (Yang et al., 2016; Climente-González et al., 2017). According to the literature, apoptosis, ubiquitin, signaling, and spliceosomes were the most disrupted protein interactions (Kahraman et al., 2020). It has been reported that isoform switching leads to the loss of the DNA sequence that encodes for protein domains, promoting functional loss. The subsequent switches have functional consequences for cancer development and progression (Vitting-Seerup and Sandelin, 2017).

The present study has identified the differential usage of transcript isoforms among ERP and TNBC. It revealed isoforms that are significantly expressed and used by shortlisted downregulated genes (STMN1, MELTF, and CXCL8). CXCL8 encodes a protein that is involved in chemotaxis (Łukaszewicz-Zając et al., 2020). It transcribes five transcripts; among them, only one isoform transcript was significantly used. Differential gene expression plotting shows that CXCL8 is downregulated in ERP (Figure 6). On the other hand, there is in isoform usage an increased use of isoform ENST00000483500 in ERP. However, this isoform is non-coding due to retained introns and the unavailability of any domain, leading to functional loss of CXCL8 and immune suppression as a consequence of the non-recruitment of macrophages and neutrophils to the TME (Xiong et al., 2022). STMN1, a cytosolic phosphoprotein, is involved in microtubule destabilization by regulating the microtubule filament system and signal transduction (Bao et al., 2017). It transcribes four known and two novel isoforms (Figure 6). The isoform usage plot indicates that a single isoform (ENST00000485226) is significantly used in ERP, as compared to TNBC. However, it lacks a domain, which results in its non-coding behavior. Furthermore, STMN1 is repressed in ERP, as shown in differential gene expression plotting which promotes ERP progression due to the disruption of microtubules and their subsequent role in the growth of immune cells such as T-cells and natural killer cells (Zhang et al., 2022). MELTF, a cell surface glycoprotein, is involved in cellular iron uptake (Sawaki et al., 2019). MELTF could transcribe six novel and two known isoforms (Figure 6); however, increased use of the single isoform MSTRG.29921.1 has been identified in ERP. Moreover, this isoform contains nonsense codons that prematurely terminate translation—nonsense-mediated decay (NMD). Differential gene expression plotting also shows that MELTF is downregulated in ERP, leading to the decreased proliferation and maturation of immune cells such as lymphocytes due to decreased iron uptake (Roemhild et al.,

2021). Gene Ontology analysis of shortlisted immune genes revealed that CXCR chemokine receptor binding, iron ion binding, and cytokine activity are the most dysregulated molecular functions (Figure 7). These functions mediate immune response by recruiting immune cells such as monocytes, T cells, lymphocytes, and natural killer cells, and assist their growth and proliferation. The downregulation of these functions promotes ERP BC due to suppressed immune response in TME (Bates et al., 2018).

This research therefore provides key insights into the genes that are differentially expressed in ERP. Moreover, DNALI1 is a novel gene that has not been previously reported and is involved in ERP BC development. Furthermore, the identification of three immune system-related genes (STMN1, MELTF, and CXCL8) reveals that the dysregulation of the immune system due to isoform switching is the major factor in ERP BC development and progression. Downregulation and isoform switching of key immune system genes suggest BC progression and possible metastasis due to the non-recruitment of cytokines in the TME.

Conclusion

ERP BC is characterized by the growth of tumor cells in response to estrogen hormone. The dysregulation of gene expression results in the development of significant biological changes that are key features of multiple human carcinomas such as prostate cancer, gastric cancer, hepatocellular carcinoma, and lung cancer. In this study, 9,771 DEGs were identified; among these, 86 genes were upregulated and 117 were downregulated. Six genes (FOXA1, RHOB, AGR2, ESR1, CBLC, and FOXC1) were found to be significantly associated with the development and progression of ERP BC. This study also identified a novel set of genes (DNALI1, TMSB15A, AR, TFF3, SYBU, CENPW, EN1, CDK6, MELTF, and CA9) not previously reported positive for estrogen receptors but that has been reported in other carcinomas. Moreover, alternative splicing and subsequent isoform expression in three downregulated immune system genes (STMN1, MELTF, and CXCL8) had been identified that were mainly responsible for ERP progression due to suppression of the immune system and the non-recruitment of cytokines against cancer cells. It was found that CXCR chemokine receptor binding, iron ion binding, and cytokine activity were the most dysregulated functions due to immune system suppression. This study reveals that dysregulation of the immune system due to isoform switching is the major factor in ERP BC development and progression. Therefore, these crucial immune system genes should be targeted as therapeutic biomarkers.

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.

Author contributions

Conceptualization, methodology, formal analysis, investigation, and validation: MK, WH, NA, BJ, IS, AhA, MA, SA, AS, DF, SA, YM, AmA, A-UR, and BH; data curation, writing—original draft preparation, and writing—review and editing: MK, WH, NA, BJ, and IS. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors extend their appreciation to the Researchers supporting project number (RSP2023R470), King Saud University, Riyadh, Saudi Arabia.

Conflict of interest

Author A-UR was employed by the company Keystone Pharmacogenomics, LLC.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1230998/full#supplementary-material

References

Ann, P., Seagle, B.-L. L., Shilpi, A., Kandpal, M., and Shahabi, S. (2018). Association of increased primary breast tumor AGR2 with decreased disease-specific survival. *Oncotarget* 9, 23114–23125. doi:10.18632/oncotarget.25225

Bao, P., Yokobori, T., Altan, B., Iijima, M., Azuma, Y., Onozato, R., et al. (2017). High STMN1 expression is associated with cancer progression and chemo-resistance in lung squamous cell carcinoma. *Ann. Surg. Oncol.* 24, 4017–4024. doi:10.1245/s10434-017-6083-0

Baralle, F. E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* 18, 437–451. doi:10.1038/nrm.2017.27

Bates, J. P., Derakhshandeh, R., Jones, L., and Webb, T. J. (2018). Mechanisms of immune evasion in breast cancer. *BMC cancer* 18, 556–614. doi:10.1186/s12885-018-4441-3

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi:10.1093/bioinformatics/bty560

Chen, Y.-L., Wang, K., Xie, F., Zhuo, Z.-L., Liu, C., Yang, Y., et al. (2022). Novel biomarkers identified in triple-negative breast cancer through RNA-sequencing. *Clin. Chim. Acta* 531, 302–308. doi:10.1016/j.cca.2022.04.990

Climente-González, H., Porta-Pardo, E., Godzik, A., and Eyras, E. (2017). The functional impact of alternative splicing in cancer. *Cell Rep.* 20, 2215–2226. doi:10.1016/j.celrep.2017.08.012

Costa-Silva, J., Domingues, D., and Lopes, F. M. (2017). RNA-seq differential expression analysis: an extended review and a software tool. *PloS one* 12, e0190152. doi:10.1371/journal.pone.0190152

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10, giab008. doi:10.1093/gigascience/giab008

Darb-Esfahani, S., Kronenwett, R., Von Minckwitz, G., Denkert, C., Gehrmann, M., Rody, A., et al. (2012). Thymosin beta 15A (TMSB15A) is a predictor of chemotherapy response in triple-negative breast cancer. *Br. J. cancer* 107, 1892–1900. doi:10.1038/bjc.2012.475

Elazezy, M., Schwentesius, S., Stegat, L., Wikman, H., Werner, S., Mansour, W. Y., et al. (2021). Emerging insights into keratin 16 expression during metastatic progression of breast cancer. *Cancers* 13, 3869. doi:10.3390/cancers13153869

Fatima, F., Saleem, S., Hameed, A., Haider, G., Ali Zaidi, S. A., Kanwal, M., et al. (2019). Association analysis and allelic distribution of deletion in CC chemokine receptor 5 gene (CCR5Δ32) among breast cancer patients of Pakistan. *Mol. Biol. Rep.* 46, 2387–2394. doi:10.1007/s11033-019-04699-6

Formaggio, N., Rubin, M. A., and Theurillat, J.-P. (2021). Loss and revival of androgen receptor signaling in advanced prostate cancer. *Oncogene* 40, 1205–1216. doi:10.1038/s41388-020-01598-0

Frazee, A. C., Pertea, G., Jaffe, A. E., Langmead, B., Salzberg, S. L., and Leek, J. T. (2015). Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat. Biotechnol.* 33, 243–246. doi:10.1038/nbt.3172

Frazee, A. C., Pertea, G., Jaffe, A. E., Langmead, B., Salzberg, S. L., and Leek, J. T. (2014). Flexible isoform-level differential expression analysis with Ballgown. *Biorxiv*, 003665. doi:10.1101/003665

Gil Del Alcazar, C. R., Alečković, M., and Polyak, K. (2020). Immune escape during breast tumor progression. *Cancer Immunol. Res.* 8, 422–427. doi:10.1158/2326-6066.CIR-19-0786

Gonzalez, H., Hagerling, C., and Werb, Z. (2018). Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes and Dev.* 32, 1267–1284. doi:10.1101/gad.314617.118

Guo, J., Gong, G., and Zhang, B. (2017). Screening and identification of potential biomarkers in triple-negative breast cancer by integrated analysis. *Oncol. Rep.* 38, 2219–2228. doi:10.3892/or.2017.5911

He, Y., Wang, L., Wei, T., Xiao, Y.-T., Sheng, H., Su, H., et al. (2021). FOXA1 overexpression suppresses interferon signaling and immune response in cancer. *J. Clin. Investigation* 131, e147025. doi:10.1172/JCI147025

Hsu, I., Yeh, C.-R., Slavin, S., Miyamoto, H., Netto, G. J., Muyan, M., et al. (2014). Estrogen receptor alpha prevents bladder cancer via INPP4B inhibited akt pathway *in vitro* and *in vivo*. *Oncotarget* 5, 7917–7935. doi:10.18632/oncotarget.1421

Hu, J., Boritz, E., Wylie, W., and Douek, D. C. (2017). Stochastic principles governing alternative splicing of RNA. *PLoS Comput. Biol.* 13, e1005761. doi:10.1371/journal.pcbi.1005761

Ianevski, A., Giri, A. K., and Aittokallio, T. (2022). Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nat. Commun.* 13, 1246. doi:10.1038/s41467-022-28803-w

Javaid, N., and Choi, S. (2020). Toll-like receptors from the perspective of cancer treatment. *Cancers* 12, 297. doi:10.3390/cancers12020297

Jiang, C.-L., He, S.-W., Zhang, Y.-D., Duan, H.-X., Huang, T., Huang, Y.-C., et al. (2017). Air pollution and DNA methylation alterations in lung cancer: A systematic and comparative study. *Oncotarget* 8, 1369–1391. doi:10.18632/oncotarget.13622

Jun, J. C., Rathore, A., Younas, H., Gilkes, D., and Polotsky, V. Y. (2017). Hypoxia-inducible factors and cancer. *Curr. Sleep. Med. Rep.* 3, 1–10. doi:10.1007/s40675-017-0062-7

Kahraman, A., Karakulak, T., Szklarczyk, D., and Von Mering, C. (2020). Pathogenic impact of transcript isoform switching in 1,209 cancer samples covering 27 cancer types using an isoform-specific interaction network. *Sci. Rep.* 10, 14453. doi:10.1038/s41598-020-71221-5

Kang, J. M., Park, S., Kim, S. J., Hong, H., Jeong, J., Kim, H., et al. (2012). CBL enhances breast tumor formation by inhibiting tumor suppressive activity of TGF-β signaling. *Oncogene* 31, 5123–5131. doi:10.1038/onc.2012.18

Kim, D., Langmead, B., and Salzberg, S. L. (2015). Hisat: A fast spliced aligner with low memory requirements. *Nat. methods* 12, 357–360. doi:10.1038/nmeth.3317

Knudsen, E. S., Nambiar, R., Rosario, S. R., Smiraglia, D. J., Goodrich, D. W., and Witkiewicz, A. K. (2020). Pan-cancer molecular analysis of the RB tumor suppressor pathway. *Commun. Biol.* 3, 158. doi:10.1038/s42003-020-0873-9

Lamb, C. A., Vanzulli, S., and Lanari, C. L. M. (2019). Hormone receptors in breast cancer: more than estrogen receptors. *Med. (B Aires)* 79, 540–545.

Lei, J. T., Gou, X., Seker, S., and Ellis, M. J. (2019). ESR1 alterations and metastasis in estrogen receptor positive breast cancer. *J. cancer metastasis Treat.* 5, 38. doi:10.20517/2394-4722.2019.12

Li, G., Hu, J., and Hu, G. (2017). Biomarker studies in early detection and prognosis of breast cancer. *Transl. Res. Breast Cancer Biomark. Diagnosis, Target. Ther. Approaches Precis. Med.* 1026, 27–39. doi:10.1007/978-981-10-6020-5_2

Li, Y., Tang, X.-Q., Bai, Z., and Dai, X. (2016). Exploring the intrinsic differences among breast tumor subtypes defined using immunohistochemistry markers based on the decision tree. *Sci. Rep.* 6, 35773. doi:10.1038/srep35773

Lieu, E. L., Nguyen, T., Rhyne, S., and Kim, J. (2020). Amino acids in cancer. *Exp. Mol. Med.* 52, 15–30. doi:10.1038/s12276-020-0375-3

Liu, X., and Liu, Y. (2022). Comprehensive analysis of the expression and prognostic significance of the CENP family in breast cancer. *Int. J. General Med.* 15, 3471–3482. doi:10.2147/IJGM.S354200

Łukaszewicz-Zając, M., Pączek, S., and Mroczko, B. (2020). The significance of chemokine CXCL-8 in esophageal carcinoma. *Archives Med. Sci.* 13. doi:10.5114/aoms.2017.71933

Marei, H. E., Althani, A., Afifi, N., Hasan, A., Caceci, T., Pozzoli, G., et al. (2021). p53 signaling in cancer progression and therapy. *Cancer Cell Int.* 21, 703–715. doi:10.1186/s12935-021-02396-8

Médale-Giamarchi, C., Lajoie-Mazenc, I., Malissein, E., Meunier, E., Couderc, B., Bergé, Y., et al. (2013). RhoB modifies estrogen responses in breast cancer cells by influencing expression of the estrogen receptor. *Breast Cancer Res.* 15, R6–R13. doi:10.1186/bcr3377

Milczarek, M. (2020). The premature senescence in breast cancer treatment strategy. *Cancers* 12, 1815. doi:10.3390/cancers12071815

Munir, M. T., Kay, M. K., Kang, M. H., Rahman, M. M., Al-Harrasi, A., Choudhury, M., et al. (2021). Tumor-associated macrophages as multifaceted regulators of breast tumor growth. *Int. J. Mol. Sci.* 22, 6526. doi:10.3390/ijms22126526

Nisar, M., Paracha, R. Z., Arshad, I., Adil, S., Zeb, S., Hanif, R., et al. (2021). Integrated analysis of microarray and RNA-Seq data for the identification of hub genes and networks involved in the pancreatic cancer. *Front. Genet.* 12, 663787. doi:10.3389/fgene.2021.663787

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. doi:10.1038/nprot.2016.095

Rodgers, K. M., Udesky, J. O., Rudel, R. A., and Brody, J. G. (2018). Environmental chemicals and breast cancer: an updated review of epidemiological literature informed by biological mechanisms. *Environ. Res.* 160, 152–182. doi:10.1016/j.envres.2017.08.045

Roemhild, K., Von Maltzahn, F., Weiskirchen, R., Knüchel, R., Von Stillfried, S., and Lammers, T. (2021). Iron metabolism: pathophysiology and pharmacology. *Trends Pharmacol. Sci.* 42, 640–656. doi:10.1016/j.tips.2021.05.001

Rostovskaya, M., Andrews, S., Reik, W., and Rugg-Gunn, P. J. (2022). Amniogenesis occurs in two independent waves in primates. *Cell Stem Cell* 29, 744–759.e6. doi:10.1016/j.stem.2022.03.014

Sajnani, K., Islam, F., Smith, R. A., Gopalan, V., and Lam, A. K.-Y. (2017). Genetic alterations in Krebs cycle and its impact on cancer pathogenesis. *Biochimie* 135, 164–172. doi:10.1016/j.biochi.2017.02.008

Sawaki, K., Kanda, M., Umeda, S., Miwa, T., Tanaka, C., Kobayashi, D., et al. (2019). Level of melanotransferrin in tissue and sera serves as a prognostic marker of gastric cancer. *Anticancer Res.* 39, 6125–6133. doi:10.21873/anticanres.13820

Segovia-Mendoza, M., and Morales-Montor, J. (2019). Immune tumor microenvironment in breast cancer and the participation of estrogen and its receptors in cancer physiopathology. *Front. Immunol.* 10, 348. doi:10.3389/fimmu.2019.00348

Shuai, C., Yang, X., Pan, H., and Han, W. (2020). Estrogen receptor downregulates expression of PD-1/PD-L1 and infiltration of CD8+ T cells by inhibiting IL-17 signaling transduction in breast cancer. *Front. Oncol.* 10, 582863. doi:10.3389/fonc.2020.582863

Shumate, A., Wong, B., Pertea, G., and Pertea, M. (2022). Improved transcriptome assembly using a hybrid of long and short reads with StringTie. *PLoS Comput. Biol.* 18, e1009730. doi:10.1371/journal.pcbi.1009730

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA a cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Thu, K., Soria-Bretones, I., Mak, T., and Cescon, D. (2018). Targeting the cell cycle in breast cancer: towards the next phase. *Cell Cycle* 17, 1871–1885. doi:10.1080/15384101.2018.1502567

Tung, K.-F., Pan, C.-Y., Chen, C.-H., and Lin, W.-C. (2020). Top-ranked expressed gene transcripts of human protein-coding genes investigated with GTEx dataset. *Sci. Rep.* 10, 16245. doi:10.1038/s41598-020-73081-5

Vitting-Seerup, K., and Sandelin, A. (2017). The landscape of isoform switches in human cancers. *Mol. Cancer Res.* 15, 1206–1220. doi:10.1158/1541-7786.MCR-16-0459

Wang, J., Xu, Y., Li, L., Wang, L., Yao, R., Sun, Q., et al. (2017). FOXC 1 is associated with estrogen receptor alpha and affects sensitivity of tamoxifen treatment in breast cancer. *Cancer Med.* 6, 275–287. doi:10.1002/cam4.990

Wang, Z., Jiang, Q., and Dong, C. (2020). Metabolic reprogramming in triple-negative breast cancer. *Cancer Biol. Med.* 17, 44–59. doi:10.20892/j.issn.2095-3941.2019.0210

Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J., Evangelista, J. E., Jenkins, S. L., et al. (2021). Gene set knowledge discovery with Enrichr. *Curr. Protoc.* 1, e90. doi:10.1002/cpz1.90

Xiong, X., Liao, X., Qiu, S., Xu, H., Zhang, S., Wang, S., et al. (2022). CXCL8 in tumor biology and its implications for clinical translation. *Front. Mol. Biosci.* 9, 723846. doi:10.3389/fmolb.2022.723846

Yang, X., Coulombe-Huntington, J., Kang, S., Sheynkman, G. M., Hao, T., Richardson, A., et al. (2016). Widespread expansion of protein interaction capabilities by alternative splicing. *Cell* 164, 805–817. doi:10.1016/j.cell.2016.01.029

Yeh, M.-H., Tzeng, Y.-J., Fu, T.-Y., You, J.-J., Chang, H.-T., Ger, L.-P., et al. (2018). Extracellular matrix–receptor interaction signaling genes associated with inferior breast cancer survival. *Anticancer Res.* 38, 4593–4605. doi:10.21873/anticanres.12764

Yin, L., Duan, J.-J., Bian, X.-W., and Yu, S.-C. (2020). Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Res.* 22, 61–13. doi:10.1186/s13058-020-01296-5

Yip, H. Y. K., and Papa, A. (2021). Signaling pathways in cancer: therapeutic targets, combinatorial treatments, and new developments. *Cells* 10, 659. doi:10.3390/cells10030659

Yuan, Z., Chen, D., Chen, X., Yang, H., and Wei, Y. (2017). Overexpression of trefoil factor 3 (TFF3) contributes to the malignant progression in cervical cancer cells. *Cancer Cell Int.* 17, 7–13. doi:10.1186/s12935-016-0379-1

Zhang, E.-D., Li, C., Fang, Y., Li, N., Xiao, Z., Chen, C., et al. (2022). STMN1 as a novel prognostic biomarker in HCC correlating with immune infiltrates and methylation. *World J. Surg. Oncol.* 20, 301. doi:10.1186/s12957-022-02768-y

Zheng, C., and Yu, S. (2021). Expression and gene regulatory network of SNHG1 in hepatocellular carcinoma. *BMC Med. Genomics* 14, 28–10. doi:10.1186/s12920-021-00878-2

Zhu, S.-Y., and Yu, K.-D. (2022). Breast cancer vaccines: disappointing or promising? *Front. Immunol.* 13, 190. doi:10.3389/fimmu.2022.828386

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 16 October 2023
DOI 10.3389/fphar.2023.1267186

OPEN ACCESS

EDITED BY
Linhui Wang,
Second Military Medical University, China

REVIEWED BY
Zhifang Zhang,
City of Hope National Medical Center, United States
Giorgia Colombo,
University of Eastern Piedmont, Italy
Xiaogang Zhong,
People's Hospital of Guangxi Zhuang Autonomous Region, China
Shuxia Ma,
Jiamusi University, China

*CORRESPONDENCE
Linghua Zhu,
3198020@zju.edu.cn

RECEIVED 26 July 2023
ACCEPTED 18 September 2023
PUBLISHED 16 October 2023

CITATION
Luo P, Chen G, Shi Z, Yang J, Wang X, Pan J and Zhu L (2023), Comprehensive multi-omics analysis of tryptophan metabolism-related gene expression signature to predict prognosis in gastric cancer.
Front. Pharmacol. 14:1267186.
doi: 10.3389/fphar.2023.1267186

Comprehensive multi-omics analysis of tryptophan metabolism-related gene expression signature to predict prognosis in gastric cancer

Peng Luo, Guojun Chen, Zhaoqi Shi, Jin Yang, Xianfa Wang, Junhai Pan and Linghua Zhu*

Department of General Surgery, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China

Introduction: The 5-year survival of gastric cancer (GC) patients with advanced stage remains poor. Some evidence has indicated that tryptophan metabolism may induce cancer progression through immunosuppressive responses and promote the malignancy of cancer cells. The role of tryptophan and its metabolism should be explored for an in-depth understanding of molecular mechanisms during GC development.

Material and methods: We utilized the Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) dataset to screen tryptophan metabolism-associated genes via single sample gene set enrichment analysis (ssGSEA) and correlation analysis. Consensus clustering analysis was employed to construct different molecular subtypes. Most common differentially expressed genes (DEGs) were determined from the molecular subtypes. Univariate cox analysis as well as lasso were performed to establish a tryptophan metabolism-associated gene signature. Gene Set Enrichment Analysis (GSEA) was utilized to evaluate signaling pathways. ESTIMATE, ssGSEA, and TIDE were used for the evaluation of the gastric tumor microenvironment.

Results: Two tryptophan metabolism-associated gene molecular subtypes were constructed. Compared to the C2 subtype, the C1 subtype showed better prognosis with increased CD4 positive memory T cells as well as activated dendritic cells (DCs) infiltration and suppressed M2-phenotype macrophages inside the tumor microenvironment. The immune checkpoint was downregulated in the C1 subtype. A total of eight key genes, EFNA3, GPX3, RGS2, CXCR4, SGCE, ADH4, CST2, and GPC3, were screened for the establishment of a prognostic risk model.

Conclusion: This study concluded that the tryptophan metabolism-associated genes can be applied in GC prognostic prediction. The risk model established in the current study was highly accurate in GC survival prediction.

KEYWORDS

tryptophan metabolism, tumor microenvironment, immune cell infiltration, prognosis, gastric cancer

Abbreviations: CNV, copy number variation; CDF, cumulative distribution function; FDR, false discovery rate; FC, fold change; GC, gastric cancer; GSEA, Gene Set Enrichment Analysis; ICB, immune checkpoint blockade; IDO, indoleamine-2,3,-dioxygenase; Kyn, kynurenine; KP, signaling pathway; MDSCs, myeloid-derived suppressor cells; ROC, receiver operating characteristic; WT, wild-type.

Introduction

Gastric cancer (GC) has been recognized as a main cancer type that leads to cancer-associated mortality worldwide, with millions of new cases being diagnosed annually (Lambert et al., 2002; Brenner et al., 2009). Inflammation is typically related to GC development with both acute and chronic inflammatory cells, resulting in aggressive damage of gastric mucosa and ultimately transformation to cancer tissue (Demaria et al., 2010; Wang et al., 2014). In most cases, the 5-year survival of late-stage GC patients remains poor although current combination therapy of chemotherapy, radiation, and surgery has improved (Li et al., 2009; Song et al., 2017). Cancer immunotherapy has emerged recently as a promising and powerful cancer therapy that drives the patient's own immune system against cancer (De Felice et al., 2015; Whiteside et al., 2016; Bruni et al., 2020). The combination therapy involving first-line Opdivo (nivolumab) immunotherapy and chemotherapy was approved in 2021 for advanced or metastatic GC patients (Twomey and Zhang, 2021; Yoneda et al., 2021). The mechanism of GC progression and its related tumor immune microenvironment should be analyzed to develop novel cancer immunotargets against GC.

As an essential amino acid, L-tryptophan serves as an indispensable material and regulates protein synthesis during cell proliferation (Conejos et al., 2021). Tryptophan and its metabolites play critical roles in various physiological processes (Hoseini et al., 2019; Conejos et al., 2021). Most free tryptophan is a biologically active substrate for the function of the kynurenine (Kyn) signaling pathway (KP) that produces several metabolites related to the immune response and neurotransmission (Platten et al., 2019; Tanaka et al., 2021). Many studies have focused on the imbalances in tryptophan metabolism by targeting the KP, especially ryptophan-2,3-dioxygenase (TDO), indoleamine-2,3,-dioxygenase 1 (IDO1), and IDO2 (Platten et al., 2019; Yao et al., 2021). It has been demonstrated that the tryptophan depletion by IDO1 and IDO2 was highly associated with cellular function and survival (Zhai et al., 2018; Souissi et al., 2022). However, phase III clinical trials of IDO inhibitors against cancers were disappointing, although they did show promising outcome in early-stage cancer immunotherapy (Günther et al., 2019; Chen et al., 2021). Some evidence has indicated that tryptophan metabolism may induce cancer development and progression through the inhibition of immune responses in the tumor site and promoting the malignancy of cancer cells (Platten et al., 2019). Although it is still unclear whether KP-related enzymes are essential for cancer progression, the role of tryptophan and its metabolism should be explored for in-depth understanding of molecular mechanism during GC development.

Thus, in this study, we used genes significantly associated with tryptophan metabolism pathway score to determine the molecular subtypes. Specifically, consensus clustering followed by subsequent comparison of clinical signatures, different signaling pathways, and immune-related properties among different subtypes will be performed. We then identified genes associated with the tryptophan metabolism phenotype by differential expression analysis and LASSO. We further predict GC patients' outcome by constructing a risk model, which is also used for personalized treatment.

Materials and methods

Data sources, collection, and preprocessing

The data related to mutation, copy number variation, and RNA-Seq profile of TCGA-STAD via TCGA GDC API were downloaded first (http://cancergenome.nih.gov/). GSE66229 expression data from the GEO database were obtained (https://www.ncbi.nlm.nih.gov/geo/). Samples need to be processed as follows (Brenner et al., 2009): excision of samples of primary tumors (Lambert et al., 2002); removal of incomplete samples with clinical characteristic information to ensure that samples have complete clinical prognostic information and transcriptome expression data. TCGA-STAD was the training set and the GSE66229 dataset was the validation set. We excluded samples without survival time or status. Finally, a sum of 350 primary tumor samples together with 32 normal matches were screened. We kept the protein-encoding genes for TCGA RNA-seq data analysis. Meanwhile, all data were log2 transformed, and RNA expression data were normalized. For the GEO dataset, 300 GC samples were finally screened. Specifically, the annotation information for each chip platform was acquired and subsequently utilized to map probes with all the detected genes. Then, we removed the probes that matched more than one gene. When one gene can be matched by more than one probes, the mean value of the gene expression was calculated and set as the value for that specific gene.

The tryptophan metabolism-related gene information is derived from a specific pathway named "KEGG_TRYPTOPHAN_METABOLISM", which can be found in the Molecular Signatures Database (MSigDB) (https://www.gsea-msigdb.org/).

Molecular subtypes of tryptophan metabolism-associated genes

We constructed a consensus matrix and clustered the samples through consensus clustering (Wilkerson and Hayes, 2010). The transcriptional expressions of prognostic genes, which are involved in the tryptophan metabolism score, were subsequently evaluated to determine the molecular subtypes. We executed 500 bootstraps employing the "hc" algorithm and "pearson" as the metric distance. Each bootstrap process involved around 80% of the training set patients. The cluster number was set within a range of 2–10, and we determined the optimal classification via cumulative distribution function (CDF). Notably, the consistency of CDF was carefully evaluated when constructing various molecular subtypes for GC samples.

Risk model

The distinctly expressed genes were identified among the molecular subtypes, and then the distinctly expressed genes associated with statistical significance for prognosis were

10.3389/fphar.2023.1267186

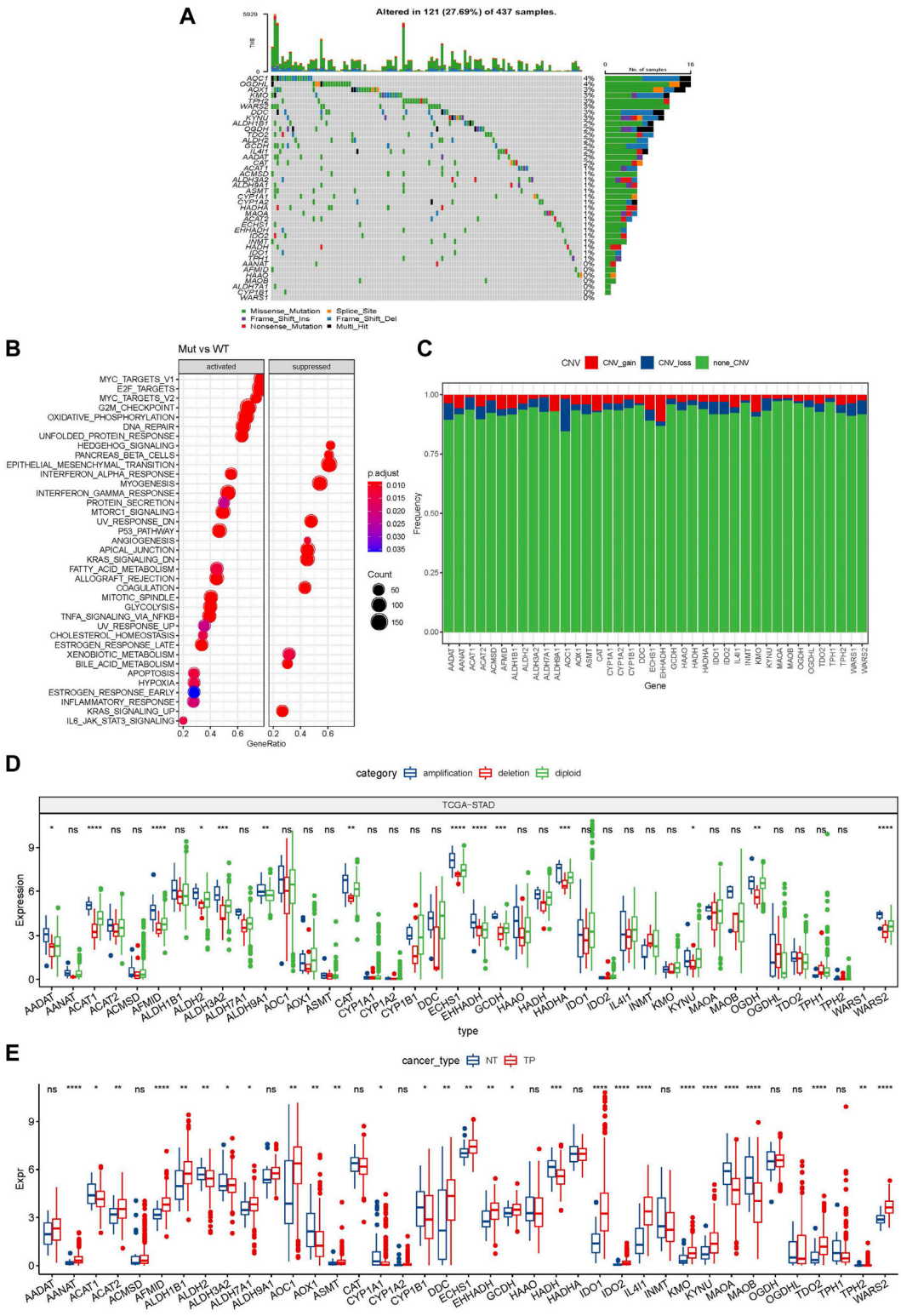


FIGURE 1
Genetic variation landscape of tryptophan metabolism-related genes in GC. **(A)** Mutational map of genes associated with tryptophan metabolism in primary tumor samples; **(B)** Primary tumor samples GSEA analysis between mutated and non-mutated groups; **(C)** CNVs of genes associated with tryptophan metabolism within primary tumor samples; **(D)** Gene expression between different types of copy number variation in primary tumor samples; **(E)** Differential expression of tryptophan metabolism-related genes between tumor and normal tissue samples. * means *p*-value less than 0.05; ** means *p*-value less than 0.01; *** means *p*-value less than 0.001, and **** means *p*-value less than 0.0001. ns means there is no significant difference between the two groups. The same statistical criteria apply to the following figures.

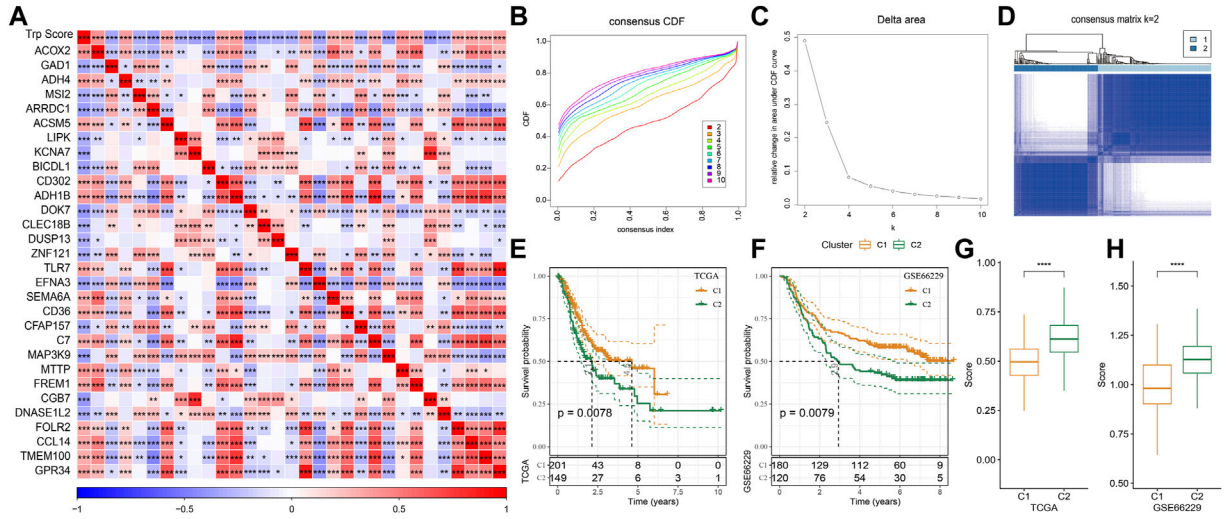


FIGURE 2
Molecular typing on genes that are associated with tryptophan metabolism. **(A)** The result of correlation analysis on genes that are significantly and prognostically associated with tryptophan metabolic pathway scores is summarized into a heatmap; **(B)** Cumulative distribution function curves for samples that are from TCGA cohorts; **(C)**. Curves for the delta area under the cumulative distribution function curves for samples that are from TCGA cohorts; **(D)**. The second sample clustering (k = 2) is displayed as a heatmap; **(E)**. The prognosis of two TCGA subtypes is displayed as a KM curve; **(F)** The prognosis of the two GSE66229 cohort subtypes is displayed as a KM curve; **(G,H)**: The statistical differences of tryptophan metabolism scores among different molecular subtypes in the TCGA cohort **(G)** and in the GSE66229 cohort **(H)** were analyzed by one-way ANOVA.

sorted out (|Log2FC|>1 and FDR<0.05). The gene numbers were further condensed by Lasso regression, and the prognostic genes potentially contributing to the tryptophan metabolism phenotype were filtered. Then, the risk score for each patient was calculated by the formula:

RiskScore = Σβi×Expi, where Expi is the expression value of each prognostic gene that determines tryptophan metabolism phenotype. The Cox regression coefficient of corresponding prognostic gene is referred to as β. Based on the calculated numbers, samples were then distributed into two subgroups, that is, RiskScore-high and RiskScore-low groups, with the threshold set as "0". The commonly used Kaplan-Meier method was utilized to analyze patient survival, and the patients' statistical significance was calculated via a log-rank test.

GSEA

In different molecular subtypes, pathways of different biological processes were investigated by performing GSEA for signaling pathway analysis based on the candidate gene sets from the KEGG/hallmark (https://www.gsea-msigdb.org/gsea/index.jsp).

Calculation of cell invasion abundance in tumor microenvironment

Relative abundance of 22 immune cells in tumor tissues and percentage of immune cells was determined using CIBERSORT algorithm (https://cibersort.stanford.edu/) and ESTIMATE software (Wilkerson and Hayes, 2010), respectively. A total of 28 immune cells were scored using ssGSEA function (Charoentong et al., 2017).

Prediction of patients' responsiveness to immunotherapy

The TIDE, as a widely used algorithm for immune checkpoint blockade (ICB) responsiveness prediction (Thorsson et al., 2018), was applied to verify the prediction of clinical responsiveness to ICB, which evaluated various cell types including tumor-related fibroblasts, which are responsible for excessive extracellular matrix deposition, immunosuppressive cells such as the M2 subtype of tumor-associated macrophages, and myeloid-derived suppressor cells that suppressed the T cell infiltration inside the tumor microenvironment, and two distinct mechanisms involved in escaping immune surveillance, including the score determining the dysfunctionality of tumor-infiltrating cytotoxic T lymphocytes (CTLs) and the score showing the rejection of CTLs by immunosuppressive factors.

Results

Genetic variation landscape of tryptophan metabolism-associated genes in GC

A total of 40 genes were involved in tryptophan metabolism. To determine the genetic alteration of tryptophan metabolism in GC, mutation frequency of cells was analyzed among the tryptophan metabolism-associated genes. Among the

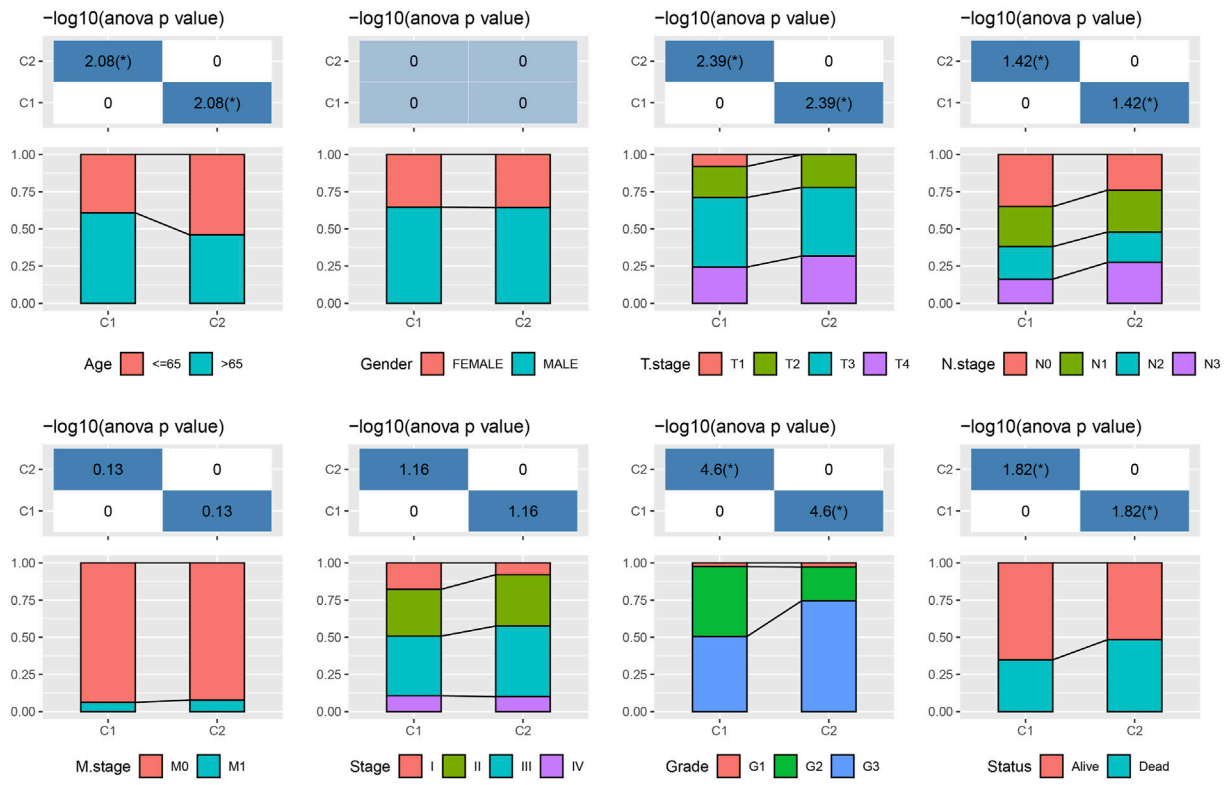


FIGURE 3
Clinical information distribution of molecular subtypes for the TCGA cohort.

437 tumor samples, 121 (27.69%) samples had tryptophan metabolism mutations. As shown in Figure 1A, AOX1 and OGDHL genes had the highest mutation frequencies, and no mutation was found in the WARS1 gene. To understand the effect of mutations on tryptophan metabolism-related genes, we analyzed the biological signaling pathways in wild-type (WT) and mutant (Mut) groups through GSEA enrichment analysis. It shows that tumor-associated pathways, including TNFA_SIGNALING_VIA_NFKB, P53_PATHWAY, and MYC_TARGET, were enriched in the mutant group (Sanchez-Vega et al., 2018) (Figure 1B). We then examined the somatic copy number variations of these tryptophan metabolism-associated genes in GC tumor samples and discovered a lower frequency of copy number variation (CNV) deletion or amplification (Figure 1C). In order to explore mRNA expression of CNV value in tumor tissue, the samples were distributed into different groups relying on CNV value, including increase and loss of CNV, as well as no obvious variation in CNV. Comparison of the expression of genes correlating with tryptophan metabolism between these groups showed that patients with CNV gain had a higher mRNA expression level compared to those with CNV loss (Figure 1D). To determine the expression of the tryptophan metabolism-associated genes between tumor tissue samples and adjacent normal tissues. As indicated in Figure 1E, most tryptophan metabolism genes were significantly differentially expressed, such as AANAT, AFMID, HADH, IDO1, IDO2, IL4I1, KMO, KYNU, MAOA, MAOB, TDO2, and WARS2.

Molecular subtyping based on genes related to tryptophan metabolism

The tryptophan metabolism score in the TCGA dataset was calculated by the ssGSEA, and then Pearson was used to estimate the relationship between the protein-encoding genes and the tryptophan metabolism score. A total, 30 prognostic genes associated with tryptophan metabolism score were screened. Figure 2A showed the correlation between the 30 genes and tryptophan metabolism scores. We classified patients based on the consensus clustering on 30 prognosis-correlated gene expression profiles and selected the optimal cluster number based on the CDF. With relatively stable clustering results shown in Figures 2B, C, we finally chose k = 2 to acquire two molecular subtypes (Figure 1D). We further performed the prognostic analysis of these two molecular subtypes. As shown in Figure 2E, we found that the overall survival of C1 was significantly better than that of C2. In addition, a sample of GC patients from the GSE66229 dataset showed similar results. This suggests that GC patients in C1 would have a better prognosis relative to the C2 subtype (Figure 2F). Meanwhile, we determined the tryptophan metabolism scores of each sample in the TCGA and GSE66229 datasets, which showed the score of C2 was higher compared to that of the C1 subtype with a good survival benefit (Figures 2G, H). Furthermore, differences in clinicopathological characteristics of TCGA molecular subtypes were analyzed. As indicated in Figure 3, we found significant differences between the two molecular

10.3389/fphar.2023.1267186

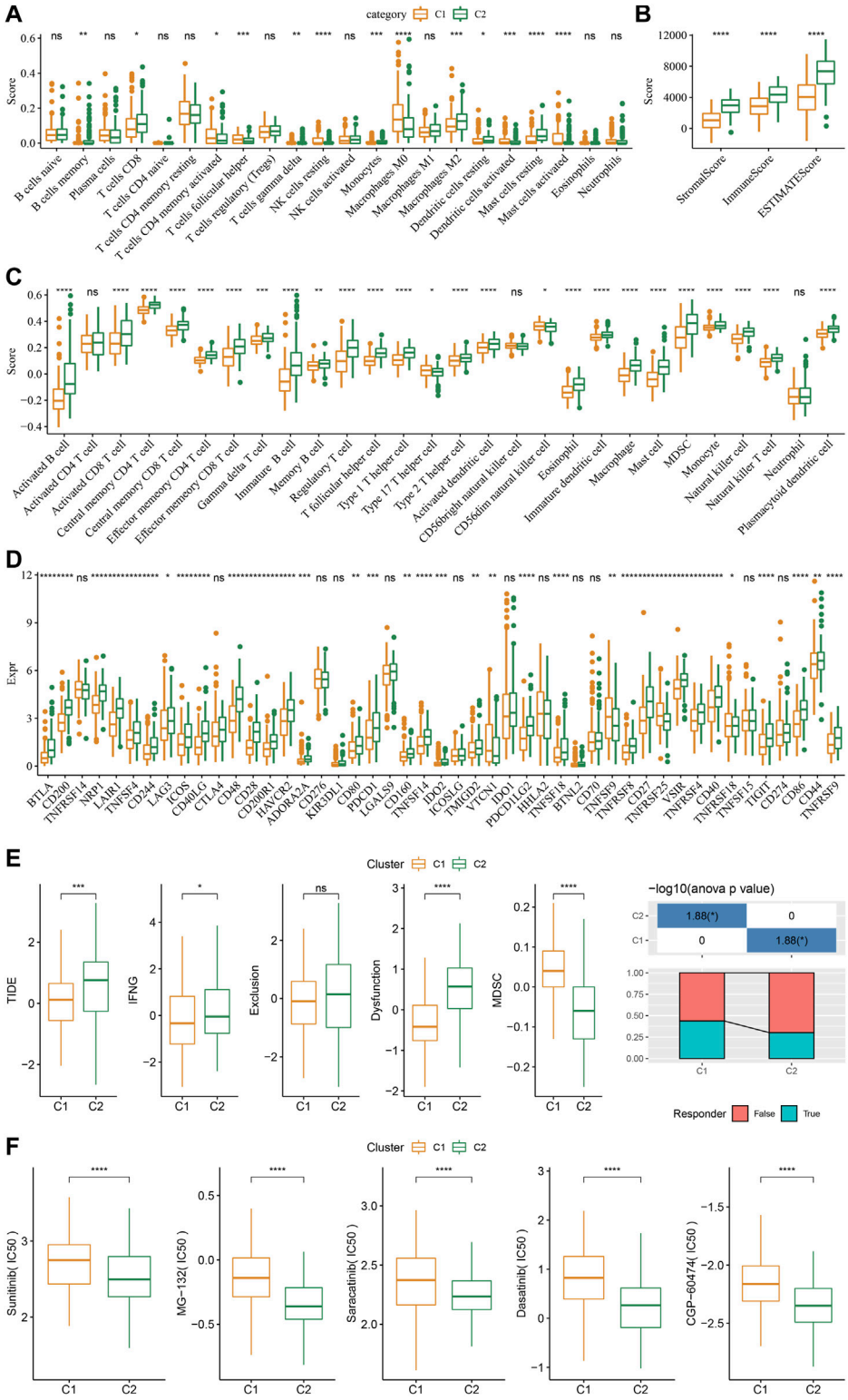


FIGURE 4
Differences in immune signatures between molecular subtypes treated with immunotherapy or chemotherapy. **(A)** The abundance of 22 immune cells shows differences among various molecular subtypes; **(B)** ESTIMATE immune infiltration differences among various molecular subtypes; **(C)** The scores for 28 immune cells vary among different molecular subtypes; **(D)** Differentially expressed immune checkpoints between different groups; **(E)** The results of TIDE analysis show significant differences comparing different TCGA cohort groups; **(F)** The estimated IC50 values for drugs in TCGA-STAD are displayed as box plots.

10.3389/fphar.2023.1267186

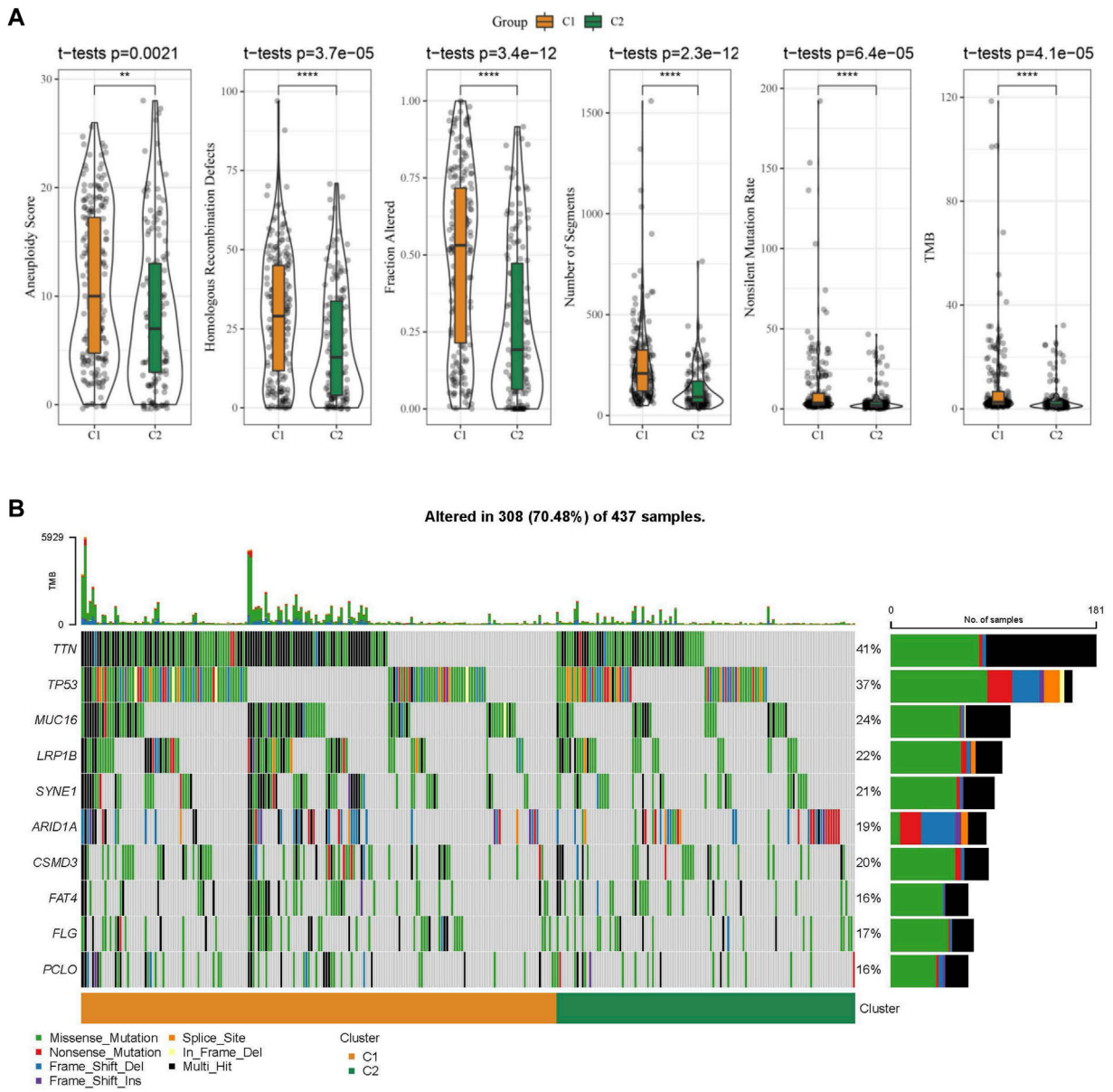


FIGURE 5
The molecular subtypes within the TCGA cohort show genomic variations. **(A)** Comparison of Homologous Recombination Defects, Aneuploidy Score, Fraction Altered, Number of Segments, Nonsilent Mutation Rate, and Tumor mutation burden in the TCGA cohort molecular subtypes; **(B)** Somatic mutations in the two molecular subtypes.

subtypes in terms of age, disease stages, grade classification, and patient survival status.

Various immune characteristics and responses to immunotherapy and chemotherapy between molecular subtypes.

As shown in Figure 4A, some immune cell types were significantly different among the subtypes. Compared to the C2 subtype, activated dendritic cells (DCs) and CD4 positive memory T cells were significantly upregulated while the M2-phenotype macrophages were suppressed in the C1 subtype. Meanwhile, ESTIMATE was applied to evaluate the level of immune cells inside tumor tissues. Figure 4B showed that the ImmuneScore of the C2 subtype was significantly higher compared to that of other groups, indicating that C2 has a high level of immune cell infiltration. In addition, ssGSEA demonstrated significant differences in most immune cell scores between different subtypes (Figure 4C). We further analyzed the responsiveness to immunotherapy between different TCGA cohort molecular subtypes. As shown in Figure 4D, compared to C1 subtype, the gene expression of

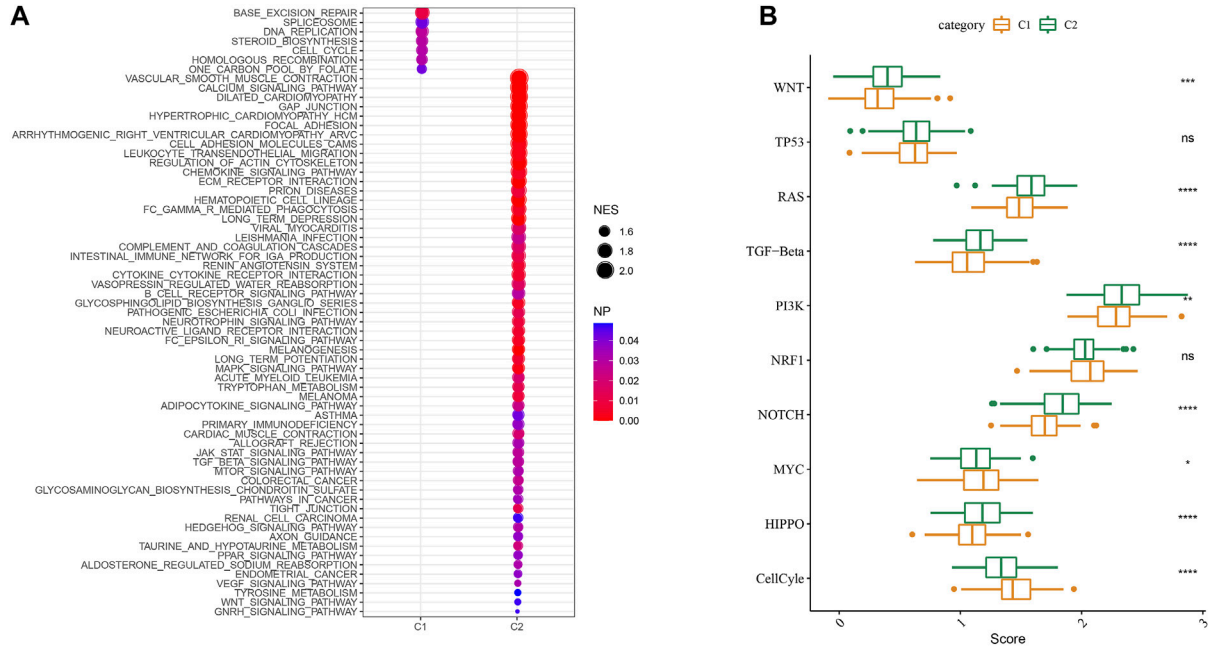


FIGURE 6
Signaling analysis between molecular subtypes. **(A)** The GSEA results of the TCGA cohort; **(B)** The scores of 10 signaling pathways associated with tumor abnormalities for various TCGA cohort molecular subtypes.

immune checkpoints such as IDO1, IDO2, and CD274 were increased dramatically in the C2 subtype.

We did this by employing the TIDE algorithm in order to assess the potential response of tumor samples to immune checkpoint inhibitors. Higher TIDE scores represent a higher likelihood of immune escape in response to immune checkpoint therapy (Jiang et al., 2018). As shown in Figure 4E, in the TCGA cohort, the C2 subtype showed a much higher TIDE score than the C1 subtype with best prognosis, indicating that the C2 subtype had greater potential of immune escape. In addition, analysis on the response of different molecular subtypes to conventional chemotherapeutic drugs showed that C2 patients were more sensitive to these drugs including sunitinib, MG-132, saracatinib, dasatinib, and CGP-60474 (Figure 4F).

Mutational signatures and pathway analysis between molecular subtypes

Further, differences in genomic alterations between two different molecular subtypes in the TCGA cohort were analyzed. Here, we obtained molecular information of TCGA-STAD collected from a pan-cancer study (Thorsson et al., 2018). The C1 subtype showed a higher aneuploidy score, fraction altered, homologous recombination defects, non-silent mutation rate, tumor mutation burden (TMB), and number of segments (Figure 5A). Gene mutation differences between different molecular subtypes were also studied. The top 10 genes were exhibited in Figure 5B, which showed significant differences of TTN, TP53, and MUC16 genes in mutation frequency between the two molecular subtypes. Similarly to our results, it has been reported that TTN, TP53 and MUC16 are the most significantly mutated driver genes in GC, which is closely related to the prognosis of cancer patients (Dong et al., 2022).

We then performed GSEA analysis which showed that DNA replication, spliceosome and base excision repair signaling pathways were highly enriched in the C1 subtype, while the C2 subtype had high enrichment of the phagocytosis, chemokine, and leukocyte transendothelial migration signaling pathways (Figure 6A). We also evaluated the 10 oncogenic pathways from the previous study (Sanchez-Vega et al., 2018). The differences showed statistical significance in the rest of the pathways except for the TP53 and NRF1 signaling pathways (Figure 6B). Compared to the C2 subtype, the oncogenic pathways, such as the Wnt, PI3K, and RAS pathways (Zhang et al., 2001; Zhang et al., 2019; BY et al., 2020), were significantly downregulated in the C1 subtype, which indicated an association between activated oncogenic pathways and tryptophan metabolism that may result in poor prognosis in C2.

Identification of key genes for the tryptophan metabolism phenotype

As shown above, we identified two different molecular subtypes with distinct prognostic, immune, mutational, and pathway signatures. Next, we screened genes which are differentially expressed in C1 subtypes compared with C2 subtypes ($|Log2FC| > 1$; FDR <0.05), and a sum of 618 distinctly expressed genes were obtained as shown in the volcano plot of differential analysis (Figure 7A) in which 36 genes

10.3389/fphar.2023.1267186

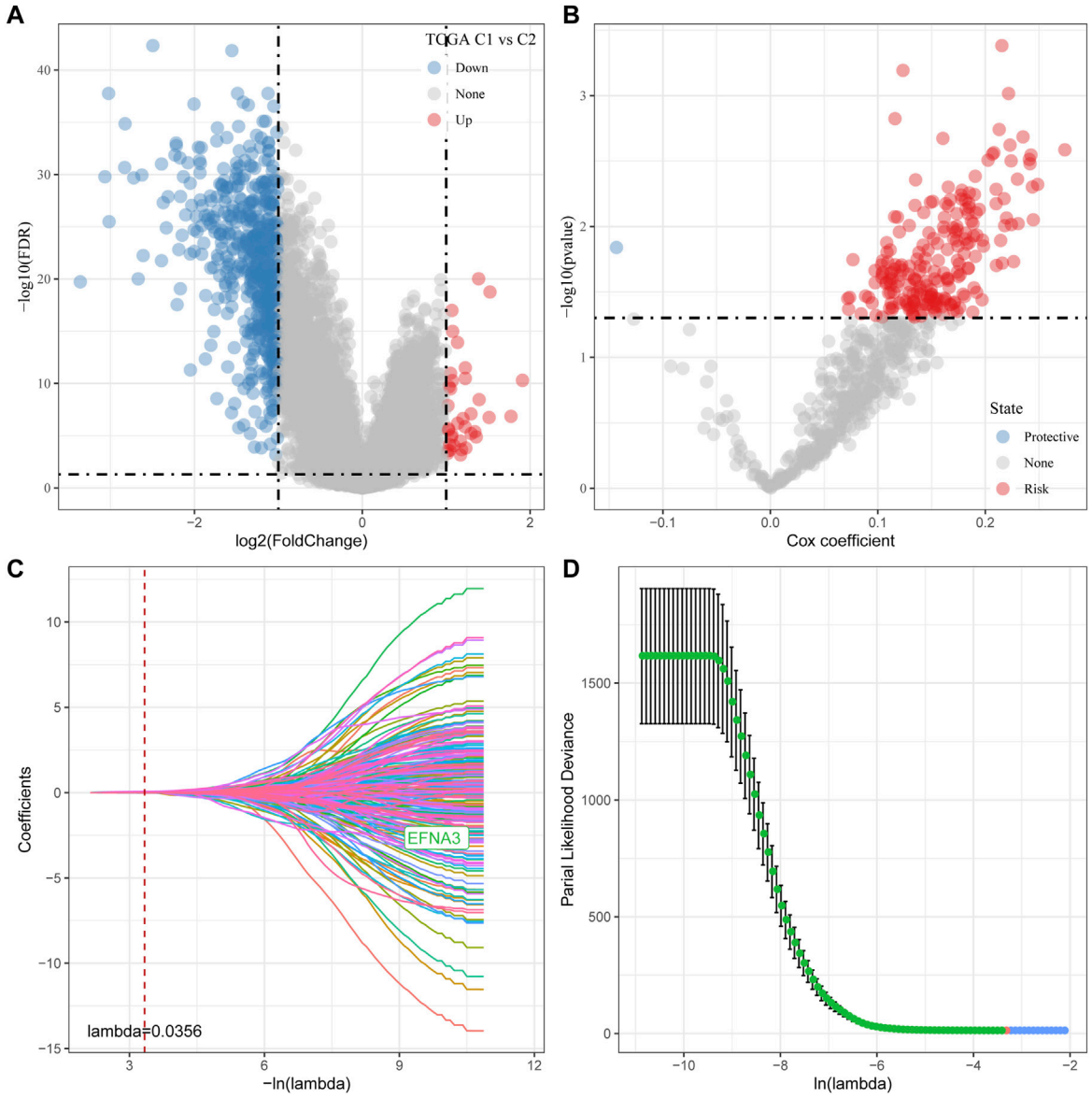


FIGURE 7
Determination of predominant genes contributing to phenotypes related to tryptophan metabolism. **(A)** Gene expression difference is displayed by a volcano plot; **(B)** A total of 218 potential candidates were determined among the differentially expressed genes; **(C)** Trajectory schemes were drawn for every independent variable associated with lambda; **(D)** Confidence interval under lambda.

were significantly increased and 582 genes were decreased. We next utilized univariate cox analysis on the 618 differentially expressed genes and identified a total of 218 genes that showed the greater impact on prognosis, including 217 Risk genes and 1 Protective gene (Figure 7B). Subsequently, we compressed these 218 significant differentially expressed genes through lasso regression. As shown in Figures 7C, D, 10-fold cross-validation was used to construct the model and eight genes at lambda = 0.0356 were screened for further analysis: EFNA3, GPX3, RGS2, CXCR4, SGCE, ADH4, CST2, GPC3. The final 8-gene signature formula was as follows:

RiskScore=(0.024*EFNA3)+0.079*GPX3+0.102*RGS2+0.051*CXCR4+0.014*SGCE+0.076*ADH4+0.067*CST2+0.066*GPC3.

Establishment and validation of clinical prognostic model

The risk score for each TCGA sample was separately calculated, followed by conducting a receiver operating characteristic (ROC) analysis on the prognostic classification of the RiskScore. The prediction classification efficiency from 1 year to 4 years was calculated to have an area under the time-dependent ROC curves (AUC) of 0.7, which validated the prediction capability of the model (Figure 8A). We then performed the zscore on the RiskScore. When the RiskScore was lower than zero, the samples were separated into the low-risk group, while those with a RiskScore higher than zero were in the high-risk group. As shown in Figure 8B, the low-risk group showed

10.3389/fphar.2023.1267186

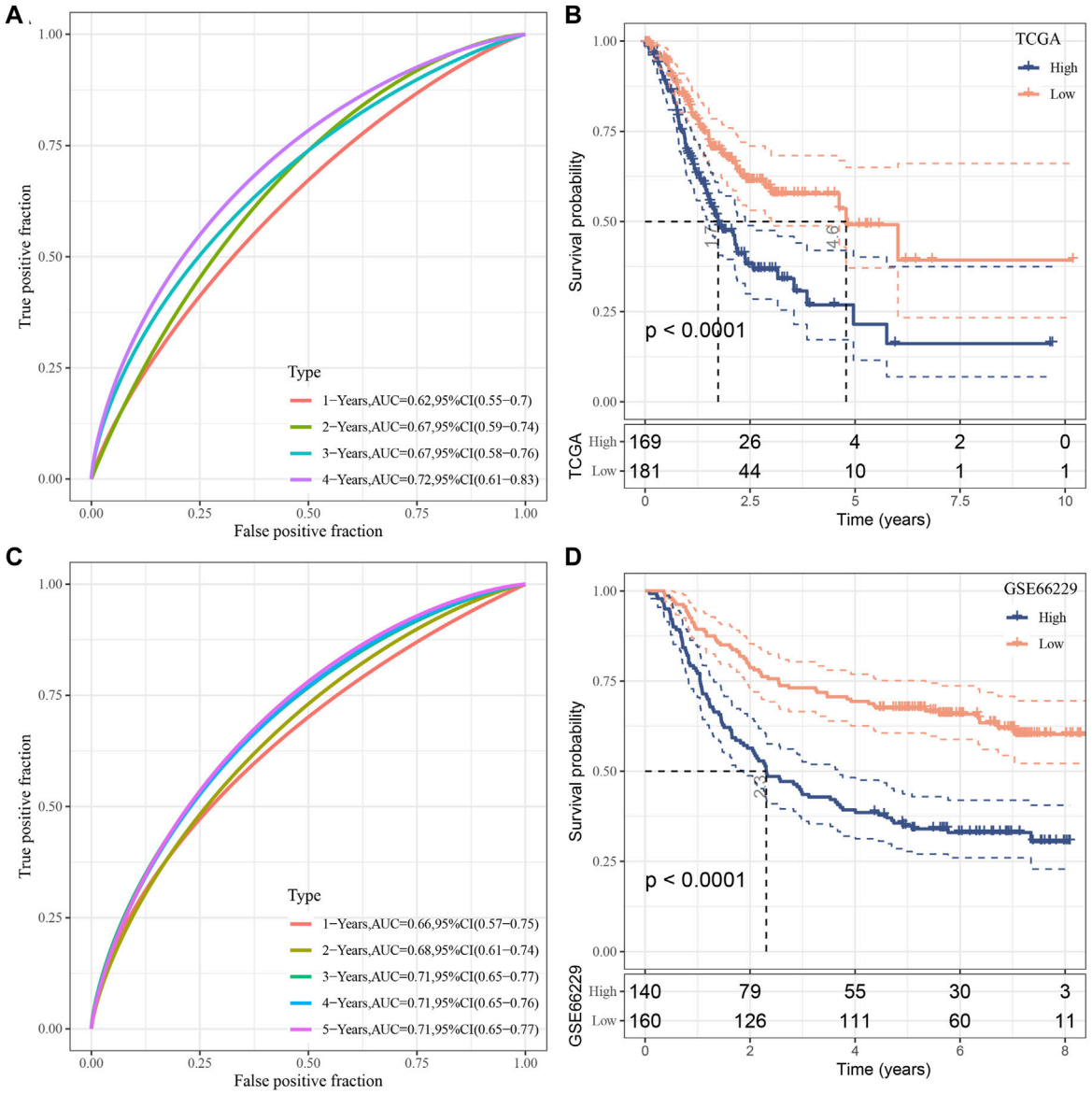


FIGURE 8
Establishment as well as validation of clinical prognostic model. **(A–B)** ROC curves and KM curves for eight genes derived from TCGA dataset; **(C–D)** ROC curves and KM curves for eight genes derived from GSE66229 dataset.

prolonged survival time, indicating the good performance of this prognostic model. Additionally, GSE66229 dataset was utilized to test the robustness and validate the risk model constructed by these eight genes. As shown in Figures 8C, D, similar results were observed, which indicated the excellent predictive ability of this model.

RiskScore performance on clinicopathological features and different molecular subtypes

We analyzed the differences in RiskScore between Gender, Age, TNM grades, Stage clinical grades, and Grade grades in the TCGA-STAD dataset to explore the relationship between RiskScore and clinical characteristics of GC. With the increase in the clinical grade, the RiskScore also increased (Figure 9A). Moreover, we analyzed the difference in RiskScores among different molecular subtypes. The RiskScore of the C2 subtype with worse prognosis was obviously higher compared to that of the C1 molecular subtype with the best prognosis. Clinicopathological characteristics between the RiskScore groups in the TCGA-STAD cohort were analyzed, and the high-risk group was found to have a higher clinical grade (Figure 9B), which was consistent with previous results.

Immune infiltration/pathway characteristics between RiskScore groups

Next, the enrichment of 22 immune cells in the high and low RiskScore groups was analyzed between RiskScore groups. As shown in Figure 10A, compared to the high-risk group, resting

10.3389/fphar.2023.1267186

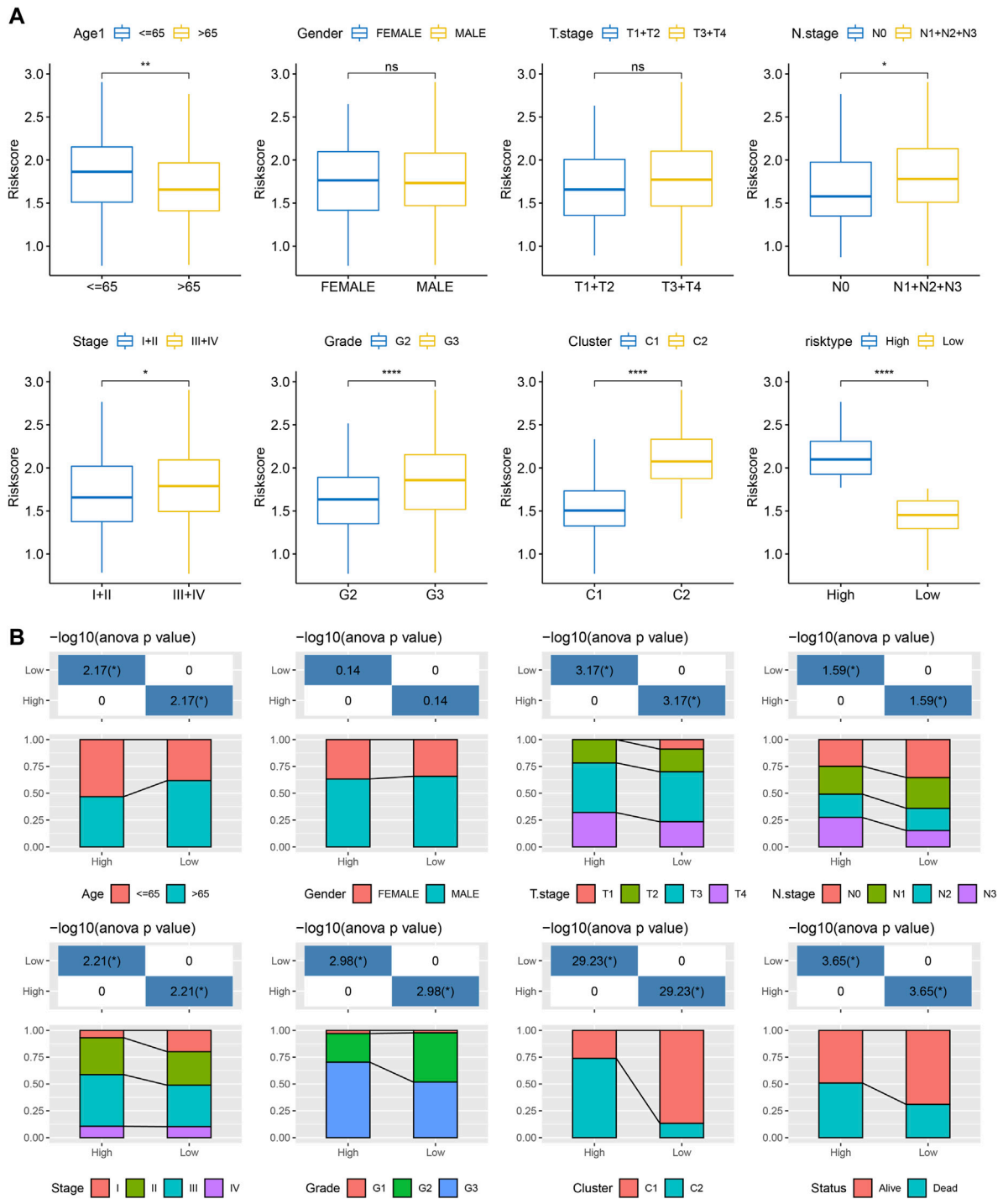


FIGURE 9
RiskScore performance on various clinicopathological features and various molecular subtypes. **(A)** Various clinicopathological groups derived from TCGA cohort are compared pairwise in the parameter of RiskScore; **(B)** Clinicopathological characteristics between RiskScore groups derived from TCGA cohort. G2 and G3 stages were selected for comparison. For M stage, there were 312 patients in M0 stage, and only 23 patients in M1 stage. Therefore, the difference in RiskScore between M stages was not compared in this study.

NK cells, activated CD4 T cells, and activated DCs were increased dramatically while M2 macrophages were substantially decreased in the low-risk group. ESTIMATE was also utilized to evaluate the level of immune cells in the tumor tissues. As shown in Figure 10B, the ImmuneScore in the high-risk group is much higher. Although it showed a higher level of immune cells in the high-risk group, the immunosuppressive cells such as MDSC and M2-phenotype macrophages also increased significantly, which might result in a poor prognosis. Further, we also used the ssGSEA function to analyze the scores of 28 types of immune

10.3389/fphar.2023.1267186

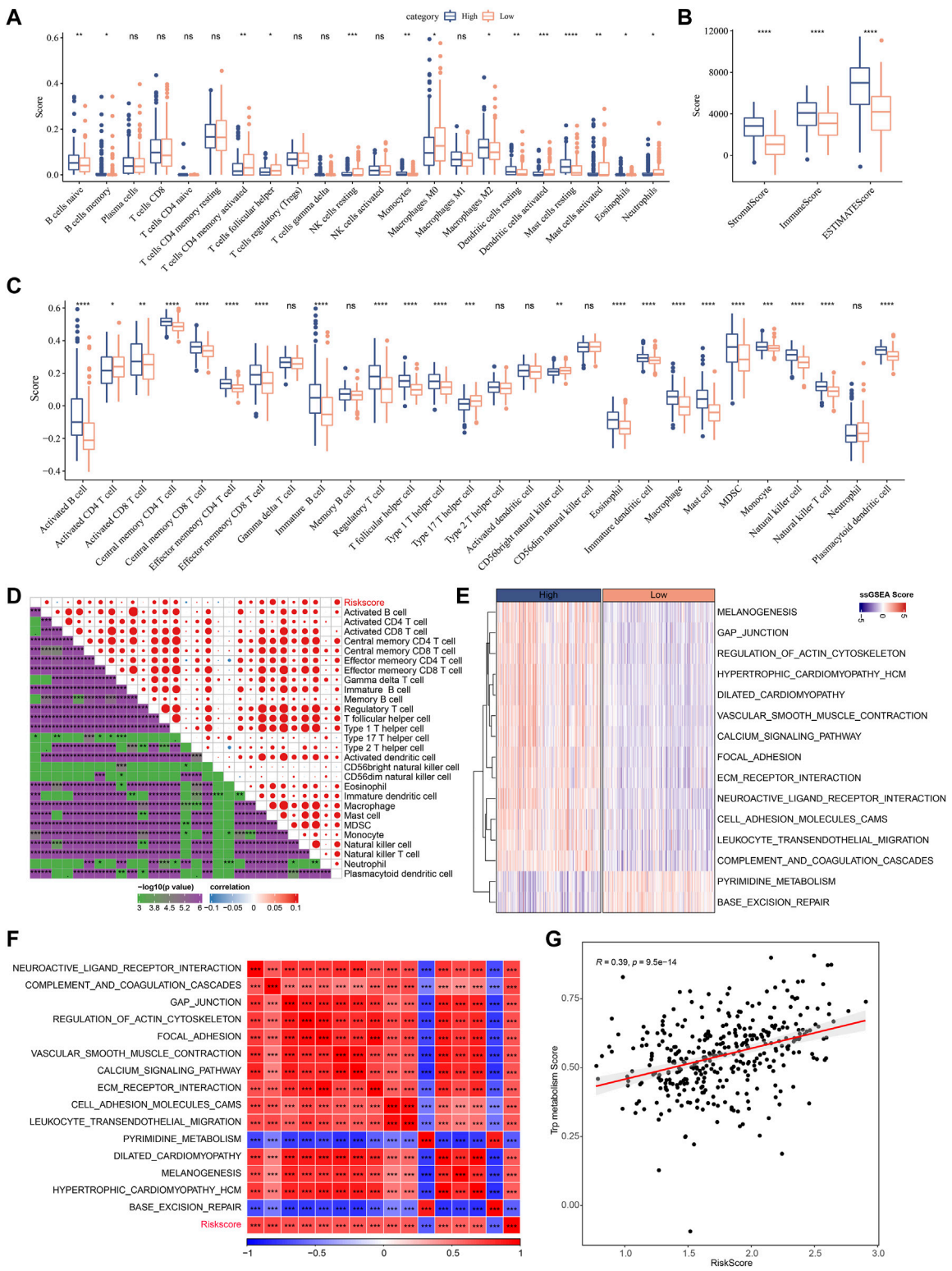


FIGURE 10
Immune infiltration/pathway characteristics between RiskScore groups. **(A)** The abundance of 22 immune cells in TCGA cohort between high and low risk groups is shown as boxplot; **(B)** Boxplot of differences in immune scores calculated by ESTIMATE software in TCGA cohort; **(C)** Boxplots of differences in 28 immune cell scores calculated by ssGSEA in TCGA cohort; **(D)** Correlation analysis between 28 immune cell scores and RiskScore in the TCGA cohort; **(E)** The enrichment scores for signaling pathways in high-risk and low-risk groups with a correlation factor larger than 0.6 are displayed as a heatmap; **(F)** Correlation relationship between GSEA pathways and RiskScore was performed and those with a correlation factor larger than 0.6 are displayed; **(G)** Correlation scatter plot between RiskScore and tryptophan scores.

10.3389/fphar.2023.1267186

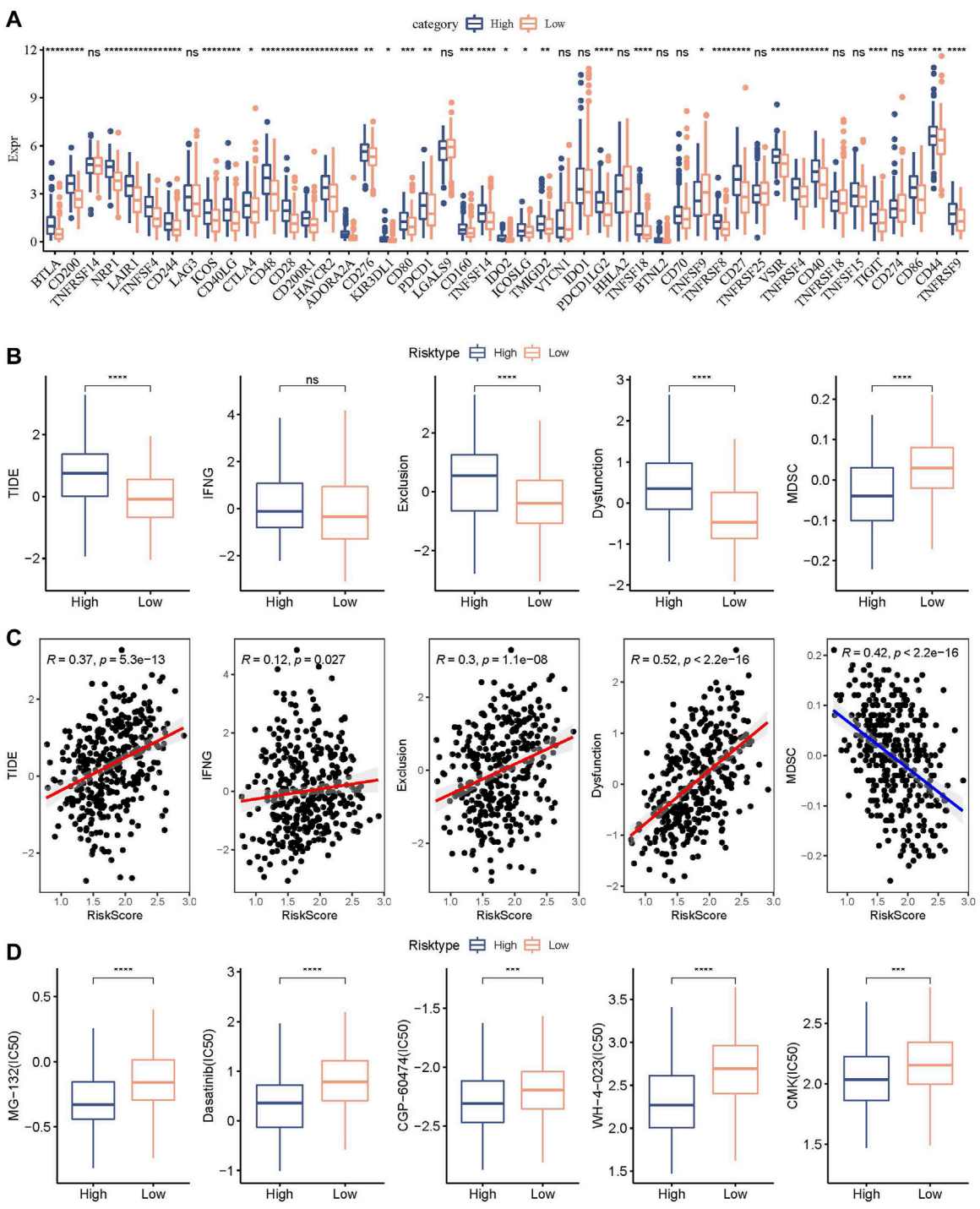


FIGURE 11
Genomic variations after immunotherapy and chemotherapy among RiskScore groups. **(A)** Differentially expressed immune checkpoint genes among different TCGA cohort groups; **(B)** The TIDE analysis results were compared between different TCGA cohort groups to show differences; **(C)** Within the TCGA cohort, a correlation relationship was performed between RiskScore and TIDE results; **(D)** The estimated IC50 values for drugs in TCGA-STAD were displayed as box plots.

cells, and most of the immune cell scores were significantly different between high- and low-risk groups (Figure 10C). The correlation between immune cells and RiskScore was further evaluated. As shown in Figure 10D, the RiskScore positively correlated with most immune cells, especially CD4 T cells, DCs, NK cells, and MDSCs, which could support the prediction of prognoses. This suggests that the infiltration of these immune cells increases as the risk score rises.

In Figures 10E–G, the most signaling pathways were positively correlated with the RiskScore of the samples and a significant positive correlation could be found between RiskScore and

10.3389/fphar.2023.1267186

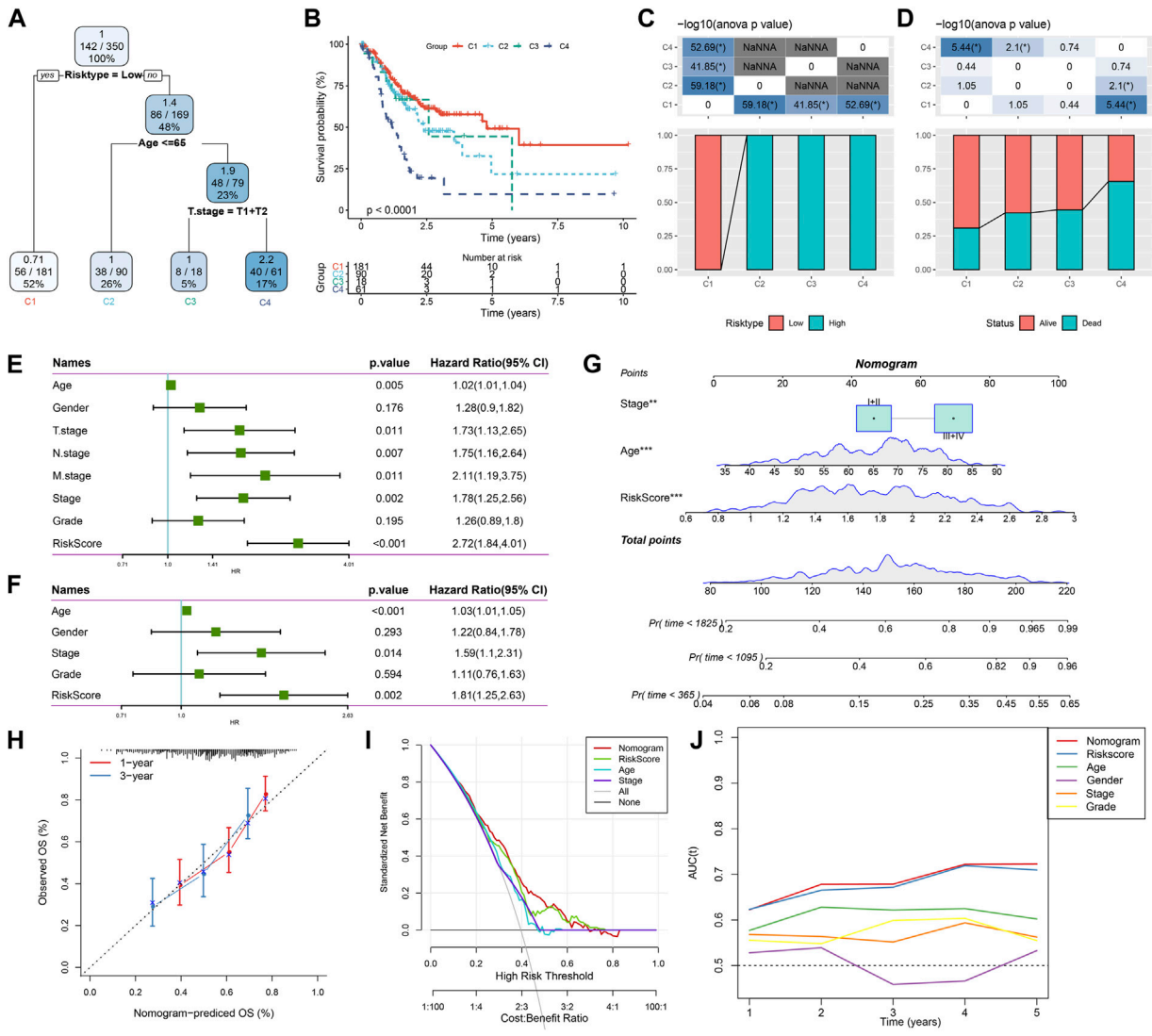


FIGURE 12
Combining RiskScore with clinicopathological features for the optimization of prognostic model and survival prediction. Patient samples with RiskScore, stage, and age were utilized to generate the survival decision tree; **(A,B)**: Overall survival among the different risk groups; **(C,D)**: Comparative analysis was performed between different groups; **(E)**: Univariate cox analysis of RiskScore and clinical features; **(F)**: Multivariate Cox analysis of RiskScore and clinical features; **(G)**: Nomogram model; **(H)**: Compared with other clinicopathological features, the nomogram showed the most powerful capacity for survival prediction; **(I)**: 1- and 3-year calibration curves of nomograms; **(J)**: Decision curves of nomograms.

tryptophan metabolism ssGSEA scores, indicating that the activated tryptophan metabolism may induce poor prognosis.

Differences in immunotherapy/ chemotherapy between RiskScore groups

We further explored the expression level of immune checkpoints between RiskScore groups. As shown in Figure 11A, some immune checkpoint genes, including NRP1, CD200, and CTLA4, were significantly downregulated in low-risk groups. We further analyzed the difference in immunotherapy among different RiskScore groups. TIDE software was applied to evaluate the immunotherapy response in the high and low RiskScore groups. We can find that in the TCGA cohort, the high-risk group showed a higher TIDE score (Figure 11B), indicating that the high-risk group showed greater potential of immune escape and less sensitive to immunotherapy. We further analyzed the relationship between RiskScore and TIDE score. Figure 11C showed a positive correlation between RiskScore and TIDE, IFNG, Exclusion, and Exclusion scores and a significant negative correlation with MDSC. In addition, we also analyzed the response of RiskScore groups to chemotherapeutic drugs in the TCGA cohort and found that the high RiskScore group was more sensitive to chemotherapeutic drugs including MG-132, dasatinib, CGP-60474, WH-4-023, and CMK (Figure 11D).

Combining RiskScore and clinicopathological features to optimize prognostic model and survival prediction.

Here, patient age, gender, TNM stage, Stage clinical grade, Grade, and RiskScore in the TCGA-STAD cohort were used to develop a decision tree. Only age, T stage, and RiskScore remained in the decision tree, where RiskScore was the most effective parameter (Figure 12A). Figure 12B showed a significant overall survival differences among the four risk subgroups. C2, C3, and C4 were all high-risk patients (Figure 12C). A significantly decreased survival benefit was found in the C2, C3, and C4 subgroups (Figure 12D). RiskScore, age, and Stage were significant prognostic factors, as confirmed by Univariate and multivariate Cox regression analysis on the clinical characteristics and RiskScore (Figures 12E, F). A nomogram combining the RiskScore with other clinicopathological characteristics was developed for the risk assessment and survival probability evaluation for patients. On survival rate prediction, the RiskScore showed the greatest impact. The accuracy of the prediction model was evaluated by the calibration curve (Figure 12G). As displayed in Figure 12H, the predicted calibration curve at 1 and 3 years(s) almost overlapped with the standard curve, indicating a strong prediction of the nomogram. Further decision curve analysis (DCA) demonstrated that the Riskscore and nomogram showed noticeably greater benefits than extreme curves. Additionally, both nomogram and RiskScore demonstrated the most powerful survival predictors compared to other clinicopathological features (Figures 12I, J).

Discussion

Inflammation is typically associated with GC development and migration in damage of gastric mucosa (Demaria et al., 2010; Wang et al., 2014). The large number of immune cells, inflammatory cells and cytokines often present in the tumor microenvironment leads to a state of immunosuppression and chronic inflammation (Li et al., 2020). Although FDA approved combination therapy for the treatment of early stage and advanced GC patients recently, the 5-year survival of GC patients with advanced stage remains poor. We still need to investigate the mechanism of GC progression and its related tumor immune microenvironment for the development of novel cancer immunotargets against GC. Cancer is a typical metabolic disease. Nie et al. developed a prognostic model based on the metabolic profile of TCGA that can well predict the prognosis of STAD patients. This new metabolism-related feature can respond to the dysregulated STAD metabolic microenvironment (Nie et al., 2021). In addition, numerous studies have proved that the tryptophan metabolism could progress the GC development (Platten et al., 2019; Tanaka et al., 2021; Yao et al., 2021). Free tryptophan is a substrate for the kynurenine signaling pathway, which produces various metabolites related to the immune response. IDO1 and IDO2 were most popular rate-limiting enzymes to catabolize tryptophan, and many studies have focused on the blockade of IDO1 to active the antitumor immunity (Günther et al., 2019). However, phase III clinical studies of IDO inhibitors against cancers were unsatisfying (Chen et al., 2021). Although it is still unclear whether IDO enzymes are essential for cancer progression, tryptophan metabolism-related genes and signaling pathways were highly related to the GC immune microenvironment. Long et al. showed that tryptophan metabolism-related genes were significantly associated with immune infiltration of different cells in hepatocellular carcinoma (Long et al., 2023). Similarly, Zhang et al. showed that tryptophan metabolism-related genes play an important role in the immune microenvironment of gliomas. They constructed a tryptophan metabolism-related predictive model and found that higher tryptophan metabolism-related gene markers were significantly associated with immune cell infiltration (Zhang et al., 2022).

We therefore explored the molecular mechanism of tryptophan metabolism in GC by using RNA-seq data from human samples to establish the risk model used for predicting clinical results. Firstly, RNA-seq data derived from patients with GC were collected, and we analyzed the expression signatures and mutation profiles of tryptophan metabolism-associated genes. Then, two tryptophan metabolism-associated molecular subtypes were constructed to investigate the role of tryptophan metabolism in tumor immune microenvironment and further developed and verified the use of the model in a clinical setting.

It has been shown that Kynurenine, a catabolic metabolite of tryptophan, is able to bind to receptors for transcription factors, which in turn induces tumor cell invasion and immunosuppression of the tumor microenvironment (Xu et al., 2021). This suggests that amino acid metabolism plays a key role in the immunoregulatory mechanisms in tumor cells and the tumor microenvironment. In this study, we obtained a total of 40 genes that were highly related to tryptophan metabolism. Around 27.69% samples were found with tryptophan metabolism gene mutations, which was consistent with previous studies (Santhanam et al., 2016; Lu et al., 2020; Pirzadeh et al., 2022). This study classified two molecular subtypes and found that the C1 subtype showed better prognosis compared with the C2 subtype. But the immune cell infiltration and the ratio of some key immune activation cells especially naïve and memory CD8 T cells in the C1 molecular subtype were significantly more suppressed than those in C2 subtype. As demonstrated, CD8 T cell infiltration in the tumor microenvironment was critical in reviving antitumor immunity (Rahir and Moser, 2012; Chen et al., 2018). The suppressed immune cell infiltration in the C1 subtype might be induced by other immunosuppressive signals such as MDSC and the M2-phenotype macrophage, which were also important to form an immunosuppressive tumor microenvironment and inhibit the migration and penetration of immune cells into tumor tissues (Sica and Massarotti, 2017; Wang et al., 2021). We also analyzed the immune checkpoint gene expression in these two subtypes. Interestingly, more immune checkpoint genes were found to be suppressed in the C1 subtype compared with the C2 subtype, indicating that immune checkpoint blockade instead of immune cell infiltration resulted in a good clinical outcome for the C1 subtype. Additionally, the tumor-related signaling pathways such as the Nuclear Factor kappa B (NF-κB)/P53 pathway and MYC targets pathways were enriched in the mutant group. As reported, the NF-κB family was considered as a key regulator of immune responses and inflammation. Some literature has demonstrated that NF-κB/P53 signaling pathway activation was associated with human cancer development, progression, and metastatic potential (Tilborghs et al., 2017; Thorsson et al., 2018; Marei et al., 2021). Moreover, several studies have indicated that the transcription factor MYC served as a proto-oncogene in multiple cancers, which can result in transcriptional activation or repression of specific genes including those involved in tumor cell growth, proliferation, and survival (Zhang et al., 2010; Hu et al., 2018). The tumor-related signaling pathways enriched in the mutant group

suggested that the mutation may result in functional changes and survival benefits for GC patients.

This study determined that eight significant genes (EFNA3, GPX3, RGS2, CXCR4, SGCE, ADH4, CST2, and GPC3) are correlated genes to construct the risk model. Among these genes, the chemokine receptor CXCR4 and its ligand CXCL12 were widely reported to be involved in cancer cell survival, proliferation, and migration (O'Boyle et al., 2013; Lee and Jo, 2012). Preliminary *in vivo* experiments suggested that CXCR4 might be essential in the development of a range of cancer malignancies (Conley-LaComb et al., 2013; Xue et al., 2017). The CXCR4/CXCL12 signaling could be considered as a therapeutic target in antitumor immunity and more in-depth exploration should be performed for the prediction of clinical outcomes. Moreover, some studies have demonstrated that glypican-3 (GPC3) was closely associated with tumor progression and acted as an oncogene in GC (Zhu et al., 2002; Ushiku et al., 2009), which was consistent with our findings. GPC3 might provide another potential therapeutic target for the treatment of GC.

We further established the risk model to predict the clinical outcome, which has been evaluated and verified with good performance and high survival prediction accuracy. The potential therapeutic targets among the tryptophan metabolism-related genes and signaling pathways could be applied in the clinical GC diagnosis and treatment. However, there are some limitations to note in this study. First, our study data were obtained from the TCGA and GEO databases, which were analyzed only by bioinformatics. In further studies, we should conduct relevant *in vivo* and *in vitro* validation experiments to verify the effect of risk modeling. In addition, the molecular mechanisms related to tryptophan metabolism in GC remain to be further verified.

Conclusion

In this study, we screened and determined eight key genes that are related to the phenotype of tryptophan metabolism through differentially expressed gene analysis between molecular subtypes and constructed the risk model based on these key genes, which showed strong robustness and stable predictive performance with independence of clinicopathological characteristics. To optimize the risk model and prognostic prediction, we combined the RiskScore with clinicopathological features, which showed high accuracy and capability for survival prediction.

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 author.

Author contributions

PL: Conceptualization, Funding acquisition, Methodology, Software, Supervision, Validation, Writing–original draft, Writing–review and editing. GC: Conceptualization, Investigation, Methodology, Project administration, Writing–review and editing. ZS: Conceptualization, Investigation, Software, Writing–original draft. JY: Conceptualization, Data curation, Project administration, Resources, Supervision, Writing–review and editing. XW: Conceptualization, Formal Analysis, Methodology, Validation, Writing–original draft. JP: Conceptualization, Investigation, Methodology, Software, Writing–review and editing. LZ: Formal Analysis, Methodology, Project administration, Resources, Writing–review and 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 the Zhejiang Provincial Natural Science Foundation of China (LQ20H160038).

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

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Brenner, H., Rothenbacher, D., and Arndt, V. (2009). Epidemiology of stomach cancer. *Cancer Epidemiol.* 472, 467–477. doi:10.1007/978-1-60327-492-0_23

Bruni, D., Angell, H. K., and Galon, J. (2020). The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. *Nat. Rev. Cancer* 20 (11), 662–680. doi:10.1038/s41568-020-0285-7

By, S., Dass, M. S., Smalley, M. J., and Pearson, H. B. (2020). The PI3K-AKT-mTOR pathway and prostate cancer: at the crossroads of AR, MAPK, and WNT signaling. *Int. J. Mol. Sci.* 21 (12), 4507. doi:10.3390/ijms21124507

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Chen, H.-Y., Xu, L., Li, L.-F., Liu, X.-X., Gao, J.-X., and Bai, Y.-R. (2018). Inhibiting the CD8+ T cell infiltration in the tumor microenvironment after radiotherapy is an important mechanism of radioresistance. *Sci. Rep.* 8 (1), 11934. doi:10.1038/s41598-018-30417-6

Chen, S., Tan, J., and Zhang, A. (2021). The ups, downs and new trends of IDO1 inhibitors. *Bioorg. Chem.* 110, 104815. doi:10.1016/j.bioorg.2021.104815

Conejos, J. R. V., Ghassemi Nejad, J., Kim, J.-E., Moon, J.-O., Lee, J.-S., and Lee, H.-G. (2021). Supplementing with l-tryptophan increases medium protein and alters expression of genes and proteins involved in milk protein synthesis and energy metabolism in bovine mammary cells. *Int. J. Mol. Sci.* 22 (5), 2751. doi:10.3390/ijms22052751

Conley-LaComb, M. K., Saliganan, A., Kandagatla, P., Chen, Y. Q., Cher, M. L., and Chinni, S. R. (2013). PTEN loss mediated Akt activation promotes prostate tumor

growth and metastasis via CXCL12/CXCR4 signaling. *Mol. Cancer* 12 (1), 85. doi:10.1186/1476-4598-12-85

De Felice, F., Marchetti, C., Palaia, I., Musio, D., Muzii, L., Tombolini, V., et al. (2015). Immunotherapy of ovarian cancer: the role of checkpoint inhibitors. *J. Immunol. Res.* 2015, 191832. doi:10.1155/2015/191832

Demaria, S., Pikarsky, E., Karin, M., Coussens, L. M., Chen, Y.-C., El-Omar, E. M., et al. (2010). Cancer and inflammation: promise for biologic therapy. *J. Immunother. Hagerst. Md 1997)* 33 (4), 335–351. doi:10.1097/CJI.0b013e3181d32e74

Dong, Y., Song, N., Wang, J., Shi, L., Zhang, Z., and Du, J. (2022). Driver Gene Alterations in Malignant Progression of Gastric Cancer. *Front. Oncol.* 12, 920207. doi:10.3389/fonc.2022.920207

Günther, J., Däbritz, J., and Wirthgen, E. (2019). Limitations and off-target effects of tryptophan-related IDO inhibitors in cancer treatment. *Front. Immunol.* 10, 1801. doi:10.3389/fimmu.2019.01801

Hoseini, S. M., Pérez-Jiménez, A., Costas, B., Azeredo, R., and Gesto, M. (2019). Physiological roles of tryptophan in teleosts: current knowledge and perspectives for future studies. *Rev. Aquac.* 11 (1), 3–24. doi:10.1111/raq.12223

Hu, Y., Yu, K., Wang, G., Zhang, D., Shi, C., Ding, Y., et al. (2018). Lanatoside C inhibits cell proliferation and induces apoptosis through attenuating Wnt/β-catenin/c-Myc signaling pathway in human gastric cancer cell. *Biochem. Pharmacol.* 150, 280–292. doi:10.1016/j.bcp.2018.02.023

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

Lambert, R., Guilloux, A., Oshima, A., Pompe-Kirn, V., Bray, F., Parkin, M., et al. (2002). Incidence and mortality from stomach cancer in Japan, Slovenia and the USA. *Int. J. Cancer* 97 (6), 811–818. doi:10.1002/ijc.10150

Lee, H. J., and Jo, D. Y. (2012). The role of the CXCR4/CXCL12 axis and its clinical implications in gastric cancer. *Histol. Histopathol.* 27 (9), 1155–1161. doi:10.14670/HH-27.1155

Li, C., Oh, S. J., Kim, S., Hyung, W. J., Yan, M., Zhu, Z. G., et al. (2009). Risk factors of survival and surgical treatment for advanced gastric cancer with large tumor size. *J. Gastrointest. Surg.* 13 (5), 881–885. doi:10.1007/s11605-009-0800-3

Li, L., Yu, R., Cai, T., Chen, Z., Lan, M., Zou, T., et al. (2020). Effects of immune cells and cytokines on inflammation and immunosuppression in the tumor microenvironment. *Int. Immunopharmacol.* 88, 106939. doi:10.1016/j.intimp.2020.106939

Long, G., Wang, D., Tang, J., and Tang, W. (2023). Development of tryptophan metabolism patterns to predict prognosis and immunotherapeutic responses in hepatocellular carcinoma. *Aging* 15 (15), 7593–7615. doi:10.18632/aging.204928

Lu, S., Wang, L. J., Lombardo, K., Kwak, Y., Kim, W. H., and Resnick, M. B. (2020). Expression of Indoleamine 2, 3-dioxygenase 1 (IDO1) and Tryptophanyl-tRNA Synthetase (WARS) in Gastric Cancer Molecular Subtypes. *Appl. Immunohistochem. Mol. Morphol. AIMM.* 28 (5), 360–368. doi:10.1097/PAI.0000000000000761

Marei, H. E., Althani, A., Afifi, N., Hasan, A., Caceci, T., Pozzoli, G., et al. (2021). p53 signaling in cancer progression and therapy. *Cancer Cell Int.* 21 (1), 703–715. doi:10.1186/s12935-021-02396-8

Nie, Y., Liu, L., Liu, Q., and Zhu, X. (2021). Identification of a metabolic-related gene signature predicting the overall survival for patients with stomach adenocarcinoma. *PeerJ* 9, e10908. doi:10.7717/peerj.10908

O'Boyle, G., Swidenbank, I., Marshall, H., Barker, C. E., Armstrong, J., White, S. A., et al. (2013). Inhibition of CXCR4-CXCL12 chemotaxis in melanoma by AMD11070. *Br. J. cancer* 108 (8), 1634–1640. doi:10.1038/bjc.2013.124

Pirzadeh, M., Khalili, N., and Rezaei, N. (2022). The interplay between aryl hydrocarbon receptor, *H. pylori*, tryptophan, and arginine in the pathogenesis of gastric cancer. *Int. Rev. Immunol.* 41 (3), 299–312. doi:10.1080/08830185.2020.1851371

Platten, M., Nollen, E. A., Röhrig, U. F., Fallarino, F., and Opitz, C. A. (2019). Tryptophan metabolism as a common therapeutic target in cancer, neurodegeneration and beyond. *Nat. Rev. Drug Discov.* 18 (5), 379–401. doi:10.1038/s41573-019-0016-5

Rahir, G., and Moser, M. (2012). Tumor microenvironment and lymphocyte infiltration. *Cancer Immunol. Immunother.* 61 (6), 751–759. doi:10.1007/s00262-012-1253-1

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W. K., Luna, A., La, K. C., et al. (2018). Oncogenic signaling pathways in the cancer genome atlas. *Cell* 173 (2), 321–337. e10. doi:10.1016/j.cell.2018.03.035

Santhanam, S., Alvarado, D. M., and Ciorba, M. A. (2016). Therapeutic targeting of inflammation and tryptophan metabolism in colon and gastrointestinal cancer. *Transl. Res.* 167 (1), 67–79. doi:10.1016/j.trsl.2015.07.003

Sica, A., and Massarotti, M. (2017). Myeloid suppressor cells in cancer and autoimmunity. *J. Autoimmun.* 85, 117–125. doi:10.1016/j.jaut.2017.07.010

Song, Z., Wu, Y., Yang, J., Yang, D., and Fang, X. (2017). Progress in the treatment of advanced gastric cancer. *Tumor Biol.* 39 (7), 1010428317714626. doi:10.1177/1010428317714626

Souissi, S., Ghedira, R., Macherki, Y., Ben-Haj-Ayed, A., Gabbouj, S., Remadi, Y., et al. (2022). Indoleamine 2, 3-dioxygenase gene expression and kynurenine to tryptophan ratio correlation with nasopharyngeal carcinoma progression and survival. *Immun. Inflamm. Dis.* 10 (9), e690. doi:10.1002/iid3.690

Tanaka, M., Tóth, F., Polyák, H., Á, S., Mándi, Y., and Vécsei, L. (2021). Immune influencers in action: metabolites and enzymes of the tryptophan-kynurenine metabolic pathway. *Biomedicines* 9 (7), 734. doi:10.3390/biomedicines9070734

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Yang, T.-H. O., et al. (2018). The immune landscape of cancer. *Immunity* 48 (4), 812–830. e14. doi:10.1016/j.immuni.2018.03.023

Tilborghs, S., Corthouts, J., Verhoeven, Y., Arias, D., Rolfo, C., Trinh, X. B., et al. (2017). The role of nuclear factor-kappa B signaling in human cervical cancer. *Crit. Rev. oncology/hematology* 120, 141–150. doi:10.1016/j.critrevonc.2017.11.001

Twomey, J. D., and Zhang, B. (2021). Cancer immunotherapy update: FDA-approved checkpoint inhibitors and companion diagnostics. *AAPS J.* 23 (2), 39–11. doi:10.1208/s12248-021-00574-0

Ushiku, T., Uozaki, H., Shinozaki, A., Ota, S., Matsuzaka, K., Nomura, S., et al. (2009). Glypican 3-expressing gastric carcinoma: distinct subgroup unifying hepatoid, clear-cell, and alpha-fetoprotein-producing gastric carcinomas. *Cancer Sci.* 100 (4), 626–632. doi:10.1111/j.1349-7006.2009.01108.x

Wang, F., Meng, W., Wang, B., and Qiao, L. (2014). Helicobacter pylori-induced gastric inflammation and gastric cancer. *Cancer Lett.* 345 (2), 196–202. doi:10.1016/j.canlet.2013.08.016

Wang, Y., Tiruthani, K., Li, S., Hu, M., Zhong, G., Tang, Y., et al. (2021). mRNA Delivery of a Bispecific Single-Domain Antibody to Polarize Tumor-Associated Macrophages and Synergize Immunotherapy against Liver Malignancies. *Adv. Mater.* 33 (23), 2007603. doi:10.1002/adma.202007603

Whiteside, T. L., Demaria, S., Rodriguez-Ruiz, M. E., Zarour, H. M., and Melero, I. (2016). Emerging opportunities and challenges in cancer immunotherapy. *Clin. Cancer Res.* 22 (8), 1845–1855. doi:10.1158/1078-0432.CCR-16-0049

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Xu, Y., Zhang, H., Sun, Q., Geng, R., Yuan, F., Liu, B., et al. (2021). Immunomodulatory Effects of Tryptophan Metabolism in the Glioma Tumor Microenvironment. *Front. Immunol.* 12, 730289. doi:10.3389/fimmu.2021.730289

Xue, L.-J., Mao, X.-B., Ren, L.-L., and Chu, X.-Y. (2017). Inhibition of CXCL12/CXCR4 axis as a potential targeted therapy of advanced gastric carcinoma. *Cancer Med.* 6 (6), 1424–1436. doi:10.1002/cam4.1085

Yao, Y., Liang, H., Fang, X., Zhang, S., Xing, Z., Shi, L., et al. (2021). What is the prospect of indoleamine 2, 3-dioxygenase 1 inhibition in cancer? Extrapolation from the past. *J. Exp. Clin. Cancer Res.* 40 (1), 60–14. doi:10.1186/s13046-021-01847-4

Yoneda, A., Kuroki, T., and Eguchi, S. (2021). Immunotherapeutic advances in gastric cancer. *Surg. Today* 51 (11), 1727–1735. doi:10.1007/s00595-021-02236-2

Zhai, L., Ladomersky, E., Lenzen, A., Nguyen, B., Patel, R., Lauing, K. L., et al. (2018). IDO1 in cancer: a Gemini of immune checkpoints. *Cell. Mol. Immunol.* 15 (5), 447–457. doi:10.1038/cmi.2017.143

Zhang, L., Hou, Y., Ashktorab, H., Gao, L., Xu, Y., Wu, K., et al. (2010). The impact of C-MYC gene expression on gastric cancer cell. *Mol. Cell. Biochem.* 344 (1), 125–135. doi:10.1007/s11010-010-0536-0

Zhang, S., Chen, S., Wang, Z., Li, J., Yuan, Y., Feng, W., et al. (2022). Prognosis prediction and tumor immune microenvironment characterization based on tryptophan metabolism-related genes signature in brain glioma. *Front. Pharmacol.* 13, 1061597. doi:10.3389/fphar.2022.1061597

Zhang, X., Gaspard, J. P., and Chung, D. C. (2001). Regulation of vascular endothelial growth factor by the Wnt and K-ras pathways in colonic neoplasia. *Cancer Res.* 61 (16), 6050–6054.

Zhang, Z., Yu, W., Zheng, M., Liao, X., Wang, J., Yang, D., et al. (2019). Pin1 inhibition potently suppresses gastric cancer growth and blocks PI3K/AKT and Wnt/β-catenin oncogenic pathways. *Mol. Carcinog.* 58 (8), 1450–1464. doi:10.1002/mc.23027

Zhu, Z., Friess, H., Kleeff, J., Wang, L., Wirtz, M., Zimmermann, A., et al. (2002). Glypican-3 expression is markedly decreased in human gastric cancer but not in esophageal cancer. *Am. J. Surg.* 184 (1), 78–83. doi:10.1016/s0002-9610(02)00884-x

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 17 October 2023
DOI 10.3389/fphar.2023.1260742

OPEN ACCESS

EDITED BY
Linhui Wang,
Second Military Medical University, China

REVIEWED BY
Ziheng Wang,
University of Macau, China
Fan Li,
Huazhong University of Science and Technology, China
Aiguo Shen,
Nantong University, China

*CORRESPONDENCE
Jun Li,
ljun01@jlu.edu.cn
Yuansong Bai,
baiys@jlu.edu.cn

†These authors have contributed equally to this work

RECEIVED 18 July 2023
ACCEPTED 02 October 2023
PUBLISHED 17 October 2023

CITATION
Cong D, Zhao Y, Zhang W, Li J and Bai Y (2023), Applying machine learning algorithms to develop a survival prediction model for lung adenocarcinoma based on genes related to fatty acid metabolism.
Front. Pharmacol. 14:1260742.
doi: 10.3389/fphar.2023.1260742

Applying machine learning algorithms to develop a survival prediction model for lung adenocarcinoma based on genes related to fatty acid metabolism

Dan Cong†, Yanan Zhao†, Wenlong Zhang, Jun Li* and Yuansong Bai*

Department of Oncology and Hematology, China-Japan Union Hospital of Jilin University, Changchun, China

Background: The progression of lung adenocarcinoma (LUAD) may be related to abnormal fatty acid metabolism (FAM). The present study investigated the relationship between FAM-related genes and LUAD prognosis.

Methods: LUAD samples from The Cancer Genome Atlas were collected. The scores of FAM-associated pathways from the Kyoto Encyclopedia of Genes and Genomes website were calculated using the single sample gene set enrichment analysis. ConsensusClusterPlus and cumulative distribution function were used to classify molecular subtypes for LUAD. Key genes were obtained using limma package, Cox regression analysis, and six machine learning algorithms (GBM, LASSO, XGBoost, SVM, random forest, and decision trees), and a RiskScore model was established. According to the RiskScore model and clinical features, a nomogram was developed and evaluated for its prediction performance using a calibration curve. Differences in immune abnormalities among patients with different subtypes and RiskScores were analyzed by the Estimation of STromal and Immune cells in MAlignant Tumours using Expression data, CIBERSORT, and single sample gene set enrichment analysis. Patients' drug sensitivity was predicted by the pRRophetic package in R language.

Results: LUAD samples had lower scores of FAM-related pathways. Three molecular subtypes (C1, C2, and C3) were defined. Analysis on differential prognosis showed that the C1 subtype had the most favorable prognosis, followed by the C2 subtype, and the C3 subtype had the worst prognosis. The C3 subtype had lower immune infiltration. A total of 12 key genes (SLC2A1, PKP2, FAM83A, TCN1, MS4A1, CLIC6, UBE2S, RRM2, CDC45, IGF2BP1, ANGPTL4, and CD109) were screened and used to develop a RiskScore model. Survival chance of patients in the high-RiskScore group was significantly lower. The low-RiskScore group showed higher immune score and higher expression of most immune

Abbreviations: ACC2, acetyl-coA carboxylase 2; BP, biological process; CC, cellular component; CDF, cumulative distribution function; CNV, copy number variant; DCA, decision curve analysis; FABP5, fatty acid-binding protein 5; FAM, fatty acid metabolism; FASN, fatty acid synthase; GEO, Gene Expression Omnibus; GSEA, gene set enrichment analysis; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; KM, Kaplan–Meier; LUAD, lung adenocarcinoma; MF, molecular function; SNV, single nucleotide variant; ssGSEA, single sample gene set enrichment analysis; and TCGA, The Cancer Genome Atlas.

checkpoint genes. Patients with a high RiskScore were more likely to benefit from the six anticancer drugs we screened in this study.

Conclusion: We developed a RiskScore model using FAM-related genes to help predict LUAD prognosis and develop new targeted drugs.

KEYWORDS

lung adenocarcinoma, fatty acid metabolism, RiskScore, machine learning, survival probability, prognosis

1 Introduction

Lung adenocarcinoma (LUAD) accounts for about 40% of primary lung tumors. LUAD is one of the tumor types that have rapid metastasis and high mortality, with a survival time shorter than 5 years (Denisenko et al., 2018; Hutchinson et al., 2019; Shi et al., 2022). LUAD, at an early stage, usually has no obvious clinical symptoms and is often diagnosed by adjuvant methods at the middle and late stages or when metastasis occurs (Patz et al., 2014; Skřičková et al., 2018; Sung et al., 2021). Although significant advances have been made in the research and clinical treatment of LUAD, the prognosis of LUAD remains dismal, despite the clinical use of chemoradiotherapy, targeted therapy, and immunotherapy. Currently, the underlying cellular and molecular mechanisms of tumor behavior remain unclear (Lin et al., 2021; Chen et al., 2022a). Therefore, molecular characteristics of LUAD should be comprehensively investigated to improve clinical therapies and the accuracy of prognosis prediction for LUAD.

Lipid metabolism is an important metabolic process for cells. Abnormal fatty acid metabolism (FAM) in cancer cells has been increasingly studied. Carcinogenesis mechanisms of various cancers vary greatly, but they often show similar abnormalities in metabolism. Reprogramming the metabolism of glucose, fatty acids, and other biomolecules could promote the progression of tumor cells (Li and Zhang, 2016). Growing evidence demonstrated that some changes occur in tumor tissues in different processes of FAM (Amiri et al., 2018), including in deciding the types, abundance, and mechanisms of action of lipid-signaling molecules with regulatory functions (Santos and Schulze, 2012). Changes of FAM also affect the proliferation, differentiation, and metastasis of tumor cells (Yu et al., 2018). However, the characteristics and functions of genes related to FAM in LUAD have not been fully explored.

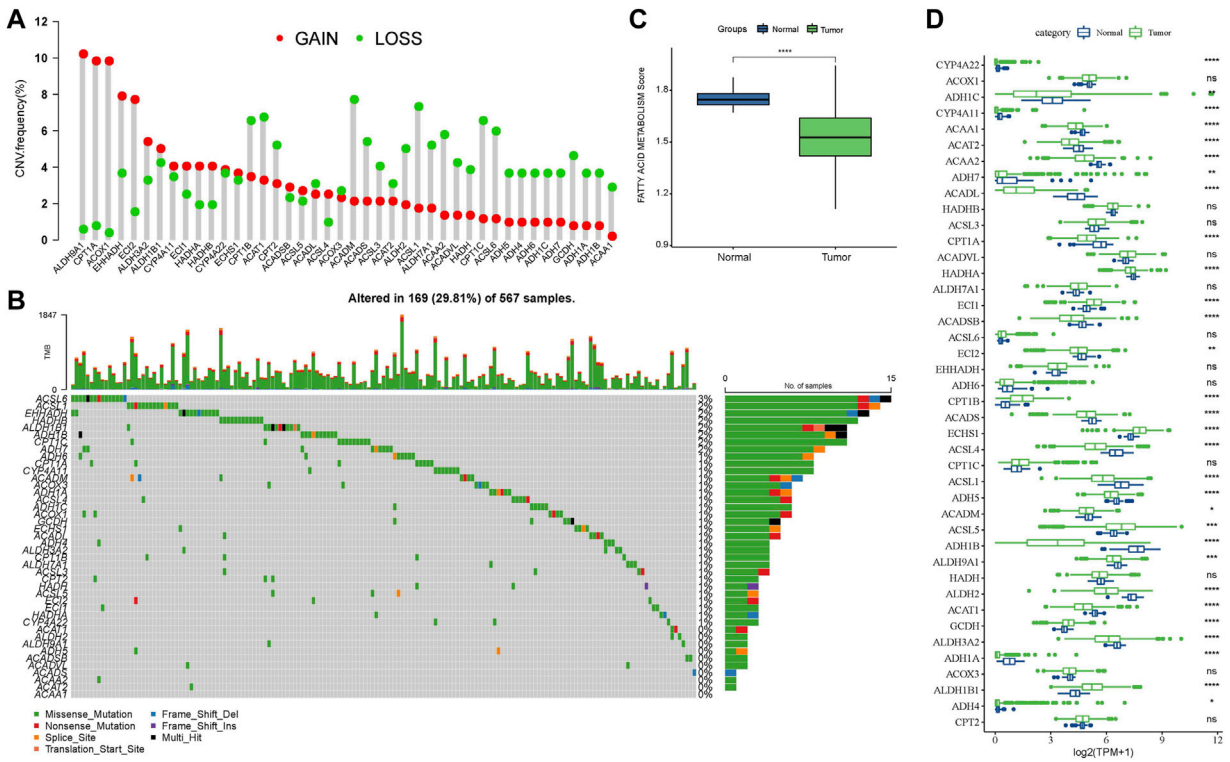


FIGURE 1
Abnormal FAM-related pathway genes in LUAD. **(A)** Proportion of gene deletion and amplification in the CNV of FAM-related pathway genes. **(B)** Waterfall diagram of mutant information in the SNV of FAM-related pathway genes. **(C)** Comparison of FAM pathway scores in LUAD and para-cancer tissues. **(D)** Comparison of expression of FAM-related pathway genes in LUAD and para-cancer tissues.

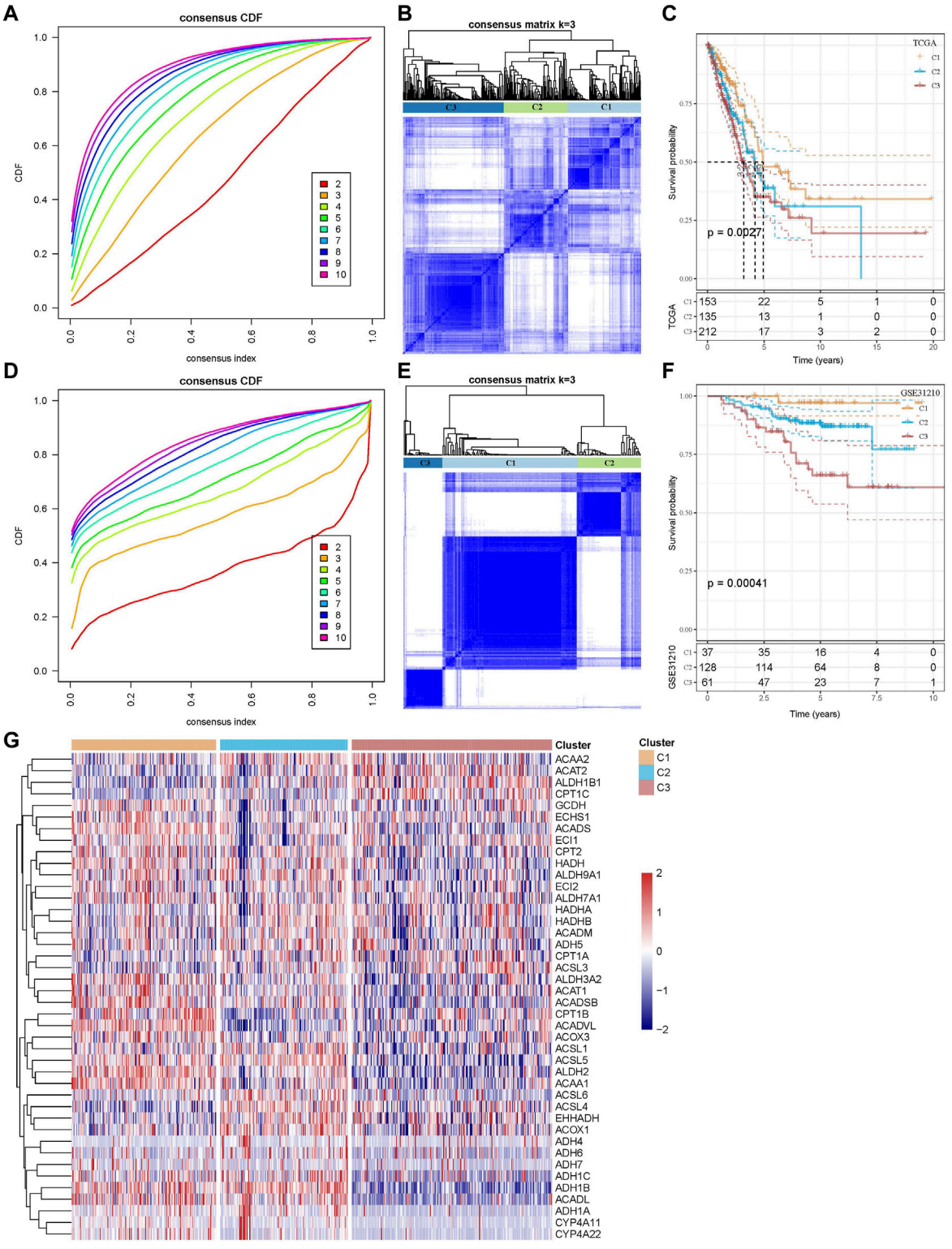
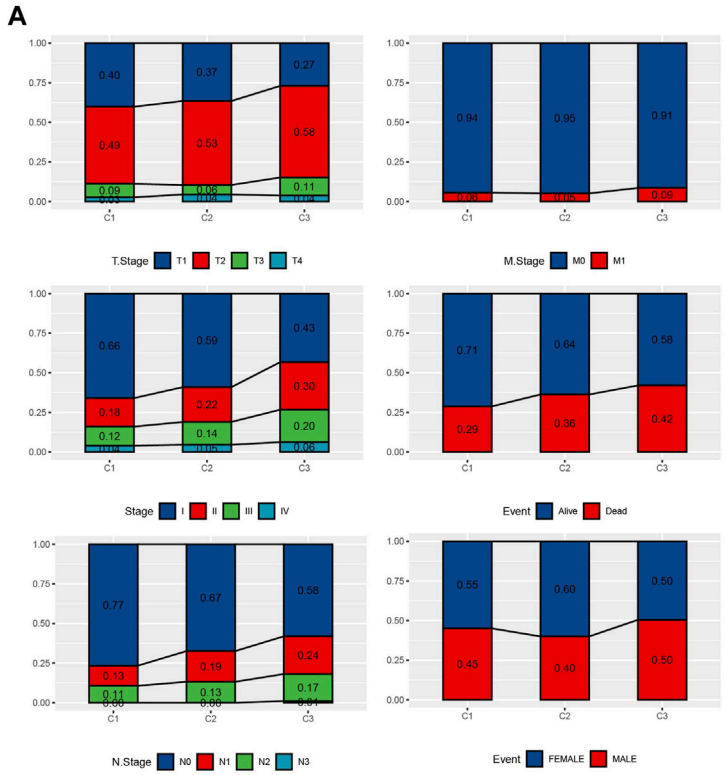


FIGURE 2
Construction of molecular subtypes based on genes related to FAM pathways. **(A)** CDF curve of TCGA cohort samples. **(B)** Clustering heatmap of samples in the TCGA cohort when consensus k = 3. **(C)** Relationship between the prognoses of three TCGA subtypes is shown by the K–M curve. **(D)** CDF curve of GSE31210 cohort samples. **(E)** Clustering heatmap of samples with consensus k = 3 in the GSE31210 cohort. **(F)** K–M curve of the relationship between the prognoses of three subtypes of GSE31210. **(G)** Heatmap of the expression of FAM-related pathway genes between three subtypes in the TCGA cohort.



B

	C1 (N=153)	C2 (N=135)	C3 (N=212)	P-value
Gender				
FEMALE	84 (54.9%)	81 (60.0%)	105 (49.5%)	0.156
MALE	69 (45.1%)	54 (40.0%)	107 (50.5%)	
Event				
Alive	109 (71.2%)	86 (63.7%)	123 (58.0%)	0.035
Dead	44 (28.8%)	49 (36.3%)	89 (42.0%)	
T.Stage				
T1	61 (39.9%)	49 (36.3%)	57 (26.9%)	
T2	74 (48.4%)	71 (52.6%)	122 (57.5%)	0.133
T3	13 (8.5%)	8 (5.9%)	24 (11.3%)	
T4	4 (2.6%)	6 (4.4%)	8 (3.8%)	
Missing	1 (0.7%)	1 (0.7%)	1 (0.5%)	
N.Stage				
N0	115 (75.2%)	87 (64.4%)	122 (57.5%)	
N1	19 (12.4%)	25 (18.5%)	50 (23.6%)	
N2	16 (10.5%)	17 (12.6%)	36 (17.0%)	0.016
N3	0 (0%)	0 (0%)	2 (0.9%)	
Missing	3 (2.0%)	6 (4.4%)	2 (0.9%)	
M.Stage				
M0	101 (66.0%)	92 (68.1%)	139 (65.6%)	0.496
M1	6 (3.9%)	5 (3.7%)	13 (6.1%)	
Missing	46 (30.1%)	38 (28.1%)	60 (28.3%)	
Stage				
I	99 (64.7%)	78 (57.8%)	91 (42.9%)	
II	27 (17.6%)	29 (21.5%)	63 (29.7%)	0.003
III	18 (11.8%)	19 (14.1%)	43 (20.3%)	
IV	6 (3.9%)	6 (4.4%)	13 (6.1%)	
Missing	3 (2.0%)	3 (2.2%)	2 (0.9%)	

FIGURE 3
Differences in clinical characteristics among molecular subtypes. **(A)** Bar chart of pair-to-pair comparison of three molecular subtypes in the TCGA cohort. **(B)** Comparative table of each clinical characteristic molecular subtype.

A previous study investigating abnormal FAM showed that the overexpression of fatty acid-binding protein 5 (FABP5) is related to the poor prognosis in LUAD and may be a new clinical target to treat LUAD (Garcia et al., 2022). The downregulation of fatty acid synthase (FASN) interferes with the progression of LUAD through regulating the glucose metabolism and inhibiting the AKT/ERK pathway (Chang et al., 2019). Related drugs could act on the FAM process in LUAD. For example, anlotinib controls LUAD progression through inhibiting FASN-mediated FAM (Shen et al., 2022). Chaoyang Liang et al. also showed that the overexpression of genes related to FAM enzymes (ACOT11) regulates the growth, differentiation, and metastasis of LUAD cells through a variety of signaling pathways (Liang et al., 2020). In addition, Wang et al. developed a fatty acid-related RiskScore model to predict the prognosis of lung cancer patients and identified 38 fatty acid-related genes. Among these 38 genes, eight genes (HGNAT, MCTP2, ENPP5, PLEKHA6, ANKRD29, CNTNAP2, SLC4A5, and ZNF738) have not been reported in previous lung cancer-related studies (Wang et al., 2022a). Therefore, the identification and verification of genes related to FAM may have great potential for developing new prognostic models and improving clinical treatment for LUAD.

In this study, we downloaded genomic information about the clinical characteristics of LUAD from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. Molecular subtypes related to FAM pathways were developed for LUAD, and we further established a risk assessment model based on FAM-related genes using six machine learning algorithms. In addition, we assessed the level of immune cell infiltration and sensitivity to common drugs in different risk groups. The current study provided a better understanding of the mechanism of abnormal FAM in LUAD cells, helping to improve the therapeutic strategies for treating LUAD patients.

2 Materials and methods

2.1 Data downloading and preprocessing

2.1.1 TCGA-LUAD dataset downloading and preprocessing

Data with clinical phenotypes were obtained from the TCGA database (Liu et al., 2020). Samples without the survival time or state were eliminated to ensure that the survival time of all the included samples was longer than 0 days. Finally, 500 tumor samples and 59 para-cancer tissues samples from the TCGA dataset were obtained.

2.1.2 GEO data download and preprocessing

A set of chip data was obtained from the GEO (Barrett et al., 2013), and the probe was converted into symbol according to the annotation file. Normal tissue samples or those without clinical information were excluded to ensure that the survival time of all the included samples was longer than 0 days, and only LUAD samples

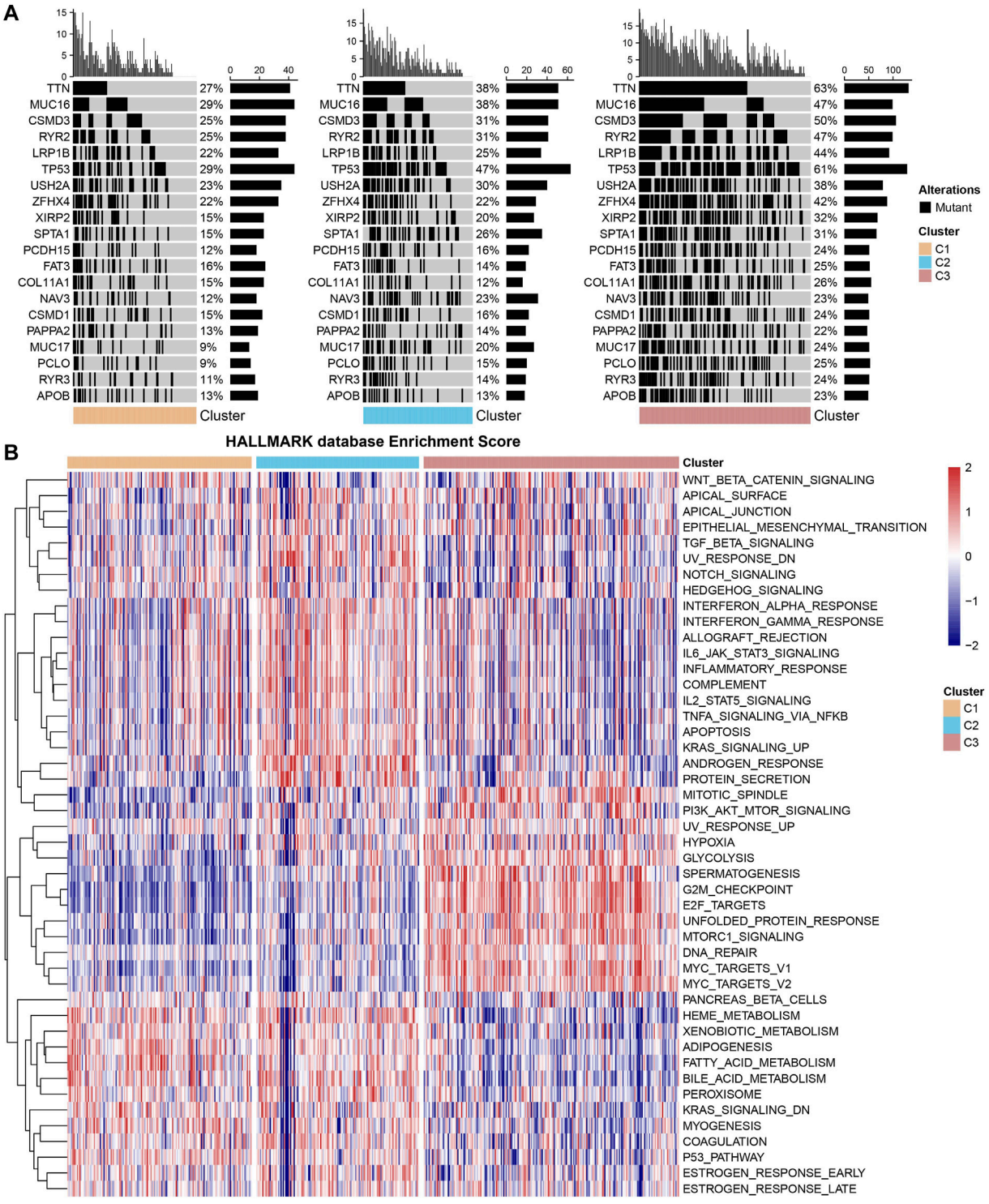


FIGURE 4
Mutant characteristics and differential activation pathways of molecular subtypes. **(A)** Different molecular subtypes in the TCGA cohort were analyzed by somatic mutation analysis. **(B)** Heatmap of functional enrichment scores of each subtype in the TCGA cohort.

were retained through data filtering. Specifically, 289 samples were obtained from the GSE30219 dataset; 226 samples were from the GSE31210 dataset; 196 samples were from the GSE37745 dataset; and 127 samples were from the GSE50081 dataset.

2.1.3 Acquisition of FAM-related genes

From the Kyoto Encyclopedia of Genes and Genomes (KEGG), a collection of 42 related genes was downloaded (Kanehisa and Goto, 2000).

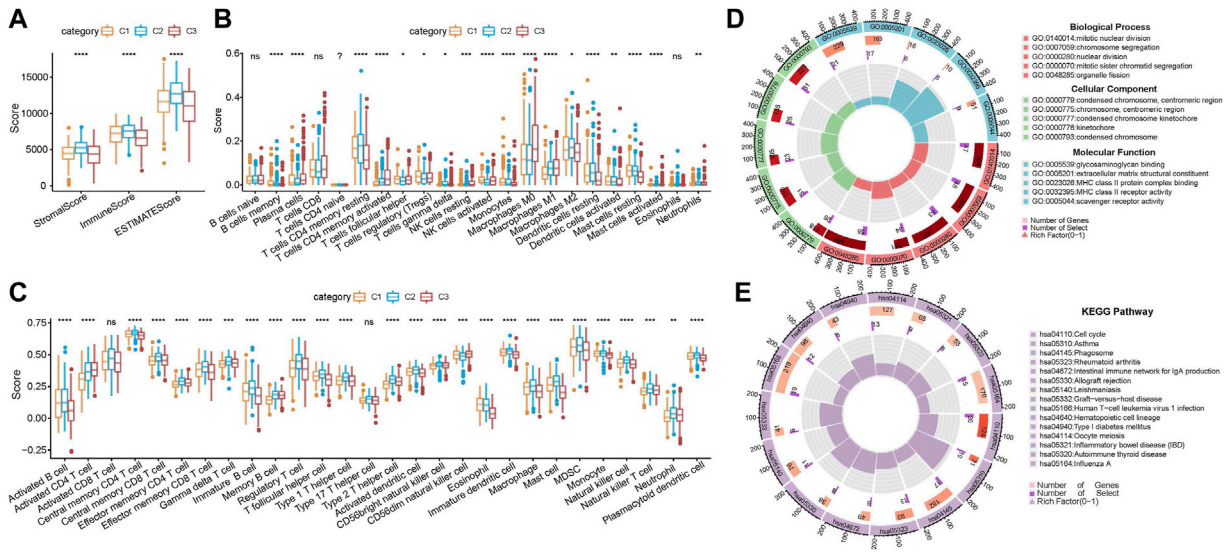


FIGURE 5
Analysis of immune abnormalities of molecular subtypes and enrichment analysis of differential genes. **(A)** Comparison of immune scores among different subtypes. **(B)** Comparison of 22 immune cell scores among different subtypes. **(C)** Comparison of the abundance of 28 immune cells among different subtypes; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, and ns represents $p > 0.05$. **(D)** Circle diagram of the top five functional enrichment analyses of differential genes GO enrichment analysis. **(E)** Circle diagram of the top 15 functional enrichment analyses for KEGG enrichment analysis of differential genes.

2.2 Classification of molecular subtypes

The ConsensusClusterPlus package was used to cluster the TCGA-LUAD and GSE31210 dataset, and the clustering heatmap of the samples was drawn (Liu et al., 2022a). Cumulative distribution function (CDF) was used to obtain the optimal clustering number and relatively stable clustering results. Three molecular subtypes (C1, C2, and C3) were then identified. To further analyze the prognosis of different molecular subtypes, Kaplan–Meier (KM) curves were drawn using the survminer package.

2.3 Filtering of differentially expressed genes and enrichment analysis

In order to further screen gene sets related to FAM subtypes, we used the limma package to analyze the differences between C1 and C2+C3, C2 and C1+C3, and C3 and C1+C2 in the TCGA-LUAD dataset under the threshold of | log2 (a Fold Change) | > 1, FDR <0.05 (Ritchie et al., 2015). GO and KEGG enrichment analysis were performed on genes showing abnormal expression using the clusterProfiler software package (Yang et al., 2023).

2.4 Construction of the RiskScore model

We used univariate Cox analysis for analyzing the differentially expressed genes (DEGs) (Peng et al., 2021). The prognostic genes with $p < 0.001$ were screened. Machine learning models can be widely used in the medical field due to their excellent performance in predicting classification problems (Choi et al., 2020). Therefore, based on six machine learning algorithms, namely, GBM (Dash et al., 2022), LASSO (Kang et al., 2021), XGBoost (Li et al., 2022a), SVM (Zhou, 2022), random forest (Utkin and Konstantinov, 2022), and decision trees (Streeb et al., 2022), the DEGs in the comparison pairs of C1 and C2+C3, C2 and C1+C3, and C3 and C1+C2 were comprehensively analyzed, and the characteristic genes were obtained by overlapping analysis. A stepwise regression method was used to further compress the characteristic genes. We calculated the β value by multivariate Cox analysis (Zhang et al., 2021a). The calculation formula of the model is as follows:

$$\text{RiskScore} = \Sigma\beta i \times \text{Expi}.$$

In the formula, *Expi* is the expression value of key FAM-related genes and β is the Cox regression coefficient of the key genes.

According to the abovementioned formula, the RiskScore of each TCGA-LUAD sample was determined and then processed by the Z-score (DeVore, 2017). Then, the samples with a RiskScore less than 0 were categorized as the low-RiskScore group, while those with a RiskScore greater than 0 were categorized as the high-RiskScore group. Five sets of chip data (GSE31210 cohort, GSE19188 cohort, GSE30219 cohort, GSE37745 cohort, and GSE50081 cohort) were used to calculate the RiskScore by the same method. A receiver operating characteristic (ROC) curve was obtained using the timeROC package (Lu et al., 2022a).

2.5 Comparison of clinical features

The clinicopathological features (Gender, Event, T. Stage, M. Stage, N. Stage, and Stage) in different molecular subtypes and different RiskScore groups in the TCGA cohort were analyzed. The pheatmap package of R software was applied to plot a heatmap to

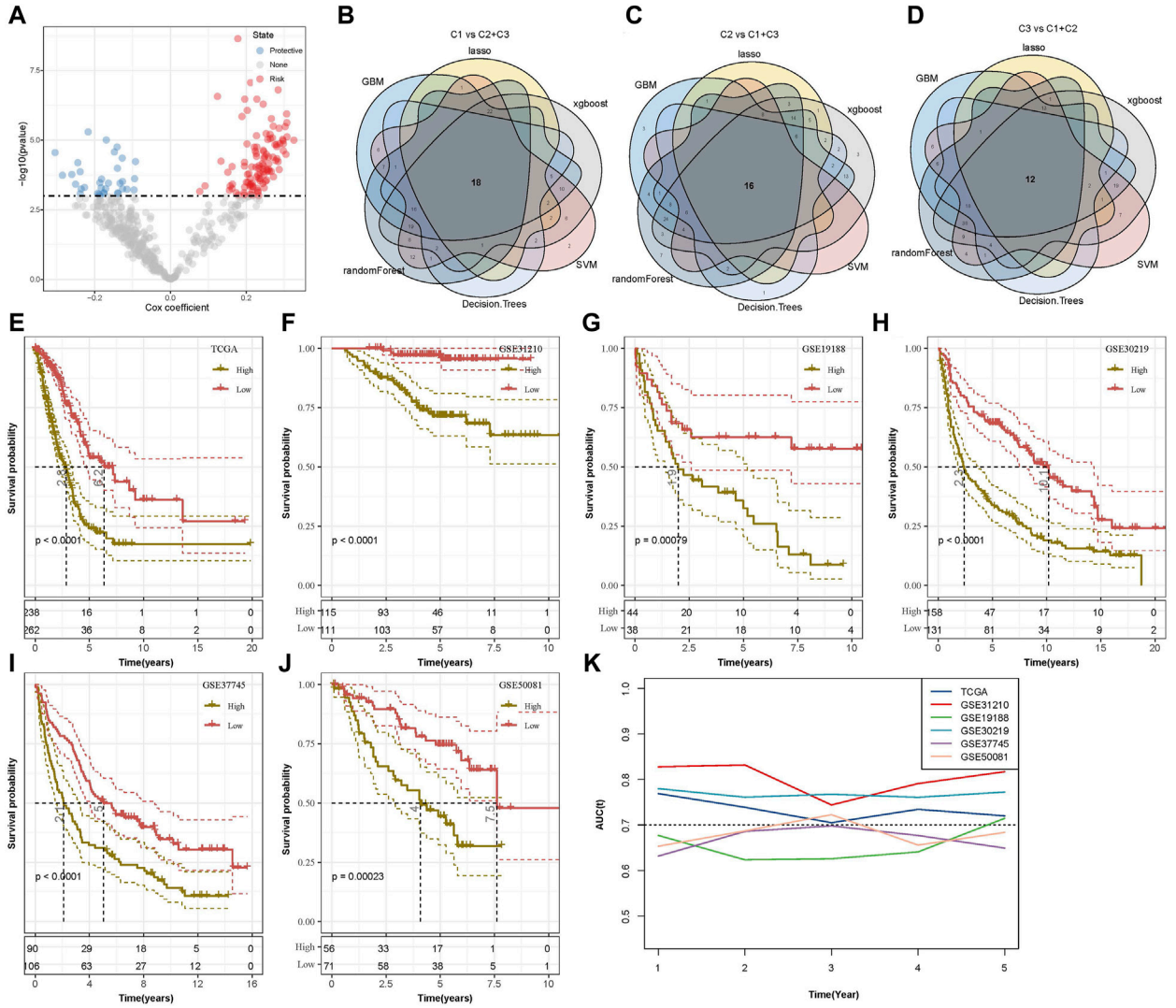


FIGURE 6
Key characteristic genes screening and K–M curves of six datasets. **(A)** Scatter plot of univariate Cox analysis of 493 genes associated with subtypes of FAM. **(B)** Venn diagram comparing C1 and C2 + C3 for characteristic gene screening of six algorithms. **(C)** Venn diagram comparing C2 and C1 + C3 for characteristic gene screening of six algorithms. **(D)** Venn diagram comparing C3 and C1 + C2 for characteristic gene screening of six algorithms. **(E)** K–M curve of the RiskScore model developed by the 12 genes in the TCGA cohort. **(F)** K–M curve of the RiskScore model developed by the 12 genes in the GSE31210 cohort. **(G)** K–M curve of the RiskScore model developed by the 12 genes in the GSE19188 cohort. **(H)** K–M curve of the RiskScore model developed by the 12 genes in the GSE30219 cohort. **(I)** K–M curve of the RiskScore model developed by the 12 genes in the GSE37745 cohort. **(J)** K–M curve of the RiskScore model developed by the 12 genes in the GSE50081 cohort. **(K)** Line graph of AUC for 1–5 years of RiskScore for six datasets.

examine the distribution of samples with different clinical features (Zhang et al., 2021b).

2.6 Establishment of a nomogram

The relationships between clinical features, RiskScore, and prognosis were assessed applying univariate and multivariate Cox analyses (van de Vijver et al., 2002). The model prediction efficiency was evaluated by developing a nomogram combining key clinicopathological features using the rms package (Liu et al., 2021a). A calibration curve was used to evaluate the predictive power of the model and to test the prediction performance of the nomogram (Van Calster et al., 2019). We also used the ggDCA package to assess the stability of the decision curve analysis (DCA) and to plot the calibration curve and DCA for the nomogram in predicting 1-, 3-, and 5-year prognosis (Van Calster et al., 2018).

2.7 Mutation analysis

Data for copy number variant (CNV) were downloaded to compare the deletion or amplification of genes associated with FAM pathways (Wu et al., 2020). Then, mutation data of single nucleotide variants (SNVs) were downloaded and a waterfall map was generated using the maftools package to display SNV mutations in FAM-related pathway genes (Li et al., 2022b).

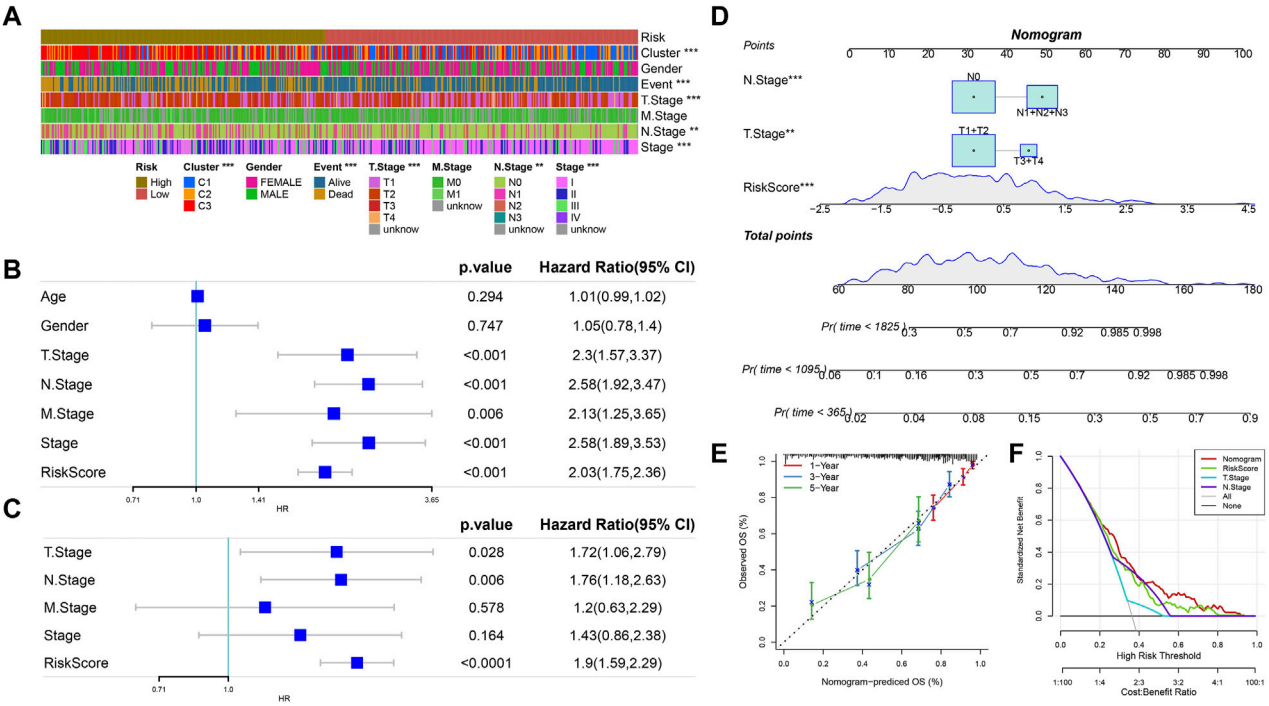


FIGURE 7
RiskScore combined with clinical characteristics to analyze the predictive performance of the model. **(A)** Heatmap of the distribution of clinical features in the high- and low-RiskScore group samples. **(B)** Univariate Cox analysis of clinical features and RiskScore. **(C)** Multivariate Cox analysis of clinical features and RiskScore (univariate prognostic correlation was selected here). **(D)** Nomogram model. **(E)** Calibration curves of the nomogram in 1, 3, and 5 years. **(F)** Decision curve of the nomogram.

Differences in genomic changes were examined in different molecular subtypes. Mutated datasets were processed using mutect2 software (Pei et al., 2021). Genes showing a mutation frequency greater than 3 were filtered. Fisher's test was applied to detect frequently mutated genes in each subtype ($p < 0.05$).

2.8 Pathway difference analysis

The FAM pathway scores were calculated by single sample gene set enrichment analysis (ssGSEA) (Zhuang et al., 2021). Differences in FAM-related pathway scores between LUAD and para-cancer tissues were compared by the Wilcoxon signed-rank test (Divine et al., 2013). We used the pheatmap package to draw heatmap to show the expression of related genes (Zheng et al., 2022).

In order to characterize biological process pathways, the GSVA software package (Hänzelmann et al., 2013) was used to analyze all the relevant gene sets in the Hallmark database. The Kruskal test was performed to examine differentially activated pathways in different molecular subtypes (Liu et al., 2021b). Significant pathways were selected under $p < 0.05$, and the heatmap of functional enrichment scores of each subtype was generated.

2.9 Comparison of immune abnormalities

Immune infiltration was evaluated by Estimation of STromal and Immune cells in MAlignant Tumors using Expression (ESTIMATE), and the differences in immune scores were compared (Yang et al., 2021). Then, CIBERSORT algorithm was used to calculate the abundance of 22 kinds of immune cells and compare the differences in immune cell scores (Zhang et al., 2022a). A variety of immune cell characteristic genes were identified (Charoentong et al., 2017). We compared the differences of 28 immune cell scores using the ssGSEA. Furthermore, the expression of gene multiple immune checkpoint genes was analyzed between different RiskScore groups (Danilova et al., 2019).

2.10 Drug sensitivity analysis

The R language pRRophetic package could be used to predict patient sensitivity to drugs. Several commonly used drugs such as erlotinib, paclitaxel, MG-132, rapamycin, sunitinib, and cisplatin were selected (Skalniak et al., 2013; Landi and Cappuzzo, 2015; Bhaoighill and Dunlop, 2019; Wang et al., 2020; Lu et al., 2022b).

2.11 Statistical analysis

This study mainly used R software for statistical analysis. A $p < 0.05$ was defined as a statistically significant difference. The Wilcoxon test was used to assess differences in immune abnormalities between RiskScore groups.

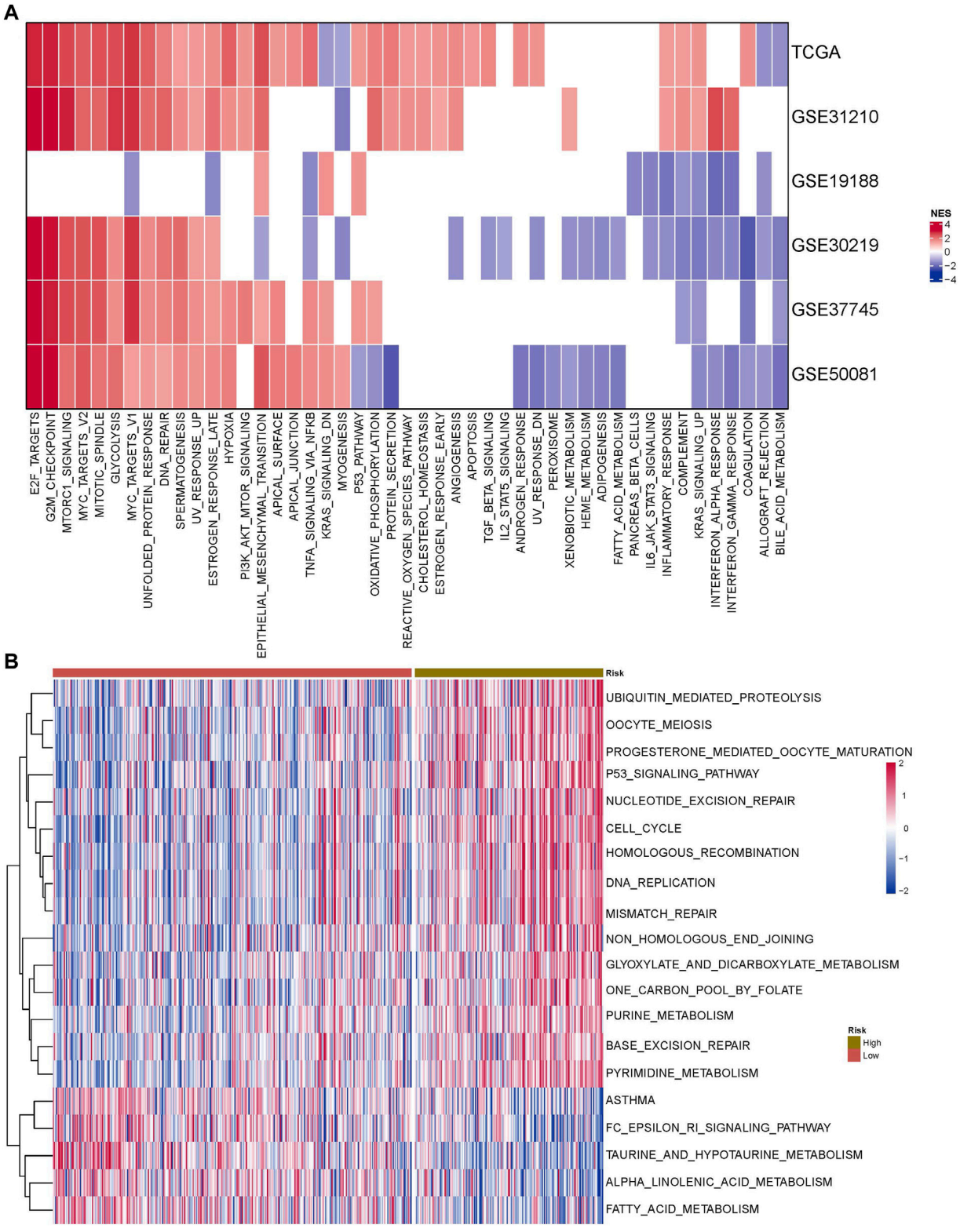


FIGURE 8
Combined with the RiskScore model, potential regulatory pathways were identified. **(A)** Pathway enrichment score heatmap obtained by six datasets in the HALLMARK gene set. **(B)** Heatmap of correlated pathway scores in the KEGG database TCGA cohort in high- and low-RiskScore groups.

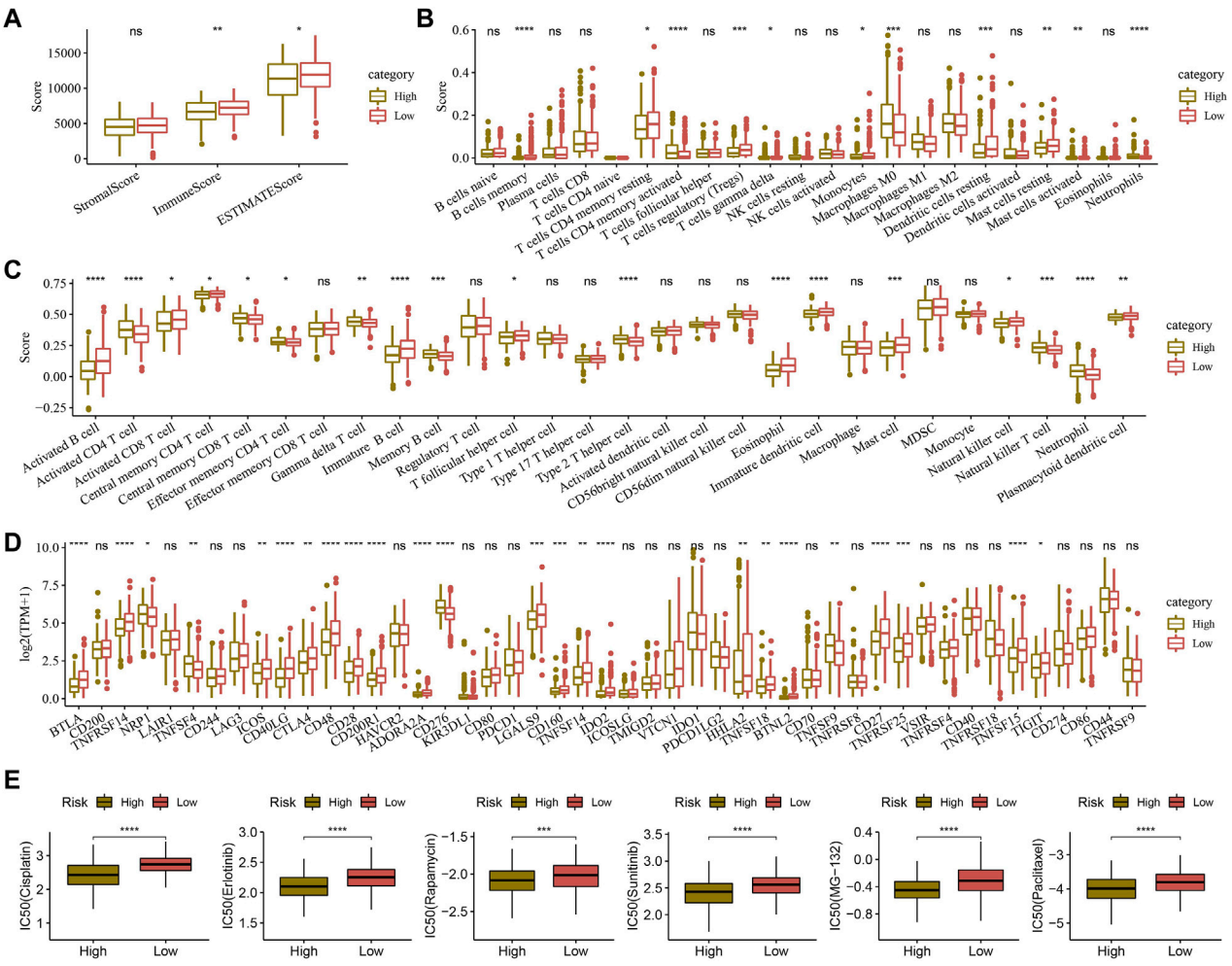


FIGURE 9
Comparison of high- and low-RiskScore groups with immune abnormalities. **(A)** Immune score was compared between low- and high-RiskScore groups. **(B)** In total, 22 immune cell scores were compared between low- and high-RiskScore groups. **(C)** In total, 28 immune cell scores were compared between low- and high-RiskScore groups. **(D)** Immune checkpoint gene expression was compared between low- and high-RiskScore groups. **(E)** Drug susceptibility was compared in low- and high-RiskScore groups. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, and ns represents $p > 0.05$.

3 Results

3.1 Abnormal FAM-related pathway genes in LUAD

After selecting FAM-related pathway genes, we found that some genes, such as ACADM, ACSL1, and CPT1C, tended to show deletion, while some genes, such as ALDH9A1, CPT1A, and ACOX1, tended to show amplification (Figure 1A). The SNV of the genes related to FAM pathways was shown in waterfall diagram, and it was found that ACSL6 had the highest mutation rate and was mostly missense mutation (Figure 1B). A comparison of scores of FAM pathways showed that LUAD tissues had a lower score of FAM pathways compared to para-cancer tissues (Figure 1C). The expression of 24 FAM-related pathway genes was more active in para-cancer tissues than that in LUAD tissues, such as ACAA1, ACAT2, and ADH1B (Figure 1D). These results suggested that FAM-related genes may have an impact on the progression of LUAD.

3.2 Classification of molecular subtypes based on genes related to FAM pathways

In the TCGA dataset, the CDF curve showed that cluster 3 was a relatively stable clustering (Figure 2A). Finally, three molecular subtypes of C1, C2, and C3 were defined based on the sample clustering heatmap (Figure 2B). Further analysis of the K–M curves for the three molecular subtypes showed significant differences in terms of prognostic survival among the three subtypes ($p = 0.0027$). Overall, C1 had the best survival outcome, followed by C2 and C3 (Figure 2C).

We used the abovementioned methods to analyze and classify the GSE31210 dataset. The CDF results showed that cluster 3 also had relatively stable clustering results (Figure 2D). At k = 3, the three molecular subtypes were significantly different (Figure 2E). There were also significant differences in prognostic survival among the three subtypes ($p = 0.00041$). The survival of C1 was found to be the most favorable, while that of C3 was the worst, and the overall results were similar to those of the TCGA dataset (Figure 2F).

The gene expression of FAM-related pathways in the three subtypes was shown in the heatmap. It was found that the gene expression of FAM-related pathways in C1 was relatively active, while that in C3 was relatively poor (Figure 2G). These results suggested that FAM-related subtypes were associated with different prognosis, and that tumors of different subtypes had large differences in the FAM status.

3.3 Significant differences in clinicopathological characteristics among the three subtypes

The clinicopathological characteristics of different subtypes in the TCGA cohort were analyzed. There were significant differences in three clinical indicators (Event, N Stage, and Stage) of the three subtypes ($p < 0.05$). Event showed that C1 had a significantly higher survival probability compared to C2 and C3, while C3 had the worst survival outcomes. As for N stage, C1 had the highest proportion of N0, while C3 had the highest proportion of N2. In Stage, C1 had the largest proportion in stage I, while C3 had the largest proportion in stage IV. Therefore, the outcome, clinical grade, and staging of C1 were relatively favorable, while those of C3 were unfavorable (Figure 3).

3.4 Mutant characteristics and differential activation pathways of the three subtypes

Differences in genomic changes among subtypes in the TCGA cohort were analyzed. The mutation dataset of TCGA was processed by mutect2 software, and a total of 9,922 genes were screened. Fisher's test was used for screening with $p < 0.05$, which filtered 770 genes. The top 20 genes were selected for further analysis on the characteristics of somatic mutations. The results showed that C1 had the lowest mutation rate and C3 had the highest mutation rate (Figure 4A). Moreover, whether differentially activated pathways were present in different subtypes were explored. Some screened pathways were found to be significantly differentially activated in different subtypes, for example, PI3K AKT MTOR SIGNALING, G2M CHECKPOINT, and FATTY ACID METABOLISM. The FATTY ACID METABOLISM pathway was actively expressed in the C1 and C2 subtypes but less expressed in the C3 subtype (Figure 4B). This could explain a poorer prognosis of C3.

3.5 Analysis of immune abnormalities of subtypes

Analysis on differences in the immune microenvironment among the three subtypes showed that C3, with a poor prognosis, had the lowest scores of StromalScore, ImmuneScore, and ESTIMATEScore, which indicated lower immune infiltration of C3 (Figure 5A). The abundance of 18 kinds of immune cells, such as B-cell memory, was different among the three subtypes (Figure 5B). The calculation results of 28 immune cell scores, such as activated B cells and activated CD4 T cells, demonstrated differences in 26 immune cell scores among the three subtypes (Figure 5C).

3.6 Screening and enrichment of FAM-related genes

To screen gene sets associated with FAM subtypes, differential analysis was performed for C1 and C1+C2, C2 and C1+C3, and C3 and C1+C2. A total of 124 upregulated genes and 123 downregulated genes were screened in the comparison between C1 and C2+C3; 59 upregulated genes and four downregulated genes were screened in the comparison between C2 and C1+C3; and 160 upregulated genes and 276 downregulated genes were screened in the comparison between C3 and C1+C2. There were 493 DEGs in total. Circle diagrams of GO enrichment analysis on the top five functional enrichment analyses were plotted, and we observed that mitotic sister chromatid segregation was the most significant biological process. The condensed chromosome centromeric region was the most prominent cellular component. The MHC class II receptor activity was the most active molecular function (Figure 5D). Asthma was found to be the most significant pathway in the circle diagram of the top 15 functional enrichment analyses on the differential genes (Figure 5E).

3.7 Construction and evaluation of a RiskScore model

Building upon these findings, univariate Cox analysis was used to perform prognostic analysis on 493 genes related to FAM subtypes, and 143 were screened to be prognostic genes relevant to LUAD ($p < 0.001$) (Figure 6A). Subsequently, we determined feature genes using six machine algorithms (including LASSO, GBM, random forest, SVN, XGBoost, and decision trees) for C1 and C2+C3, C2 and C1+C3, and C3 and C1+C2. Through the Venn diagram, we found a total of 18 characterized genes between C1 and C2+C3 (Figure 6B), a total of 16 genes between C2 and C1+C3 (Figure 6C), and a total of 12 genes between C3 and C1+C2 (Figure 6D). Further comparison and screening showed 34 important genes for subsequent studies. Finally, the number of important genes was reduced to 12 key genes (SLC2A1, PKP2, FAM83A, TCN1, MS4A1, CLIC6, UBE2S, RRM2, CDC45, IGF2BP1, ANGPTL4, and CD109) by the stepwise regression method.

The calculation formula is as follows:

RiskScore = −0.293*SLC2A1+0.145*PKP2+0.113*FAM83A+0.092*TCN1-0.142*MS4A1-0.081*CLIC6+0.24*UBE2S+0.217*RRM2-0.286*CDC45+0.162*IGF2BP1+0.093*ANGPTL4+0.112*CD109.

The RiskScore of the sample in TCGA was calculated by the abovementioned model formula, and high- and low-RiskScore groups were classified. The K–M curve results showed that the survival probability of the high-RiskScore group was lower ($p < 0.0001$, Figure 6E). At the same time, five sets of chip data (GSE31210, GSE19188, GSE30219, GSE37745, and GSE50081) were used to draw K–M curves, and the survival probability of the high-RiskScore group was still lower (Figures 6F–J). The AUC values of the 1-, 3-, and 5-year RiskScore of six datasets were observed, and it was found that the values of six datasets were all around 0.7 and that the AUC values of three datasets were

consistently above 0.7, indicating that the model had a strong predictive performance (Figure 6K).

3.8 Testing the predictive performance of the RiskScore model combined with clinical characteristics

Combined with the RiskScore, the distribution of samples with multiple clinical characteristics was presented in the form of a heatmap. The results showed that the distribution of five clinical characteristics (Cluster, Event, T Stage, N Stage, and Stage) was closely correlated with that of the RiskScore (Figure 7A). Meanwhile, the univariate Cox analysis between each clinical feature and the RiskScore showed that the *p*-values of T Stage, N Stage, Stage, and RiskScore were all less than 0.001. The multivariate Cox analysis showed that the *p*-values of RiskScore, T Stage, and N Stage were all less than 0.05. Therefore, T Stage, N Stage, and RiskScore were independent prognostic factors (Figures 7B, C). A nomogram was established by the abovementioned factors. According to the results, the RiskScore had the strongest survival prediction ability (Figure 7D). The slope and distance between all the calibration curves and standard ones were similar, which verified the nomogram's prediction capability (Figure 7E). The benefit rates of the RiskScore and nomogram were significantly higher than the extremum curves, which proved that the nomogram and RiskScore had the greatest power for survival prediction (Figure 7F).

3.9 Potential regulatory pathways identified by the RiskScore model

The HALLMARK gene set was enriched in six datasets by GSEA, and it was found that E2F target and G2M checkpoint pathways had the highest scores, while bile acid metabolism and other pathways had lower scores. The FAM pathway score was also generally low (Figure 8A). We compared the predicted pathway scores of RiskScore groups in the TCGA cohort and observed that the scores of 15 pathways including ubiquitin-mediated proteolysis were higher in the high-RiskScore group, while the scores of five pathways including FAM were lower in the high-RiskScore group (Figure 8B).

3.10 Immune status and immunotherapy preference predicted by the RiskScore

The ImmuneScore and ESTIMATEScore were higher in the high-RiskScore group compared to those in the low-RiskScore group (Figure 9A). Among the 22 immune cell scores predicted, 11 immune cell scores, such as T-cell CD4 memory resting, showed significant differences between the two RiskScore groups (Figure 9B). Among the 28 immune cell scores predicted, 18 immune cell scores, such as activated CD4 T cells, showed differences between the two RiskScore groups (Figure 9C). The expression of 48 different immune checkpoint genes was compared between the two RiskScore groups, and higher expression of 19 immune checkpoint genes, such as BTLA, TNFRSF14, ICOS, and CD48, was found in the low-RiskScore group (Figure 9D).

Cisplatin, erlotinib, rapamycin, sunitinib, MG-132, and paclitaxel were all found to be more sensitive to the high-RiskScore group, suggesting that patients in high-RiskScore groups might respond better to these drugs (Figure 9E).

4 Discussion

Lung cancer is one of the most deadly malignant tumors worldwide (Wang et al., 2022b). The metabolic reprogramming of cancer cells, particularly the modification of FAM, is firmly connected with tumor growth (Maan et al., 2018). Abnormal FAM is associated with the growth, differentiation, and metastasis of LUAD cells. Acetyl-coA carboxylase 2 (ACC2) is a key FAM enzyme. Fei-Yuan Yu et al. found that ACC2 is low-expressed in tumor cells, and its expression is negatively correlated with tumor progression (Yu et al., 2022). FASN is a homodimeric multienzymatic protein that inhibits and blocks the adipogenic pathway and hinders fatty acid synthesis. This causes apoptosis in tumor cells to overexpress FASN without affecting non-malignant cells (Relat et al., 2012). Recent studies showed that FASN expression is upregulated and overactivated in LUAD, which may be related to the progression of LUAD (Relat et al., 2012). Drug targeting FAM pathways in LUAD has been designed. For example, AZ12756122, a novel FASN inhibitor, can induce cell apoptosis, downregulate FASN expression and activity, and reduce EGFR and Akt/mTOR pathway activation (Polonio-Alcalá et al., 2022). Although the relationship between the gene expression of FAM pathways and the prognosis of LUAD has been explored, this study introduced a variety of machine learning analysis algorithms to more comprehensively analyze the genetic characteristics of FAM pathways (Ganggayah et al., 2019), which can help establish a more effective risk prediction model for LUAD based on the FAM pathway genes.

Using LUAD data from the TCGA dataset, GEO dataset, and FAM-related gene sets obtained by KEGG analysis, we found that LUAD had lower scores of FAM-related pathways. Molecular subtypes were classified using the genes related to FAM pathways, and six machine learning methods were applied to select key genes related to the three LUAD subtypes. A total of 12 key genes (SLC2A1, PKP2, FAM83A, TCN1, MS4A1, CLIC6, UBE2S, RRM2, CDC45, IGF2BP1, ANGPTL4, and CD109) were determined to be closely related to LUAD prognosis. The downregulation of SLC2A1-AS1 can inhibit LUAD cell growth and expansion, and its overexpression increases tumor cell proliferation and differentiation. PKP2 promotes the growth, division, and migration of cancer cells through activating the EGFR signaling pathway in LUAD cells (Hao et al., 2019). FAM83A-AS1 knockdown can suppress the proliferation of LUAD cells, can inhibit the expression of HIF-1α and glycolytic genes, and also plays a role in FAM (Chen et al., 2022b). High expression of TCN1, a vitamin B12-binding protein, is positively associated with cancer aggressiveness and a poor prognosis (Li et al.,

2022c). The expression level of MS4A1 in colorectal cancer is positively correlated with patients' prognosis. CLIC6 is upregulated in most obese patients with endometrial cancer (López-Ozuna et al., 2021; Mudd et al., 2021). Mengjun Zhang et al. observed that UBE2S can promote PI3K or mTOR signaling pathway, block the regulation of cell cycle, inhibit cell apoptosis, and promote the proliferation, migration, and prognosis of ovarian cancer (Zhang et al., 2022b). RRM2 is upregulated in LUAD, and high RRM2 expression is associated with a poorer survival and lower immune infiltration (Ma et al., 2020). In addition, Zhou et al. demonstrated that RRM2 is overexpressed in the cell lines and clinical samples of bladder cancer and that blocking RRM2 inhibits the growth and proliferation of cancer cells (Zhou et al., 2022). CDC45, a key protein involved in the initiation of DNA replication, is upregulated in many cancers, and its expression is significantly negatively correlated with patient prognosis (Lu et al., 2022c). JinFeng Liu et al. found that IGF2BP1 is significantly abnormally expressed in LUAD samples. Moreover, ANGPTL4 is also significantly upregulated in LUAD samples, which are all closely related to the development and a poor prognosis of LUAD (Liu et al., 2022b; Yang et al., 2022). Tetsuro Taki et al. demonstrated the biological significance of regulating the TGF-β signal in the cancer cell matrix through the correlation verification of CD109 and LTBP1, and they indicated that the expression level of CD109 plays an important role in promoting the proliferation and diffusion of LUAD cells (Taki et al., 2020). Furthermore, Lee et al. demonstrated that C109 expression is correlated with the invasiveness and metastasis of LUAD. They observed that CD109 expression is mechanistically mediated by binding to EGFR to regulate AKT/mTOR signaling (Lee et al., 2020). Therefore, in this study, the selected FAM-related genes may all be involved in the progression of LUAD and can serve as biomarkers for the clinical diagnosis and treatment of cancer. Therefore, the 12 key genes were used to construct a RiskScore model, laying a foundation for the survival prediction and further study of LUAD.

LUAD tissues had a lower score of the FAM-related pathway compared to para-cancer tissues. This also indicated that abnormal FAM was involved in the progression and prognosis of LUAD, which is consistent with the characterization results of FAM in LUAD by Wang et al. (2022a). As a new treatment method, immunotherapy has become an effective strategy to treat cancers (Riley et al., 2019). The level of immune cell infiltration in the tumor microenvironment has also been used as an important indicator for assessing lung cancer (Liu et al., 2021c). There are many studies investigating the effect of FAM on immunotherapy in various cancers (Bleve et al., 2020). The differential expression of FASN is closely correlated with immune cell infiltration, and patients with a low expression of FASN have active response to immune checkpoint inhibitor treatment (Xiong et al., 2022). The analysis and comparison of various tumor-related studies showed that the upregulated FASN gene expression and activity is negatively correlated with tumor immune infiltration. The methylation of the FASN promoter in DNA can be used to serve as a new biomarker for cancer (Zhang et al., 2022c). Hence, a better understanding of the correlation between FAM and the immunological signature of the tumor microenvironment could facilitate the identification of new therapeutic targets for improving clinical cancer therapies.

However, this study also had certain limitations. The tumor and gene sample data collected were all from the database with a small number of samples, which demanded further *in vivo* or *in vitro* validation experiments to verify the predictive performance of the prognostic model. At the same time, the specific mechanism of abnormal FAM in LUAD was not clearly studied, and its interactions and regulatory mechanisms should be explored in depth.

5 Conclusion

In summary, this study determined 12 key genes (SLC2A1, PKP2, FAM83A, TCN1, MS4A1, CLIC6, UBE2S, RRM2, CDC45, IGF2BP1, ANGPTL4, and CD109) using six machine learning methods. A RiskScore model was constructed based on the 12 key genes mentioned previously. The model can accurately predict the survival of LUAD patients. We demonstrated that a specific model based on FAM could provide significant benefits for the precision treatment of LUAD and was effective in improving the prediction of patients' prognoses.

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.

Author contributions

DC: conceptualization and writing–original draft. YZ: data curation and writing–original draft. WZ: methodology and writing–original draft. JL: software and writing–original draft. YB: conceptualization, data curation, and writing–review and editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by grants from the Jilin Province Natural Science Foundation (YDZJ202301ZYTS039), Jilin Province Health Research Talent Special Project (2022SCZ20), Beijing Medical Award Foundation (YXJL-20W-0625-0295), and Wu Jieping Foundation (320.6750).

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

Amiri, M., Yousefnia, S., Seyed Forootan, F., Peymani, M., Ghaedi, K., and Nasr Esfahani, M. H. (2018). Diverse roles of fatty acid binding proteins (FABPs) in development and pathogenesis of cancers. *Gene* 676, 171–183. doi:10.1016/j.gene.2018.07.035

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets--update. *Nucleic acids Res.* 41, D991–D995. (Database issue). doi:10.1093/nar/gks1193

Bhaoighill, M. N., and Dunlop, E. A. (2019). Mechanistic target of rapamycin inhibitors: successes and challenges as cancer therapeutics. *Cancer drug Resist. (Alhambra, Calif.)* 2 (4), 1069–1085. doi:10.20517/cdr.2019.87

Bleve, A., Durante, B., Sica, A., and Consonni, F. M. (2020). Lipid metabolism and cancer immunotherapy: immunosuppressive myeloid cells at the crossroad. *Int. J. Mol. Sci.* 21 (16), 5845. doi:10.3390/ijms21165845

Chang, L., Fang, S., Chen, Y., Yang, Z., Yuan, Y., Zhang, J., et al. (2019). Inhibition of FASN suppresses the malignant biological behavior of non-small cell lung cancer cells via deregulating glucose metabolism and AKT/ERK pathway. *Lipids health Dis.* 18 (1), 118. doi:10.1186/s12944-019-1058-8

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell. Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Chen, J., Fu, Y., Hu, J., and He, J. (2022a). Hypoxia-related gene signature for predicting LUAD patients' prognosis and immune microenvironment. *Cytokine* 152, 155820. doi:10.1016/j.cyto.2022.155820

Chen, Z., Hu, Z., Sui, Q., Huang, Y., Zhao, M., Li, M., et al. (2022b). LncRNA FAM83A-AS1 facilitates tumor proliferation and the migration via the HIF-1α/glycolysis axis in lung adenocarcinoma. *Int. J. Biol. Sci.* 18 (2), 522–535. doi:10.7150/ijbs.67556

Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., and Campbell, J. P. (2020). Introduction to machine learning, neural networks, and deep learning. *Transl. Vis. Sci. Technol.* 9 (2), 14. doi:10.1167/tvst.9.2.14

Danilova, L., Ho, W. J., Zhu, Q., Vithayathil, T., De Jesus-Acosta, A., Azad, N. S., et al. (2019). Programmed cell death ligand-1 (PD-L1) and CD8 expression profiling identify an immunologic subtype of pancreatic ductal adenocarcinomas with favorable survival. *Cancer Immunol. Res.* 7 (6), 886–895. doi:10.1158/2326-6066.CIR-18-0822

Dash, T. K., Chakraborty, C., Mahapatra, S., and Panda, G. (2022). Gradient boosting machine and efficient combination of features for speech-based detection of COVID-19. *IEEE J. Biomed. health Inf.* 26 (11), 5364–5371. doi:10.1109/JBHI.2022.3197910

Denisenko, T. V., Budkevich, I. N., and Zhivotovsky, B. (2018). Cell death-based treatment of lung adenocarcinoma. *Cell. death Dis.* 9 (2), 117. doi:10.1038/s41419-017-0063-y

DeVore, G. R. (2017). Computing the Z Score and centiles for cross-sectional analysis: a practical approach. *J. ultrasound Med. official J. Am. Inst. Ultrasound Med.* 36 (3), 459–473. doi:10.7863/ultra.16.03025

Divine, G., Norton, H. J., Hunt, R., and Dienemann, J. (2013). Statistical grand rounds: a review of analysis and sample size calculation considerations for Wilcoxon tests. *Anesth. analgesia* 117 (3), 699–710. doi:10.1213/ANE.0b013e31827f53d7

Ganggayah, M. D., Taib, N. A., Har, Y. C., Lio, P., and Dhillon, S. K. (2019). Predicting factors for survival of breast cancer patients using machine learning techniques. *BMC Med. Inf. Decis. Mak.* 19 (1), 48. doi:10.1186/s12911-019-0801-4

Garcia, K. A., Costa, M. L., Lacunza, E., Martinez, M. E., Corsico, B., and Scaglia, N. (2022). Fatty acid binding protein 5 regulates lipogenesis and tumor growth in lung adenocarcinoma. *Life Sci.* 301, 120621. doi:10.1016/j.lfs.2022.120621

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

Hao, X. L., Tian, Z., Han, F., Chen, J. P., Gao, L. Y., and Liu, J. Y. (2019). Plakophilin-2 accelerates cell proliferation and migration through activating EGFR signaling in lung adenocarcinoma. *Pathology, Res. Pract.* 215 (7), 152438. doi:10.1016/j.prp.2019.152438

Hutchinson, B. D., Shroff, G. S., Truong, M. T., and Ko, J. P. (2019). Spectrum of lung adenocarcinoma. *Seminars ultrasound, CT, MR* 40 (3), 255–264. doi:10.1053/j.sult.2018.11.009

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids Res.* 28 (1), 27–30. doi:10.1093/nar/28.1.27

Kang, J., Choi, Y. J., Kim, I. K., Lee, H. S., Kim, H., Baik, S. H., et al. (2021). LASSO-based machine learning algorithm for prediction of lymph node metastasis in T1 colorectal cancer. *Cancer Res. Treat.* 53 (3), 773–783. doi:10.4143/crt.2020.974

Landi, L., and Cappuzzo, F. (2015). Experience with erlotinib in the treatment of non-small cell lung cancer. *Ther. Adv. Respir. Dis.* 9 (4), 146–163. doi:10.1177/1753465815588053

Lee, K. Y., Shueng, P. W., Chou, C. M., Lin, B. X., Lin, M. H., Kuo, D. Y., et al. (2020). Elevation of CD109 promotes metastasis and drug resistance in lung cancer via activation of EGFR-AKT-mTOR signaling. *Cancer Sci.* 111 (5), 1652–1662. doi:10.1111/cas.14373

Li, B., Yu, L., and Gao, L. (2022b). Cancer classification based on multiple dimensions: SNV patterns. *Comput. Biol. Med.* 151, 106270. doi:10.1016/j.compbiomed.2022.106270

Li, H., Guo, L., and Cai, Z. (2022c). TCN1 is a potential prognostic biomarker and correlates with immune infiltrates in lung adenocarcinoma. *World J. Surg. Oncol.* 20 (1), 83. doi:10.1186/s12957-022-02556-8

Li, Y., Xu, Y., Ma, Z., Ye, Y., Gao, L., and Sun, Y. (2022a). An XGBoost-based model for assessment of aortic stiffness from wrist photoplethysmogram. *Comput. methods programs Biomed.* 226, 107128. doi:10.1016/j.cmpb.2022.107128

Li, Z., and Zhang, H. (2016). Reprogramming of glucose, fatty acid and amino acid metabolism for cancer progression. *Cell. Mol. life Sci. CMLS* 73 (2), 377–392. doi:10.1007/s00018-015-2070-4

Liang, C., Wang, X., Zhang, Z., Xiao, F., Feng, H., Ma, Q., et al. (2020). ACOT11 promotes cell proliferation, migration and invasion in lung adenocarcinoma. *Transl. lung cancer Res.* 9 (5), 1885–1903. doi:10.21037/tlcr-19-509

Lin, W., Chen, Y., Wu, B., Chen, Y., and Li, Z. (2021). Identification of the pyroptosis-related prognostic gene signature and the associated regulation axis in lung adenocarcinoma. *Cell. death Discov.* 7 (1), 161. doi:10.1038/s41420-021-00557-2

Liu, J., Gu, M., Xue, Y., Wang, Q., Ren, Y., and Huang, W. (2021c). Clinical significance of PD-L1 expression and CD8-positive tumor-infiltrating lymphocytes in patients with cavitary lung adenocarcinoma. *Oncologie* 23 (3), 439–452. doi:10.32604/oncologie.2021.017220

Liu, J., Li, Z., Cheang, I., Li, J., and Zhou, C. (2022b). RNA-binding protein IGF2BP1 associated with prognosis and immunotherapy response in lung adenocarcinoma. *Front. Genet.* 13, 777399. doi:10.3389/fgene.2022.777399

Liu, J., Sun, G., Pan, S., Qin, M., Ouyang, R., Li, Z., et al. (2020). The Cancer Genome Atlas (TCGA) based m(6)A methylation-related genes predict prognosis in hepatocellular carcinoma. *Bioengineered* 11 (1), 759–768. doi:10.1080/21655979.2020.1787764

Liu, T. T., Li, R., Huo, C., Li, J. P., Yao, J., Ji, X. L., et al. (2021a). Identification of CDK2-related immune forecast model and ceRNA in lung adenocarcinoma, a pan-cancer analysis. *Front. Cell. Dev. Biol.* 9, 682002. doi:10.3389/fcell.2021.682002

Liu, X., Li, J., Wang, Q., Bai, L., Xing, J., Hu, X., et al. (2022a). Analysis on heterogeneity of hepatocellular carcinoma immune cells and a molecular risk model by integration of scRNA-seq and bulk RNA-seq. *Front. Immunol.* 13, 1012303. doi:10.3389/fimmu.2022.1012303

Liu, Z., Zhang, W., Cheng, X., Wang, H., Bian, L., Wang, J., et al. (2021b). Overexpressed XRCC2 as an independent risk factor for poor prognosis in glioma patients. *Mol. Med. Camb. Mass.)* 27 (1), 52. doi:10.1186/s10020-021-00316-0

López-Ozuna, V. M., Kogan, L., Hachim, M. Y., Matanes, E., Hachim, I. Y., Mitric, C., et al. (2021). Identification of predictive biomarkers for lymph node involvement in obese women with endometrial cancer. *Front. Oncol.* 11, 695404. doi:10.3389/fonc.2021.695404

Lu, H., Wu, J., Liang, L., Wang, X., and Cai, H. (2022b). Identifying a novel defined pyroptosis-associated long noncoding RNA signature contributes to predicting prognosis and tumor microenvironment of bladder cancer. *Front. Immunol.* 13, 803355. doi:10.3389/fimmu.2022.803355

Lu, L., Wang, H., Fang, J., Zheng, J., Liu, B., Xia, L., et al. (2022a). Overexpression of OAS1 is correlated with poor prognosis in pancreatic cancer. *Front. Oncol.* 12, 944194. doi:10.3389/fonc.2022.944194

Lu, Y., Chen, X., Liu, F., Yu, H., Zhang, Y., Du, K., et al. (2022c). Systematic pan-cancer analysis identifies CDC45 as having an oncogenic role in human cancers. *Oncol. Rep.* 48 (4), 185. doi:10.3892/or.2022.8400

Ma, C., Luo, H., Cao, J., Gao, C., Fa, X., and Wang, G. (2020). Independent prognostic implications of RRM2 in lung adenocarcinoma. *J. Cancer* 11 (23), 7009–7022. doi:10.7150/jca.47895

Maan, M., Peters, J. M., Dutta, M., and Patterson, A. D. (2018). Lipid metabolism and lipophagy in cancer. *Biochem. biophysical Res. Commun.* 504 (3), 582–589. doi:10.1016/j.bbrc.2018.02.097

Mudd, T. W., Jr., Lu, C., Klement, J. D., and Liu, K. (2021). MS4A1 expression and function in T cells in the colorectal cancer tumor microenvironment. *Cell. Immunol.* 360, 104260. doi:10.1016/j.cellimm.2020.104260

Patz, E. F., Jr., Pinsky, P., Gatsonis, C., Sicks, J. D., Kramer, B. S., Tammemägi, M. C., et al. (2014). Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern. Med.* 174 (2), 269–274. doi:10.1001/jamainternmed.2013.12738

Pei, S., Liu, T., Ren, X., Li, W., Chen, C., and Xie, Z. (2021). Benchmarking variant callers in next-generation and third-generation sequencing analysis. *Briefings Bioinforma.* 22 (3), bbaa148. doi:10.1093/bib/bbaa148

Peng, Y., Liu, C., Li, M., Li, W., Zhang, M., Jiang, X., et al. (2021). Identification of a prognostic and therapeutic immune signature associated with hepatocellular carcinoma. *Cancer Cell. Int.* 21 (1), 98. doi:10.1186/s12935-021-01792-4

Polonio-Alcalá, E., Porta, R., Ruiz-Martínez, S., Vásquez-Dongo, C., Relat, J., Bosch-Barrera, J., et al. (2022). AZ12756122, a novel fatty acid synthase inhibitor, decreases resistance features in EGFR-TKI resistant EGFR-mutated NSCLC cell models. *Biomed. Pharmacother. = Biomedecine Pharmacother.* 156, 113942. doi:10.1016/j.biopha.2022.113942

Relat, J., Blancafort, A., Oliveras, G., Cufí, S., Haro, D., Marrero, P. F., et al. (2012). Different fatty acid metabolism effects of (-)-epigallocatechin-3-gallate and C75 in adenocarcinoma lung cancer. *BMC cancer* 12, 280. doi:10.1186/1471-2407-12-280

Riley, R. S., June, C. H., Langer, R., and Mitchell, M. J. (2019). Delivery technologies for cancer immunotherapy. *Nat. Rev. Drug Discov.* 18 (3), 175–196. doi:10.1038/s41573-018-0006-z

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Santos, C. R., and Schulze, A. (2012). Lipid metabolism in cancer. *FEBS J.* 279 (15), 2610–2623. doi:10.1111/j.1742-4658.2012.08644.x

Shen, J., Huang, J., Huang, Y., Chen, Y., Li, J., Luo, P., et al. (2022). Anlotinib suppresses lung adenocarcinoma growth via inhibiting FASN-mediated lipid metabolism. *Ann. Transl. Med.* 10 (24), 1337. doi:10.21037/atm-22-5438

Shi, J., Chen, Y., Peng, C., Kuang, L., Zhang, Z., Li, Y., et al. (2022). Advances in targeted therapy against driver mutations and epigenetic alterations in non-small cell lung cancer. *Oncologie* 24 (4), 613–648. doi:10.32604/oncologie.2022.027545

Skalniak, L., Koj, A., and Jura, J. (2013). Proteasome inhibitor MG-132 induces MCPIP1 expression. *FEBS J.* 280 (11), 2665–2674. doi:10.1111/febs.12264

Skřičková, J., Kadlec, B., Venclíček, O., and Merta, Z. (2018). Lung cancer. *Cas. Lek. ceskych* 157 (5), 226–236.

Streeb, D., Metz, Y., Schlegel, U., Schneider, B., El-Assady, M., Neth, H., et al. (2022). Task-based visual interactive modeling: decision trees and rule-based classifiers. *IEEE Trans. Vis. Comput. Graph.* 28 (9), 3307–3323. doi:10.1109/TVCG.2020.3045560

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA a cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Taki, T., Shiraki, Y., Enomoto, A., Weng, L., Chen, C., Asai, N., et al. (2020). CD109 regulates *in vivo* tumor invasion in lung adenocarcinoma through TGF-β signaling. *Cancer Sci.* 111 (12), 4616–4628. doi:10.1111/cas.14673

Utkin, L. V., and Konstantinov, A. V. (2022). Attention-based random forest and contamination model. *Neural Netw. official J. Int. Neural Netw. Soc.* 154, 346–359. doi:10.1016/j.neunet.2022.07.029

Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., Steyerberg, E. W., and Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative (2019). Calibration: the Achilles heel of predictive analytics. *BMC Med.* 17 (1), 230. doi:10.1186/s12916-019-1466-7

Van Calster, B., Wynants, L., Verbeek, J. F. M., Verbakel, J. Y., Christodoulou, E., Vickers, A. J., et al. (2018). Reporting and interpreting decision curve analysis: a guide for investigators. *Eur. Urol.* 74 (6), 796–804. doi:10.1016/j.eururo.2018.08.038

van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* 347 (25), 1999–2009. doi:10.1056/NEJMoa021967

Wang, D., Xiao, F., Feng, Z., Li, M., Kong, L., Huang, L., et al. (2020). Sunitinib facilitates metastatic breast cancer spreading by inducing endothelial cell senescence. *Breast cancer Res. BCR* 22 (1), 103. doi:10.1186/s13058-020-01346-y

Wang, H., Fang, J., Wang, Y., Li, S., Wang, Z., He, W., et al. (2022b). Gene editing in non-small cell lung cancer: current application and future perspective. *Oncologie* 24 (1), 65–83. doi:10.32604/oncologie.2022.021863

Wang, S., Chen, A., Zhu, W., Feng, D., Wei, J., Li, Q., et al. (2022a). Characterization of fatty acid metabolism in lung adenocarcinoma. *Front. Genet.* 13, 905508. doi:10.3389/fgene.2022.905508

Wu, L., Wang, H., Xia, Y., Xi, R., and Cnv-Bac, (2020). CNV-BAC: copy number variation detection in bacterial circular Genome. *Bioinforma. Oxf. Engl.* 36 (12), 3890–3891. doi:10.1093/bioinformatics/btaa208

Xiong, Q., Feng, D., Wang, Z., Ying, Y., Xu, C., Wei, Q., et al. (2022). Fatty acid synthase is the key regulator of fatty acid metabolism and is related to immunotherapy in bladder cancer. *Front. Immunol.* 13, 836939. doi:10.3389/fimmu.2022.836939

Yang, W. X., Gao, H. W., Cui, J. B., Zhang, A. A., Wang, F. F., Xie, J. Q., et al. (2023). Development and validation of a coagulation-related genes prognostic model for hepatocellular carcinoma. *BMC Bioinforma.* 24 (1), 89. doi:10.1186/s12859-023-05220-4

Yang, Y., Liu, Y., Gao, P., Liu, K., Zhao, K., Ying, R., et al. (2022). Prognostic significance of ANGPTL4 in lung adenocarcinoma: a meta-analysis based on integrated TCGA and GEO databases. *Evidence-based complementary Altern. Med. eCAM* 2022, 3444740. doi:10.1155/2022/3444740

Yang, Z., Wei, X., Pan, Y., Xu, J., Si, Y., Min, Z., et al. (2021). A new risk factor indicator for papillary thyroid cancer based on immune infiltration. *Cell. death Dis.* 12 (1), 51. doi:10.1038/s41419-020-03294-z

Yu, F. Y., Xu, Q., Wei, Q. Y., Mo, H. Y., Zhong, Q. H., Zhao, X. Y., et al. (2022). ACC2 is under-expressed in lung adenocarcinoma and predicts poor clinical outcomes. *J. cancer Res. Clin. Oncol.* 148 (11), 3145–3162. doi:10.1007/s00432-021-03910-1

Yu, X. H., Ren, X. H., Liang, X. H., and Tang, Y. L. (2018). Roles of fatty acid metabolism in tumourigenesis: beyond providing nutrition (Review). *Mol. Med. Rep.* 18 (6), 5307–5316. doi:10.3892/mmr.2018.9577

Zhang, M., Liu, Y., Yin, Y., Sun, Z., Wang, Y., Zhang, Z., et al. (2022b). UBE2S promotes the development of ovarian cancer by promoting PI3K/AKT/mTOR signaling pathway to regulate cell cycle and apoptosis. *Mol. Med. Camb. Mass.)* 28 (1), 62. doi:10.1186/s10020-022-00489-2

Zhang, M. Y., Huo, C., Liu, J. Y., Shi, Z. E., Zhang, W. D., Qu, J. J., et al. (2021b). Identification of a five autophagy subtype-related gene expression pattern for improving the prognosis of lung adenocarcinoma. *Front. Cell. Dev. Biol.* 9, 756911. doi:10.3389/fcell.2021.756911

Zhang, M., Yu, L., Sun, Y., Hao, L., Bai, J., Yuan, X., et al. (2022c). Comprehensive analysis of FASN in tumor immune infiltration and prognostic value for immunotherapy and promoter DNA methylation. *Int. J. Mol. Sci.* 23 (24), 15603. doi:10.3390/ijms232415603

Zhang, W., Ji, L., Wang, X., Zhu, S., Luo, J., Zhang, Y., et al. (2021a). Nomogram predicts risk and prognostic factors for bone metastasis of pancreatic cancer: a population-based analysis. *Front. Endocrinol.* 12, 752176. doi:10.3389/fendo.2021.752176

Zhang, Y. P., Wang, X., Jie, L. G., Qu, Y., Zhu, X. T., Wu, J., et al. (2022a). Osteoarticular involvement-associated biomarkers and pathways in psoriasis: the shared pathway with ankylosing spondylitis. *Front. Immunol.* 13, 836533. doi:10.3389/fimmu.2022.836533

Zheng, X., Ma, Y., Bai, Y., Huang, T., Lv, X., Deng, J., et al. (2022). Identification and validation of immunotherapy for four novel clusters of colorectal cancer based on the tumor microenvironment. *Front. Immunol.* 13, 984480. doi:10.3389/fimmu.2022.984480

Zhou, S. (2022). Sparse SVM for sufficient data reduction. *IEEE Trans. pattern analysis Mach. Intell.* 44 (9), 5560–5571. doi:10.1109/TPAMI.2021.3075339

Zhou, Z., Song, Q., Yang, Y., Wang, L., and Wu, Z. (2022). Comprehensive landscape of RRM2 with immune infiltration in pan-cancer. *Cancers* 14 (12), 2938. doi:10.3390/cancers14122938

Zhuang, W., Sun, H., Zhang, S., Zhou, Y., Weng, W., Wu, B., et al. (2021). An immunogenomic signature for molecular classification in hepatocellular carcinoma. *Mol. Ther. Nucleic acids* 25, 105–115. doi:10.1016/j.omtn.2021.06.024

Frontiers in Pharmacology

Explores the interactions between chemicals and living beings

The most cited journal in its field, which advances access to pharmacological discoveries to prevent and treat human disease.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact