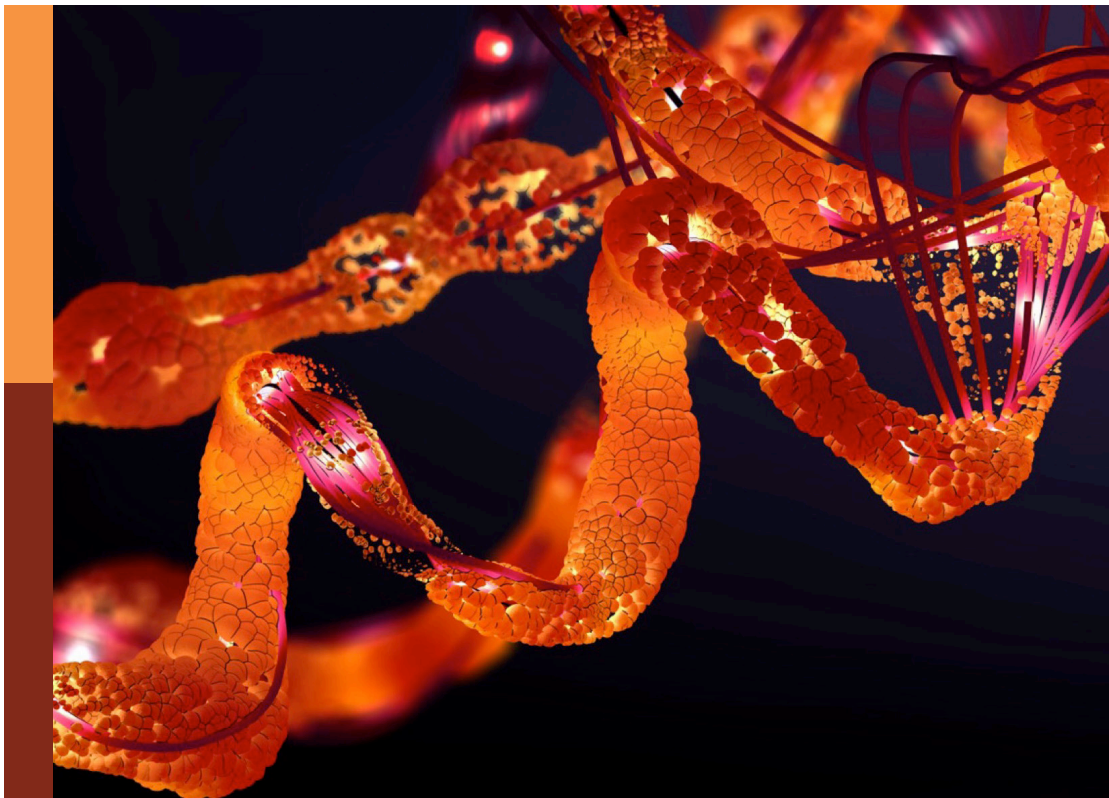
frontiers | Research Topics

Novel biomarkers and big data-based biomedical studies in cancer diagnosis and management

Edited by
Lin Zhang and Qingyu Luo

Published in
Frontiers in Molecular Biosciences



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-6421-9
DOI 10.3389/978-2-8325-6421-9

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

Novel biomarkers and big data-based biomedical studies in cancer diagnosis and management

Topic editors
Lin Zhang — Monash University, Australia
Qingyu Luo — Dana–Farber Cancer Institute, United States

Citation
Zhang, L., Luo, Q., eds. (2025). *Novel biomarkers and big data-based biomedical studies in cancer diagnosis and management*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-6421-9

Table of contents

06 Editorial: Bridging discovery and translation in novel biomarkers and big data-based biomedical studies for cancer management
Lin Zhang, Weiye Qian and Qingyu Luo

09 Identification of a novel CNV at the *APC* gene in a Chinese family with familial adenomatous polyposis
Juyi Li, Chengzhi He, Jing Gong, Xiufang Wang, Chao Liu, Aiping Deng and Lin Zhu

18 Three E2F target-related genes signature for predicting prognosis, immune features, and drug sensitivity in hepatocellular carcinoma
Baozhu Zhang, Boyang Chang, Lu Wang and Yuzhong Xu

32 Clinical analysis of prolonged viral clearance time in patients with lymphoma combined with novel coronavirus infection
Ying Li, Chao Wu, Liming Fei, Qin Xu, Xianru Shao, Bangjie Chen and Gengyun Sun

39 AVEN: a novel oncogenic biomarker with prognostic significance and implications of AVEN-associated immunophenotypes in lung adenocarcinoma
Dengxia Fan, Moses Yang, Hye Jung Lee, Jeong Hee Lee and Hong Sook Kim

53 Role of aggressive locoregional surgery in treatment strategies for ipsilateral supraclavicular lymph node metastasis of breast cancer: a real-world cohort study
Kexin Feng, Zeyu Xing, Qichen Dai, Han Cheng and Xiang Wang

65 Longitudinal dynamic contrast-enhanced MRI radiomic models for early prediction of response to neoadjuvant systemic therapy in triple-negative breast cancer
Bikash Panthi, Rania M. Mohamed, Beatriz E. Adrada, Medine Boge, Rosalind P. Candelaria, Huiqin Chen, Kelly K. Hunt, Lei Huo, Ken-Pin Hwang, Anil Korkut, Deanna L. Lane, Huong C. Le-Petross, Jessica W. T. Leung, Jennifer K. Litton, Sanaz Pashapoor, Frances Perez, Jong Bum Son, Jia Sun, Alastair Thompson, Debu Tripathy, Vicente Valero, Peng Wei, Jason White, Zhan Xu, Wei Yang, Zijian Zhou, Clinton Yam, Gaiane M. Rauch and Jingfei Ma

76 Relationships among breast, gut, and oral microbiota across diverse pathological types of breast cancer, a Chinese cohort study
Kexin Feng, Fei Ren and Xiang Wang

88 Association between oral microbiome and seven types of cancers in East Asian population: a two-sample Mendelian randomization analysis
Kexin Feng, Fei Ren and Xiang Wang

103 Period2 is associated with immune cell infiltration and is a potential diagnostic and prognostic marker for hepatocellular carcinoma
Xiaolin Chen, Qiaosu Zhao, Haibiao Wang, Kaijie Qiu, Xi Deng and Feng Xu

116 A risk signature based on endoplasmic reticulum stress-associated genes predicts prognosis and immunity in pancreatic cancer
Haofei Chen, Ning Xu, Jia Xu, Cheng Zhang, Xin Li, Hao Xu, Weixiong Zhu, Jinze Li, Daoming Liang and Wence Zhou

133 Efficacy and safety of neoadjuvant immunotherapy protocols and cycles for non-small cell lung cancer: a systematic review and meta-analysis
Huaiyong Wang, Song Liang, Yue Yu and Yun Han

147 Elucidating the susceptibility to breast cancer: an in-depth proteomic and transcriptomic investigation into novel potential plasma protein biomarkers
Yang Wang, Kexin Yi, Baoyue Chen, Bailin Zhang and Gao Jidong

160 Gene expression in metastatic breast cancer—patterns in primary tumors and metastatic tissue with prognostic potential
Julia Tutzauer, Anna-Maria Larsson, Kristina Aaltonen, Caroline Bergenfelz, Pär-Ola Bendahl and Lisa Rydén

179 CT-based radiomics for predicting lymph node metastasis in esophageal cancer: a systematic review and meta-analysis
Liangsen Liu, Hai Liao, Yang Zhao, Jiayu Yin, Chen Wang, Lixia Duan, Peihan Xie, Wupeng Wei, Meihai Xu and Danke Su

189 A prospective study on using shear wave elastography to predict the ypT0 stage of rectal cancer after neoadjuvant therapy: a new support for the watch-and-wait approach?
Mengjia Liu, Ningyi Cui, Chao Sun, Xuantong Gong, Bo Wang, Di Yang and Yong Wang

200 *MIR4435-2HG* as a possible novel predictive biomarker of chemotherapy response and death in pediatric B-cell ALL
Yulieth Torres-Llanos, Jovanny Zabaleta, Nataly Cruz-Rodriguez, Sandra Quijano, Paula Carolina Guzmán, Iliana de los Reyes, Nathaly Poveda-Garavito, Ana Infante, Liliana Lopez-Kleine and Alba Lucía Combita

218 Unraveling breast cancer prognosis: a novel model based on coagulation-related genes
Chuqi Lei, Yuan Li, Huaiyu Yang, Ke Zhang, Wei Lu, Nianchang Wang and Lixue Xuan

236 TUBB, a robust biomarker with satisfying abilities in diagnosis, prognosis, and immune regulation via a comprehensive pan-cancer analysis
Zaifu Zhu, Wei Zhang, Shaohu Huo, Tiantuo Huang, Xi Cao and Ying Zhang

256 Comprehensive landscape of junctional genes and their association with overall survival of patients with lung adenocarcinoma
Bin Xie, Ting Wu, Duiguo Hong and Zhe Lu

273 The prognostic value of Dickkopf-3 (Dkk3), TGFB1 and ECM-1 in prostate cancer
Zainab Al Shareef, Mahmood Y. Hachim, Amal Bouzid, Iman M. Talaat, Natheer Al-Rawi, Rifat Hamoudi and Ibrahim Y. Hachim

284 A deep learning method for classification of HNSCC and HPV patients using single-cell transcriptomics
Akanksha Jarwal, Anjali Dhall, Akanksha Arora, Sumeet Patiyal, Aman Srivastava and Gajendra P. S. Raghava

frontiers | Frontiers in Molecular Biosciences

TYPE Editorial
PUBLISHED 22 May 2025
DOI 10.3389/fmolb.2025.1620512

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Matteo Becatti,
University of Firenze, Italy

*CORRESPONDENCE
Lin Zhang,
tony1982110@gmail.com

RECEIVED 29 April 2025
ACCEPTED 12 May 2025
PUBLISHED 22 May 2025

CITATION
Zhang L, Qian W and Luo Q (2025) Editorial: Bridging discovery and translation in novel biomarkers and big data-based biomedical studies for cancer management.
Front. Mol. Biosci. 12:1620512.
doi: 10.3389/fmolb.2025.1620512

COPYRIGHT
© 2025 Zhang, Qian and Luo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Bridging discovery and translation in novel biomarkers and big data-based biomedical studies for cancer management

Lin Zhang[1,2]*, Weiye Qian[1,2] and Qingyu Luo[3,4]

[1]The School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, Australia, [2]Department of Medical Research, Suzhou Industrial Park Monash Research Institute of Science and Technology, Suzhou, China, [3]Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, United States, [4]One Patient One Cure, Boston, MA, United States

KEYWORDS

noval biomarkers, big data, biomedical studies, cancer diagnosis, cancer management

Editorial on the Research Topic
Bridging discovery and translation in novel biomarkers and big data-based biomedical studies for cancer management

Cancer remains one of the most formidable challenges in modern medicine. The discovery of novel biomarkers and the application of big data-based biomedical studies have emerged as powerful tools to enhance our understanding, diagnosis, and management of this complex disease. As a testament to this progress, we present a collection of 21 articles published for the *Novel Biomarkers and Big Data-Based Biomedical Studies in Cancer Diagnosis and Management* research topic in Frontiers in Molecular Biosciences.

This collection of papers highlights the critical role of biomarker discovery and research in advancing cancer treatment. Al Shareef et al. investigated the prognostic value of Dickkopf-3 (Dkk3), TGFB1, and ECM-1 in prostate cancer and found that the decreased expression of DKK3 and the increased expression of TGFB1 were closely related to disease progression and poor prognosis, highlighting their potential as prognostic indicators. Zhu et al. identified TUBB as a robust biomarker through comprehensive pan-cancer analysis. TUBB was consistently overexpressed in multiple tumor types and was significantly associated with prognosis, immune invasion, and chemosensitivity. Meanwhile, Torres-Llanos et al. proposed the MIR4435-2HG as a regulatory biomarker predicting treatment response and survival in pediatric B-cell ALL, with its overexpression being linked to positive minimal residual disease and increased risk of poor outcomes. These findings highlight the prevalence and specificity of molecular biomarkers across different cancer types.

This paper collection also spans a broad range of cancer types, demonstrating the diversity of current biomarker discoveries. Xie et al. discovered novel junctional genes associated with the survival of patients with lung adenocarcinoma (LUAD), establishing a risk model that integrates clinical features for improved prognosis prediction. Fan et al. confirmed that elevated AVEN expression was associated with tumor progression and poor survival rate in LUAD, and they established a robust AVEN-derived prognostic model, which was validated in an external cohort. Chen et al. found that low expression of circadian gene period2 (PER2) in hepatocellular carcinoma (HCC) was associated with immune infiltration and adverse clinical features, highlighting its diagnostic and prognostic potential.

Zhang et al. constructed a three-gene marker based on the E2F, which can predict the prognosis, immune evasion, and drug sensitivity of HCC, suggesting its clinical application in personalized treatment. Chen et al. identified risk markers of endoplasmic reticulum stress-related genes in pancreatic cancer, revealing associations with poor survival rate and changes in the immune microenvironment.

The integration of big data analytics and artificial intelligence into cancer research is another cornerstone of this research topic. Jarwal et al. combined single-cell transcriptomics and deep learning to accurately classify head and neck squamous cell carcinoma (HNSCC) and human papillomavirus (HPV) status, providing a non-invasive diagnostic tool. Lei et al. developed a six-gene prognostic model based on coagulation-related genes, which can stratify breast cancer patients according to survival outcomes and guide treatment strategies, including immunotherapy and chemotherapy. Panthi et al. established radiological models using longitudinal DCE-MRI features to predict the early treatment response of triple-negative breast cancer and promote personalized treatment decisions. Liu et al. conducted a meta-analysis demonstrating that CT-based radiomics provides reliable diagnostic performance in predicting lymph node metastasis of esophageal cancer, highlighting its clinical value in staging.

Several studies emphasize the importance of translational research and clinical applications. Liu et al. explored the application of shear wave elastography in 60 rectal cancer patients to predict the pathological complete response (ypT0) stage after neoadjuvant therapy in rectal cancer, demonstrating that corrected elasticity parameters significantly improved prediction of the ypT0 stage compared with traditional ultrasound, thereby supporting a watch-and-wait strategy. Similarly, Wang et al. conducted a systematic review and meta-analysis of 44 studies (n = 2430) on neoadjuvant immunotherapy protocols in non-small cell lung cancer. Their research results show that compared with immunotherapy alone, chemotherapy combined with immunotherapy significantly increased the rates of major and complete pathological responses, and the three treatment cycles provided a good balance between efficacy and adverse events.

This topic also includes new advances in breast cancer research and treatment through multiomics and clinical analyses. Feng et al. conducted a Mendelian randomization (MR) analysis to study the potential causal relationship between the oral microbiome and the risk of seven major cancers, including breast cancer, and identified microbial genera that have different effects on cancer susceptibility. In a supplementary study, Feng et al. conducted 16S rRNA gene sequencing on oral, intestinal, and breast tissue samples from patients with different pathological subtypes of breast cancer and found that microbial patterns varied across sample types and cancer subtypes. Tutzauer et al. investigated gene expression profiles in primary tumors and distant metastases in metastatic breast cancer patients. Five distinct gene expression subtypes were identified among metastases, indicating that the expression of androgen receptors at the primary and metastatic sites is a powerful prognostic marker for progression-free survival. Wang et al. conducted an comprehensive proteomic and transcriptomic analysis using proteome/transcriptome-wide association studies and MR, identifying several plasma proteins (such as PEX14, CTSF, and SNUPN) with causal links to breast cancer risk, which also showed subtype-specific patterns. In another study, Feng et al. evaluated the actual impact of aggressive locoregional surgery, including infraclavicular and supraclavicular lymph node dissection, in breast cancer patients with ipsilateral supraclavicular lymph node metastasis. They found that more extensive axillary dissection improved overall and disease-free survival, although supraclavicular dissection alone did not add further benefit. Finally, Li et al. reported a novel copy number variant in the APC gene in a family with familial adenomatous polyposis, while another study by Li et al. examined lymphoma patients with COVID-19, highlighting the delayed viral clearance and its potential association with compromised immune function in malignancy.

The articles collected in this research feature represent a comprehensive and impactful effort to advance cancer diagnosis, prognosis, and treatment through biomarker discovery and big data-driven approaches. By integrating molecular analysis, artificial intelligence, imaging technologies, and translational research, these works highlight how modern research shapes personalised cancer treatment. From the identification of prognostic genes and immune-related markers to the application of deep learning and radiomics, each study has provided new insights with potential clinical implications. These include rigorous clinical evaluations, such as shear wave elastography in rectal cancer and meta-analyses of immunotherapy cycles in lung cancer, highlighting the translational focus. In conclusion, these contributions not only deepen our understanding of cancer biology but also lay the foundation for more precise, individualised, and effective treatment strategies for various tumour types.

Author contributions

LZ: Conceptualization, Writing – review and editing, Supervision, Writing – original draft. WQ: Conceptualization, Writing – review and editing, Writing – original draft. QL: Supervision, Conceptualization, Writing – review and editing, Resources.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

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

The author(s) 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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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.

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 27 July 2023
DOI 10.3389/fmolb.2023.1234296

Check for updates

OPEN ACCESS

EDITED BY
Lin Zhang,
Hefei Comprehensive National Science Center, China

REVIEWED BY
Farah Ballout,
University of Miami Health System, United States
Marco Bono,
University of Palermo, Italy

*CORRESPONDENCE
Lin Zhu,
zhulinhust@hust.edu.cn
Aiping Deng,
dapyxb@163.com

†These authors have contributed equally to this work

RECEIVED 04 June 2023
ACCEPTED 10 July 2023
PUBLISHED 27 July 2023

CITATION
Li J, He C, Gong J, Wang X, Liu C, Deng A and Zhu L (2023), Identification of a novel CNV at the *APC* gene in a Chinese family with familial adenomatous polyposis. *Front. Mol. Biosci.* 10:1234296. doi: 10.3389/fmolb.2023.1234296

COPYRIGHT
© 2023 Li, He, Gong, Wang, Liu, Deng and Zhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identification of a novel CNV at the *APC* gene in a Chinese family with familial adenomatous polyposis

Juyi Li[1†], Chengzhi He[2†], Jing Gong[3†], Xiufang Wang[4†], Chao Liu[5], Aiping Deng[1*] and Lin Zhu[6*]

[1]Department of Pharmacy, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China, [2]Department of Gastrointestinal Surgery, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [3]Department of Gastroenterology, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China, [4]Department of Pain, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China, [5]Hubei Key Laboratory of Diabetes and Angiopathy, Hubei University of Science and Technology, Xianning, Hubei, China, [6]Department of Pediatrics, Tongji Hospital, Huazhong University of Science and Technology, Wuhan, Hubei, China

Introduction: Familial adenomatous polyposis (FAP) is the second most commonly inherited colorectal cancer (CRC) predisposition caused by germline mutations within the *adenomatous polyposis coli* (*APC*) gene. The molecular defects and clinical manifestations of two FAP families were analyzed, and individual prevention strategies suitable for mutation carriers in different families were proposed.

Methods and results: The pathogenic gene mutations were identified among the two families using whole-exome sequencing and verified with Sanger sequencing or quantitative polymerase chain reaction (qPCR). One novel (GRCh37:Chr5: 112145676–112174368, del, 28,692 bp) and a known (c.C847T:p.R283X) mutation in the *APC* gene were pathogenic mutations for FAP, according to the sequencing data and tumorigenesis pattern among the family members. The two mutations led to a premature translational stop signal, synthesizing an absent or disrupted protein product.

Conclusion: Our findings expand the known germline mutation spectrum of the *APC* gene among the Chinese population. This reaffirms the importance of genetic testing in FAP. Genetic consultation and regular follow-ups are necessary for the individualized treatment of cancer-afflicted families with APC expression deficiency. Additional work is required to develop safe and effective chemotherapy and immunotherapy for FAP based on the mutation type.

KEYWORDS

familial adenomatous polyposis, adenomatous polyposis coli gene, whole-exome sequencing, copy number variations, genetic counseling

1 Introduction

Familial adenomatous polyposis (FAP) is a hereditary colorectal disease subtype with a poor prognosis. Colorectal cancer (CRC) predisposition syndrome is rare, characterized by 100 s–1,000 s of adenomas developing in the colon and rectum with their onset in childhood and adolescence. Moreover, CRC possesses associated extracolonic manifestations, such as desmoid tumors, dental and skin abnormalities, retinal spots, and malignant tumors from

other organs (Carr and Kasi, 2022). Surgery is an effective therapy for FAP patients with colonic disorders, and regular chemotherapy has been shown to benefit FAP patients with the postoperative pathological diagnosis of adenocarcinoma (Tougeron et al., 2020). Recent colorectal cancer guidelines indicate that the type of pathogenic gene mutation is related to the patient's prognosis and responses to chemotherapy and immunotherapy (Stjepanovic et al., 2019).

FAP follows an autosomal dominant inheritance pattern caused by the monoallelic mutation in the *adenomatous polyposis coli* (*APC*) gene (Recio-Boiles and Cagir, 2022). APC is a tumor suppressor gene located on chromosome 5q21, encoding a large scaffolding protein with functions in cell cycle regulation, apoptosis, transcription, and cell migration (Kim and Bodmer, 2021). *APC* comprises 16 exons; the last exon encodes nearly 70% of the APC protein. FAP frequency is approximately 1:8,000 in the general population with almost complete penetrance, affecting multiple family generations (Bisgaard et al., 1994).

Currently, more than 2,000 pathogenic variations in the *APC* gene are found, the majority of which are observed in the 5′-ends of exon 15, also called the mutation cluster region (MCR) (Schirosi et al., 2013; Jung et al., 2016), and are situated

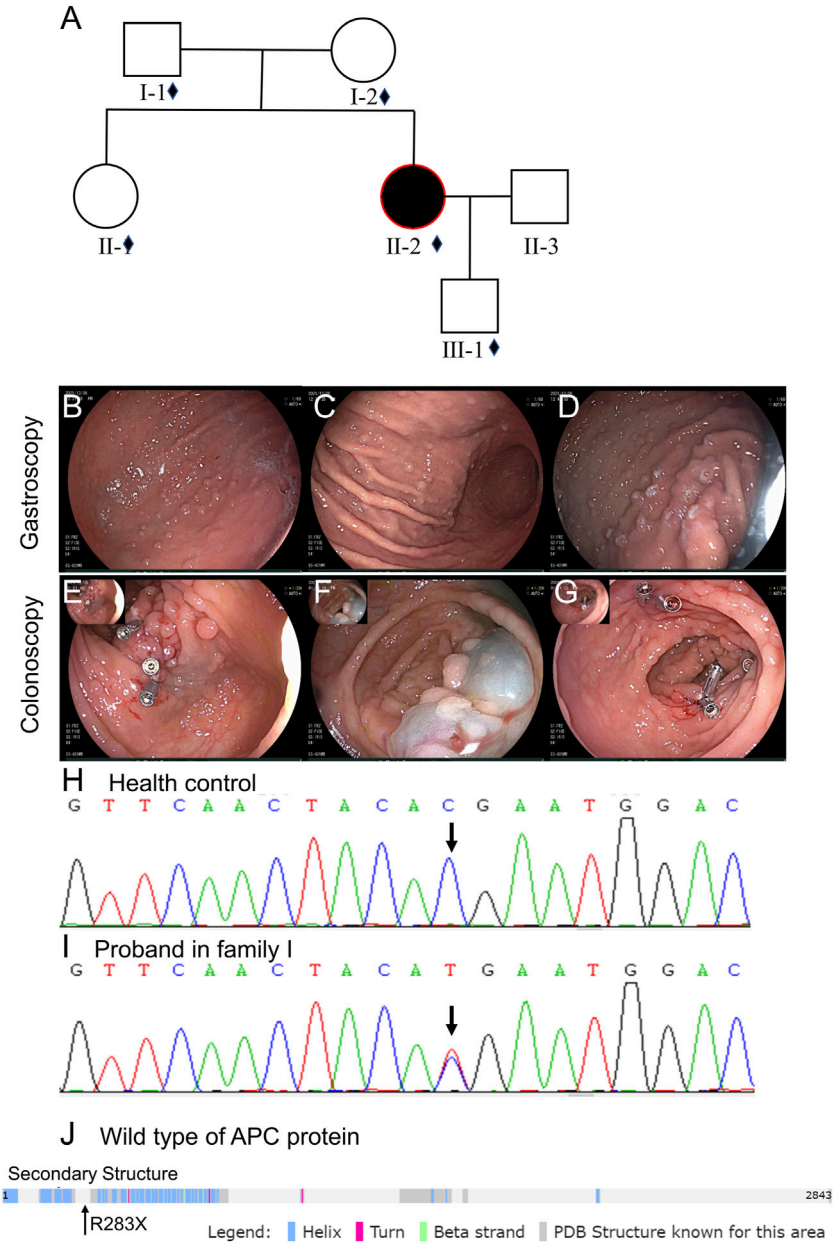


FIGURE 1
Clinical data of family I. **(A)** Pedigree structure of family I: the red ring depicts the proband, squares indicate male members, and circles represent female members. Black shading represents FAP individuals. The small diamonds denote the relatives whose DNA was available for testing. **(B–D)** Gastroscopy of the proband. **(E–G)** Colonoscopy of the proband. **(H,I)** Sanger sequencing analysis of the proband's *APC* gene (c.C847T:p.R283X): the C base at position 847 of the *APC* gene is replaced by the T base. **(J)** Location of the p.R283X mutation in the secondary structures of the APC protein.

10.3389/fmolb.2023.1234296

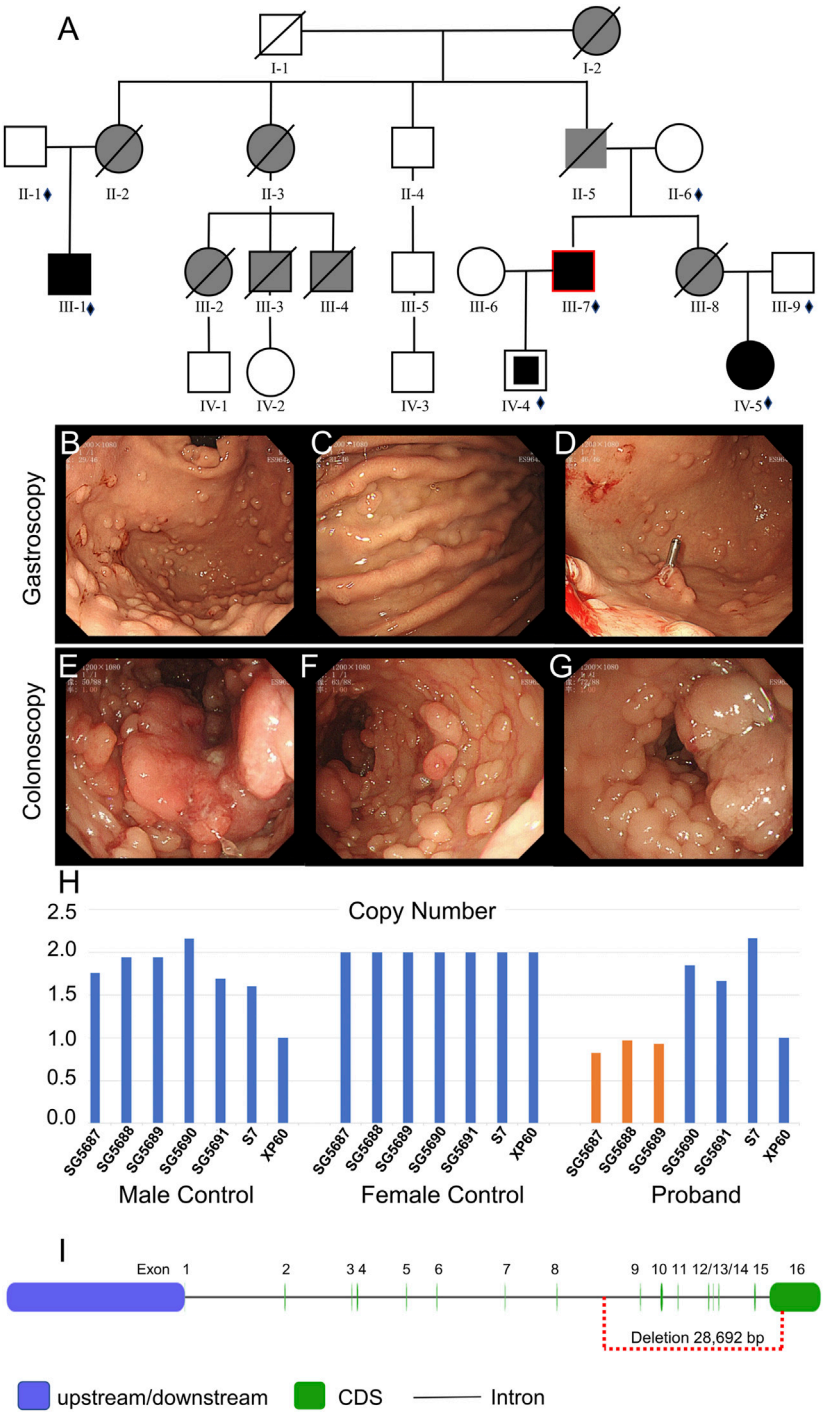


FIGURE 2
Clinical data of family II. **(A)** Pedigree structure of family II: the red ring indicates the proband, squares indicate male members, circles indicate female members, and crosses indicate deceased individuals. Dark shading represents individuals with FAP, partly dark shading represents mutation carriers, and gray shading represents CRC without the gene test. The small diamonds denote the relatives whose DNA was available for testing. **(B–D)** Gastroscopy of the proband. **(E–G)** Colonoscopy of the proband. **(H)** qPCR of the proband's peripheral blood DNA to verify the CNV in *APC* (GRCh37:Chr5: 112145676–112174368, del, 28,692 bp). **(I)** Location of the GRCh37:Chr5: 112145676–112174368 delete mutation in the *APC* gene.

between 1286 and 1513 codons (Cetta and Dhamo, 2007). Codons 1309 and 1061 are hotspots for mutation, accounting for nearly 17% and 11% of all germline *APC* mutations, respectively (Leoz et al., 2015). Compared to mutation patients outside the MCR, those with MCR mutations typically have a worse prognosis, manifesting their condition early (Ghadamyari et al., 2021). In a family, identifying the specific *APC* mutation can help targeted sequencing testing for

presymptomatic at-risk family members (Cruz-Correa et al., 2017).

Identifying patients and families at an extremely high risk of developing cancer can help reduce cancer occurrence and mortality. Several studies described an association between *APC* mutation localization and the phenotype among FAP patients (Heinen, 2010). The diagnosis and patient follow-up could be improved by connecting the genotypes to the phenotypes. However, there is a paucity of information on the genotypic spectrum and clinical characterization of FAP in China. One novel and one known *APC* mutation were reported in two Chinese families with FAP. Our study aimed to analyze the molecular defects and clinical manifestations in the two families, for appropriate personalized prevention strategies against all mutation carriers.

2 Results

2.1 Clinical characteristics

The detailed pedigree of family I is given in Figure 1A. The proband was a 34-year-old female, whose first symptom of altered bowel movement appeared at 31 years (July 2019). She underwent her first gastrointestinal endoscopy and treatment that year. This involved the endoscopic dissection of colonic polyps and gastric polypectomy. Biopsy indicated multiple tubular adenomas inside the ascending colon and rectum. The patient underwent three more gastrointestinal endoscopy examinations and treatment in the following 2 years. The last endoscopy examination (6 December 2021, Figure 1) showed multiple glandular polyps inside the gastric body and fundus, with tubular adenoma and adenomatous polyp inside the ascending colon and tubular adenoma in the transverse colon with mild-to-moderate heteroplasia of the focal glandular epithelium. However, no extracolonic manifestations appeared, such as desmoid tumors or dental and skin abnormalities. The possibility of classic FAP was highly suspected with her clinical manifestations, so whole-exome sequencing was conducted.

Family II is a typical and interesting large family. The detailed pedigree is given in Figure 2A, and the clinical characteristics of the family members are listed in Supplementary Table S2. The proband of family II sought medical attention for bloody stools at 41 years (July 2019). Gastrointestinal endoscopy showed multiple new organisms in the stomach and colon (Figure 2), and biopsy suggested adenocarcinoma of the transverse colon. The glandular epithelium of the mucosa showed moderate-to-severe dysplasia of approximately 40 cm from the anal margin. A fragmentary villous or serrated glandular epithelium was observed with low-grade intraepithelial neoplasia of approximately 10 cm from the anal margin. In August 2019, laparoscopic total colectomy and ileostomy were conducted. The pathological examination suggested the following: 1. multiple neoplasm and ulcerated areas were carcinomatous, with medium well-differentiated adenocarcinoma. The deepest cancer infiltration reached the outer membrane (the largest mass was 4 × 3.5 cm), with tumor plugs in the vascular area. 2. There were no cancer cells in the colon tissues of the two broken ends, and the local glandular epithelium of the broken ends indicated low-grade intraepithelial neoplasia. 3. Metastatic cancer was found in 3/14 of the peri-intestinal lymph nodes. Oxaliplatin chemotherapy was performed several times after the operation. Over the next year or so, multiple gastroscopies and colonoscopies were performed.

The father (II-5), aunts (II-2 and II-3), sister (III-8), and cousins (III-2 and III-3) of the proband were all diagnosed with colon or rectal cancer and were deceased. The brother of the proband (III-1) was hospitalized at 32 years (December 2020) due to irregular stools, diarrhea, and stool blood. Five gastrointestinal endoscopy examinations and treatments were performed, among which the gastrointestinal endoscopy examination performed in December 2020 is depicted in Figure 3. Gastroscopy (Figures 3A–C) indicated duodenal bulb ulcer (stage A2) and erosive gastritis (grade 2). Biopsy suggested severe chronic inflammation of the gastric antrum mucosa accompanied by erosive gastritis with mild activity. Colonoscopy (Figure 3D–F) revealed multiple neoplasms of the large intestine. EMR and argon coagulation were performed. Biopsy of the ascending colon revealed adenomatous polyps with multiple adenomatous polyps at the rectum and sigmoid colon junction. Rectal eminence was observed under a rectoscope (Figures 3G–I), and a biopsy revealed low-grade intraepithelial neoplasia of superficial mucosal glands with polyps inside the anal canal.

The 22-year-old niece (IV-5) of the proband underwent gastrointestinal polyp extraction at 17 years. After surgery, gastrointestinal endoscopy was performed several times. Gastroscopy indicated multiple gastric polyps treated with argon coagulation and erosive gastritis (grade 1) (Figures 4A–C). Biopsy depicted mild-to-moderate chronic inflammation of the gastric antrum mucosa showing mild activity. A colonoscopy (March 2021) showed multiple large intestine polyps, and EMR and argon coagulation were conducted (Figures 4D–F). Biopsy revealed adenomatous polyps inside the transverse colon. The 6-year-old son (IV-4) of the proband is currently asymptomatic; he is too young to develop clinical symptoms. Whole-exome sequencing was performed for proband II and quantitative polymerase chain reaction (qPCR) for the other family members associated with the disease characteristics of family II members.

2.2 Genetic test results

The whole-exome sequencing data of the proband in family I are described in Supplementary Table S3A. Approximately 117,621 mutations were identified with 106,580 SNPs and 11,041 indels. The number of synonymous and missense mutations, new SNPs, and indels was 11,015, 9,886, 628, and 676, respectively. We identified a known mutation NM_000038: exon9:c.C847T:p.R283X of *APC* (rs786201856; accession: VCV000836957.3) (Hong et al., 2021a; Hong et al., 2021b). In this mutation, the C base at position 847 of the *APC* gene was replaced by the T base (Figure 1H/I). This led to the premature emergence of a stop codon, premature APC protein translation termination, and protein function loss (Figure 1J), showing the specific clinical significance of this mutation.

Sanger sequencing helped detect the APC mutation (NM_000038:exon9:c.C847T:p.R283X) in the proband and family

10.3389/fmolb.2023.1234296

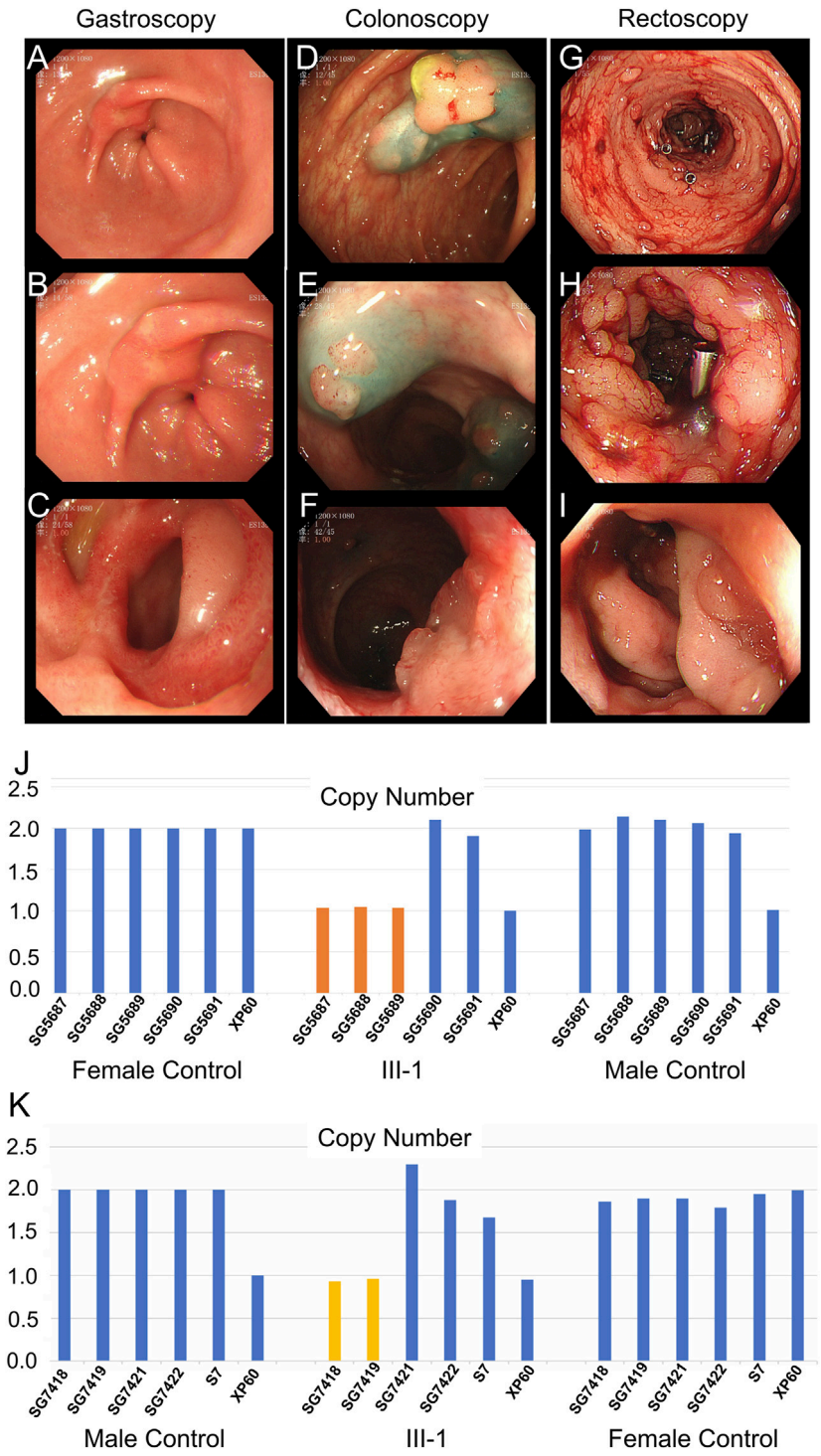


FIGURE 3
Clinical data of III-1 (the proband's brother) of family II. **(A–C)** Gastroscopy of III-1. **(D–F)** Colonoscopy of III-1. **(G–I)** Rectoscopy of III-1. **(J)** qPCR of III-1's peripheral blood DNA to verify the CNV in *APC* (GRCh37:Chr5: 112145676–112174368, del, 28,692 bp). **(K)** qPCR of III-1's DNA extracted from newly developed polyps to verify the CNV in *APC* (GRCh37:Chr5: 112145676–112174368, del, 28,692 bp).

members of the patient to provide genetic counseling. The germline mutation in *APC* was not observed in other family members (I-1, I-2, II-1, and III-1). Thus, the APC mutation of the proband is spontaneous and not passed on to her son.

Whole-exome sequencing was also performed on the proband in family II (Supplementary Table S3B). Approximately 114,508 mutations with 104,007 SNPs, 10,485 indels, and 16 copy number variations (CNVs) were identified. The number

10.3389/fmolb.2023.1234296

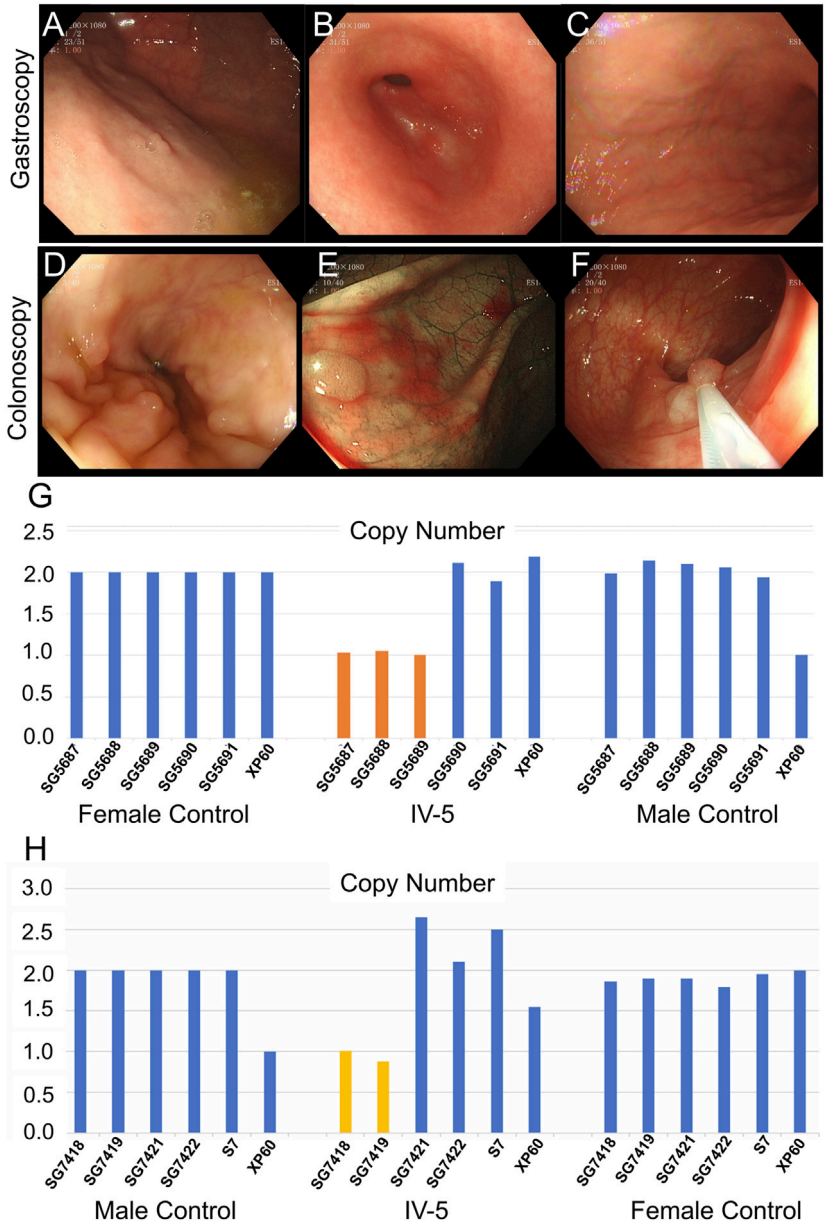


FIGURE 4
Clinical data of IV-5 (the proband's niece) in family II. **(A–C)** Gastroscopy of IV-5. **(D–F)** Colonoscopy of IV-5. **(G)** qPCR of IV-5's peripheral blood DNA to verify the CNV in *APC* (GRCh37:Chr5: 112145676–112174368, del, 28,692 bp). **(H)** qPCR of IV-5's DNA extracted from newly developed polyps to verify the CNV in *APC* (GRCh37:Chr5: 112145676–112174368, del, 28,692 bp).

of new SNPs and indels was 656 and 656, respectively. We verified the proband harboring a new CNV loss within the *APC* tumor suppressor gene (GRCh37:Chr5: 112145676–112174368, del, 28,692 bp) (Figure 2I) using qPCR (Figure 2H). The brother (III-1), niece (IV-5), and son (IV-4) of the proband possessed the same mutation (Figure 3J; Figure 4G; Supplementary Figure S1A). The mutation was detected in the first two polypous tissues (Figures 3K, 4H). In contrast, this mutation was not observed in the other family members (Supplementary Figures S1B–D). Therefore, the substantial deletion of the *APC* gene in this family is the causative mutation for polyposis. FAP is a dominant syndrome caused by truncating mutations or large deletions.

3 Discussion

We identified one novel mutation (GRCh37:Chr5: 112145676–112174368, del) and a known mutation (NM_000038:exon9: c.C847T:p.R283X, rs786201856) within the *APC* gene in two typical FAP families, which were pathogenic for FAP.

APC is a tumor suppressor gene, promoting the rapid degradation of CTNNB1 (cadherin-associated protein, beta 1). *APC* participates in Wnt signaling as a negative regulator. Currently, more than 10,000 *APC* mutations have been identified. There are several important observations regarding *APC* mutations: 1) the vast majority of mutations discovered would lead to *APC*

product truncation; 2) most mutations occurred in the first half of the coding sequence, and somatic mutations in colorectal tumors were concentrated in the MCR; 3) transition of cytosine to other nucleotides was the most common point mutations in the *APC* gene; and 4) there was a correlation between the germline mutations of FAP1 patients and colorectal polyps. For most adenomas and carcinomas in the colon and rectum and some in the stomach, inactivating both alleles of the *APC* gene is necessary as an early event.

During the diagnosis of colorectal cancer in the 34-year-old female patient, our patient from family I had more than 100 adenomatous colorectal polyps with gastric polyps. However, there were no additional extracolonic signs linked to FAP (desmoids, osteomas, cutaneous soft-tissue tumors, dental abnormalities, and CHRPE). The *APC* mutation is denoted by *APC* c.847C>T in the proband at the cDNA level and p.Arg283Ter (R283X) at the protein level. It has been reported in sporadic and syndromic adenocarcinoma patients within the COSMIC database (Tate et al., 2019) and is considered pathogenic by ClinVar (Landrum et al., 2018). Additionally, this mutation has been clarified in FAP patients as having adamantinomatous craniopharyngiomas.

The c.847C>T pathogenic mutation, also called p.R283X, affects the coding exon 8 of the *APC* gene and is caused by a C-to-T substitution at the nucleotide position 847. This mutation converts the amino acid within coding exon 8 from an arginine to a stop codon. This harmful variation exists in many families across several ethnic groups with familial adenomatous polyposis (Owen et al., 2018; de Oliveira et al., 2019). Other than the clinical information in the literature, this change is anticipated to cause function loss by nonsense-mediated mRNA decay (NMD) or premature protein truncation. This variant has been reported in 15 probands, meeting eight phenotype points, and in 50 probands with FAP not otherwise specified, meeting more than 16 phenotype points in total. It has been reported to segregate with FAP in five meioses from one family and in 31 members from one large FAP family (Mohamed et al., 2003). In summary, this variant meets the criteria to be classified as pathogenic for FAP based on the ACMG/AMP criteria applied. The mutation of the proband was not passed on to her son, and she is expecting her second child. However, the second child probably carried the disease-causing mutation. Genetic counseling was recommended according to her genetic mutation.

Family II is a typical FAP family. The genomic region of the *APC* gene is deleted out-of-frame in family II (GRCh37:Chr5: 112145676–112174368, del, 28,692 bp). The consequence could be a missing or disrupted protein product with a premature translational stop signal. This variation has not been documented among people with *APC*-related diseases. Pathogenic loss-of-function mutations in *APC* exist (Lagarde et al., 2010), which are labeled pathogenic based on these factors. The elder brother (III-1) and niece (IV-5) of the proband were mutant gene carriers; both were of marriageable age, but neither was married. They were concerned that their future children would carry the same gene mutation. Therefore, according to the principle of eugenics, genetic counseling is necessary, for the third generation of test-tube babies before pregnancy.

Nearly all FAP patients develop CRC if not identified and treated early (Half et al., 2009). Adjuvant chemotherapy and immunotherapy are required in most instances of FAP after total colectomy, similar to sporadic colorectal cancer and Lynch syndrome (Zhang et al., 2019; Koskenvuo et al., 2020; Kemp Bohan et al., 2021; Rohani et al., 2022). Cancer research significantly emphasizes the therapeutic targeting of abnormal beta-catenin activity. Creating medicines that target cancer cells, while having acceptable safety profiles, has remained a pharmacological challenge since Wnt signaling is a highly conserved system in normal cellular physiology. Early clinical trials of several medicines that inhibit upstream effectors in the Wnt signaling pathway have been undertaken, albeit with worries about off-target effects. In the phase 1 clinical trial of paclitaxel and vantictumab, a monoclonal antibody against Fzd receptors, was used to suppress Wnt signaling by preventing binding with all Wnt ligands. The trial indicated moderate efficacy among Wnt-upregulated metastatic breast (Diamond et al., 2020) and pancreatic (Davis et al., 2020) cancer. However, concerns have been raised about the safety of the drug concerning the bones. Furthermore, porcupine inhibitors, an enzyme that processes Wnt signaling proteins, have been created to combat Wnt-driven cancers. Most notably, WNT974 has demonstrated safety but limited efficacy in advanced solid tumors (Rodon et al., 2021). However, the effectiveness of focusing on upstream targets is still debatable due to tumor resistance with more downstream alterations, including CTNNB1- or *APC*-mutated malignancies. There are serious concerns about off-target consequences, such as damage to normal intestinal tissues and directly inhibiting beta-catenin or *APC* (Kahn, 2014; Zhang et al., 2016; Lai and Kahn, 2021; Wang et al., 2021). Hence, much research on cancer is required to create safe and efficient Wnt signaling inhibitors.

Our study contributes to identifying FAP genotypes in China, with significant implications for genetic counseling, diagnosis, cancer prevention, and treatment. Whole-exome sequencing is a rapid, accurate, and reliable technique to identify genetic variants in suspected FAP patients. It has various potential applications in the genetic testing of FAP-related tumors. Thus, abnormal Wnt signaling, including *APC* mutations, can be a promising target for the development of chemotherapy and immunotherapy against FAP.

4 Materials and methods

4.1 Patients

The Ethics Committee of the Central Hospital of Wuhan approved the study, with the ethical approval code of 2020-192. All subjects included in this analysis were informed in person, and their written informed consent was obtained. The two probands, diagnosed with FAP, were recruited from the Department of Gastrointestinal Surgery at the Central Hospital of Wuhan. The clinical diagnosis of FAP was confirmed by a gastroenterologist using multiple gastroscopic and colonoscopic biopsy reports, clinical investigations, and a detailed family pedigree.

4.2 Mutation analysis

We extracted the genomic DNA from the peripheral blood or newly developed polyps of each proband and part of the family

members. DNA fragments were sequenced using a high-throughput sequencer (Illumina HiSeq 2500 Analyzer; Illumina, CA, United States). As previously demonstrated, single-nucleotide variant, insertion, and deletion queries were performed (Li et al., 2017). The reference genome for whole-exome sequencing was UCSC hg19, NCBI build 37. PCR amplification and Sanger sequencing confirmed the detected variants, which helped in segregation analysis.

Samples were detected using next-generation sequencing for suspected CNVs. The quantitative polymerase chain reaction system using Roche LightCycler® II 480 Probes Master mix, as per the manufacturer's instructions, was applied to verify the result accuracy and determine the breakpoint positions. Primer sequences used are listed in Supplementary Table S1.

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 Wuhan Central Hospital Ethics Committee. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

AD and LZ conceived and designed the experiments. JL, CH, JG, and XW performed the experiments. JL, CH, JG, XW, and CL analyzed the data. JL and LZ wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Wuhan Municipal Health Commission (No. WX18M02), Central Guiding Local Science and Technology Development Special Project (No. 2022BGE272), and Guangdong Yiyang Healthcare Charity Foundation (JZ2022011).

Acknowledgments

The authors thank the patients, their families, referral physicians, and investigators for their cooperation and contribution.

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/fmolb.2023.1234296/full#supplementary-material

References

Bisgaard, M. L., Fenger, K., Bülow, S., Niebuhr, E., and Mohr, J. (1994). Familial adenomatous polyposis (FAP): Frequency, penetrance, and mutation rate. *Hum. Mutat.* 3 (2), 121–125. doi:10.1002/humu.1380030206

Carr, S., and Kasi, A. (2022). "Familial adenomatous polyposis," in *StatPearls* (Treasure Island (FL: StatPearls Publishing LLC).*StatPearls publishing copyright © 2022.*

Cetta, F., and Dhamo, A. (2007). Inherited multitumoral syndromes including colorectal carcinoma. *Surg. Oncol.* 16 (Suppl. 1), S17–S23. doi:10.1016/j.suronc.2007.10.013

Cruz-Correa, M., Pérez-Mayoral, J., Dutil, J., Echenique, M., Mosquera, R., Rivera-Román, K., et al. (2017). Hereditary cancer syndromes in latino populations: Genetic characterization and surveillance guidelines. *Hered. Cancer Clin. Pract.* 15, 3. doi:10.1186/s13053-017-0063-z

Davis, S. L., Cardin, D. B., Shahda, S., Lenz, H. J., Dotan, E., O'Neil, B. H., et al. (2020). A phase 1b dose escalation study of Wnt pathway inhibitor vantictumab in combination with nab-paclitaxel and gemcitabine in patients with previously untreated metastatic pancreatic cancer. *Invest. New Drugs* 38 (3), 821–830. doi:10.1007/s10637-019-00824-1

de Oliveira, J. C., Viana, D. V., Zanardo, C., Santos, E. M. M., de Paula, A. E., Palmero, E. I., et al. (2019). Genotype-phenotype correlation in 99 familial adenomatous polyposis patients: A prospective prevention protocol. *Cancer Med.* 8 (5), 2114–2122. doi:10.1002/cam4.2098

Diamond, J. R., Becerra, C., Richards, D., Mita, A., Osborne, C., O'Shaughnessy, J., et al. (2020). Phase Ib clinical trial of the anti-frizzled antibody vantictumab (OMP-18R5) plus paclitaxel in patients with locally advanced or metastatic HER2-negative breast cancer. *Breast Cancer Res. Treat.* 184 (1), 53–62. doi:10.1007/s10549-020-05817-w

Ghadamyari, F., Heidari, M. M., Zeinali, S., Khatami, M., Merat, S., Bagherian, H., et al. (2021). Mutational screening through comprehensive bioinformatics analysis to detect novel germline mutations in the APC gene in patients with familial adenomatous polyposis (FAP). *J. Clin. Lab. Anal.* 35 (5), e23768. doi:10.1002/jcla.23768

Half, E., Bercovich, D., and Rozen, P. (2009). Familial adenomatous polyposis. *Orphanet J. Rare Dis.* 4, 22. doi:10.1186/1750-1172-4-22

Heinen, C. D. (2010). Genotype to phenotype: Analyzing the effects of inherited mutations in colorectal cancer families. *Mutat. Res.* 693 (1-2), 32–45. doi:10.1016/j.mrfmmm.2009.09.004

Hong, C. S., Elsamadicy, A. A., Fisayo, A., Inzucchi, S. E., Gopal, P. P., Vining, E. M., et al. (2021b). Comprehensive genomic characterization of A case of granular cell tumor of the posterior pituitary gland: A case report. *Front. Endocrinol. (Lausanne)* 12, 762095. doi:10.3389/fendo.2021.762095

Hong, C. S., Omuro, A., An, Y., Inzucchi, S. E., Kohli, A. A., McGuone, D., et al. (2021a). Sporadic adamantinomatous craniopharyngioma with double-hit somatic APC mutations. *Neurooncol Adv.* 3 (1), vdab124. doi:10.1093/noajnl/vdab124

Jung, S. M., Yoon, Y. S., Lim, S. B., Yu, C. S., and Kim, J. C. (2016). Clinicopathological features of familial adenomatous polyposis in Korean patients. *World J. Gastroenterol.* 22 (17), 4380–4388. doi:10.3748/wjg.v22.i17.4380

Kahn, M. (2014). Can we safely target the WNT pathway? *Nat. Rev. Drug Discov.* 13 (7), 513–532. doi:10.1038/nrd4233

Kemp Bohan, P. M., Mankaney, G., Vreeland, T. J., Chick, R. C., Hale, D. F., Cindass, J. L., et al. (2021). Chemoprevention in familial adenomatous polyposis: Past, present and future. *Fam. Cancer* 20 (1), 23–33. doi:10.1007/s10689-020-00189-y

Kim, J. C., and Bodmer, W. F. (2021). Genotypic and phenotypic characteristics of hereditary colorectal cancer. *Ann. Coloproctol.* 37 (6), 368–381. doi:10.3393/ac.2021.00878.0125

Koskenvuo, L., Ryynänen, H., and Lepistö, A. (2020). Timing of prophylactic colectomy in familial adenomatous polyposis. *Colorectal Dis.* 22 (11), 1553–1559. doi:10.1111/codi.15151

Lagarde, A., Rouleau, E., Ferrari, A., Noguchi, T., Qiu, J., Briaux, A., et al. (2010). Germline APC mutation spectrum derived from 863 genomic variations identified through a 15-year medical genetics service to French patients with FAP. *J. Med. Genet.* 47 (10), 721–722. doi:10.1136/jmg.2010.078964

Lai, K. K. Y., and Kahn, M. (2021). Pharmacologically targeting the WNT/β-Catenin signaling cascade: Avoiding the sword of damocles. *Handb. Exp. Pharmacol.* 269, 383–422. doi:10.1007/164_2021_523

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., et al. (2018). ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 46 (D1), D1062–d1067. doi:10.1093/nar/gkx1153

Leoz, M. L., Carballal, S., Moreira, L., Ocaña, T., and Balaguer, F. (2015). The genetic basis of familial adenomatous polyposis and its implications for clinical practice and risk management. *Appl. Clin. Genet.* 8, 95–107. doi:10.2147/TACG.S51484

Li, C. G., Jin, P., Yang, L., Zang, W. C., Kang, Q., et al. (2017). Germline mutations in patients with multiple colorectal polyps in China. *J. Gastroenterol. Hepatol.* 32 (10), 1723–1729. doi:10.1111/jgh.13776

Mohamed, Z., Ahmad, R., Yoke, N. S., Zakaria, Z., Ahmad, H., and Yew, T. H. (2003). A nonsense mutation in exon 8 of the APC gene (Arg283Ter) causes clinically variable FAP in a Malaysian Chinese family. *Cancer Sci.* 94 (8), 725–728. doi:10.1111/j.1349-7006.2003.tb01509.x

Owen, D. R., Wong, H. L., Bonakdar, M., Jones, M., Hughes, C. S., Morin, G. B., et al. (2018). Molecular characterization of ERBB2-amplified colorectal cancer identifies potential mechanisms of resistance to targeted therapies: A report of two instructive cases. *Cold Spring Harb. Mol. Case Stud.* 4 (2), a002535. doi:10.1101/mcs.a002535

Recio-Boiles, A., and Cagir, B. (2022). "Colon cancer," in *StatPearls* (Treasure Island (FL: StatPearls Publishing LLC). *StatPearls publishing copyright © 2022.*

Rodon, J., Argilés, G., Connolly, R. M., Vaishampayan, U., de Jonge, M., Garralda, E., et al. (2021). Phase 1 study of single-agent WNT974, a first-in-class Porcupine inhibitor, in patients with advanced solid tumours. *Br. J. Cancer* 125 (1), 28–37. doi:10.1038/s41416-021-01389-8

Rohani, S., Suhada, N., Mohammad, Z., and Azhar, S. (2022). Traditional herbal medicine as adjunctive therapy for colorectal cancer: A scoping review. *Traditional Med. Res.* 7, 15–18. doi:10.53388/tmr20220127260

Schirosi, L., Pellegrino, M., Tarantino, P., Mauro, S., Tinelli, A., and Greco, M. (2013). A new germline stop codon mutation in exon 15 of the APC gene predisposing to familial adenomatous polyposis. *Int. J. Biol. Markers* 28 (4), e405–e408. doi:10.5301/jbm.5000042

Stjepanovic, N., Moreira, L., Carneiro, F., Balaguer, F., Cervantes, A., Balmaña, J., et al. (2019). Hereditary gastrointestinal cancers: ESMO clinical practice guidelines for diagnosis, treatment and follow-up†. *Ann. Oncol.* 30 (10), 1558–1571. doi:10.1093/annonc/mdz233

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., et al. (2019). Cosmic: The catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 47 (D1), D941–d947. doi:10.1093/nar/gky1015

Tougeron, D., Sueur, B., Zaanan, A., de la Fouchardiére, C., Sefrioui, D., Lecomte, T., et al. (2020). Prognosis and chemosensitivity of deficient MMR phenotype in patients with metastatic colorectal cancer: An AGEO retrospective multicenter study. *Int. J. Cancer* 147 (1), 285–296. doi:10.1002/ijc.32879

Wang, Z., Li, Z., and Ji, H. (2021). Direct targeting of β-catenin in the Wnt signaling pathway: Current progress and perspectives. *Med. Res. Rev.* 41 (4), 2109–2129. doi:10.3390/foods10092109

Zhang, L., Jiang, B., Zhu, N., Tao, M., Jun, Y., Chen, X., et al. (2019). Mitotic checkpoint kinase Mps1/TTK predicts prognosis of colon cancer patients and regulates tumor proliferation and differentiation via PKCα/ERK1/2 and PI3K/Akt pathway. *Med. Oncol.* 37 (1), 5. doi:10.1007/s12032-019-1320-y

Zhang, L., Theodoropoulos, P. C., Eskiocak, U., Wang, W., Moon, Y. A., Posner, B., et al. (2016). Selective targeting of mutant adenomatous polyposis coli (APC) in colorectal cancer. *Sci. Transl. Med.* 8 (361), 361ra140. doi:10.1126/scitranslmed.aaf8127

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 03 October 2023
DOI 10.3389/fmolb.2023.1266515

Check for updates

OPEN ACCESS

EDITED BY
Qingyu Luo,
Dana–Farber Cancer Institute,
United States

REVIEWED BY
Hong Yu,
The University of Texas Health Science Center at San Antonio, United States
Kezhi Yan,
Cystic Fibrosis Foundation, United States
Yabing Nan,
Chinese Academy of Medical Sciences, China

*CORRESPONDENCE
Yuzhong Xu,
254501651@qq.com

†These authors have contributed equally to this work

RECEIVED 25 July 2023
ACCEPTED 22 September 2023
PUBLISHED 03 October 2023

CITATION
Zhang B, Chang B, Wang L and Xu Y (2023), Three E2F target-related genes signature for predicting prognosis, immune features, and drug sensitivity in hepatocellular carcinoma.
Front. Mol. Biosci. 10:1266515.
doi: 10.3389/fmolb.2023.1266515

COPYRIGHT
© 2023 Zhang, Chang, Wang and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Three E2F target-related genes signature for predicting prognosis, immune features, and drug sensitivity in hepatocellular carcinoma

Baozhu Zhang[1†], Boyang Chang[2†], Lu Wang[3] and Yuzhong Xu[3*]

[1]Department of Radiation Oncology, The People's Hospital of Baoan Shenzhen, The Second Affiliated Hospital of Shenzhen University, Shenzhen, China, [2]Department of Interventional Radiology, The Third Affiliated Hospital of Sun Yat-Sen University, Guangzhou, China, [3]Department of Clinical Laboratory, The People's Hospital of Baoan Shenzhen, The Second Affiliated Hospital of Shenzhen University, Shenzhen, China

Background: Hepatocellular carcinoma (HCC) is extremely malignant and difficult to treat. The adenoviral early region 2 binding factors (E2Fs) target pathway is thought to have a major role in tumor growth. This study aimed to identify a predictive E2F target signature and facilitate individualized treatment for HCC patients.

Methods: We constructed an E2F target-related gene profile using univariate COX and LASSO regression models and proved its predictive efficacy in external cohorts. Furthermore, we characterized the role of the E2F target pathway in pathway enrichment, immune cell infiltration, and drug sensitivity of HCC.

Results: Lasso Cox regression created an E2F target-related gene signature of GHR, TRIP13, and CDCA8. HCC patients with high risk were correlated with shorter survival time, immune evasion, tumor stem cell characteristics and high sensitivity to Tipifarnib and Camptothecin drugs.

Conclusion: Hepatocellular carcinoma prognosis was predicted by an E2F target signature. This finding establishes the theoretical usefulness of the E2F target route in customized identification and treatment for future research.

KEYWORDS

E2F target, hepatocellular carcinoma, prognosis, immune landscape, drug sensitivity

Introduction

Liver cancer is the third leading cause of cancer-related death worldwide, with an expected 830,000 fatalities in 2020 (Sung et al., 2021). Hepatocellular carcinoma (HCC) accounts for over three-quarters of liver cancer cases. Two-thirds of patients are diagnosed when they cannot undergo curative surgery. Therefore, early diagnosis is necessary to improve the therapeutic effects and prognosis of HCC. Exploring novel genes and pathways is urgent to promote early diagnosis and individualized treatment.

The cyclin-dependent kinase (CDK), retinoblastoma transcriptional corepressor 1 (RB1), and the adenoviral early region 2 binding factors (E2Fs) form a sophisticated mechanism to control cell cycle progression (Kent and Leone, 2019). E2Fs detach from the E2F-RB1 complex and promote cell cycle-dependent gene transcription when RB1 is

altered or phosphorylated (van den Heuvel and Dyson, 2008; Zheng et al., 2023). Upregulation of E2F target genes is correlated with poor outcomes of neuroblastoma, breast cancer, colorectal cancer, ovarian cancer, and prostate cancer (Molenaar et al., 2012; Dahl et al., 2019; Oshi et al., 2021; Xia et al., 2022; Xu et al., 2022). Moreover, the E2F score serves as a predictive biomarker of response to neoadjuvant chemotherapy in estrogen receptor (ER)-positive/human epidermal growth factor receptor 2 (HER2)-negative breast cancer patients (Oshi et al., 2020).

A gene signature of the E2F target pathway was created to predict HCC patients' prognoses. First, using single-sample gene set enrichment analysis (ssGSEA), we identified that the E2F target pathway affects HCC prognosis. Weighted gene co-expression network analysis (WGCNA) and differential expression gene (DEG) analysis were used to uncover the E2F target-related gene set related to HCC prognosis. Using The Cancer Genome Atlas (TCGA) data as a training group, we found an E2F target-related gene signature using univariate analysis and least absolute shrinkage and selection operator (LASSO) Cox regression analyses. Each patient's E2F target pathway risk score was calculated based on the established gene signature.

Additionally, two external cohorts from the International Cancer Genome Consortium (ICGC) and Gene Expression Omnibus (GEO) were used to validate the predictive power of the E2F target gene signature. We also compared clinical characteristics, route enrichment, immune cell infiltration, and medication sensitivity across risk groups. The novel E2F target gene signature could contribute to the judgment of patients' prognosis and guidelines of clinical therapy for HCC patients.

Materials and methods

Data acquisition and processing

Six hundred forty-three HCC patients from TCGA, GEO, and ICGC were enrolled. A total of 370 HCC patients with comprehensive transcriptional and clinical data from the TCGA database (https://portal.gdc.cancer.gov/) were used as a training cohort. In this study, two external cohorts of HCC were used as validation groups. In validation group 1, 243 HCC patients were selected from the ICGC database (https://dcc.icgc.org/projects/). The validation group 2 consists of 30 HCC patients from the GSE107943 database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE107943).

Construction of E2F target signature

Reference genes were chosen from the Molecular Signatures Database (MSigDB) (http://www.gsea-msigdb.org/gsea/msigdb/search.jsp) (Table 1).

We calculated GSEA in the training group using the R package "GSVA" based on the above-mentioned gene set (Lee et al., 2008). R package for WGCNA was used to evaluate TCGA database mRNA matrix data (Langfelder and Horvath, 2008). We generated an adjacency matrix and then converted it into a topological overlap matrix (TOM) to assess the correlation strength between the nodes. By performing hierarchical clustering, we ensured that each module contained at least 50 genes. Finally, we integrated similar modules to find module genes having a good correlation with ssGSEA E2F target scores. By combining these genes with the E2F target gene set in HALLMARKS, we got a new gene set named gene set A.

Univariate Cox regression was used to identify gene set B's prognosis-related genes from gene set A. We used R package "limma" to determine DEGs between TCGA-LIHC tumors and healthy controls with a false discovery rate [FDR] < 0.05 and a log|fold change [FC]|> 1) (Ritchie et al., 2015). E2F-related DEGs and gene set B were intersected to obtain candidate genes to construct the gene signature. To avoid overfitting, LASSO Cox regression (R package "glmnet") was conducted to exclude collinear genes (T and ibshirani, 1997). Finally, the three best model genes were selected, and their coefficients were recorded to create an E2F target gene signature.

Subtypes based on etiology of the TCGA HCC cohorts

According to the contents of the "hist_hepato_carc_fact" record in the clinical information provided in the UCSC database, liver cancer samples were divided into HBV, HCV, ALD, and NASH. Finally, 104 samples with Hepatitis B, 56 samples with Hepatitis C, and 20 samples with Non-Alcoholic Fatty Liver Disease were retrieved. Alcoholic liver disease (ALD) related content was not obtained. Therefore, we divided the remaining 200 samples into others. Further, we compared the distribution of samples of different subgroups in the high-low-risk group, and no significant difference was observed (Table 2, $p = 0.533$).

TABLE 1 HALLMARK_E2F_TARGETS (Number:200).

Gene name
AK2 ANP32E ASF1A ASF1B ATAD2 AURKA AURKB BARD1 BIRC5 BRCA1 BRCA2 BRMS1L BUB1B CBX5 CCNB2 CCNE1 CCP110 CDC20 CDC25A CDC25B CDCA3 CDCA8 CDK1 CDK4 CDKN1A CDKN1B CDKN2A CDKN2C CDKN3 CENPE CENPM CHEK1 CHEK2 CIT CKS1B CKS2 CNOT9 CSE1L CTCF CTPS1 DCK DCLRE1B DCTPP1 DDX39A DEK DEPDC1 DIAPH3 DLGAP5 DNMT1 DONSON DSCC1 DUT E2F8 EED
EIF2S1 ESPL1 EXOSC8 EZH2 GINS1 GINS3 GINS4 GSPT1 H2AX H2AZ1 HALLMARK_E2F_TARGETS HELLS HMGA1 HMGB2 HMGB3 HMMR HNRNPD
HUS1 ILF3 ING3 IPO7 JPT1 KIF18B KIF22 KIF2C KIF4A KPNA2 LBR LIG1 LMNB1 LUC7L3 LYAR MAD2L1 MCM2 MCM3 MCM4 MCM5 MCM6 MCM7 MELK
MKI67 MLH1 MMS22L MRE11 MSH2 MTHFD2 MXD3 MYBL2 MYC NAA38 NAP1L1 NASP NBN
NCAPD2 NME1 NOLC1 NOP56 NUDT21 NUP107 NUP153 NUP205 ORC2 ORC6 PA2G4 PAICS PAN2 PCNA PDS5B PHF5A PLK1 PLK4 PMS2 PNN
POLA2 POLD1 POLD2 POLD3 POLE POLE4 POP7 PPM1D PPP1R8 PRDX4 PRIM2 PRKDC PRPS1 PSIP1 PSMC3IP
PTTG1 RACGAP1 RAD1 RAD21 RAD50 RAD51AP1 RAD51C RAN RANBP1 RBBP7 RFC1 RFC2 RFC3 RNASEH2A RPA1 RPA2 RPA3 RRM2 SHMT1 SLBP SMC1A
SMC3 SMC4 SMC6 SNRPB SPAG5 SPC24 SPC25 SRSF1 SRSF2 SSRP1 STAG1 STMN1 SUV39H1 SYNCRIP TACC3 TBRG4 TCF19 TFRC TIMELESS TIPIN TK1 TMPO
TOP2A TP53 TRA2B TRIP13 TUBB TUBG1 UBE2S UBE2T UBR7 UNG USP1 WDR90 WEE1 XPO1 XRCC6 ZW10

TABLE 2 Subtypes based on etiology of the TCGA HCC cohorts.

	High_risk	Low_risk	*p*-value
HBV	54	50	
HCV	24	25	0.533
NASH	6	12	
Others	101	99	

Calculation of mRNAsi and DNAsi

We use one Class Linear Regression (OCLR) to quantify the stemness of tumor samples. Two stemness indices were constructed from stem cell transcriptome, methylation group, and multi-platform data: mRNA expression-based stemness index (mRNAsi) represents gene expression, and epigenetically regulated-mRNAsi (DNAsi) measures epigenetics characteristics of stem cells. A stemness index (si) is a measure of stemness ranging from low (0) to high (1).

Evaluation of drug sensitivity

By developing regression models from cell line and gene expression profiles of Genomics of Drug Sensitivity in Cancer (GDSC), the pRRophetic algorithm predicted drug maximum 50% inhibitory concentration (IC50): (www.cancerrxgene.org/).

Statistical and bioinformatics analyses

All analyses and graphs were constructed using R v4.1.1. We calculated the gene signature's inter-gene correlations using the Pearson correlation test. The training set (TCGA cohort) and two test sets (ICGC and GEO cohorts) were divided into high- and low-risk groups based on the median value of the E2F target-related risk score. Group differences were calculated using the *t*-test and Chi-square test. The survival distribution and gene expression patterns of HCC patients were visualized using scatter plots and heat maps. Kaplan-Meier survival analysis determined whether high- and low-risk groups had distinct prognoses. A time-dependent ROC curve was calculated with R package time ROC. To test the independent predictability of the gene signature, the R package "survival" was used to run multivariate Cox stepwise regression models.

We used GSEA v4.1.0 to examine whether the E2F target pathway was activated in different risk categories. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases were used to explore the functions of the DEGs in high- and low-risk groups (Subramanian et al., 2005).

The CIBERSORT algorithm (https://cibersortx.stanford.edu/) and the immunophenoscore (IPS) were used to evaluate immune cells infiltrated in high- and low-risk groups (Newman et al., 2015; Charoentong et al., 2017). Finally, we used CellMiner to find new medications and biological targets based on gene-drug sensitivity (https://discover.nci.nih.gov/cellminer/home.do) (Reinhold et al., 2012).

Results

Prognostic role of E2F target pathway in HCC

Figure 1A illustrates the overall workflow. Using univariate Cox regression, we calculated 50 pathways ssGSEA scores in 370 TCGA HCC samples. The top 20 pathways with statistical significance are listed in Figure 1B, among which MYC_TARGET_V1, G2M_CHECKPOINT, and E2F_TARGETs are the first three items ($p < 0.0001$). Next, we compared the expression scores of these pathways in tumor tissue and paracancer tissue. E2F target showed the most significant score difference between tumor and paracancer tissue (Figure 1C). The score of G2M_CHECKPOINT in the tumor was also higher than that in para-cancer tissue, and the difference between tumor and para-cancer was the second largest (Figure 1C). The score of MYC_TARGET_V1 in tumors was significantly lower than in adjacent tissues, as shown in Figure 1C. By applying Decision Curve Analysis (DCA), we found that MYC_TARGET_V1 performed slightly better than the other two pathways, and the evaluation effects of the E2F target and G2M_CHECKPOINT were very similar, as shown in Figure 1D. Then, we performed ROC curve analysis, and the results showed that G2M_CHECKPOINT had the highest AUC value, but the AUC values of the three channels were all above 0.71 (Figure 1E). There was no significant difference in ROC diagnostic performance among the three channels. Based on the above results, we finally choose the E2F target as the main content of this study. According to the median value (0.4072497) of the E2F target pathway, we classified HCC patients as low- and high-score groups. Kaplan–Meier survival curves showed that high-score patients have a shorter overall survival time than low-score individuals ($p < 0.001$, Figure 1F).

Screen the candidate genes and build the E2F target signature

WGCNA co-expression algorithm identified co-expressed coding genes and modules of 370 HCC patients from the TCGA database. As a first step, we clustered the samples using hierarchical clustering and calculated the distance between each gene using the Pearson correlation coefficient. We used WGCNA to build a weight co-expression network with a soft threshold 10 to identify co-expression modules. To ensure that the network is scale-free, β= 10 is selected (Supplementary Figure S1). The expression matrix was converted into an adjacency matrix, which was then converted into a topological matrix. A hierarchical clustering method based on average linkage was employed to cluster genes, and each gene network module had a minimum of 50 genes per mixed dynamic shear tree standard. Cluster analysis was performed on the gene modules after the eigengenes of each module were calculated. Modules with relatively close distances were grouped as new modules.

When setting minModuleSize = 50, deepSplit = 3, and height = 0.25, we got a sum of 10 modules (Figure 2A). Grey modules are unaggregatable gene sets. Moreover, we analyzed the correlation between each module and each subtype and identified 609 genes with high correlation coefficients (above 0.5) with E2F, including the red (84), pink (82), and brown modules (443). These 609 genes and MSigDB E2F target pathway genes yielded 809 E2F target pathway

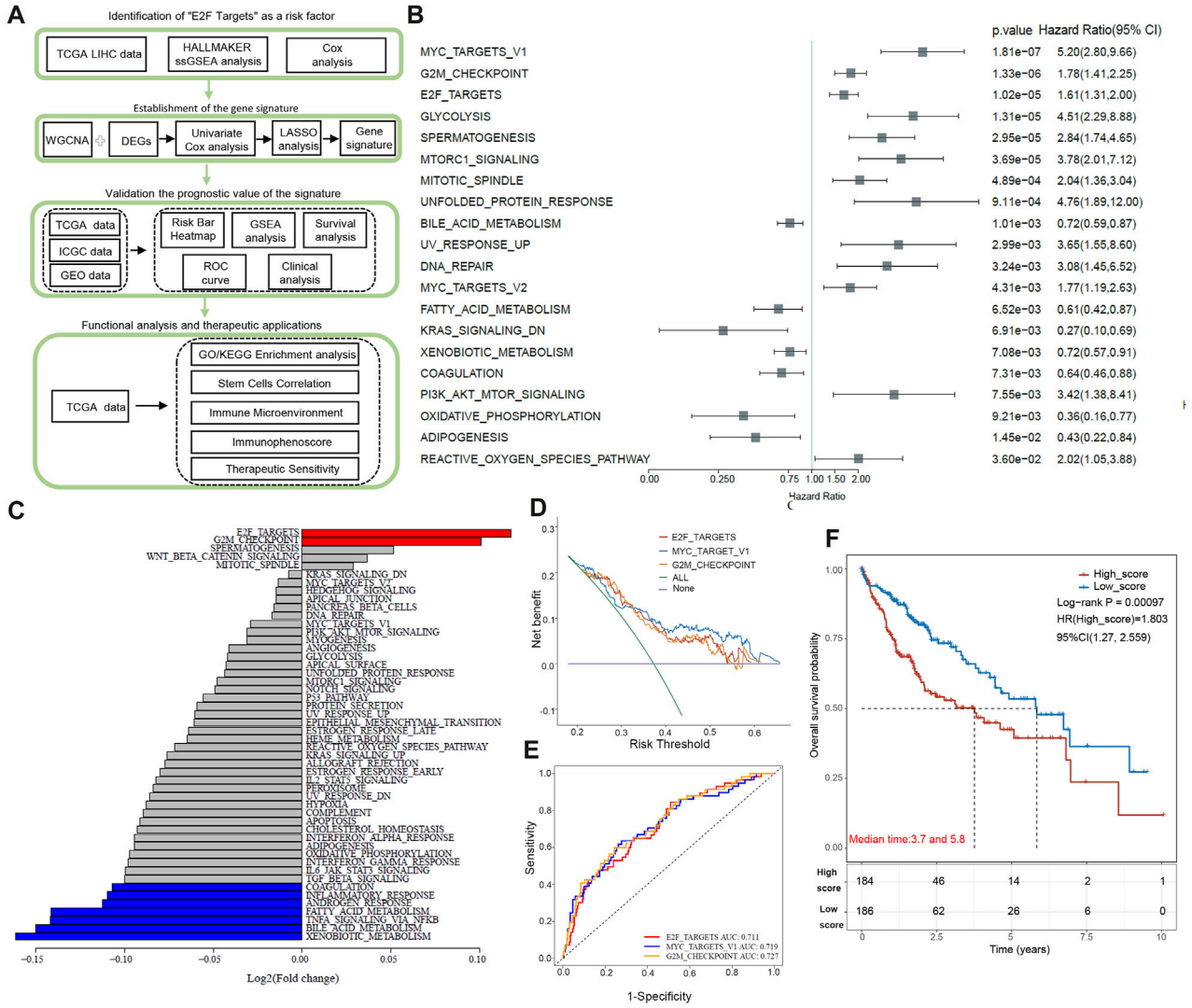


FIGURE 1
Diagram and identification of the "E2F target" pathway as a high-risk factor. **(A)** This study's flow chart and design. **(B)** The top 20 pathways are closely associated with poor prognosis of HCC patients based on Univariate Cox regression analysis. **(C)** Comparison of the expression scores of the top 20 pathways in tumor tissue and paracancer tissue. **(D)** Decision Curve Analysis of top 3 pathways: MYC_TARGET_V1, G2M_CHECKPOINT, and E2F_TARGETs. **(E)** ROC curve analysis of top 3 pathways: MYC_TARGET_V1, G2M_CHECKPOINT, and E2F_TARGETs. **(F)** By Kaplan-Meier analysis, patients with high E2Fs scores have shorter overall survival than those with low E2Fs scores.

genes. Using univariate Cox regression analysis, 576 E2F-related prognostic genes were identified ($p < 0.05$). Meanwhile, edgR and Deseq2 were used to analyze the differences between liver and paracancer samples. A total of 1,400 DEGs (1,222 upregulated and 178 downregulated in the tumor) were identified by edgR, while 1,313 DEGs (1,100 up-expressed and 213 down-expressed) were obtained by Deseq2. These differential genes and 576 E2F-related prognostic genes revealed 52 differential prognostic genes (Figure 2C). Figure 2D exhibited the univariate Cox regression results of these 52 genes. Figure 2E shows the expression heatmap of 52 genes in HCC and adjacent liver tissue. Using the R software package glmnet, we conducted a lasso cox regression analysis. In Figures 2F,G, we analyzed the confidence interval for each lambda. Figure 2H showed that when lambda = 0.07227, there are three genes remained in the model, namely, growth hormone receptor (GHR), thyroid hormone receptor interactor 13 (TRIP13), and cell division cycle associated 8 (CDCA8). The final model formula is:

$$\text{RiskScore} = -0.0302^{*}exp^{\,GHR} + 0.0694^{*}exp^{\,TRIP13} + 0.1701^{*}exp^{\,CDCA8}$$

The formula suggested that increased GHR expression is a protective factor in HCC and related to low risk, while TRIP13 and CDCA8 were associated with high RiskScore and poor outcomes. TRIP13 accelerates the mitotic process and leads to chromosome instability via playing a role in spindle assembly checkpoint and DNA repair pathways (Lu et al., 2019). High TRIP13 gene expression is found in HCC tissues, which promotes cell growth and metastasis via activating AKT/mTOR and silencing TGF-β1/smad3 pathway (Yao et al., 2018; Zhu et al., 2019). Many

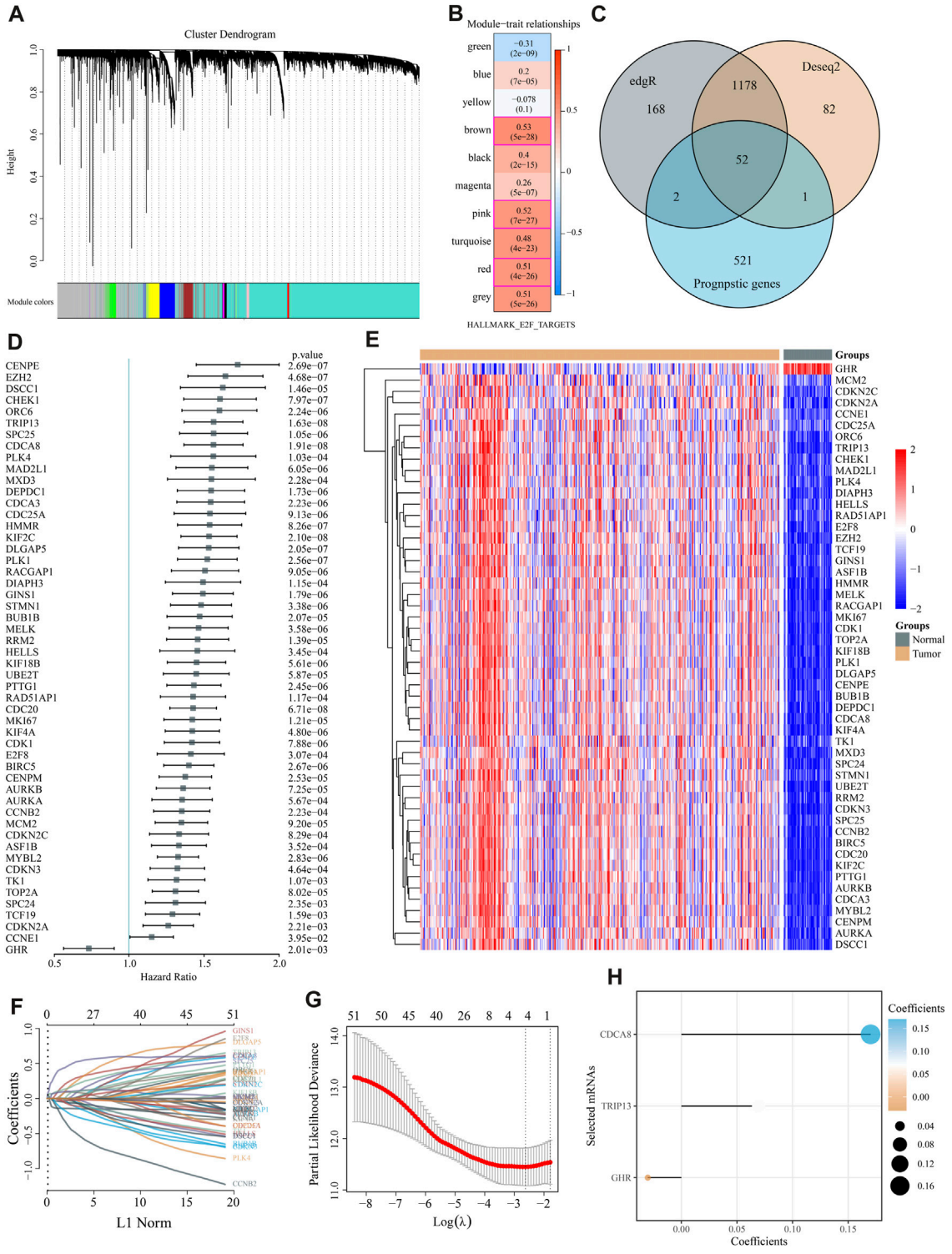


FIGURE 2
Modeling of an "E2F target" risk score. **(A)** Gene dendrograms and module colors for the WGCNA analysis. **(B)** The red, pink, and brown modules were significantly and consistently upregulated in gene ontology (GO) analysis. **(C)** Venn diagrams depicted 52 prognostic differential genes obtained by fusing DESeq, edgR, and prognostic genes. **(D)** Univariate Cox regression results for 52 genes are shown in the forest map. **(E)** The heat map reveals 52 DEGs\ between liver cancer and adjacent normal liver tissue. **(F,G)** We used LASSO Cox regression to identify the signature; the best log (λ) value was −2.63, and 3 indicators remained. **(H)** LASSO coefficients for the three genes in the signature.

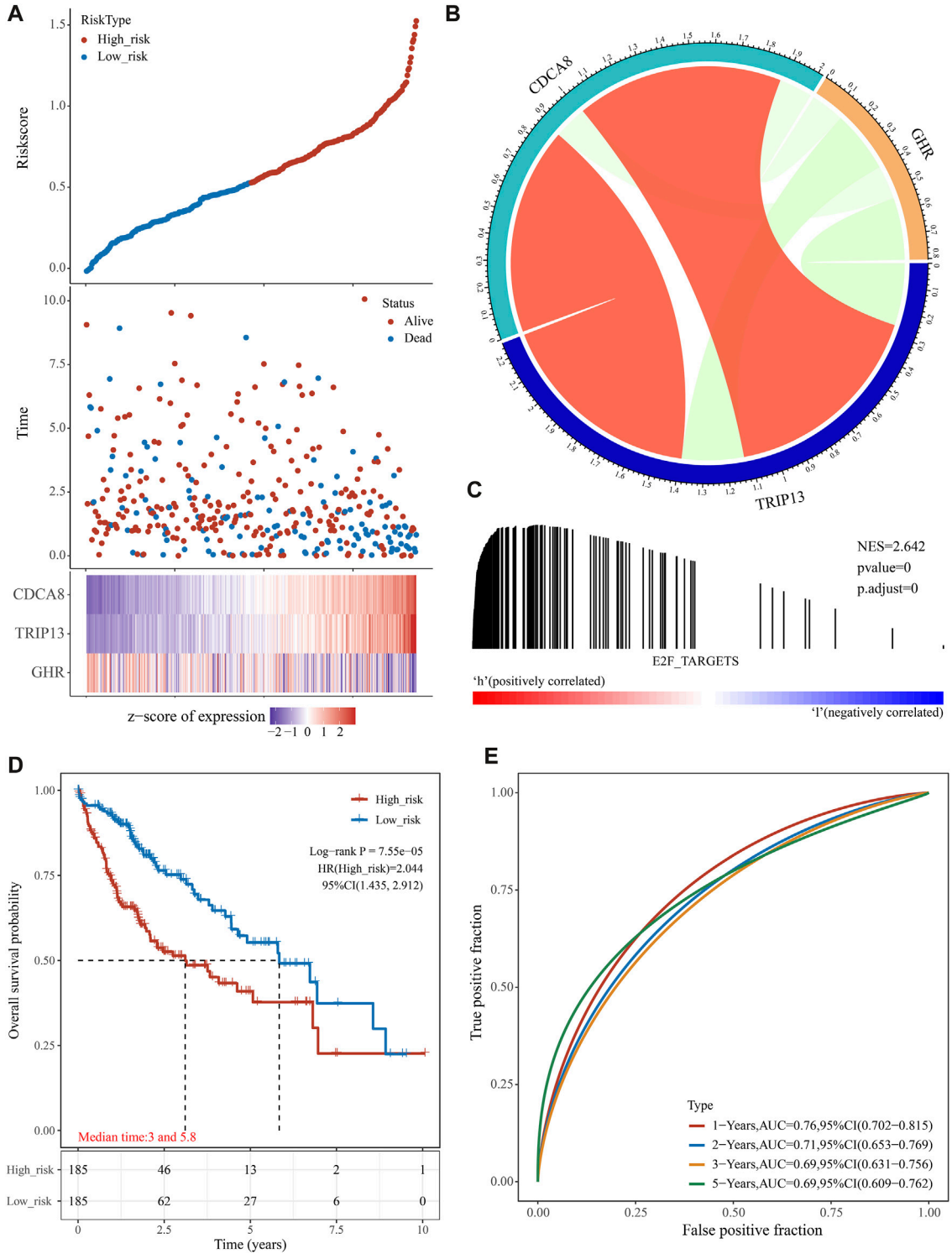


FIGURE 3
The prognostic analysis of the E2F target-related signature on TCGA cohort. **(A)** The distribution of patient risk score-survival and heat map of "E2F target"-related genes. **(B)** According to the gene correlation heat map, three genes were not highly correlated. **(C)** In the high-risk group, GSEA detected the "E2F Target" pathway activation. **(D)** A Kaplan-Meier survival analysis showed that the overall survival time was shorter in patients in high-risk groups. **(E)** ROC analysis showed that the gene signature had a high 5-year AUC, indicating a powerful predictive ability.

researchers have identified CDCA8 as a novel oncogene, which predicts a poor prognosis in HCC as well as promotes tumor proliferation via MEK/ERK, AKT/β-Catenin, and CDK1/cyclin B1 signaling (Jeon et al., 2021; Cui and Jiang, 2023). In the terms of the GHR gene, some researchers have observed its upregulation promoted HCC development (Garcia-Caballero et al., 2000; Haque et al., 2022). On

the other hand, its downregulation has been noted in HCV-induced HCC and related to an unfavorable outcome (Lin et al., 2021; Abu El-Makarem et al., 2022).

We searched the clinical information of the TCGA HCC cohort (n = 380) from the UCSC database. As a result, 104 samples with Hepatitis B, 56 samples with Hepatitis C, and 20 samples with Non-Alcoholic Fatty Liver Disease were retrieved. Alcoholic liver disease (ALD) related content was not obtained. We divided the remaining 200 samples into others. Next, we analyzed the prognostic KM curves of the three genes in the analysis model in different subpopulations, and the results showed that GHR showed low expression and poor prognosis in HCV samples and total samples, which was consistent with the trend presented in our model (Supplementary Figures S2A–D). TRIP13 and CDCA8 only showed high expression and poor prognosis in total samples, which was also consistent with the trend shown in our model (Supplementary Figures S2A–D). Finally, we observed the expression of the three genes in different subgroups, and the results showed that the expression of GHR in different subgroups was significantly lower than that in adjacent tissues. TRIP13 and CDCA8 were significantly higher than paracancer tissues, as shown in Supplementary Figure S2E.

A high E2F risk score predicts a poor prognosis in training group

We generated the risk score for each training group sample and presented the RiskScore distribution (Figure 3A). Patients with a high RiskScore have a worse prognosis than those with a low RiskScore for HCC (Figure 3A). TRIP13 and CDCA8 were risk factors whose expression altered with increasing risk value, whereas high GHR expression was a protective factor linked with low risk (Figure 3B). Using GSEA, it was confirmed that genes associated with the E2F target pathway were significantly enriched in high-risk individuals (Figure 3C). We divided TCGA HCC samples into high- and low-risk groups based on the median risk score (0.528115). The KM curve displayed that the OS of the high-risk group was significantly lower than the low-risk group [Figure 3D, log-rank $p < 0.0001$, HR = 2.044 (1.435–2.912)]. Using R software package timeROC, we assessed prediction power over one, two, three, and 5 years. This risk model has a very high AUC area below the line (AUC >0.69, Figure 3E).

Two published papers have constructed E2F-related models in liver cancer, including E2F target gene characteristics of five genes constructed by Hu et al. (2022), and two gene models constructed by Wang et al. (2023), respectively. To figure out whether our new E2F target signature has an advantage, we calculated the risk score of each sample according to the formulas provided in the paper. 371 TCGA HCC samples and corresponding prognostic follow-up information provided in TCGA were used for the KM curve and AUC analysis. The results showed that the Kaplan-Meier curves of all models showed significant differences, as shown in Supplementary Figure S3A. The AUC value of the model constructed by Hu et al. (2022) was above 0.658. The AUC value of the model constructed by Wang et al. (2023) is above 0.658, as shown in Supplementary Figure S3B; Compared with the above results, we found that the AUC values of the two models are lower than ours. Also, through c-index analysis, the c-index value of our model is 0.68 (Cl 95%:0.63–0.73), and the c-index value of the model constructed by Hu et al. (2022) is 0.65 (Cl 95%: 0.63–0.71). The c-index value of the model Wang et al. (2023) constructed was 0.64 (Cl 95%: 0.59–0.69). Finally, through the DCA decision curve analysis, we found that the return rate of the E2F model we built was also higher than that of the other two models, as shown in Supplementary Figure S3C.

The prognostic value of the E2F target-related risk score was validated in the test cohort

ICGC and GEO HCC cohorts were used as external validation sets to evaluate this risk model's predictive value. First, the risk score of each sample was calculated in two datasets, and the RiskScore distribution of the samples was plotted as shown in Figures 4A, B. Next, we separated patients in two validation cohorts into high- and low-risk groups according to the median risk score (ICGC cohort: 0.3101617, GEO cohort: 0.7349518). Kaplan-Meier plotter revealed that patients in high-risk groups have poor prognoses within 6 years in the ICGC cohort and 8 years in the GEO cohort (Figures 4C, D). The prediction power was evaluated for one, two, three, and 5 years using the R software package timeROC. Figures 4E, F showed that the AUC area below the line is very high for the risk model, with 0.76 for the ICGC cohort and 0.657 for the GEO cohort.

Gene signature and clinical characteristics

Pearson correlation analysis assessed whether risk groups and pathological characteristics were correlated. According to our findings, advanced T Stage, advanced stage, and high grade were significantly correlated with high-risk groups ($p < 0.05$, as shown in Figure 5).

Different functions of two risk groups

We calculated the DEGs between high- and low-risk groups using the "limma" algorithm, and the threshold was set as $p < 0.05$. Figure 6A shows 473 upregulated and 115 downregulated genes in the high-risk group. To analyze functional enrichment in DEGs, we used R's clusterProfiler package and parameterized the threshold to an FDR of 0.05. According to KEGG analysis, the high-risk group showed activation of the DNA replication pathway, cell cycle pathway, and p53 signaling pathway (Figures 6B, C). Based on the GO analysis, E2F targets are primarily involved in DNA replication and cell cycle checkpoints (Figure 6D). The association between risk ratings and tumour stem cell features were also evaluated. Risk score has no relationship with DNA stem index (mDNAsi) but is significantly correlated with RNA stem index (mRNAsi), as shown in Figures 6E, F.

Immune microenvironment and therapeutic sensitivity

CIBERSORT was utilized as an online tool to deconvolute expression matrices of human immune cell subtypes using linear support vector regression (Newman et al., 2015). We characterized the differences of 22 immune cell infiltration between high- and low-

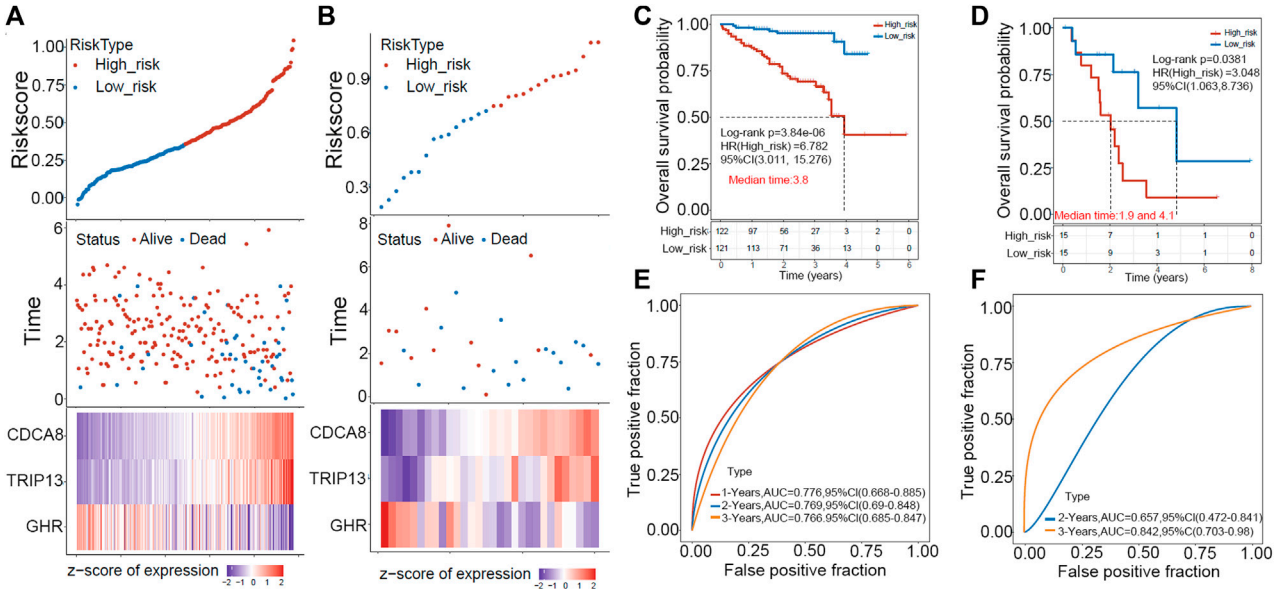


FIGURE 4
The GEO and ICGC cohorts were used to validate the prognostic signature externally. **(A,B)** A distribution map of risk score survival and a heat map of expression of "E2F target"-related genes in ICGC (left) and GEO (right) cohorts are shown. **(C,D)** In ICGC (left) and GEO (right), patients in high-risk groups displayed shorter overall survival time. **(E,F)** ROC analysis indicated a high 3-year AUC for the gene signature, suggesting its power as a predictive tool in the ICGC (left) and GEO (right) cohorts.

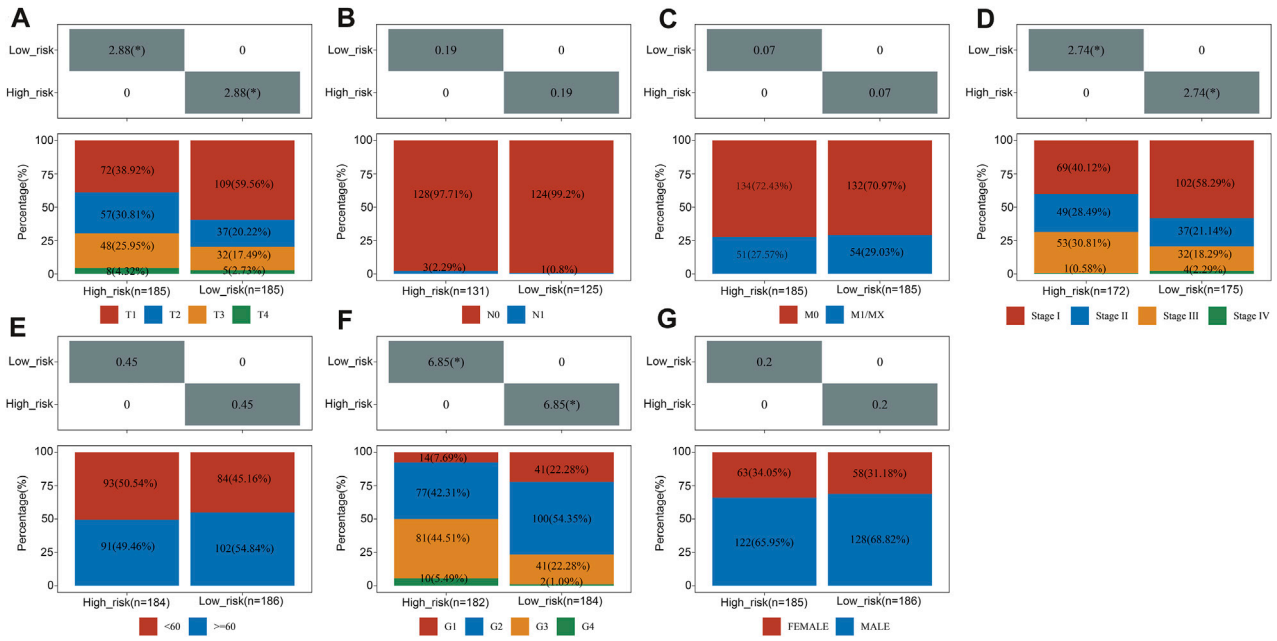


FIGURE 5
The relationship between clinical characteristics and risk groups. The correlation between the risk score and T Stage **(A)**, N stage **(B)**, M stage **(C)**, TNM stage **(D)**, sex **(E)**, differentiated degree **(F)** and sex **(G)**.

risk groups. In the high-risk group, there were more infiltrating B cell plasma, T cell follicular helper, T cell regulatory (Tregs), macrophage M0, neutrophils, and T cell CD4$^+$ memory activated within tumor tissues than in the low-risk group (Figure 7A). In contrast, the high-risk group had fewer T cell CD4$^+$ memory resting, monocyte, NK cell resting, mast cell activated, B cell naïve infiltrated compared to the low-risk group (Figure 7A). According to the CIBERSORT algorithm, the high-risk group displayed immunosuppressive infiltration and a lack of immunoactive cells. Moreover, we analyzed the expression levels of 41 immune suppressor genes from the tumor immunophenotype

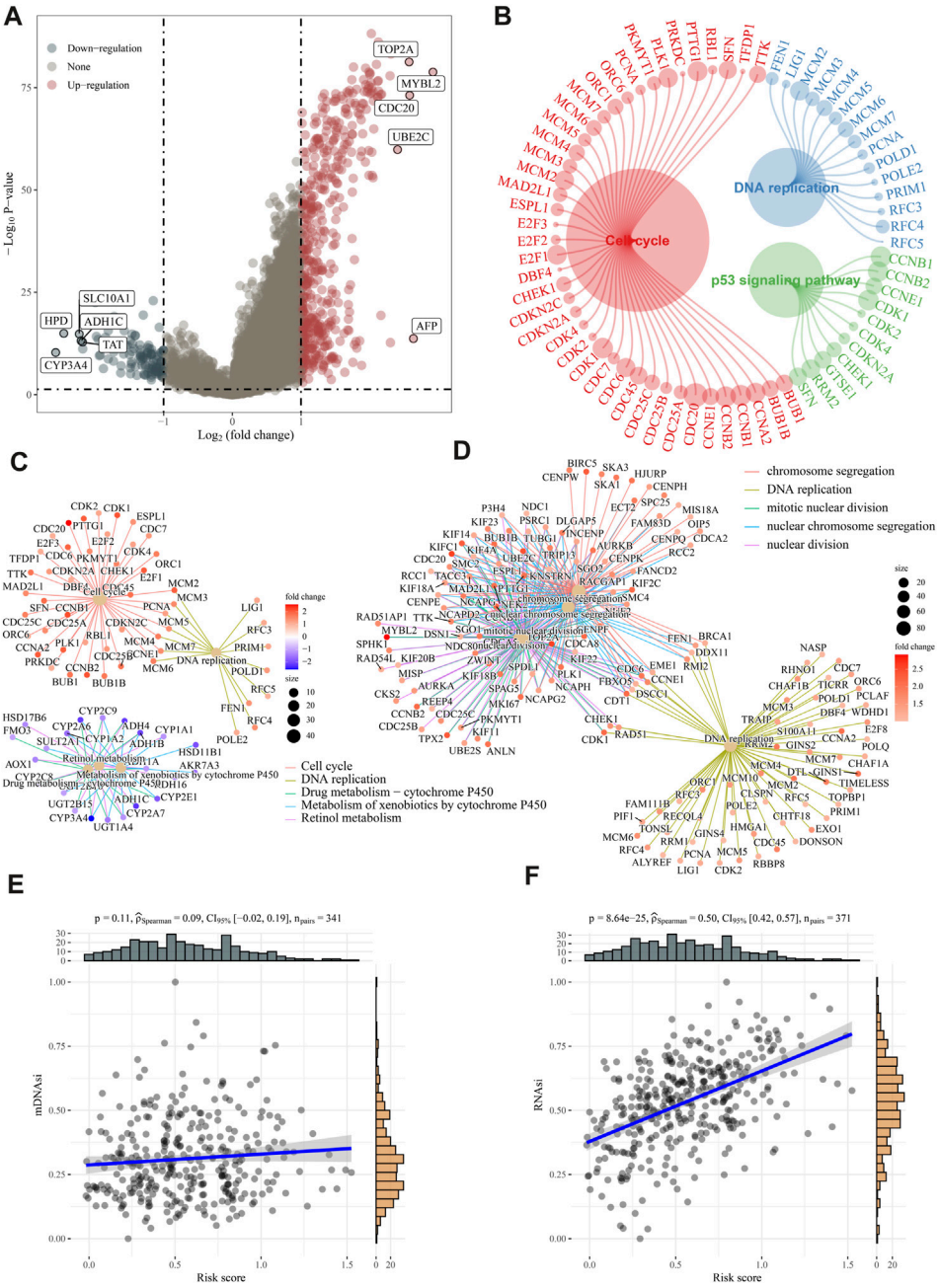


FIGURE 6
Analysis of biological function and stem cell characteristics. **(A)** DEG volcano diagrams between high- and low-risk groups. **(B–D)** According to KEGG and GO enrichment analyses, DNA replication, and cell cycle pathways were active in high-risk patients. **(E–F)** A positive correlation was found between the "E2F target" risk score and DNAsi and RNAsi.

database (TIP, http://biocc.hrbmu.edu.cn/TIP/index.jsp). Results revealed that 31 genes out of 41 (75.6%) were significantly upregulated in the high-risk group, including programmed cell death 1 (PDCD1), hepatitis A virus cellular receptor 1/2 (HAVCR1/2), T cell immunoreceptor with Ig and ITIM domains (TIGIT), indoleamine 2,3-dioxygenase 1 (IDO1), enhancer of zeste 2 polycomb repressive complex 2 subunit (EZH2), and DNA methyltransferase 1 (DNMT1) (Figure 7B).

ESTIMATE predicts tumor purity and infiltrated matrix/immune cells in tumor tissue based on ssGSEA. Estimate score (tumor purity), stromal score (matrix in tumor tissue), and immune score (infiltration of immune cells in tumor tissue) are the three main scores derived from ESTIMATE. Figure 7C shows the proportion of stromal scores is significantly lower in a high-risk group than in the low-risk group. Immune and ESTIMATE Scores had no obvious difference between the two groups (Figures 7D, E). The results suggest that the stromal component of TME is more suitable for distinguishing high-risk from low-risk patients.

The TCGA data was used by Thorsson et al. (2018) to identify six immune subtypes: C1 (Wound Healing), C2 (IFN-γ Dominant), C3

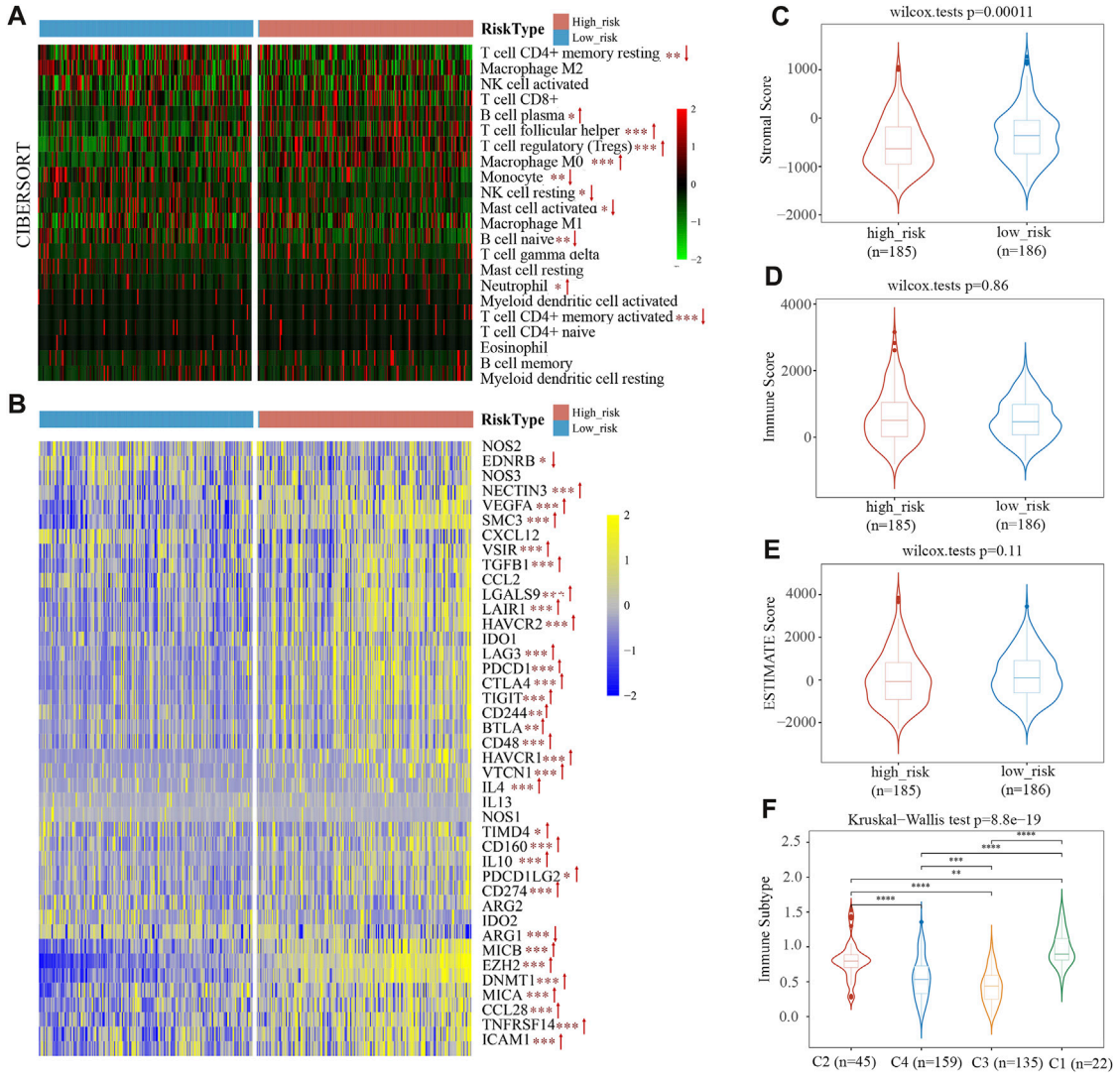


FIGURE 7
The immune cell infiltration in TCGA cohort. **(A)** Infiltration of 22 immune cells in high- and low-risk groups. **(B)** Heat map of immunosuppressor genes' expression in two risk groups. **(C–E)** Stromal score **(C)**, immune score **(D)**, and ESTIMATE Score **(E)** between the low- and high-risk groups. **(F)** Immunotyping results indicated that the risk scores of C1, C2, and C4 types were higher than C3.

(Inflammatory), C4 Lymphocyte Depleted), C5 (Immunologically Quiet), and C6 (TGF-β Dominant). The risk score of C1, C2, and C4 types with poor prognoses was higher, and C3 types with good prognoses were primarily found among patients with low-risk scores (Figure 7F). This phenomenon indicated that the immune subtype and risk score are related.

High-risk and low-risk groups' drug resistance and sensitivity

Drug resistance often emerges during cancer treatment, leading to poor efficacy and unfavorable outcome of HCC. To test E2F risk models in chemotherapy, we predicted the IC50 of 138 medicines in high- and low-risk patients using the pRRophetic algorithm. A total of 65 drugs showed significant variations between high-risk and low-risk patients, with Tipifarnib, Camptothecin, Salubrinal, and Nilotinib being more sensitive to high-risk groups. (Figure 8A, $p < 0.05$). Moreover, we found that most 65 drugs showed a significant correlation between their IC50s and risk scores (Figure 8B). The above results suggest that the E2F risk model can help HCC patients choose chemotherapy medicines based on clinical drug effects.

Discussion

The Retinoblastoma gene (Rb) was the first tumor suppressor gene cloned and sequenced (Goodrich et al., 1991). The transcription product of the Rb gene is about 4.7 kb, and the expression product is P105-Rb, a 928 amino acid protein with a molecular weight of about 105 kDa (Classon and Harlow, 2002). P105-Rb has two states: phosphorylation and dephosphorylation. Phosphorylation is inactive, and dephosphorylation is active (Gubern et al., 2016). Dephosphorized p105 inhibits cell proliferation by binding to the transcription factor

10.3389/fmolb.2023.1266515

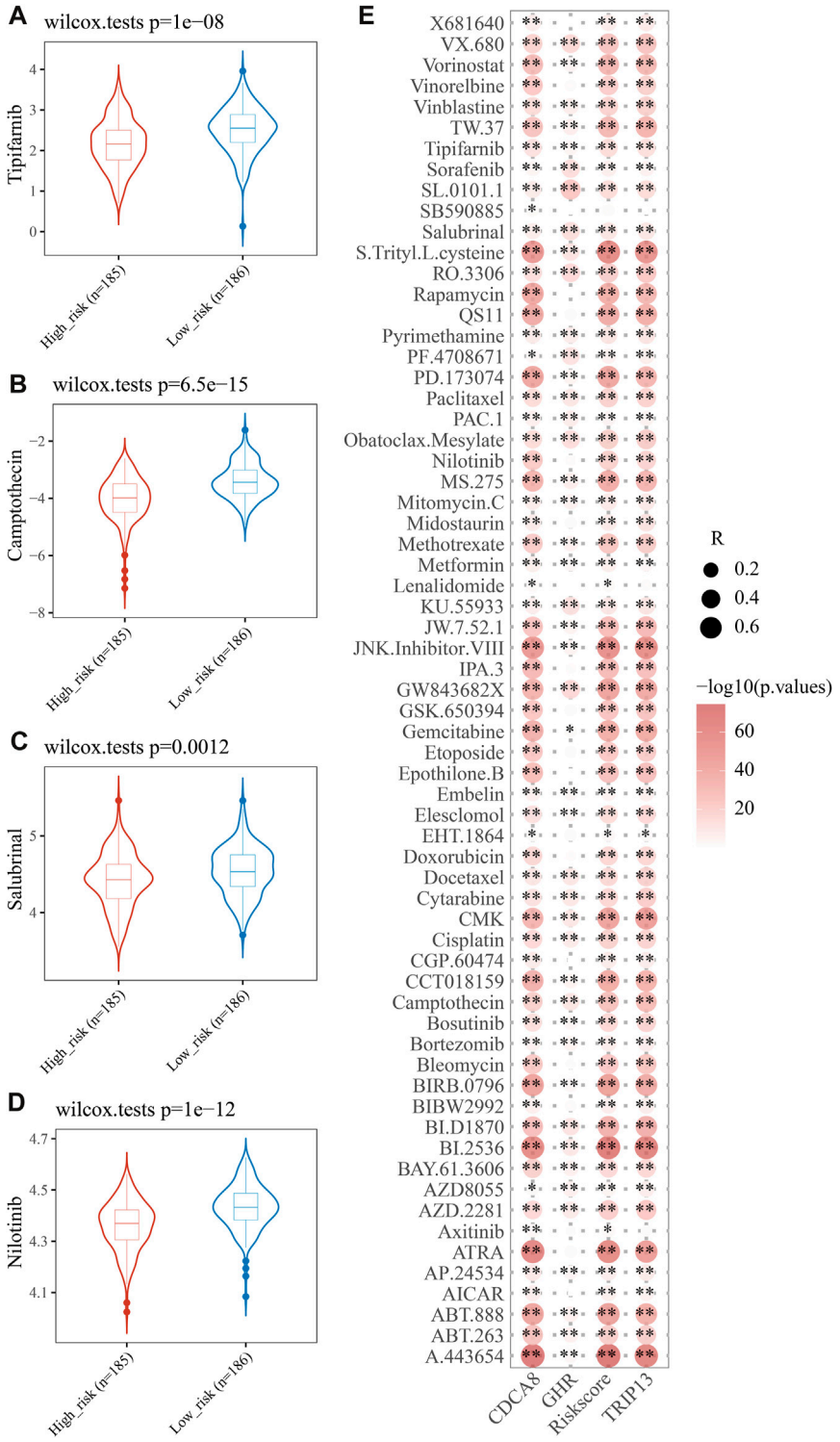


FIGURE 8
Analysis of therapeutic effect and drug sensitivity. IC50 distributions of four drugs in the high- and low-risk group, Tipifarnib **(A)**, Camptothecin **(B)**, Salubrinal **(C)**, and Nilotinib **(D)**. **(E)** Correlation analysis of three genes in the signature and common drugs used to treat HCC.

adenoviral early region 2 binding factors (E2Fs) (Dimaras et al., 2015). E2F stimulates DNA replication enzyme gene transcription. When dephosphorized p105 binds to E2F, it inactivates E2F. Rb protein is phosphorylated by Cyclin-CDKs (CyclinD/CDK4 or CyclinE/CDK2) during the G1 phase. p-Rb releases its binding E2F to increase the transcription of Cyclin (cyclin) and CDK proteins, which causes the cell to enter the S phase from G1 phase (Zhou et al., 2022).

Researchers have characterized eight members of the mammalian E2F family, namely, E2F1-E2F8 (DeGregori and Johnson, 2006). E2F1, E2F2, E2F3, E2F4, and E2F8 are shown to

be upregulated human HCC and promote cancer progression (Palaiologou et al., 2012; Liu et al., 2003a; Liu et al., 2003b; Deng et al., 2010; S et al., 2021). Cell cycle-dependent E2F transcription factors govern target gene transcription. Several E2F target genes are tightly connected, which may help predict tumor prognosis. An E2F target gene signature, formed by MDX3, PLK1, EPHA10, and KIF4A, exhibited a stronger predictive power than existing signatures in prostate cancer (Xia et al., 2022). Hu et al. (2022) established an E2F target gene signature composed of five genes (HN1, KIF4A, CDCA3, CDCA8, and SSRP1) and found it is significantly related to the prognosis of hepatocellular carcinoma. Two-E2F (E2F2 and E2F5) prognostic signature was built by Wang et al. (2023), and they estimated immune infiltration levels for patients in different risk groups. In the present study, we verified that our E2F risk model is superior to the two signatures mentioned above in predicting the overall survival of HCC patients.

Three E2F target-related genes provide a new gene signature for predicting HCC prognosis and aiding clinical decision-making. The risk score model composed of GHR, TRIP13, and CDCA8 could predict the prognosis of HCC accurately. Two external cohorts of patients from ICGC and GEO databases confirmed its prediction power. Functional enrichment analysis showed that high-risk groups were more active in DNA replication, cell cycle, p53 signaling pathway, and stem cell features.

Nowadays, anti-programmed death receptor-1 (PD-1), anti-programmed death ligand 1 (PD-L1), and anti-cytotoxic T-lymphocyte antigen-4 (CTLA-4) mAbs are among the most used immune checkpoint inhibitors in advanced HCC cases (Herbst et al., 2014). To our comfort, combining immunotherapy strategies, like anti-PD-1/PD-L1 mAbs plus anti-VEGF mAbs, TKIs, or anti-CTLA-4 mAbs, can overcome drug resistance and extend overall survival (Pinter et al., 2023). ICIs are innately resistant in 30% of HCC patients. Novel and effective biomarkers are needed to identify immunotherapy candidates. Our study used CIBERSORT to estimate immune cell infiltration. There were more immunosuppressive cells in the high-risk group than in the low-risk group, like infiltrating B cell plasma, T cell follicular helper, T cell regulatory (Tregs), macrophage M0, neutrophils, and T cell $CD4^+$ memory activated. Similarly, immunosuppressive genes were upregulated. A well-designed and combined immunotherapy strategy is needed for the high-risk group due to their immunosuppressive microenvironment (Rimassa et al., 2023).

The present study found increased GHR expression to be a protective factor in HCC and related to low risk. However, according to the concrete biological context, GHR may play the dual role of inhibiting and promoting cancer. GHR downregulation was considered an independent predictor for worse outcomes in HCC (Liu et al., 2003c; Abu El-Makarem et al., 2022; He et al., 2022). In contrast, several researchers considered GHR a tumor promoter and a potential therapeutic target in HCC. Using gene knockout mice, researchers found that most Ghr+/+ and Ghr+/-mice developed HCC in response to DEN, but not the Ghr−/− mice (5.6%) (Haque et al., 2022). Scholars from Texas MD Anderson Cancer observed that tumor cells exhibited slower growth and overcame sorafenib resistance by blocking GHR with pegvisomant *in vitro* (Kaseb et al., 2022). GH inhibition downregulates ABC transporters and sensitizes HCC allografts to sorafenib (Basu et al., 2022). We noted that GHR's downregulation is associated with HCV-induced HCC (Lin et al., 2021; Abu El-Makarem et al., 2022). We believe that rhGH and antagonists should be cautiously used in HCC patients before we fully understand the relationship between GHR and HCV virus-related HCC. High TRIP13 and CDCA8 expression predicts a favorable outcome in HCC (Zhang et al., 2009; Yao et al., 2018). Knockdown of TRIP13 and CDCA8 inhibited HCC growth and metastasis by impeding cell cycle and proliferation (Zhu et al., 2019; Chen et al., 2023).

There are several limitations to this study. Despite modern bioinformatic analysis tools, E2F target genes' role in HCC growth and metastasis has not been explored. To answer this question, solid experiments are to unveil potential molecular pathways and their intracellular effect.

Nevertheless, new inspirations and ideas were introduced by the present study. First, we created an E2F target-related gene signature that accurately predicted HCC prognosis. Second, the E2F target pathway may have tumor-promoting effects on HCC progression based on function-enriched and immune infiltration analyses. In conclusion, we furnished HCC patients with possibly treatable target genes and responsive medicines.

Conclusion

This study aimed to generate a signature associated with the E2F target that may be utilized to predict the prognosis of HCC. Moreover, possible pathway and mechanisms to understand how the E2F target pathway promotes tumor growth and progression of HCC have been analyzed. With the advances provided in this study, a molecular diagnosis and tiered treatment of HCC patients may be feasible. Our discovery also provides a theoretical basis for researchers to examine E2F target genes as potential HCC treatment targets.

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

BZ: Conceptualization, Data curation, Investigation, Methodology, Writing–original draft. BC: Conceptualization, Data curation, Formal Analysis, Methodology, Writing–original draft. LW: Formal Analysis, Conceptualization, Writing–original draft. YX: Funding acquisition, Supervision, Writing–review and editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the National Natural Science Foundation of China (Grant Number: 81802732) and Shenzhen Science and Technology Innovation Commission (Grant Number: RCBS20200714114918239).

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/fmolb.2023.1266515/full#supplementary-material

References

Abu El-Makarem, M. A., Kamel, M. F., Mohamed, A. A., Ali, H. A., Mohamed, M. R., Mohamed, A. E. M., et al. (2022). Down-regulation of hepatic expression of GHR/STAT5/IGF-1 signaling pathway fosters development and aggressiveness of HCV-related hepatocellular carcinoma: crosstalk with snail-1 and type 2 transforming growth factor-beta receptor. *PLoS One* 17 (11), e0277266. doi:10.1371/journal.pone.0277266

Basu, R., Qian, Y., Mathes, S., Terry, J., Arnett, N., Riddell, T., et al. (2022). Growth hormone receptor antagonism downregulates ATP-binding cassette transporters contributing to improved drug efficacy against melanoma and hepatocarcinoma *in vivo*. *Front. Oncol.* 12, 936145. doi:10.3389/fonc.2022.936145

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, E., He, Y., Jiang, J., Yi, J., Zou, Z., Song, Q., et al. (2023). CDCA8 induced by NF-YA promotes hepatocellular carcinoma progression by regulating the MEK/ERK pathway. *Exp. Hematol. Oncol.* 12 (1), 9. doi:10.1186/s40164-022-00366-y

Classon, M., and Harlow, E. (2002). The retinoblastoma tumour suppressor in development and cancer. *Nat. Rev. Cancer* 2 (12), 910–917. doi:10.1038/nrc950

Cui, Y., and Jiang, N. (2023). CDCA8 facilitates tumor proliferation and predicts a poor prognosis in hepatocellular carcinoma. *Appl. Biochem. Biotechnol.* doi:10.1007/s12010-023-04603-w

Dahl, E. S., Buj, R., Leon, K. E., Newell, J. M., Imamura, Y., Bitler, B. G., et al. (2019). Targeting IDH1 as a prosenescent therapy in high-grade serous ovarian cancer. *Mol. Cancer Res.* 17 (8), 1710–1720. doi:10.1158/1541-7786.MCR-18-1233

DeGregori, J., and Johnson, D. G. (2006). Distinct and overlapping roles for E2F family members in transcription, proliferation and apoptosis. *Curr. Mol. Med.* 6 (7), 739–748. doi:10.2174/156652401060607039

Deng, Q., Wang, Q., Zong, W. Y., Zheng, D. L., Wen, Y. X., Wang, K. S., et al. (2010). E2F8 contributes to human hepatocellular carcinoma via regulating cell proliferation. *Cancer Res.* 70 (2), 782–791. doi:10.1158/0008-5472.CAN-09-3082

Dimaras, H., Corson, T. W., Cobrinik, D., White, A., Zhao, J., Munier, F. L., et al. (2015). Retinoblastoma. *Nat. Rev. Dis. Prim.* 1, 15021. doi:10.1038/nrdp.2015.21

Garcia-Caballero, T., Mertani, H. M., Lambert, A., Gallego, R., Fraga, M., Pintos, E., et al. (2000). Increased expression of growth hormone and prolactin receptors in hepatocellular carcinomas. *Endocrine* 12 (3), 265–271. doi:10.1385/ENDO:12:3:265

Goodrich, D. W., Wang, N. P., Qian, Y. W., Lee, E. Y., and Lee, W. H. (1991). The retinoblastoma gene product regulates progression through the G1 phase of the cell cycle. *Cell.* 67 (2), 293–302. doi:10.1016/0092-8674(91)90181-w

Gubern, A., Joaquin, M., Marques, M., Maseres, P., Garcia-Garcia, J., Amat, R., et al. (2016). The N-terminal phosphorylation of RB by p38 bypasses its inactivation by CDKs and prevents proliferation in cancer cells. *Mol. Cell.* 64 (1), 25–36. doi:10.1016/j.molcel.2016.08.015

Haque, A., Sahu, V., Lombardo, J. L., Xiao, L., George, B., Wolff, R. A., et al. (2022). Disruption of growth hormone receptor signaling abrogates hepatocellular carcinoma development. *J. Hepatocell. carcinoma* 9, 823–837. doi:10.2147/JHC.S368208

He, S., Qiao, J., Wang, L., and Yu, L. (2022). A novel immune-related gene signature predicts the prognosis of hepatocellular carcinoma. *Front. Oncol.* 12, 955192. doi:10.3389/fonc.2022.955192

Herbst, R. S., Soria, J. C., Kowanetz, M., Fine, G. D., Hamid, O., Gordon, M. S., et al. (2014). Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* 515 (7528), 563–567. doi:10.1038/nature14011

Hu, W., Shi, Y., Han, T., Liu, C., Cao, X., Shi, G., et al. (2022). A panel of E2F target gene signature predicting the prognosis of hepatocellular carcinoma. *Front. Genet.* 13, 879299. doi:10.3389/fgene.2022.879299

Jeon, T., Ko, M. J., Seo, Y. R., Jung, S. J., Seo, D., Park, S. Y., et al. (2021). Silencing CDCA8 suppresses hepatocellular carcinoma growth and stemness via restoration of ATF3 tumor suppressor and inactivation of AKT/β-Catenin signaling. *Cancers* 13 (5), 1055. doi:10.3390/cancers13051055

Kaseb, A. O., Haque, A., Vishwamitra, D., Hassan, M. M., Xiao, L., George, B., et al. (2022). Blockade of growth hormone receptor signaling by using pegvisomant: A functional therapeutic strategy in hepatocellular carcinoma. *Front. Oncol.* 12, 986305. doi:10.3389/fonc.2022.986305

Kent, L. N., and Leone, G. (2019). The broken cycle: E2F dysfunction in cancer. *Nat. Rev. Cancer* 19 (6), 326–338. doi:10.1038/s41568-019-0143-7

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, E., Chuang, H. Y., Kim, J. W., Ideker, T., and Lee, D. (2008). Inferring pathway activity toward precise disease classification. *PLoS Comput. Biol.* 4 (11), e1000217. doi:10.1371/journal.pcbi.1000217

Lin, C. C., Liu, T. W., Yeh, M. L., Tsai, Y. S., Tsai, P. C., Huang, C. F., et al. (2021). Significant down-regulation of growth hormone receptor expression revealed as a new unfavorable prognostic factor in hepatitis C virus-related hepatocellular carcinoma. *Clin. Mol. hepatology* 27 (2), 313–328. doi:10.3350/cmh.2020.0247

Liu, L. X., Jiang, H. C., Liu, Z. H., Zhu, A. L., Zhang, W. H., Wu, L. F., et al. (2003a). Expression of cell cycle/growth regulator genes in human hepatocellular carcinoma and adjacent normal liver tissues. *Oncol. Rep.* 10 (6), 1771–1775. doi:10.3892/or.10.6.1771

Liu, L. X., Liu, Z. H., Jiang, H. C., Zhang, W. H., Qi, S. Y., Hu, J., et al. (2003b). Gene expression profiles of hepatoma cell line HLE. *World J. Gastroenterol.* 9 (4), 683–687. doi:10.3748/wjg.v9.i4.683

Liu, J. P., Wang, H. T., Ou, Q. J., Lu, Y. T., Lu, H. P., Gao, J. H., et al. (2003c). Expression of growth hormone receptor in hepatocellular carcinoma and its significance. *Ai Zheng* 22 (3), 298–301.

Lu, S., Qian, J., Guo, M., Gu, C., and Yang, Y. (2019). Insights into a crucial role of TRIP13 in human cancer. *Comput. Struct. Biotechnol. J.* 17, 854–861. doi:10.1016/j.csbj.2019.06.005

Molenaar, J. J., Koster, J., Ebus, M. E., van Sluis, P., Westerhout, E. M., de Preter, K., et al. (2012). Copy number defects of G1-cell cycle genes in neuroblastoma are frequent and correlate with high expression of E2F target genes and a poor prognosis. *Genes. Chromosom. Cancer* 51 (1), 10–19. doi:10.1002/gcc.20926

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

Oshi, M., Takahashi, H., Tokumaru, Y., Yan, L., Rashid, O. M., Nagahashi, M., et al. (2020). The E2F pathway score as a predictive biomarker of response to neoadjuvant therapy in ER+/HER2- breast cancer. *Cells* 9 (7), 1643. doi:10.3390/cells9071643

Oshi, M., Angarita, F. A., Tokumaru, Y., Yan, L., Matsuyama, R., Endo, I., et al. (2021). A novel three-gene score as a predictive biomarker for pathologically complete response after neoadjuvant chemotherapy in triple-negative breast cancer. *Cancers* 13 (10), 2401. doi:10.3390/cancers13102401

Palaiologou, M., Koskinas, J., Karanikolas, M., Fatourou, E., and Tiniakos, D. G. (2012). E2F-1 is overexpressed and pro-apoptotic in human hepatocellular carcinoma. *Virchows Archiv Int. J. pathology* 460 (5), 439–446. doi:10.1007/s00428-012-1220-4

Pinter, M., Pinato, D. J., Ramadori, P., and Heikenwalder, M. (2023). NASH and hepatocellular carcinoma: immunology and immunotherapy. *Clin. Cancer Res.* 29 (3), 513–520. doi:10.1158/1078-0432.CCR-21-1258

Reinhold, W. C., Sunshine, M., Liu, H., Varma, S., Kohn, K. W., Morris, J., et al. (2012). CellMiner: A web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res.* 72 (14), 3499–3511. doi:10.1158/0008-5472.CAN-12-1370

Rimassa, L., Finn, R. S., and Sangro, B. (2023). Combination immunotherapy for hepatocellular carcinoma. *J. Hepatol.* 79, 506–515. doi:10.1016/j.jhep.2023.03.003

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

Shen, S., and Wang, Y. (2021). Expression and prognostic role of E2F2 in hepatocellular carcinoma. *Int. J. general Med.* 14, 8463–8472. doi:10.2147/IJGM.S334033

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 (43), 15545–15550. doi:10.1073/pnas.0506580102

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

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

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 (4), 812–830.e14. doi:10.1016/j.immuni.2018.03.023

van den Heuvel, S., and Dyson, N. J. (2008). Conserved functions of the pRB and E2F families. *Nat. Rev. Mol. Cell. Biol.* 9 (9), 713–724. doi:10.1038/nrm2469

Wang, L., Chen, Y., Chen, R., Mao, F., Sun, Z., and Liu, X. (2023). Risk modeling of single-cell transcriptomes reveals the heterogeneity of immune infiltration in hepatocellular carcinoma. *J. Biol. Chem.* 299 (3), 102948. doi:10.1016/j.jbc.2023.102948

Xia, H., Wang, M., Su, X., Lv, Z., Yan, Q., Guo, X., et al. (2022). A novel gene signature associated with "E2F target" pathway for predicting the prognosis of prostate cancer. *Front. Mol. Biosci.* 9, 838654. doi:10.3389/fmolb.2022.838654

Xu, Z., Qu, H., Ren, Y., Gong, Z., Ri, H. J., Zhang, F., et al. (2022). Systematic analysis of E2F expression and its relation in colorectal cancer prognosis. *Int. J. general Med.* 15, 4849–4870. doi:10.2147/IJGM.S352141

Yao, J., Zhang, X., Li, J., Zhao, D., Gao, B., Zhou, H., et al. (2018). Silencing TRIP13 inhibits cell growth and metastasis of hepatocellular carcinoma by activating of TGF-β1/smad3. *Cancer Cell. Int.* 18, 208. doi:10.1186/s12935-018-0704-y

Zhang, J. B., Li, X. H., Ning, F., and Guo, X. S. (2009). Relationship between expression of GYPC and TRIP3 genes and prognosis of acute lymphoblastic leukemia in children. *Chin. J. Contemp. Pediatr.* 11 (1), 29–32.

Zheng, Y., Jia, H., Wang, P., Liu, L., Chen, Z., Xing, X., et al. (2023). Silencing TRAIP suppresses cell proliferation and migration/invasion of triple negative breast cancer via RB-E2F signaling and EMT. *Cancer gene Ther.* 30 (1), 74–84. doi:10.1038/s41417-022-00517-7

Zhou, L., Ng, D. S., Yam, J. C., Chen, L. J., Tham, C. C., Pang, C. P., et al. (2022). Post-translational modifications on the retinoblastoma protein. *J. Biomed. Sci.* 29 (1), 33. doi:10.1186/s12929-022-00818-x

Zhu, M. X., Wei, C. Y., Zhang, P. F., Gao, D. M., Chen, J., Zhao, Y., et al. (2019). Elevated TRIP13 drives the AKT/mTOR pathway to induce the progression of hepatocellular carcinoma via interacting with ACTN4. *J. Exp. Clin. cancer Res. CR* 38 (1), 409. doi:10.1186/s13046-019-1401-y

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 04 October 2023
DOI 10.3389/fmolb.2023.1240175

Check for updates

OPEN ACCESS

EDITED BY
Lin Zhang,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

REVIEWED BY
Jinyu Zhang,
East Tennessee State University, United States
Li Chen,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

*CORRESPONDENCE
Gengyun Sun,
sungengy@126.com

†These authors have contributed equally to this work

RECEIVED 14 June 2023
ACCEPTED 22 September 2023
PUBLISHED 04 October 2023

CITATION
Li Y, Wu C, Fei L, Xu Q, Shao X, Chen B and Sun G (2023), Clinical analysis of prolonged viral clearance time in patients with lymphoma combined with novel coronavirus infection.
Front. Mol. Biosci. 10:1240175.
doi: 10.3389/fmolb.2023.1240175

COPYRIGHT
© 2023 Li, Wu, Fei, Xu, Shao, Chen and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Clinical analysis of prolonged viral clearance time in patients with lymphoma combined with novel coronavirus infection

Ying Li†, Chao Wu†, Liming Fei†, Qin Xu†, Xianru Shao, Bangjie Chen and Gengyun Sun*

First Clinical Medical College (First Affiliated Hospital), Anhui Medical University, Hefei, China

Objective: To compare the period of viral clearance and its influencing factors after severe acute respiratory syndrome coronavirus (SARS-CoV-2) infection between patients with lymphoma and lung cancer.

Methods: We retrospectively collected the clinical data of patients with lymphoma and lung cancer (118 cases) diagnosed with SARS-CoV-2 infection and hospitalized in the First Affiliated Hospital of Anhui Medical University between 1 December 2022, and 15 March 2023. Finally, 87 patients with prolonged virus clearance times were included and divided into lymphoma (40 cases) and lung cancer (47 cases) groups. We used the Kaplan-Meier method to draw a negative turn curve. We performed a univariate analysis of the prolongation of virus clearance time and a Cox regression model for multivariate analysis.

Results: The median times for viral clearance in the lung cancer and lymphoma groups were 18 (95% confidence interval [CI] 15.112–20.888) and 32 (95%CI 27.429–36.571) days, respectively. Log-rank analysis showed a statistically significant difference (p = 0.048), and the lymphocyte count in the lymphoma group was lower than that in the lung cancer group (p = 0.044). We used the Cox regression model to conduct a multivariate analysis, which revealed that in lymphoma patients, the interval between the time of diagnosis and the time of SARS-CoV-2 infection <24 months (hazard ratio [HR]: 0.182, 95%CI: 0.062–0.535, p = 0.02), an interval between the last anti-CD20 monoclonal antibody treatment and the time of SARS-CoV-2 infection of <2 months (HR: 0.101, 95%CI: 0.029–0.358, p < 0.001), and a decrease in peripheral blood lymphocyte levels (HR: 0.380, 95%CI: 0.179–0.808, p = 0.012) were independent risk factors for prolonged viral clearance time.

Conclusion: Patients with lymphoma combined with SARS-CoV-2 infection had a longer virus clearance time than did patients with lung cancer. Moreover, the lymphocyte count in the lymphoma group was lower than that in the lung cancer group; therefore, the immune status of patients with lymphoma is lower than that of patients with lung cancer. An interval between lymphoma diagnosis and SARS-CoV-2 infection of <2 years, anti-CD20 monoclonal antibody treatment within the past 2 months, and a decrease in lymphocyte levels in the peripheral blood prolonged the virus clearance time in the patients in this study.

KEYWORDS

lymphoma, lung cancer, SARS-CoV-2, viral clearance time, immunity

1 Introduction

In March 2020, the World Health Organization declared the disease caused by severe acute respiratory syndrome coronavirus (SARS-CoV-2) infection a global pandemic known as coronavirus disease 2019 (COVID-19) (Gopalaswamy and Subbian, 2021). Several risk factors for increased morbidity and mortality in COVID-19 patients have been identified, including male sex, hypertension, chronic lung disease, diabetes, immunodeficiency, and cancer. Therefore, identifying the potential risk factors for COVID-19 is of significant value for public health and healthcare policies (Jun et al., 2022). Patients with solid tumors or hematological malignancies often experience a more severe and rapid disease course requiring high-level intensive care. These patients are at higher risk of SARS-CoV-2 infection-related mortality than the general population. Lymphoma, a hematological malignancy characterized by severe immunosuppression, has been sparsely reported in combination with COVID-19; however, patients with lymphoma diagnosed with COVID-19 have a poor prognosis (Visco et al., 2022). In patients without lymphoma, the time to test negative for SARS-CoV-2 via reverse transcription-polymerase chain reaction (RT-PCR) is usually approximately 10 days (Li et al., 2021). However, patients with lymphoma are immunosuppressed and the virus may persist in the body for extended periods, affecting disease prognosis and survival outcomes. Notably, most individuals with COVID-19 exhibit mild to moderate infection but can rapidly progress from asymptomatic to acute respiratory distress syndrome, multiple organ dysfunction syndrome, and even death (Ballow and Haga, 2021). Although COVID-19 vaccination has effectively reduced the incidence of severe or critical illnesses, patients with lymphoma receiving the vaccine may not achieve effective protection because they often fail to generate sufficient antiviral immune responses (Kumar et al., 2017). Patients with lymphoma may have a higher risk of death owing to their inability to effectively clear the virus from their bodies. Therefore, assessing the continued presence of the novel coronavirus in these patients can assist in evaluating the risk of death. This study compared the duration of viral clearance after SARS-CoV-2 infection and its influencing factors between patients with lymphoma and those with lung cancer.

2 Materials and methods

2.1 Clinical information

We retrospectively collected the clinical data and performed follow-up of 49 patients with advanced lymphoma and 69 patients with advanced lung cancer who were diagnosed with SARS-CoV-2 infection and hospitalized at the First Affiliated Hospital of Anhui Medical University between 1 December 2022, and 15 March 2023.

2.2 Inclusion criteria, diagnostic criteria, and curative effect evaluation

We included patients who met the following criteria. **Inclusion criteria:** (Gopalaswamy and Subbian, 2021): confirmed prolonged viral clearance time (defined as a turnaround time >10 days); (Jun et al., 2022); confirmed lymphoma or lung cancer diagnosis through pathological, cytological, and relevant imaging examinations; (Visco et al., 2022); late-stage (III–IV) lymphoma or lung cancer in a stable disease (SD) state; (Li et al., 2021); diagnosed with SARS-CoV-2 infection; (Ballow and Haga, 2021); complete medical record documentation; and (Kumar et al., 2017) age≥18 years. This study included 87 patients (lymphoma group: 40 patients; lung cancer group: 47 patients. **Diagnostic criteria**: The diagnostic criteria for lymphoma staging followed the 2014 Lugano staging system, as follows: Stage I: involvement of a single lymph node area (I) or localized involvement of a single extranodal organ (I.E.,); Stage II: involvement of ≥2 lymph node areas on the same side of the diaphragm (II), possibly with limited involvement of localized extranodal organs in the same lymphatic drainage area (IIE); Stage III: involvement of lymph node areas above and below the diaphragm or involvement of the spleen in addition to above diaphragm involvement (IIIS); Stage IV: involvement of extranodal organs beyond the lymphatic drainage area (IV). Lung cancer was staged according to the 8th edition of the International Association for the Study of Lung Cancer (IASLC TNM) staging manual. **Curative effect evaluation:** The patients included in the study were categorized as having SD for the evaluation of therapeutic efficacy. Lymphoma response assessment was performed using the Lugano 2014 criteria: 1) complete response (CR), typically indicated by complete resolution on positron emission tomography-computed tomography (PET-CT) and computed tomography (CT) imaging, with the disappearance of lesions, normalization of organ size, absence of new lesions, and normal bone marrow morphology on CT. 2) Partial response (PR): fewer lesions, up to a maximum of six target lesions, and generally small existing lesions (often less than 5 mm × 5 mm), with some lesions gradually disappearing. Organ involvement in these patients is usually only minimally increased, and bone marrow remains unchanged compared to baseline. 3) Stable disease (SD): patients typically show no metabolic activity, with a five-point scale (5PS) score for target and extranodal lesions. Metabolic activity remains relatively unchanged compared to baseline. The bone marrow remains unchanged, the tumor size shrinks by <50% or increases by <25%, and no new lesions appear. 4) Disease progression: individual target lesions or extranodal lesions with a 5PS score of 4–5, tumor size increases by >25% or new lymphoma-related hypermetabolic lesions appear, and new or recurrent FDG avidity in the bone marrow. Lung cancer response assessment followed the Response Evaluation Criteria in Solid Tumors (RECIST) 1.1: CR, disappearance of all tumor lesions; PR, ≥30% decrease in the sum of the longest diameters of the target lesions; PD, ≥20% increase in the sum of the longest diameters of the target lesions, the appearance of new lesions, and/or definite progression of non-target lesions; and SD, lesion changes between PR and PD.

2.3 Study design

We performed a retrospective analysis by following up and comparing the clinical data of patients with advanced lymphoma and those with advanced lung cancer and defined positive results as viral infection status and negative results as viral clearance in

patients via reverse transcription polymerase chain reaction (RT-PCR) testing. The time from positive to negative RT-PCR results in patients with lymphoma and lung cancer was recorded. The two groups of patients with prolonged conversion times were compared to investigate whether viral clearance time was longer in patients with lymphoma than in those with lung cancer. Multifactorial regression models have been used to investigate factors influencing prolonged viral clearance time in patients with lymphoma.

2.4 Study procedures, definitions, and follow-up

Follow-up was conducted by reviewing the outpatient and inpatient hospital information systems (HISs) and telephone follow-ups with a follow-up cutoff date of 1 April 2023. The collected data included demographic characteristics at enrollment, including sex, age, comorbidities (such as hypertension, diabetes, smoking, and history of hepatitis B), various laboratory test indices during hospitalization (such as peripheral blood leukocyte count, lymphocyte count, neutrophil count, albumin level, lactate dehydrogenase level, and C-reactive protein level), pathological staging and treatment of patients with lymphoma/lung cancer, the primary disease status at the time of SARS-CoV-2 infection (progression, remission, or stable), clinical symptoms (such as persistent fever ≥10 days and cough), the time interval between positive and negative RT-PCR test results in patients with lymphoma/lung cancer, the time interval between the diagnosis of the primary disease and SARS-CoV-2 infection in patients with lymphoma/lung cancer patients (more or less than 24 months) and the time interval between the last antitumor treatment and SARS-CoV-2 infection in such patients (more or less than 2 months).

2.5 Statistical treatment

Data management and statistical analyses were performed using IBM SPSS Statistics for Windows, version 27.0 (IBM Corp., Armonk, NY, USA). Measures that conformed to a normal distribution were expressed as $x \pm s$, while those that did not were expressed as median (M) and upper and lower quartiles (Q1, Q3), using the Mann–Whitney U test. Count data were expressed as frequencies (percentages), and categorical variables in the study were tested using the X^2 or Fisher's exact tests. The Kaplan-Meier method was used to plot the transition curves. The log-rank test and Cox regression model were used for univariate and multivariate analyses, respectively. The hazard ratios (HRs) and the corresponding 95% confidence intervals (CIs) were calculated. Statistical significance was defined as $p < 0.05$.

3 Results

3.1 Analysis of novel coronavirus clearance times

The median times to viral clearance in the lung cancer and lymphoma groups were 18 days (95%CI 15.11–20.89) and 32 days (95%CI 27.43–36.57), respectively. Log-rank analysis showed a significant difference ($p = 0.048$). The Kaplan–Meier curve is shown in (Figure 1).

3.2 Univariate analysis of prolonged clearance time of the novel coronavirus

Univariate analysis showed lower peripheral blood indicators, including white blood cell, neutrophil, and lymphocyte counts, in patients with lymphoma than in those with lung cancer. In contrast, the lactate dehydrogenase levels were significantly higher in patients with lymphoma than in those with lung cancer. The lymphoma group had 18 patients with persistent fever (≥10 days), and the intergroup difference was significant ($p < 0.05$). In the lymphoma group, 25 patients received the anti-CD20 treatment regimen (Table 1).

3.3 Multivariate analysis of prolonged clearance time of novel coronavirus

The results of multiple collinearity analyses of the statistically significant indicators of the univariate analysis suggested no collinearity among these indicators. These indicators were included in the multivariate Cox regression analysis. The results showed that the risk factors for prolonged viral clearance time were a time between lymphoma diagnosis and diagnosis of SARS-CoV-2 infection of <24 months (25 patients) ($p = 0.002$, HR: 0.182, 95% CI: 0.062–0.535); time between anti-CD20 monoclonal antibody treatment and diagnosis of SARS-CoV-2 infection of <2 months (18 patients) ($p < 0.001$, HR: 0.101, 95%CI: 0.029–0.358); and decreased peripheral blood lymphocyte levels in patients with lymphoma ($p = 0.012$, HR: 0.380, 95%CI: 0.179–0.808) were risk factors for prolonged viral clearance time (Table 2).

4 Discussion

The clinical manifestations of individuals infected with SARS-CoV-2 vary widely and range from mild flu-like symptoms to life-threatening respiratory failure. Acute respiratory distress syndrome (ARDS), caused by the release of proinflammatory mediators, an intense immune response, and endothelial damage is the leading cause of disease exacerbation or death (Batah and Fabro, 2021). Lymphoma is a group of non-solid tumors characterized by various immune dysfunctions, including congenital and acquired immune deficiencies, such as low serum immunoglobulin levels and impaired cellular and humoral immunity (Moyon et al., 2022), which place patients at risk of various infections, including that by SARS-CoV-2. Consistent with our findings, a previous study reported that a decreased lymphocyte count resulting from active antitumor treatments, such as anti-CD20 therapy, was a significant factor contributing to mortality in patients with lymphoma with concurrent SARS-CoV-2 infection (Lee et al., 2022). B and $CD4^+$ T cells play essential roles in viral clearance; thus, patients with lymphoma patients, who have B and $CD4^+$ T cell deficiencies, are unable to effectively and rapidly clear SARS-CoV-2, leading to a

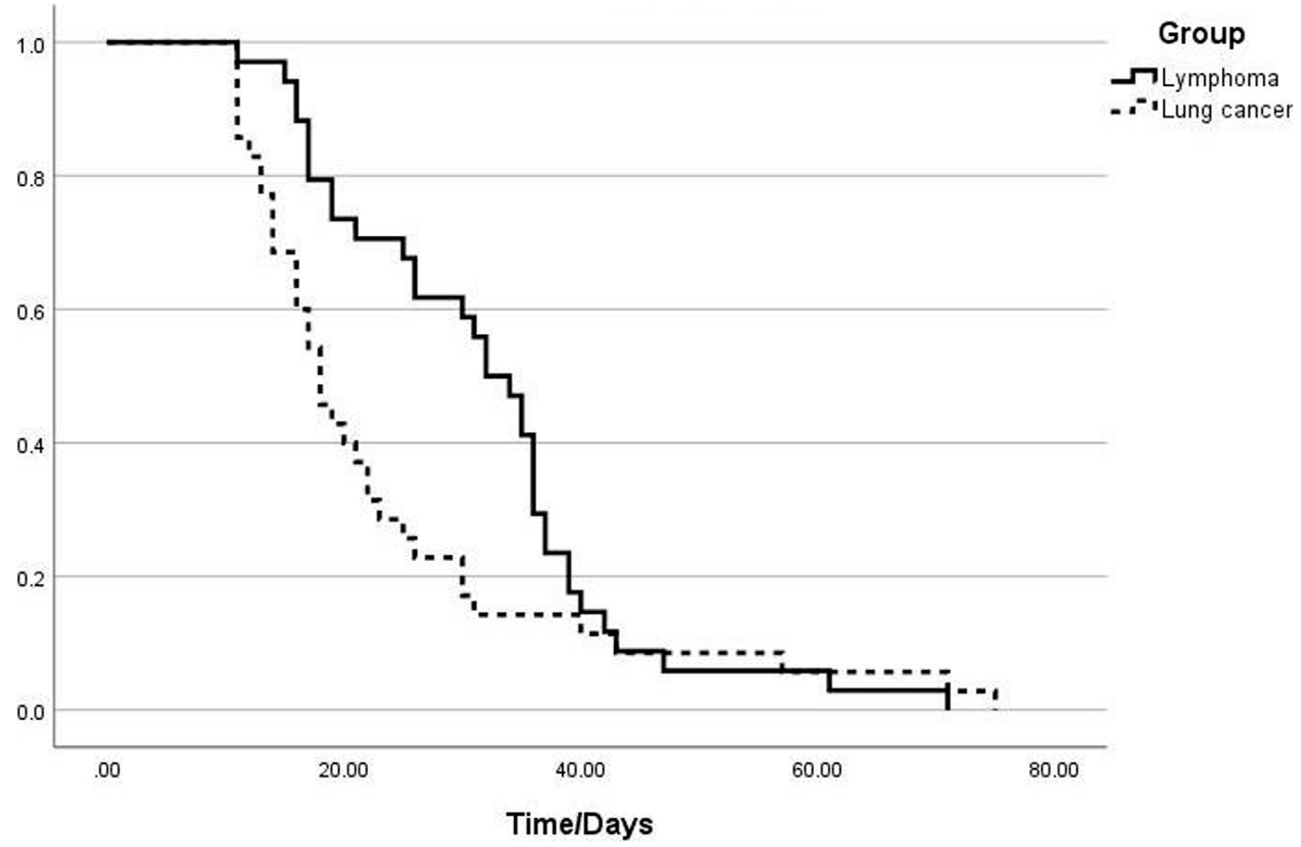


FIGURE 1
Analysis of the viral clearance time curves in the groups of patients with lymphoma and those with lung cancer.

higher risk of disease exacerbation or death. Therefore, assessing the clearance of SARS-CoV-2 in patients can assist in evaluating patient prognosis. Hence, we conducted this retrospective analysis to study the prolongation of viral clearance time and its influencing factors to guide future disease treatment. For immunocompromised hosts, delayed viral clearance should be considered and the duration of antiviral treatment should be appropriately extended. Combination therapy with antiviral drugs can help achieve rapid viral clearance and control symptoms.

In our cohort, patients with lymphoma had a longer duration for SARS-CoV-2 clearance compared to those with lung cancer. The median times to seronegativity were 18 and 32 days in the lung cancer and lymphoma groups, respectively. Patients with lymphoma had lower levels of peripheral blood leukocytes (p = 0.004), neutrophils (p = 0.045), and lymphocytes (p = 0.044) than did those in patients with lung cancer, suggesting that patients with lymphoma have a reduced immune status compared to patients with lung cancer (Heo et al., 2019). The higher lactate dehydrogenase (LDH) levels in the lymphoma group could be related to the systemic characteristics of lymphoma. Most patients with lymphoma patients experience multiorgan and tissue damage, leading to elevated LDH levels in the peripheral blood. In contrast, patients with lung cancer typically exhibit increased LDH levels in pleural effusion (Wang et al., 2019). We defined persistent fever as fever symptoms lasting for >10 days. The proportion of patients with lymphoma with persistent fever was significantly higher than that of patients with lung cancer, with as many as half of the patients experiencing fever symptoms lasting for >10 days. However, high fever (>38.5°C) is rare, likely because of the continuous presence of the virus and stimulation of the immune response in the body. However, owing to the compromised immune system of patients with lymphoma, fever symptoms tended to persist compared to that in patients with lung cancer.

Decreased peripheral blood lymphocyte levels, an interval between lymphoma diagnosis and SARS-CoV-2 infection of <24 months, and an interval of anti-CD20 monoclonal antibody treatment of <2 months were high-risk factors for prolonged clearance time among patients with lymphoma with SARS-CoV-2 infection. The decrease in peripheral blood leukocyte and lymphocyte levels prevents patients with lymphoma from rapidly clearing the new coronavirus infection, similar to the general population, through CD4/CD8 and other mechanisms. As a result, patients show continuous viral carriage, leading to positive results on throat swab PCR nucleic acid testing.

In this study, we included a new indicator, namely, the time between the first definite diagnosis of lymphoma and the time of the first confirmed SARS-CoV-2 infection via nucleic acid testing. We classified this time interval according to a threshold of 24 months. Among the patients in this study, 15 and 25 patients with lymphoma showed intervals of ≥24 months and <24 months, respectively. The results of the multivariate regression analysis showed that an interval

10.3389/fmolb.2023.1240175

TABLE 1 Univariate analysis of prolonged viral clearance time.

Indicators	Lymphoma group (n = 40)	Lung cancer group (n = 47)	$t/X^2/U$ value	p Value
Gender [cases (%)]				
Male	26 (65.0)	40 (85.1)	0.101	0.751
Female	14 (35.0)	7 (14.9)		
Age [cases (%)]				
≥65 years	19 (47.5)	32 (68.1)	3.775	0.052
<65 years old	21 (52.5)	15 (31.9)		
Underlying disease [cases (%)]				
Hypertension	11 (27.5)	9 (19.1)	0.851	0.356
Diabetes	7 (17.5)	9 (19.1)	0.039	0.843
Smoking history [cases (%)]	9 (22.5)	15 (31.9)	0.959	0.327
Laboratory indicators				
WBC [×109/L, M (Q1, Q3)]	4.1 (2.4, 5.4)	6.14 (4.1, 7.9)	2.900	0.004
NEUT%[×109/L, s]	62.820.3	70.615.1	2.037	0.045
LYMPH%[×109/L, M(Q1, Q3)]	13.0 (7.9, 19.0)	20.5 (8.7, 27.0)	2.014	0.044
ALB[g/L, M(Q1, Q3)]	35.6 (32.1, 40.7)	35.9 (31.7, 41.5)	0.797	0.426
LDH[U/L, M(Q1, Q3)]	252.5 (189.0, 313.8)	206.0 (178.0, 269.0)	2.099	0.036
CRP[mg/L, M(Q1, Q3)]	23.6 (6.7, 48.1)	24.6 (5.9, 46.1)	0.130	0.896
COVID-19 severity				
Mild and moderate	37 (92.5)	40 (85.1)	0.548	0.459
Severe	3 (7.5)	7 (14.9)		
Symptoms [cases (%) (%)[[Example(%)[例](%)]				
Fever (≥10days)	20 (50)	12 (25.5)	5.564	0.018
Cough	19 (47.5)	16 (55.3)	0.232	0.675
Treatment[cases (%)]				
Anti-CD20	25 (62.5)	0	—	—
Chemotherapy	38 (95)	45 (95.7)	0.027	0.869

TABLE 2 Cox regression analysis of independent risk factors for prolongation of virus clearance time in patients with lymphoma.

Indicator	HR value	95% confidence interval (CI)	p Value
WBC	1.044	0.541~2.012	0.898
NEUT%	1.930	0.954~3.904	0.067
LYMPH%	0.380	0.179~0.808	0.012
Diagnosis interval (<24 months)	0.182	0.062~0.535	0.002
Anti-CD20 treatment interval (<2 months)	0.101	0.029~0.358	<0.001
Aggressive tumor	1.121	0.897~3.421	0.125
Fever (≥10 days)	1.754	0.944~3.259	0.076

of <24 months between the diagnoses of lymphoma and COVID-19 was an independent risk factor for prolonged viral clearance. Generally, it is believed that the earlier the diagnosis of the underlying disease and the longer the course of the disease, the worse the patient's physical condition, leading to a longer course of COVID-19 and an extended time to negative viral

status. Interestingly, in the present study, the interval between lymphoma diagnosis and SARS-CoV-2 infection was negatively correlated with the viral clearance time. In other words, the earlier a patient was diagnosed with lymphoma, the shorter the viral clearance time after diagnosis with SARS-CoV-2 infection. One study also reported a linear decrease in the time for viral negativity via PCR with increasing time since lymphoma diagnosis. This may be because although patients have long-term lymphoma without disease progression or death, the disease is considered stable, and their immune status tends to be stable and closer to that of healthy individuals. Therefore, the viral clearance time was not prolonged.

We also included another research indicator, namely, the correlation between active antitumor chemotherapy within the past 2 months in patients with lymphoma and the viral clearance time. In this study, all 40 patients received lymphoma treatment, 25 of whom underwent treatment with the anti-CD20 regimen. We further divided these 25 patients into subgroups. Among them, 18 patients received anti-CD20 monoclonal antibody treatment within 2 months of confirmed SARS-CoV-2 infection, while seven patients had completed their last dose of anti-CD20 monoclonal antibody treatment >2 months previously. The results of the multivariate regression analysis revealed a negative between correlation anti-CD20 monoclonal antibody treatment >2 months before infection and the viral clearance time, demonstrating a prolonged clearance time in patients who had recently undergone chemotherapy. Rogado et al. (Rogado et al., 2020) reported that most patients with cancer have active disease or are undergoing aggressive treatment, which may result in higher mortality rates and may be related to cellular and humoral immune levels. Therefore, in clinical treatment, assessing the patient's medical history and determining the time from the initial diagnosis and the most recent use of anti-CD20 monoclonal antibody treatment in patients infected with SARS-CoV-2 can guide the continuation or postponement of the next course of treatment. A risk-benefit assessment must be conducted for patients with cancer. If the benefits outweigh the risks, cancer treatment should continue (Linehan et al., 2022).

In clinical practice, aggressive tumors are generally associated with poor survival and prognosis. However, in this study, we confirmed no significant association between tumor invasiveness or noninvasiveness and the SARS-CoV-2 clearance time.

Our study has some limitations, including the small number of hospitalized patients with lymphoma with SARS-CoV-2 infection in our hospital. Additionally, some data were missing owing to the lack of CD+4/CD+8 examinations during the hospitalization of patients with lymphoma. Furthermore, because of the limited testing criteria at the time, patients with mild symptoms or asymptomatic individuals may not have been tested for COVID-19, and we only included patients who were active or planning to receive treatment. Moreover, while collecting data from the electronic medical records, we observed an increase in pulmonary fungal infections among some patients, which warrants further investigation in future studies. Future studies with larger numbers of patients and with additional data on CD+4/CD+8 levels and in patients with different levels of symptoms are also needed to verify our findings.

Data availability statement

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

Ethics statement

The studies involving humans were approved by the Ethics Committee of the First Affiliated Hospital of Anhui Medical University approved this study. 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. 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

All authors contributed to the study conception and design. YL, CW, and LF contributed to the conception of the study. YL, CW, and QX collected the clinical data. CW and YL performed the data analyses. YL, CW, LF, QX, XS, BC, and GS contributed significantly in writing 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/fmolb.2023.1240175/full#supplementary-material

References

Ballow, M., and Haga, C. L. (2021). Why do some people develop serious covid-19 disease after infection, while others only exhibit mild symptoms? *J. allergy Clin. Immunol. Pract.* 9 (4), 1442–1448. doi:10.1016/j.jaip.2021.01.012

Batah, S. S., and Fabro, A. T. (2021). Pulmonary pathology of ards in covid-19: A pathological review for clinicians. *Respir. Med.* 176, 106239. doi:10.1016/j.rmed.2020.106239

Gopalaswamy, R., and Subbian, S. (2021). Corticosteroids for covid-19 therapy: potential implications on tuberculosis. *Int. J. Mol. Sci.* 22 (7), 3773. doi:10.3390/ijms22073773

Heo, J. Y., Park, C., Keam, B., Ock, C. Y., Kim, M., Kim, T. M., et al. (2019). The efficacy of immune checkpoint inhibitors in anaplastic lymphoma kinase-positive non-small cell lung cancer. *Thorac. cancer* 10 (11), 2117–2123. doi:10.1111/1759-7714.13195

Jun, T., Mathew, D., Sharma, N., Nirenberg, S., Huang, H. H., Kovatch, P., et al. (2022). Multiethnic investigation of risk and immune determinants of covid-19 outcomes. *Front. Cell. Infect. Microbiol.* 12, 933190. doi:10.3389/fcimb.2022.933190

Kumar, S. K., Vij, R., Noga, S. J., Berg, D., Brent, L., Dollar, L., et al. (2017). Treating multiple myeloma patients with oral therapies. *Clin. lymphoma, myeloma and leukemia* 17 (5), 243–251. doi:10.1016/j.clml.2017.02.024

Lee, C. Y., Shah, M. K., Hoyos, D., Solovyov, A., Douglas, M., Taur, Y., et al. (2022). Prolonged sars-cov-2 infection in patients with lymphoid malignancies. *Cancer Discov.* 12 (1), 62–73. doi:10.1158/2159-8290.Cd-21-1033

Li, T. Z., Cao, Z. H., Chen, Y., Cai, M. T., Zhang, L. Y., Xu, H., et al. (2021). Duration of sars-cov-2 rna shedding and factors associated with prolonged viral shedding in patients with covid-19. *J. Med. virology* 93 (1), 506–512. doi:10.1002/jmv.26280

Linehan, A., Fitzpatrick, O., Cowzer, D., Hennessy, M. A., Coyne, Z. L., Nolan, A., et al. (2022). Covid-19-Related mortality in cancer patients in an Irish setting. *Ir. J. Med. Sci.* 191 (5), 2013–2018. doi:10.1007/s11845-021-02815-9

Moyon, Q., Sterlin, D., Miyara, M., Anna, F., Mathian, A., Lhote, R., et al. (2022). Bnt162b2 vaccine-induced humoral and cellular responses against sars-cov-2 variants in systemic lupus erythematosus. *Ann. rheumatic Dis.* 81 (4), 575–583. doi:10.1136/annrheumdis-2021-221097

Rogado, J., Pangua, C., Serrano-Montero, G., Obispo, B., Marino, A. M., Pérez-Pérez, M., et al. (2020). Covid-19 and lung cancer: A greater fatality rate? *Lung cancer (Amsterdam, Neth.* 146, 19–22. doi:10.1016/j.lungcan.2020.05.034

Visco, C., Marcheselli, L., Mina, R., Sassone, M., Guidetti, A., Penna, D., et al. (2022). A prognostic model for patients with lymphoma and covid-19: A multicentre cohort study. *Blood Adv.* 6 (1), 327–338. doi:10.1182/bloodadvances.2021005691

Wang, J., Fu, J., Shen, Q., Zhang, F., Wang, Y., and Wu, L. L. (2019). Identification and diagnostic value of pleural fluid periostin and serum periostin of malignant pleural effusions in patients with non-small-cell lung cancer. *J. Clin. laboratory analysis* 33 (7), e22943. doi:10.1002/jcla.22943

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 16 October 2023
DOI 10.3389/fmolb.2023.1265359

Check for updates

OPEN ACCESS

EDITED BY
Qingyu Luo,
Dana–Farber Cancer Institute,
United States

REVIEWED BY
Li Chen,
Chinese Academy of Medical Sciences and Peking Union Medical College, China
Jeonghee Cho,
Dankook University, Republic of Korea

*CORRESPONDENCE
Hong Sook Kim,
hong2kim2@gmail.com

RECEIVED 22 July 2023
ACCEPTED 02 October 2023
PUBLISHED 16 October 2023

CITATION
Fan D, Yang M, Lee HJ, Lee JH and Kim HS (2023), AVEN: a novel oncogenic biomarker with prognostic significance and implications of AVEN-associated immunophenotypes in lung adenocarcinoma.
Front. Mol. Biosci. 10:1265359.
doi: 10.3389/fmolb.2023.1265359

COPYRIGHT
© 2023 Fan, Yang, Lee, Lee and Kim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

AVEN: a novel oncogenic biomarker with prognostic significance and implications of AVEN-associated immunophenotypes in lung adenocarcinoma

Dengxia Fan, Moses Yang, Hye Jung Lee, Jeong Hee Lee and Hong Sook Kim*

Department of Biological Sciences, Sungkyunkwan University, Suwon, Republic of Korea

Introduction: AVEN, an apoptosis and caspase activation inhibitor, has been associated with adverse clinical outcomes and poor prognosis in Acute myeloid leukemia (AML). Targeting AVEN in AML improves apoptosis sensitivity and chemotherapy efficacy, making it a promising therapeutic target. However, AVEN's role has not been studied in solid tumors. Therefore, our study investigated AVEN as a prognostic biomarker in a more comprehensive manner and developed an AVEN-derived prognostic model in Lung adenocarcinoma (LUAD).

Method: Pan-cancer analysis was performed to examine AVEN expression in 33 cancer types obtained from the TCGA database. GEPIA analysis was used to determine the predictive value of AVEN in each cancer type with cancer-specific AVEN expression. Lung Adenocarcinomas (LUAD) patients were grouped into $AVEN^{high}$ and $AVEN^{low}$ based on AVEN expression level. Differentially expressed genes (DEGs) and pathway enrichment analysis were performed to gain insight into the biological function of AVEN in LUAD. In addition, several deconvolution tools, including Timer, CIBERSORT, EPIC, xCell, Quanti-seq and MCP-counter were used to explore immune infiltration. AVEN-relevant prognostic genes were identified by Random Survival Forest analysis via univariate Cox regression. The AVEN-derived genomic model was established using a multivariate-Cox regression model and GEO datasets (GSE31210, GSE50081) were used to validate its prognostic effect.

Results: AVEN expression was increased in several cancer types compared to normal tissue, but its impact on survival was only significant in LUAD in the TCGA cohort. High AVEN expression was significantly correlated with tumor progression and shorter life span in LUAD patients. Pathway analysis was performed with 838 genes associated with AVEN expression and several oncogenic pathways were altered such as the Cell cycle, VEGFA-VEGFR2 pathway, and epithelial-mesenchymal-transition pathway. Immune infiltration was also analyzed, and less infiltrated B cells was observed in $AVEN^{high}$ patients. Furthermore, an AVEN-derived genomic model was established, demonstrating a reliable and improved prognostic value in TCGA and GEO databases.

Conclusion: This study provided evidence that AVEN is accumulated in LUAD compared to adjacent tissue and is associated with poor survival, high tumor

progression, and immune infiltration alteration. Moreover, the study introduced the AVEN-derived prognostic model as a promising prognosis tool for LUAD.

KEYWORDS

AVEN, lung adenocarcinoma, immune infiltration, prognostic biomarker, prognostic model

Introduction

LUAD (Lung Adenocarcinoma) is a subtype of non-small cell lung cancer (NSCLC) that arises from the glandular tissue of the lungs. It is the most common type of lung cancer, accounting for approximately 40% of all cases of NSCLC (Gelatti et al., 2019). In recent years, there have been significant advances in diagnosing and treating LUAD. For example, targeted therapies have been developed to specifically target the genetic alterations that drive the growth of cancer cells in individual patients (Chan and Hughes, 2015; Skoulidis and Heymach, 2019). Immunotherapy has also emerged as a promising treatment option for LUAD. Despite the advances in diagnostic and therapeutic methods implicated in clinical studies, these treatments have been shown to benefit a limited pool of patients. Thus, it is essential and urgent to find the potential and valuable biomarkers for diagnosis, prognosis, and targets for therapy in cancers.

AVEN (Apoptosis, caspase activation inhibitor) is a protein that plays a crucial role in inhibiting apoptosis and promoting cell survival. It binds to anti-apoptotic Bcl-2 family member, B-cell lymphoma-extra-large (Bcl-xL) specifically that retain anti-apoptotic activity. It also interacts with caspase regulator, apoptotic protease activating factor 1 (Apaf-1) (Chau et al., 2000) and prevents Apaf-1 mediated caspases activation (Chau et al., 2000). *In vivo* experiments showed that AVEN knockdown reduced tumor growth and in turn increased apoptosis of hematopoietic neoplasms (Eißmann et al., 2013). In clinical studies, it is reported that AVEN is overexpressed in acute lymphoblastic leukemias/lymphoma patients and associated with poor prognosis (Melzer et al., 2012; Eißmann et al., 2013). Indeed, AVEN expression is significantly higher in recurrent patients (Choi et al., 2006). With previous findings being limited to cancer of blood and bone, the correlation of AVEN expression with prognosis and immune infiltration in different cancers remain unclear.

We first screened the oncogenic role of AVEN in pan-cancer and found that AVEN is highly expressed in Colon adenoma (COAD), Kidney renal clear cell carcinoma (KIRC), Kidney renal papillary cell carcinoma (KIRP), Lung adenocarcinoma (LUAD), Lung squamous cell carcinoma (LUSC), and Thyroid carcinoma (THCA) compared to normal tissue. Interestingly, AVEN overexpression was associated with poor survival only in LUAD patients. Consistently, high tumor progression and reduced levels of B cell infiltration was observed in AVEN overexpressing LUAD patients. AVEN-associated genes and pathways were studied to gain valuable insight of the characteristics and function of AVEN. Furthermore, we developed an AVEN-derived prognostic model in an attempt to provide a promising prognosis tool for LUAD.

Materials and methods

Pan-cancer analysis and TCGA data processing

The AVEN mRNA expression in pan-cancer was analyzed by the GSCA web tool (http://bioinfo.life.hust.edu.cn/GSCA). GEPIA (Li et al., 2021) (http://gepia.cancer-pku.cn/) web tool was applied for the survival analysis in COAD, KIRC, KIRP, LUAD, LUSC and THCA. In order to conduct a more detailed investigation into the role of AVEN in LUAD, RNA-seq data and clinical data in LUAD patients derived from TCGA database were obtained from the UCSC Xena website (http://xena.ucsc.edu/). Log2(FPKM+1) value obtained from RNA-seq data was converted to TPM (Transcripts Per Million) value. Subsequently, patients with the highest 25% of AVEN expression were categorized into $AVEN^{high}$ group, and patients with the lowest 25% of AVEN expression were classified into the $AVEN^{low}$ groups. Overall survival analysis was performed in $AVEN^{high}$ and $AVEN^{low}$ groups.

LUAD patient characteristic analysis

To compare the characteristics between $AVEN^{high}$ and $AVEN^{low}$ patients, clinical data was downloaded from the UCSC Xena website, including age, TNM classification, gender, radiation therapy status, race, AVEN expression, and smoking status. The R package moonBook was exploited to visualize characteristics of patients between $AVEN^{high}$ and $AVEN^{low}$ groups.

Driver genes alteration analysis

The whole exome data of LUAD patients was downloaded from the cBioportal (Cerami et al., 2012; Gao et al., 2013) to examine genetic alterations (https://www.cbioportal.org/). Driver genes such as TP53, EGFR, KRAS, ERBB2, BRAF, ALK, RET, FGFR3, NTRK3 and ROS1 were selected, and their alteration pattern was examined in $AVEN^{high}$ and $AVEN^{low}$ LUAD patients. The R package ComplexHeatmap was used to generate an oncoprint plot.

Identification of DEGs and functional enrichment analysis

Spearman's rank correlation test, which works with rank-order variables instead of raw data value of the variables, was used to obtain differentially expressed genes (DEGs) between $AVEN^{high}$ and $AVEN^{low}$ group. Genes with an absolute R-value>0.4 were used for pathway

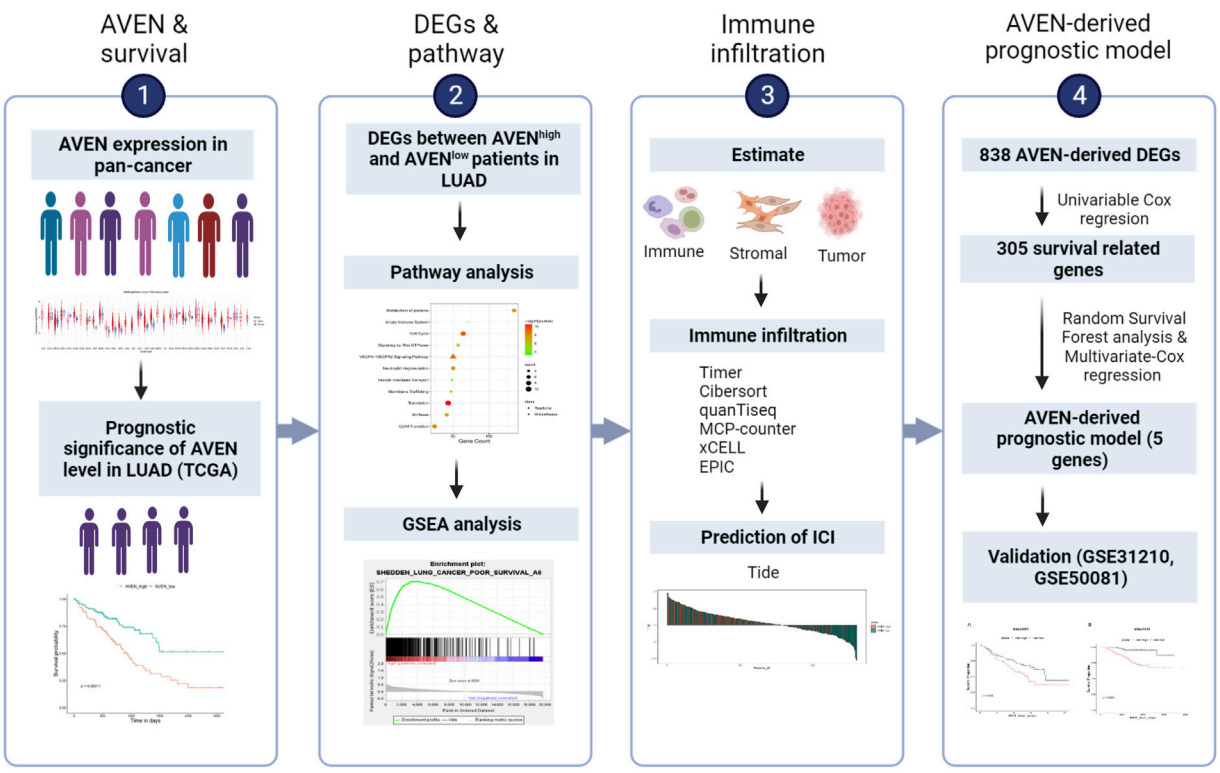


FIGURE 1
Brief overview of this study. LUAD: Lung adenocarcinoma. DEGs: Differentially expressed genes. TIDE: Tumor Immune Dysfunction and Exclusion. ICI: Immune checkpoint inhibitor.

analysis by ConsensusPathDB (Kamburov et al., 2009) (http://cpdb.molgen.mpg.de/MCPDB). Significantly altered pathways were selected with the criteria of $p < 0.05$ and were visualized using SRplot (https://www.bioinformatics.com.cn/srplot). Genes in Cell cycle and VEGFA-VEGFR2 pathways were further visualized by using the R package ComplexHeatmap. GSEA analysis was performed to compare the pathway enrichment between AVEN^{high} and AVEN^{low} groups by using these gene sets as references: "SHEDDEN_LUNG_CANCER_POOR_SURVIVAL_A6″, "HallMARK_MTORC1_SIGNALINF", "HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION", and "VEGF_A_UP.V1_DN".

Immune infiltration analysis

Immune score, stromal score, and estimate score representing immune infiltration, stromal cell level, and purity of tumor, respectively were obtained from Estimate website (https://bioinformatics.mdanderson.org/estimate/index.html) and compared between AVEN^{high} and AVEN^{low} patients. Furthermore, multiple deconvolution tools including Timer, CIBERSORT, EPIC, xCEll, Quanti-seq and MCP-counter (Li et al., 2017; Racle et al., 2017; Sturm et al., 2019) were utilized to examine various types of immune cells in tumor tissue. Immune infiltration data was obtained from Timer (https://cistrome.shinyapps.io/timer/) and was visualized by boxplot, using the R package via ggplot2.

Prediction of immunotherapy response

To evaluate the prediction value of AVEN in immunotherapy response, Tumor Immune Dysfunction and Exclusion (TIDE) score (Jiang et al., 2018) was calculated (http://tide.dfci.harvard.edu/). Consequently, the expression levels of immune checkpoint genes and functional genes associated with cytotoxic T cells were analyzed in AVEN^{high} and AVEN^{low} LUAD. Furthermore, we extended our investigation to encompass additional immunotherapy response markers indicative of B cells, testing these markers in AVEN^{high} and AVEN^{low} LUAD patients.

Establishment of an AVEN-derived prognostic genes model

To determine the correlation between each gene from the DEGs and the overall survival of LUAD patients, Univariate Cox Regression model was employed. AVEN-derived genes with p-value < 0.01 were regarded as AVEN-derived prognostic factors. Followed by Random Survival Forest analysis, the

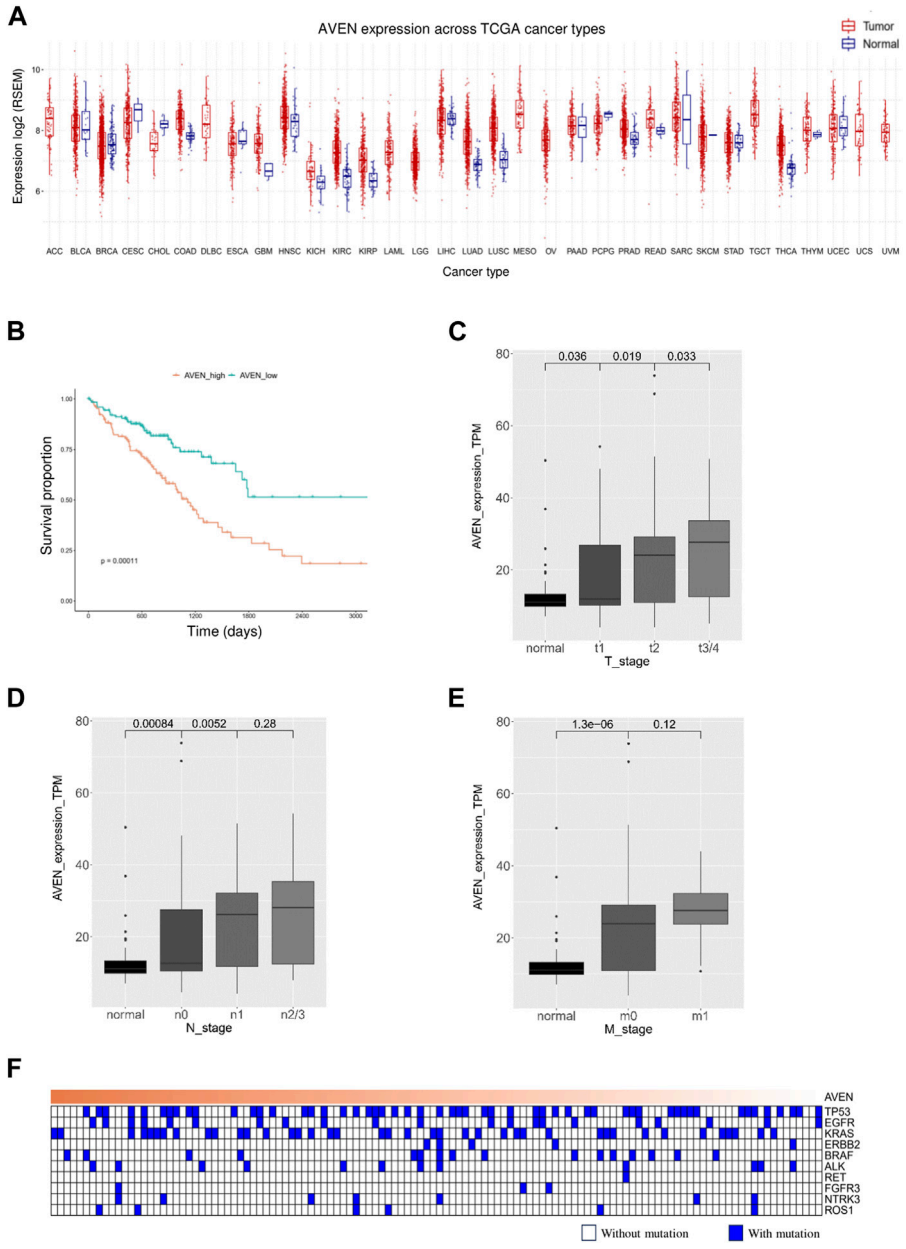


FIGURE 2
A high level of AVEN is associated with poor survival **(A)** mRNA level of AVEN in thirty-three different types of cancer and normal tissues. **(B)** Survival analysis in AVENhigh and AVENlow LUAD patients derived from TCGA database. **(C)** AVEN mRNA expression level in different T stages in LUAD and normal tissue. **(D)** AVEN mRNA expression level in different N stages in LUAD and normal tissue. **(E)** AVEN mRNA expression level in different M stages in LUAD and normal tissue. **(F)** Mutation landscape of driver genes associated with AVEN expression.

relative importance of each gene was calculated. Genes with relative importance >0.5 were used in the Multivariate Cox Regression model. Step forward Cox regression was utilized to optimize the model. The AVEN-derived genomic model was formulated as follows (Abd ElHafeez et al., 2021):

$$\begin{aligned} \textit{Risk score} = h0(t) \times \exp(&\textit{KRT6A} \times 0.0002919 \\ &+ \textit{SLC16A3} \times 0.0045 + \textit{CTSL} \times 0.0008009 \\ &+ \textit{LDHA} \times 0.007940 + \textit{CDC42EP2} \times 0.0147) \end{aligned}$$

The h0(t) represents the baseline hazard at time t, which denotes the hazard of an individual when all predictor variables are set to 0. Subsequently, based on the calculated risk scores, patients were categorized into high-risk and low-risk groups using the mean risk score. The survival and survminer packages were utilized to determine the survival state of LUAD patients in both the TCGA and GEO cohorts (GSE50081, GSE31210). ROC curves were generated to test the specificity and sensitivity of the AVEN-derived prognostic gene model by using the survivalROC package.

TABLE 1 Characteristics of AVEN^{high} and AVEN^{low} patients in LUAD.

	AVEN^{high} (N = 130)	AVEN^{low} (N = 130)	*p*
Age (years)	65.1 ± 9.7	64.5 ± 10.5	0.624
T stage			0.03
t1	31 (23.1%)	55 (41.9.0%)	
t2	75 (57.2%)	66 (50.3%)	
t3	15 (11.5%)	8 (6.1%)	
t4	9 (6.9%)	1 (0.8%)	
uncharacterized	1 (0.8%)	1 (0.8%)	
N stage			0.001
n0	67 (51.1%)	93 (71.5%)	
n1	33 (25.2%)	17 (13.1%)	
n2	27 (20.6%)	11 (8.5%)	
n3	0 (0.0%)	1 (0.8%)	
uncharacterized	4 (3.1%)	8 (6.2%)	
M stage			0.061
m0	92 (70.8%)	84 (65.6%)	
m1	11 (8.4%)	5 (3.8%)	
uncharacterized	27 (20.8%)	39 (30.5%)	
Gender			0.047
female	62 (47.3%)	79 (60.3%)	
male	69 (52.7%)	52 (39.7%)	
Radiation therapy			0.731
no	98 (83.8%)	106 (86.2%)	
yes	19 (16.2%)	17 (13.8%)	
Race			0.72
Asian	3 (2.7%)	3 (2.6%)	
black or African American	9 (8.0%)	13 (11.2%)	
white	100 (89.3%)	100 (86.2%)	
AVEN expression (TPM)	31.9 ± 8.5	10.5 ± 2.0	0
Smoke			0.788
no	90 (68.7%)	93 (71.0%)	
yes	41 (31.3%)	38 (29.0%)	

Web-based bioinformatic analysis

PrognoScan (Mizuno et al., 2009) was employed to assess the prognostic significance of AVEN across multiple LUAD data cohorts. Additionally, AVEN protein abundance was examined using the cProSite (Wang et al., 2023).

Statistical analysis

Statistical analysis was performed with R software (v4.2.1) and its suitable packages. In this study, group comparisons were performed using Student's t-test, and the interaction between variables were examined using the Spearman correlation test.

Result

A high level of AVEN is associated with poor survival in LUAD

The role of AVEN in cancer is systematically studied according to the workflow shown in Figure 1. First, the GSCA web tool was exploited (Liu et al., 2023) to identify the AVEN mRNA level in various cancer types. AVEN was highly expressed in six types of cancer compared to normal tissue: colon adenoma (COAD), Kidney renal clear cell carcinoma (KIRC), Kidney renal papillary cell carcinoma (KIRP), Lung adenocarcinoma (LUAD), Lung squamous cell carcinoma (LUSC), and Thyroid carcinoma (THCA) (Figure 2A). On the other hand, a slight downregulation of AVEN was observed in two types of cancer, Cholangiocarcinoma (CHOL), and Cervical squamous cell carcinoma (CESC). Besides mRNA, AVEN protein level was analyzed using the cProSite database, which showed a higher AVEN protein abundance in Breast cancer, Colon cancer, Kidney cancer, Liver cancer, Lung adenocarcinoma, Lung squamous cell carcinoma, and ovarian cancer compared to its adjacent normal tissue (Supplementary Figure S1). Given our interest in the oncogenic role of AVEN, we performed GEPIA analysis to determine the overall survival in the six tumor types with elevated AVEN mRNA expression: COAD, KIRC, KIRP, LUAD, LUSC and THCA (Supplementary Figure S2). Notably, AVEN showed a significant prognostic effect only in LUAD. To further validate this finding and investigate the oncogenic features of AVEN, we acquired the mRNA sequencing data and clinical information of LUAD patients from the TCGA dataset. Based on the expression level of AVEN, LUAD patients were classified into two groups: $AVEN^{high}$ (top 25%) and $AVEN^{low}$ (bottom 25%), and overall survival was investigated using Kaplan-Meier analysis (Figure 2B). Consistent with GEPIA results, high AVEN expression was markedly associated with poor survival in LUAD. Before studying details of underlying molecular and cellular mechanism of AVEN, patient characteristics of $AVEN^{high}$ and $AVEN^{low}$ gruop were analyzed (Table 1). As shown in Table 1, AVEN expression did not show significant differences in age, race, and smoking but only in gender. Interestingly, the value of the T stage and N stage showed significant differences between the two groups. TNM classification is a system that defines tumor size (T), regional lymph amount (N), and spread of cancer (M) in a patient's body (Rosen and Sapra, 2022). Thus, we wondered whether AVEN contributes to tumor progression according to TNM classification. Notably, AVEN showed a gradual increase with tumor progression when AVEN expression was seen in normal tissue and different T stages (Figure 2C). AVEN expression also increased in the N1 stage compared to N0 (Figure 2D). We further analyzed AVEN expression in different M stages, however, there were no significant differences in the number of patients in different M stages between $AVEN^{high}$ and $AVEN^{low}$ patients (Table 1). Consistently, AVEN expression was similar between M0 and M1 stages (Figure 2E). Since genetic alterations can lead to the activation of various signaling pathways that promote the growth and survival of cancer cells (Skoulidis and Heymach, 2019), we next examined genetic alterations in driver genes that were well defined in LUAD. However, no significant associations between AVEN expression and genetic mutations in EGFR, KRAS, ALK, ROS1, BRAF, *etc.*, Were observed (Figure 2F).

Functional signaling pathways associated with AVEN

We next investigated molecular and cellular pathways associated with AVEN. First, DEGs between $AVEN^{high}$ and $AVEN^{low}$ groups were analyzed by using Spearman's rank correlation test with the criteria of the absolute R-value over 0.4. A total of 838 genes were obtained (Supplementary Table S1), followed by pathway enrichment analysis via the ConsensusPathDB website tool. We found that those DEGs were mainly involved in the biological process pathways (such as metabolism of proteins, and membrane trafficking), oncogenic pathways (such as the cell cycle and VEGFA-VEGFR2 Signaling Pathway), and immune regulation process (such as neutrophil degranulation, and innate immune system) (Figure 3A).

Given the fact that VEGF plays an important role in tumor progression and angiogenesis (Jiang et al., 2020), we suggested that AVEN might enhance tumor aggressiveness by promoting the cell cycle and angiogenesis. To better demonstrate a correlation between AVEN and the gene expression profile in the cell cycle or VEGFA-VEGFR2 Signaling Pathway, heatmaps were introduced for visualization. Genes in these pathways were distinctly upregulated by AVEN overexpression (Figure 3B). Gene set enriched analysis (GSEA) was applied to further investigate the oncogenic role of AVEN. Genes in MTORC1 signaling, epithelial-mesenchymal-transition (EMT), and VEGF pathways were remarkably enriched in $AVEN^{high}$ patients (Figure 3C). MTOR pathway is a key pathway to regulate cell growth. EMT and VEGFA also play a critical role in tumor progression (Brabletz et al., 2018; Zou et al., 2020). Furthermore, we found that AVEN overexpression was associated with a poor survival genomic signature (Figure 3D) in NSCLC (NES = 3.43, p-value = 0.0), consistent with results shown in Figure 2B.

AVEN-associated immune infiltration landscape in LUAD

The tumor microenvironment (TME) consists of various cells such as normal epithelial, vascular cells, stromal cells, and immune cells. Immune cells, such as T cells, B cells, natural killer cells, dendritic cells, macrophages, and others are attracted to the site of a tumor by various signals, including chemokines and cytokines produced by tumor cells and stromal cells and regulate tumor cells proliferation/apoptosis (Galli et al., 2020). Considering our finding of AVEN-associated signaling pathways such as innate immune system, neutrophil degranulation, and EMT (Figures 3A, C), it suggested unique TME features associated with AVEN expression.

Thus, we first analyzed immune cells and stromal cells using the ESTIMATE algorithm in $AVEN^{high}$ and $AVEN^{low}$ groups. Despite the fact that there was no significant difference in stromal score, $AVEN^{high}$ patients showed lower immune scores compared to $AVEN^{low}$ patients, which

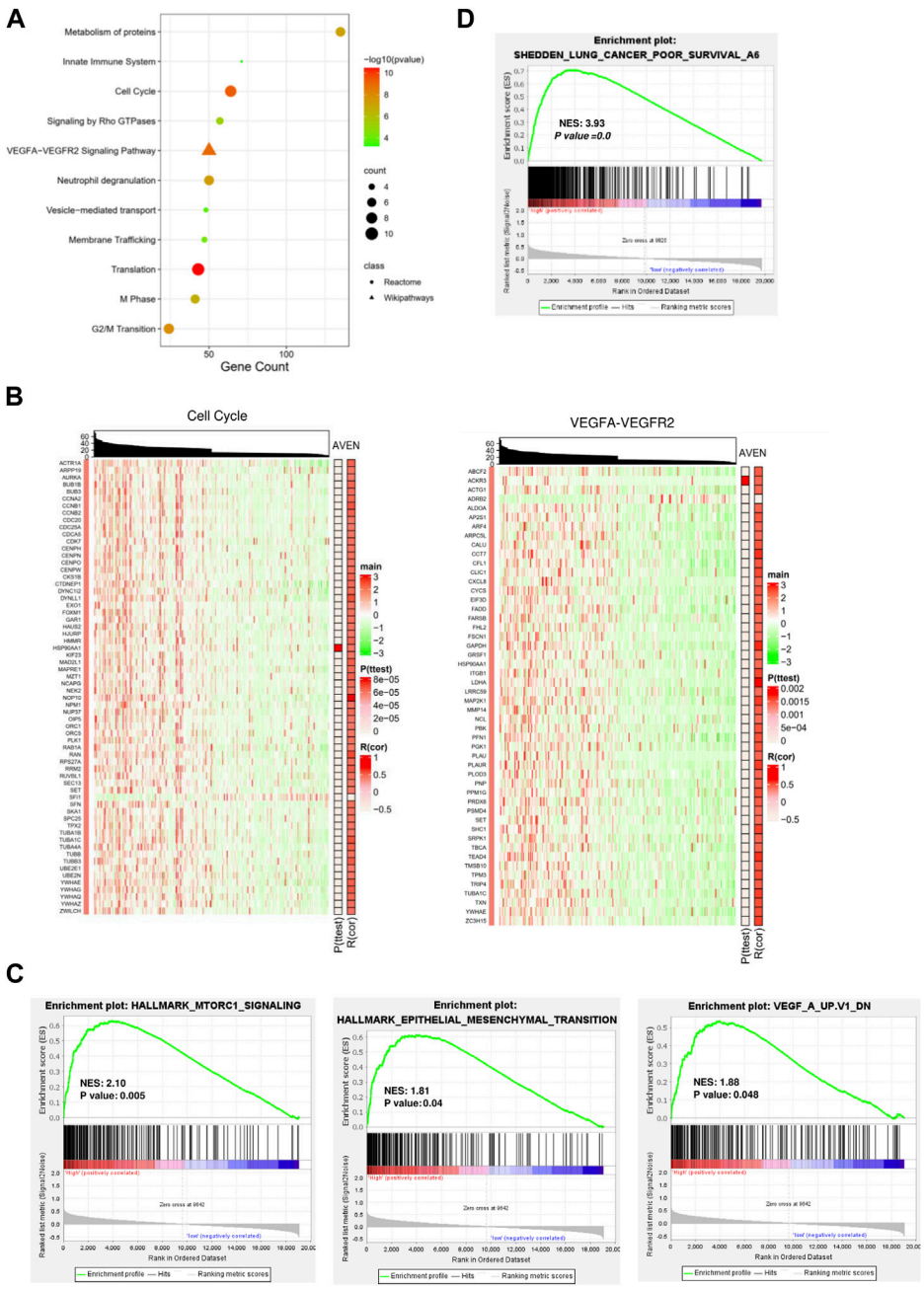


FIGURE 3
Functional signaling pathways associated with AVEN expression. **(A)** Pathway enrichment analysis with DEGs. **(B)** Heatmaps visualized the association of AVEN expression with genes involved in cell cycle and VEGF-VEGFR2 pathways. **(C)** GSEA analysis in AVENhigh and AVENlow groups using MTORC1, Epithelial-mesenchymal-transition (EMT), and VEGF signature. **(D)** GSEA analysis in AVENhigh and AVENlow groups using a lung cancer poor survival signature.

suggests the functional possibility of AVEN to inhibit the abundance of immune cells in the TME. Based on the stromal score and immune score, the Estimate score was calculated to infer the purity of tumor tissue (Yoshihara et al., 2013), and the result showed AVEN^{high} and AVEN^{low} groups shared a similar tumor purity (Figure 4A). To explore the diverse types of immune cells, multiple deconvolution tools were employed, including Timer, CIBERSORT, EPIC, xCEll, Quanti-seq, and MCP-counter (Sturm et al., 2019). These tools utilize different methodologies to analyze the RNA-seq data and provide insights into the composition and abundance of immune cell populations (Im and Kim, 2023). The analysis using various deconvolution tools revealed that B cells were significantly more abundant in AVEN^{low} groups. However, different patterns were observed for other immune cell types depending on the specific tool used. For instance, Timer and EPIC indicated that CD4 T cells were significantly higher in AVEN^{low} groups, while Quanti-Seq suggested higher levels in AVEN^{high} groups. Additionally, CIBERSORT analysis showed that activating memory CD4 T cells were more abundant in AVEN^{high} groups, whereas resting memory CD4 T cells were higher in

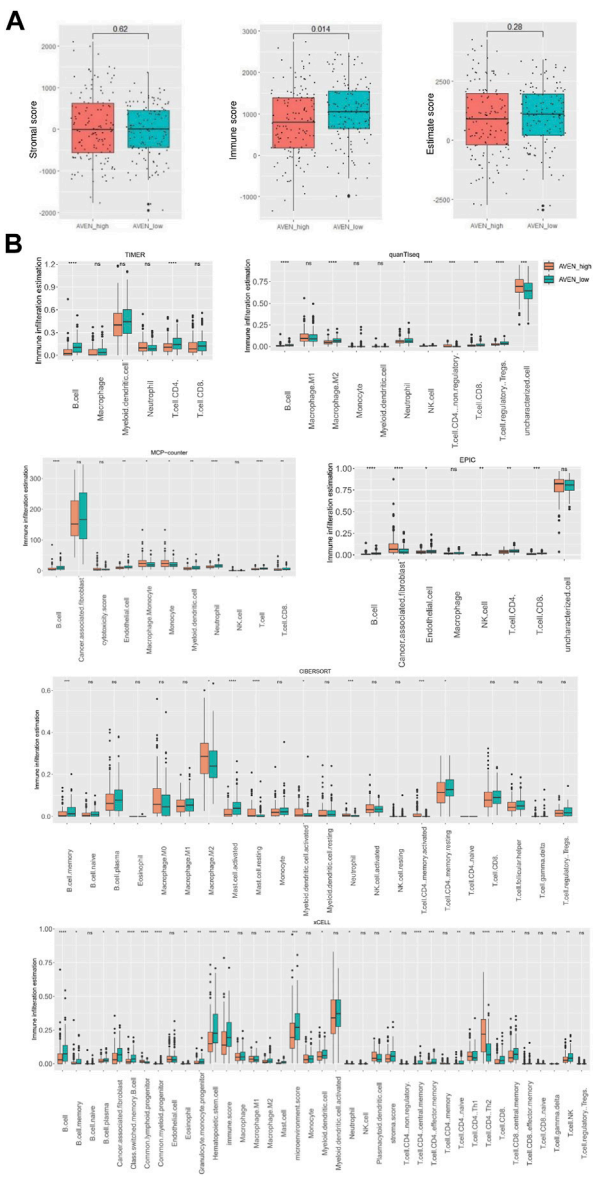


FIGURE 4
Immune infiltration landscape associated with AVEN. **(A)** Stromal score, immune score and estimate score in AVENhigh and AVENlow groups. **(B)** immune infiltration evaluation by diverse deconvolution tools including TIMER, MCP-Counter, EPIC, quanTiseq, xCELL, and CIBERSORT.

AVEN^{low} groups (Figure 4B). EPIC, MCP-counter and xCEll subtracted cancer associated fibroblast (CAF) from tumor. Specifically, EPIC and MCP-counter indicated that CAF were significantly more abundant in AVEN^{high} groups. However, in contrast to these findings, xCEll tool showed lower levels of CAF in AVEN^{high} groups (Figure 4B).

A high level of AVEN expression predicts a poor immunotherapy response

Immune checkpoint inhibitors (ICIs) in cancer treatments have demonstrated significant potential in enhancing antitumor immune responses and improving patient outcomes (Bondhopadhyay et al., 2020; Robert, 2020). However, their responsiveness remains limited, necessitating the identification of predictive biomarkers (Bai et al., 2020; Błach et al., 2021). Therefore, we investigated whether AVEN could serve as a predictive biomarker for ICIs response using the TIDE algorithm, in addition to assessing the expression of immune checkpoint genes. To explore the AVEN-associated immunotherapy response, RNA-seq data from 260 patients in AVEN^{high} and AVEN^{low} groups were processed in TIDE tool. We found that among the 158 non-responders, 62% (n = 98) belonged to the AVEN^{high} group, while 38% (n = 60) were in the AVEN^{low} group. Specifically, among the AVEN^{high} patients (n = 130), only 24% (n = 32) were predicted to be responders based on the TIDE tool (Figure 5A, Supplementary Table S2). Subsequently, expression of immune checkpoint genes was examined. Immune checkpoint genes such as CD274, CTLA4, and LAG3 (Qin et al., 2019; Wang et al., 2019; Hu et al., 2021) had no significant differences between AVEN^{high} and AVEN^{low} groups (Figure 5B). Genes associated with anti-tumor CD8 T cells such as CD8A and GZMA (van der Leun et al., 2020) were also examined, but no differences were observed (Figure 5C). Immune checkpoints and CD8 T cell function are regarded as the essential indicators for immunotherapy respond, however the immunotherapy resistance is far more complicated than that. For example, a lower mutation burden (TMB), dysfunctional MHCs complex, or immunosuppressive tumor microenvironment (Lei et al., 2020) can result in a lower responsiveness in immunotherapy treatment without altering the expression level of immune checkpoint and CD8 T cell markers. Moreover, in our current study, the observation of diminished B cell infiltration (Figure 4B) in AVEN^{high} patients have spurred us to propose additional possibilities that B cell mediates immunotherapy resistance. With the knowledge that B cells also express the receptors of PD1/PDL1/CTLA4 (Kim et al., 2021) and have been established as a favorable prognostic marker in NSCLC (Germain et al., 2014), there exists a theoretical basis for B cells to respond to immune checkpoint inhibitors. We extended our analysis to assess the expression of B cell markers in LUAD. As depicted in Figure 5D, there was a significant reduction in mRNA levels of CD19, CD20, and CD22 in patients with elevated AVEN expression, further suggesting a distinct immune phenotype in AVEN^{high} patients, characterized by reduced B cell infiltration.

Developing an AVEN-derived genomic model for LUAD prognosis

As AVEN displayed significant predictive value in terms of LUAD survival within the TCGA database, we were motivated to investigate its prognostic utility across other cohorts. Employing PrognoScan, we delved into the influence of AVEN expression on different datasets. While a substantial correlation with LUAD survival was observed in some cohorts, it is important to note that certain cohorts exhibited limitations in significant prognostic effect of AVEN (Supplementary Table S3). Since AVEN-derived DEGs showed remarkable effect in pivotal oncogenic pathways, we further investigated the prognosis effect of 838 DEGs by using Univariate Cox Regression (Ma et al., 2022; Yu et al., 2022).

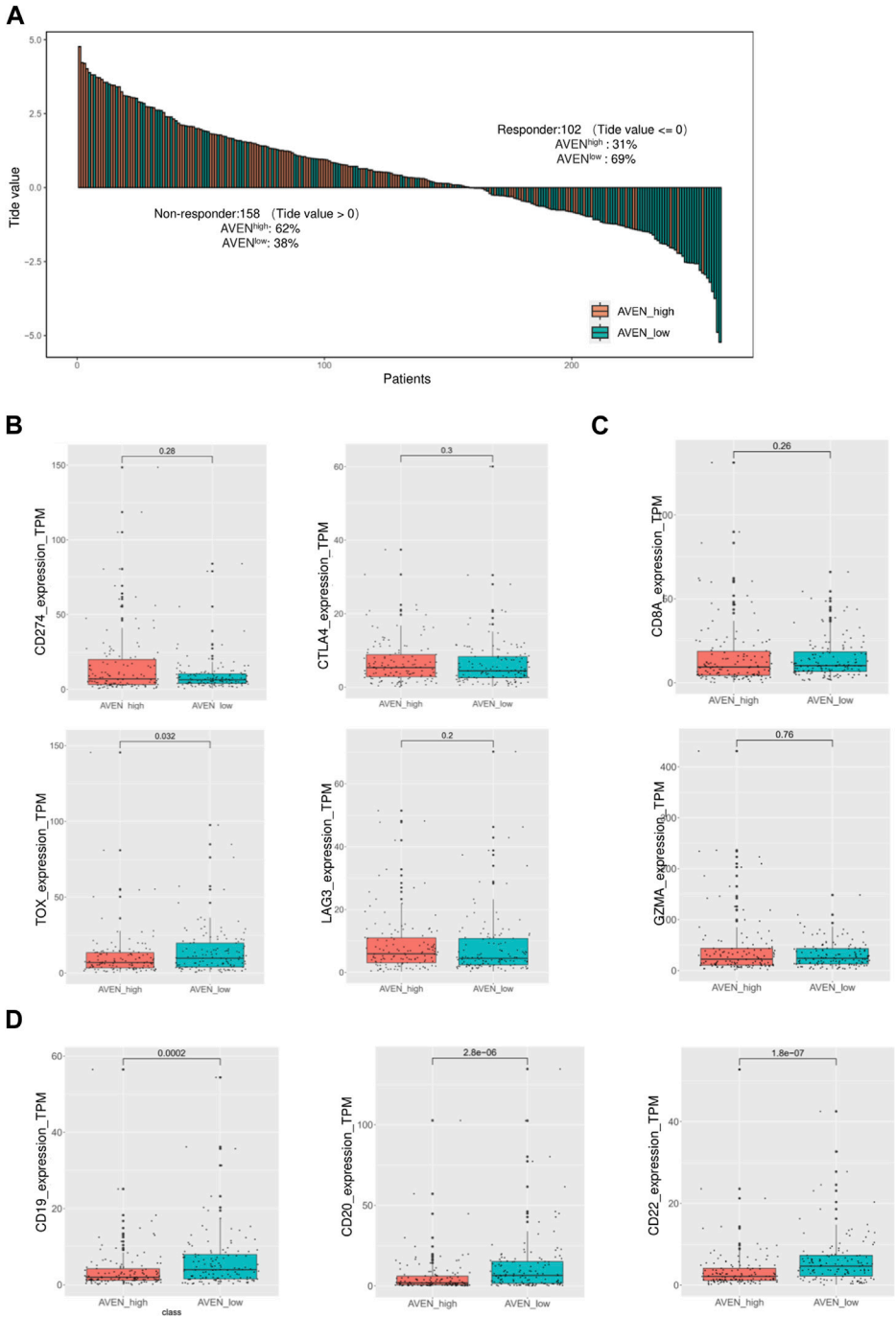


FIGURE 5
Immunotherapy response prediction based on AVEN expression. **(A)** Bar-plot of TIDE score among 260 patients with different AVEN expression levels. **(B)** Expression of immune checkpoint genes in AVENhigh and AVENlow patients. **(C)** Genes expression related to the cytotoxicity activity of tumor-killing T cells in AVENhigh and AVENlow patients. **(D)** mRNA expression of B cell markers was compared between AVEN^{high} and AVEN^{low} patients.

305 of survival-related genes with p-value < 0.01 (Supplementary Table S4) were shown to contribute to LUAD patient survival and determined as prognostic factors, which were then used as input for the Random Survival Forest analysis (Figure 6A). Following the Random Survival Forest analysis, the relative importance of the above prognostic genes was calculated and ranked. Eight genes, KRT6A, SLC16A3, AHNAK2, CTSL, FAM83A, LDHA, CDC42EP2, and SPHK1, were selected with a relative importance value of over 0.5 (Figure 6B), and Multivariate-Cox regression model was developed with those genes. In order to optimize survival prediction model, step-forward Multivariate Cox analysis was used to further screen the valuable prognostic genes containing KRT6A, SLC16A3, CTSL, LDHA, and CDC42EP2. According to the model, the risk score of each patient was identified as follows:

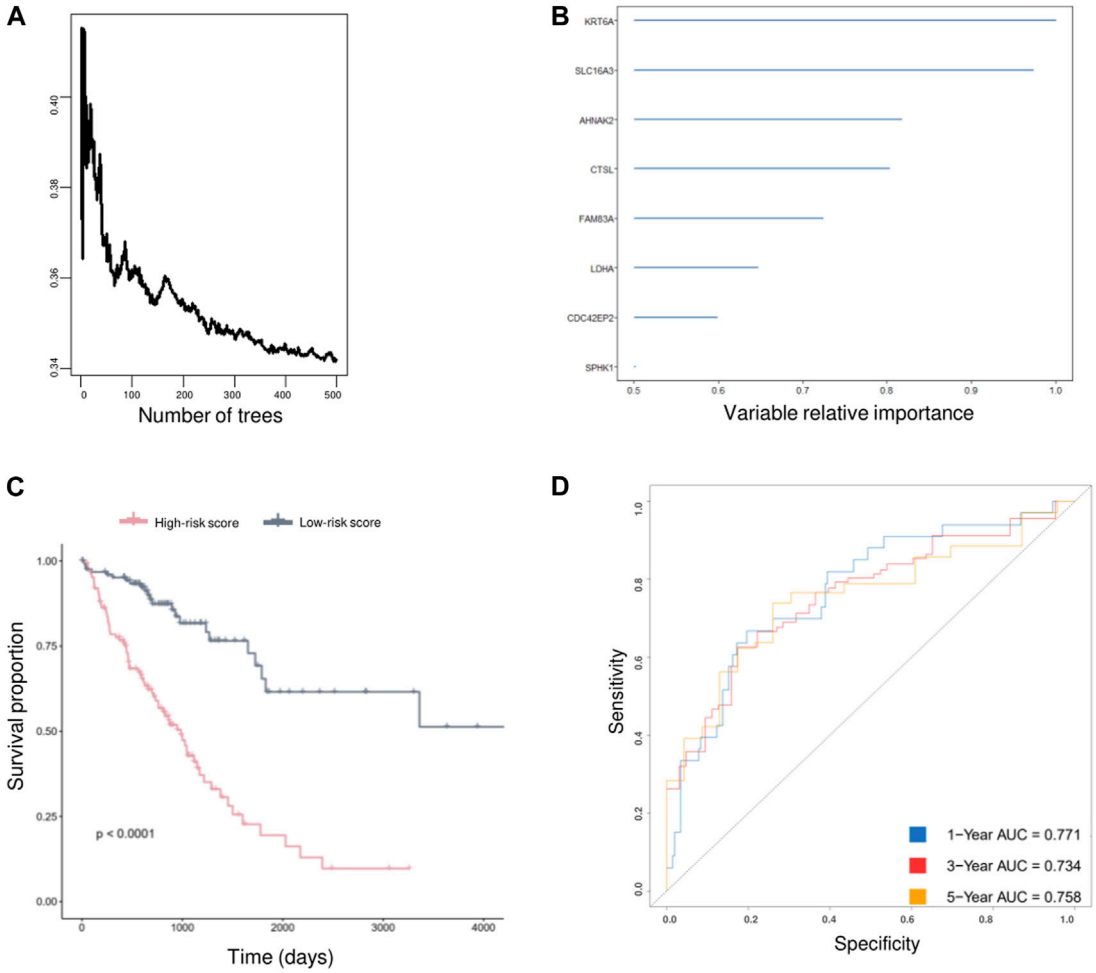


FIGURE 6
AVEN-derived prognostic model in LUAD. **(A)** Random survival forest analysis with DEGs. **(B)** AVEN-derived genes with a relative importance value over 0.5. **(C)** Survival analysis between LUAD patients with high-risk score and low-risk score using TCGA LUAD data. **(D)** Time-dependent ROC curves based on optimized AVEN-derived prognostic model.

$h0(t) \times \exp(KRT6A \times 0.0002919 + SLC16A3 \times 0.0045 + CTSL \times 0.0008009 + LDHA \times 0.007940 + CDC42EP2 \times 0.0147)$. In accordance with the mean risk score, LUAD patients were classified into two subpopulations, high and low risk score. Kaplan-Meier survival analysis showed LUAD individuals with low-risk scores had a greater benefit in survival compared with a high-risk score population, which revealed that the AVEN-derived prognostic model was able to estimate patients' prognosis according to risk score (Figure 6C). Additionally, the time-dependent ROC curves were performed and confirmed that AVEN derived prognostic model showed potency to predict patient prognosis (Figure 6D).

Validation of AVEN-derived prognostic model in LUAD

Two independent external datasets, GSE50081 and GSE31210, were used to evaluate AVEN-derived prognostic model. As shown in Figures 7A, B, LUAD patients with low-risk score had better survival, which indicated that the AVEN-derived prognostic model is effective and sufficient in predicting LUAD survival.

Discussion

Even though AVEN has been found as an apoptosis, caspase activation inhibitor, its role in solid tumor including lung cancer was rarely studied. To analyze the role of AVEN systematically, we examined AVEN mRNA in 33 cancer types and found high AVEN expression in COAD, KIRC, KIRP, LUAD, LUSC and THCA compared to normal tissue, but interestingly high AVEN expression was associated with poor survival only in LUAD (Supplementary Figure S2). The impact of AVEN on tumor progression was analyzed in the LUAD patients grouped by the TNM classification, and we found that AVEN expression was significantly associated with the tumor burden of the main tumor and the number of lymph nodes contained in cancer (Figures 2C–E). Specifically, AVEN expression was positively correlated with the size and/or extent of the main tumor, suggesting the oncogenic role of AVEN. Before delving into the cellular

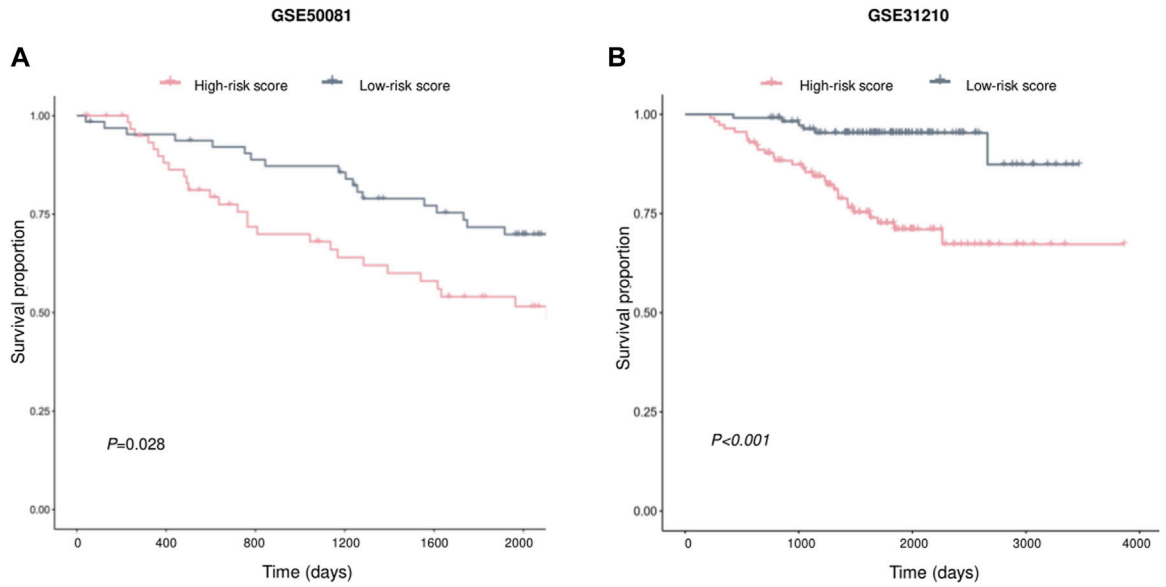


FIGURE 7
Validation of prognostic effect of AVEN-derived prognostic model in LUAD. **(A)** External validation of AVEN-derived prognostic model using GSE50081. **(B)** External validation of AVEN-derived prognostic model using GSE31210.

and molecular events in AVEN-associated LUAD, we questioned how AVEN is upregulated in LUAD. Interestingly, we found that AVEN upregulation in LUAD was not due to genomic amplification (Supplementary Table S5). Although it is unclear which factor regulates AVEN expression in the cancer cells in the absence of genomic amplification, previous studies have suggested potential regulatory mechanism. MiR-30 family (Ouzounova et al., 2013) has been reported to negatively regulate the AVEN expression in the breast cancer cell line, while Foxo1 (Cai et al., 2017) has a positive regulatory effect on AVEN expression in regulatory T cells. Hence, we propose that AVEN overexpression in cancer cells may result from the interplay of miRNA, transcription factor, and epigenetic modifications.

We furthermore analyzed signaling pathways altered by AVEN expression to understand the underlying mechanism of AVEN in tumor and tumor microenvironment. First, 838 DEGs were obtained between $AVEN^{high}$ and $AVEN^{low}$ groups by using Spearman's rank correlation test using the criteria of absolute R-value over 0.4, followed by pathway enrichment analysis (Figure 3A). The results were explained by three different functional cellular pathways, biological process pathway with metabolism of proteins and membrane trafficking, oncogenic pathways with cell cycle and VEGFA-VEGFR2 signaling pathway, and immune regulation process with neutrophil degranulation and innate immune system.

The majority of genes under cell cycle and VEGFA-VEGFR2 pathways were dramatically upregulated in $AVEN^{high}$ patients (Figure 3B). Although AVEN is known as an apoptosis inhibitor, further studies are needed to confirm whether AVEN boosts the cell cycle in cancer by inhibiting apoptosis. Apart from losing control of cell division, cancer cells have the ability to use various immune escape mechanisms from our immune surveillance, such as migrating to other parts of the organs via the blood vessel. One of the critical signaling pathways involved in angiogenesis is the VEGFA-VEGFR2 pathway. Activation of the VEGFA-VEGFR2 pathway is commonly observed in many types of cancer and is associated with poor prognosis. Furthermore, GSEA analysis also showed that $AVEN^{high}$ LUAD patients have more active oncogenic pathways including mTOR signaling (Figure 3C). These oncogenic pathways are targeted by drugs such as Palbociclib (CDK4/6 inhibitors) (Mills et al., 2017), Aflibercept (anti-VEGF agent) (Zirlik and Duyster, 2018) and Rapamycin (mTOR signaling) (*Rapamycin hits the target | Nature Reviews Cancer*, no date), all of which are in a clinical use. This further highlights the potential significance of AVEN as a promising therapeutic target in cancer treatment.

As tumor and their neighboring cells such as fibroblast and immune cells receive and send proliferation/apoptosis signals from each other (Galli et al., 2020), we examined the immune and stromal scores using ESTIMATE algorithm. $AVEN^{high}$ patients showed lower immune scores compared to $AVEN^{low}$ patients, while stromal scores had no significant difference (Figure 4A). In order to have a better understanding of immune cell infiltration correlated with AVEN, Timer, CIBERSORT, EPIC, xCEll, Quanti-seq, and MCP-counter were exploited to study the infiltration of diverse cell types (Figure 4B). These tools are available to analyze infiltrated immune cells using bulk RNA-seq, and each tool has been developed using a different methodology. B cells were significantly lower in $AVEN^{high}$ expressing group by all the algorithms indicating a weakened anti-tumor immunity (Figure 4B). The compromised B cell phenotype in $AVEN^{high}$ patients was further confirmed by examining expression level of B cells marker such as CD19, CD20 and CD22 (Figure 5D). B cell belongs to antigen-presenting cells (APC), which is able to gain, process, and present tumor-associated antigens for T cells activation (*Nature Reviews Immunology*, no date). In one study, the depletion of B cells decreased the production of type I T cells

(Th1) and caused scarcity of IFN-γ, supporting the crucial role that B cells may play in T cell immunity (DiLillo et al., 2010). In addition to their role in facilitating T cell activation, B cells also play a pivotal role in the tumor microenvironment, such as tumor-specific antibodies mediated antibody-dependent cell cytotoxicity (ADCC) (Germain et al., 2014; Sarvaria et al., 2017), complement activation (Dunkelberger and Song, 2010), and the formation and maintenance of tertiary lymphoid structures (TLS). Furthermore, independent cohort studies have highlighted the association of B cells with immune therapy response (Cabrita et al., 2020; Helmink et al., 2020; Petitprez et al., 2020). Therefore, even though no difference was observed in CD8 T cell markers (Figure 5C), the AVEN-mediated downregulation of B cell could potentially contribute to insufficient immunity, possibly resulting in an immunotherapy resistance phenotype through mechanisms that bypass T cell. This observation is further supported by the result from TIDE tool, indicating the association of AVEN in immunotherapy resistant effects (Figure 5A). Besides, CAF was upregulated in $AVEN^{high}$ patients as shown by EPIC and MCP-counter even though xCELL showed the opposite results (Figure 4B). CAF is involved in the matrix remodeling process and soluble factor secretion including VEGF, Exosomes, HGF, etc., which are the underlying mechanisms of tumor metastasis (Sarvaria et al., 2017). Because of the fundamental role of CAF in tumor progression, the correlation between AVEN and CAF needs to be further investigated.

In recent years, many prognostic biomarkers have been reported, however their applications are often limited due to the variations observed between different cohorts. Similarly, AVEN demonstrated a notable correlation with LUAD survival in some cohorts, but its significant prognostic effects were limited (Supplementary Table S3). To address the limitations associated with relying on a single prognostic marker, we developed a more comprehensive approach. We focused on the identification of AVEN-related genes and their association with patient survival. Initially, we identified 838 DEG that exhibited a correlation with AVEN expression. Subsequently, we performed Univariate Cox analysis, filtering down to 305 survival-related genes that displayed significant prognostic effects in LUAD patients within the TCGA cohort. To further refine our prognostic model, we employed Random Survival Forest analysis and Multiple Cox Regression analysis. The resulting model, consisting of five AVEN-related genes, demonstrated robust predictive capabilities for patient survival in both the TCGA cohort and two separate GEO cohorts, highlighting the potential of our AVEN-derived prognostic model as a valuable tool for LUAD prognosis and offering enhanced accuracy and applicability across diverse datasets.

Further investigation was conducted to explore the relationship between the identified five genes (KRT6A, SLC16A3, CTSL, LDHA, CDC42EP2) and AVEN expression. Supplementary Table S6 revealed a high correlation between these genes and AVEN expression. However, in-depth analysis using Protein-Protein Interaction (PPI) network analysis did not uncover any direct interactions among the proteins encoded by these genes (Supplementary Figure S3). It is worth noting that the lack of observed interactions in the PPI network may be attributed to insufficient available information or limitations in the current understanding of these interactions. Further studies are warranted to delve into this aspect and gather more comprehensive data in the future. In addition, previous studies have identified Bcl-xl and Apaf-1 as AVEN interacting proteins. These interactions are known to retain Bcl-xl mediated anti-apoptotic activity and prevent Apaf-1 mediated caspase activation (Chau et al., 2000). However, in this study, we observed that the expressions of Bcl-xl and Apaf-1 do not show significant association with AVEN expression (Supplementary Table S6). In our efforts to predict the underlying molecular mechanisms of AVEN, the results from this study suggest that AVEN may exert control over oncogenic events through mechanisms involving cell cycle regulation, VEGFR pathway. mTOR signaling, EMT and immune pathways. However, to gain a deeper understanding of molecular mechanism underlying AVEN's involvement in LUAD, further investigations are imperative to elucidate the specific molecular details associated with AVEN's function in the oncogenic process in LUAD.

In summary, our study elucidated the significance of AVEN in LUAD, showing its association with poor survival and its potential role in tumor progression and immune evasion. We developed an AVEN-derived prognostic model, incorporating five AVEN-related genes, which exhibited robust predictive capabilities for patient survival in diverse cohorts, highlighting the potential of our AVEN-derived prognostic model as a valuable tool for LUAD prognosis.

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

DF: Formal Analysis, Methodology, Visualization, Writing–original draft. MY: Data curation, Validation, Visualization, Writing–original draft. HL: Data curation, Writing–original draft. JL: Visualization, Writing–original draft. HK: Funding acquisition, Supervision, Writing–original draft, 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 is supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (NRF-2021R1A2C2008271, NRF-2022R1A4A3024551).

Acknowledgments

We thank the Oversea Study Program of Guangzhou lite Project (GEP) from Guangzhou, China, for their support to Fan Dengxia.

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/fmolb.2023.1265359/full#supplementary-material

References

Abd ElHafeez, S., D'Arrigo, G., Leonardis, D., Fusaro, M., Tripepi, G., and Roumeliotis, S. (2021). Methods to analyze time-to-event data: the Cox regression analysis'. *Oxidative Med. Cell. Longev.* 2021, 1302811. Available at:. doi:10.1155/2021/1302811

Bai, R., Lv, Z., Xu, D., and Cui, J. (2020). Predictive biomarkers for cancer immunotherapy with immune checkpoint inhibitors. *Biomark. Res.* 8 (1), 34. Available at:. doi:10.1186/s40364-020-00209-0

Błach, J., Wojas-Krawczyk, K., Nicoś, M., and Krawczyk, P. (2021). Failure of immunotherapy-the molecular and immunological origin of immunotherapy resistance in lung cancer. *Int. J. Mol. Sci.* 22 (16), 9030. Available at:. doi:10.3390/ijms22169030

Bondhopadhyay, B., Sisodiya, S., Chikara, A., Khan, A., Tanwar, P., Afroze, D., et al. (2020). Cancer immunotherapy: a promising dawn in cancer research. *Am. J. Blood Res.* 10 (6), 375–385.

Brabletz, T., Kalluri, R., Nieto, M. A., and Weinberg, R. A. (2018). EMT in cancer. *Nat. Rev. Cancer* 18 (2), 128–134. Available at:. doi:10.1038/nrc.2017.118

Cabrita, R., Lauss, M., Sanna, A., Donia, M., Skaarup Larsen, M., Mitra, S., et al. (2020). Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* 577 (7791), 561–565. Available at:. doi:10.1038/s41586-019-1914-8

Cai, Z., Liu, H., and Wu, X. (2017). Forkhead-box transcription factor 1 affects the apoptosis of natural regulatory T cells by controlling Aven expression. *BMC Immunol.* 18 (1), 16. Available at:. doi:10.1186/s12865-017-0198-8

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 (5), 401–404. Available at:. doi:10.1158/2159-8290.CD-12-0095

Chan, B. A., and Hughes, B. G. M. (2015). Targeted therapy for non-small cell lung cancer: current standards and the promise of the future. *Transl. Lung Cancer Res.* 4 (1), 36–54. Available at:. doi:10.3978/j.issn.2218-6751.2014.05.01

Chau, B. N., Cheng, E. H., Kerr, D. A., and Hardwick, J. M. (2000). Aven, a novel inhibitor of caspase activation, binds bcl-xL and apaf-1. *Mol. Cell.* 6 (1), 31–40. Available at:. doi:10.1016/S1097-2765(05)00021-3

Choi, J., Hwang, Y. K., Sung, K. W., Kim, D. H., Yoo, K. H., Jung, H. L., et al. (2006). Aven overexpression: association with poor prognosis in childhood acute lymphoblastic leukemia. *Leukemia Res.* 30 (8), 1019–1025. Available at:. doi:10.1016/j.leukres.2005.11.001

Dendritic cells in cancer immunology (2023). Dendritic cells in cancer immunology and immunotherapy | Nature Reviews Immunology. Available at: https://www.nature.com/articles/s41577-019-0210-z (Accessed: March 20, 2023).

DiLillo, D. J., Yanaba, K., and Tedder, T. F. (2010). B cells are required for optimal CD4+ and CD8+ T cell tumor immunity: therapeutic B cell depletion enhances B16 melanoma growth in mice. *J. Immunol.* 184 (7), 4006–4016. Available at:. doi:10.4049/jimmunol.0903009

Dunkelberger, J. R., and Song, W.-C. (2010). Complement and its role in innate and adaptive immune responses. *Cell. Res.* 20 (1), 34–50. Available at:. doi:10.1038/cr.2009.139

Eißmann, M., Melzer, I. M., Fernández, S. B. M., Michel, G., Hrabě de Angelis, M., Hoefler, G., et al. (2013). Overexpression of the anti-apoptotic protein AVEN contributes to increased malignancy in hematopoietic neoplasms. *Oncogene* 32 (20), 2586–2591. Available at:. doi:10.1038/onc.2012.263

Galli, F., Aguilera, J. V., Palermo, B., Markovic, S. N., Nisticò, P., and Signore, A. (2020). Relevance of immune cell and tumor microenvironment imaging in the new era of immunotherapy. *J. Exp. Clin. Cancer Res.* 39 (1), 89. Available at:. doi:10.1186/s13046-020-01586-y

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6 (269), pl1. Available at:. doi:10.1126/scisignal.2004088

Gelatti, A., Drilon, A., and Santini, F. C. (2019). Optimizing the sequencing of tyrosine kinase inhibitors (TKIs) in epidermal growth factor receptor (EGFR) mutation-positive non-small cell lung cancer (NSCLC). *Lung cancer* 137, 113–122. Available at:. doi:10.1016/j.lungcan.2019.09.017

Germain, C., Gnjatic, S., Tamzalit, F., Knockaert, S., Remark, R., Goc, J., et al. (2014). Presence of B cells in tertiary lymphoid structures is associated with a protective immunity in patients with lung cancer. *Am. J. Respir. Crit. Care Med.* 189 (7), 832–844. Available at:. doi:10.1164/rccm.201309-1611OC

Helmink, B. A., Reddy, S. M., Gao, J., Zhang, S., Basar, R., Thakur, R., et al. (2020). B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* 577 (7791), 549–555. Available at:. doi:10.1038/s41586-019-1922-8

Hu, F.-F., Liu, C. J., Liu, L. L., Zhang, Q., and Guo, A. Y. (2021). Expression profile of immune checkpoint genes and their roles in predicting immunotherapy response. *Briefings Bioinforma.* 22 (3), bbaa176. Available at:. doi:10.1093/bib/bbaa176

Im, Y., and Kim, Y. (2023). A comprehensive overview of RNA deconvolution methods and their application. *Mol. Cells* 46 (2), 99–105. Available at:. doi:10.14348/molcells.2023.2178

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. Available at:. doi:10.1038/s41591-018-0136-1

Jiang, X., Wang, J., Deng, X., Xiong, F., Zhang, S., Gong, Z., et al. (2020). The role of microenvironment in tumor angiogenesis. *J. Exp. Clin. Cancer Res.* 39 (1), 204. Available at:. doi:10.1186/s13046-020-01709-5

Kamburov, A., Wierling, C., Lehrach, H., and Herwig, R. (2009) 'ConsensusPathDB—a database for integrating human functional interaction networks', *Nucleic Acids Res.*, 37(Database issue), pp. D623, D628. doi:10.1093/nar/gkn698

Kim, S. S., Sumner, W. A., Miyauchi, S., Cohen, E. E. W., Califano, J. A., and Sharabi, A. B. (2021). Role of B Cells in responses to checkpoint blockade immunotherapy and overall survival of cancer patients. *Clin. Cancer Res.* 27 (22), 6075–6082. Available at:. doi:10.1158/1078-0432.CCR-21-0697

Lei, Q., Wang, D., Sun, K., Wang, L., and Zhang, Y. (2020). Resistance mechanisms of anti-PD1/PDL1 therapy in solid tumors. *Front. Cell. Dev. Biol.* 8, 672. doi:10.3389/fcell.2020.00672

Li, C., Tang, Z., Zhang, W., Ye, Z., and Liu, F. (2021). GEPIA2021: integrating multiple deconvolution-based analysis into GEPIA. *Nucleic Acids Res.* 49 (1), W242–W246. Available at:. doi:10.1093/nar/gkab418

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. Available at:. doi:10.1158/0008-5472.CAN-17-0307

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. *Briefings Bioinforma.* 24 (1), bbac558. Available at:. doi:10.1093/bib/bbac558

Ma, B., Wang, K., Liang, Y., Meng, Q., and Li, Y. (2022). Molecular characteristics, oncogenic roles, and relevant immune and pharmacogenomic features of EVA1B in colorectal cancer. *Front. Immunol.* 13, 809837. doi:10.3389/fimmu.2022.809837

Melzer, I. M., Fernández, S. B. M., Bösser, S., Lohrig, K., Lewandrowski, U., Wolters, D., et al. (2012). The Apaf-1-binding protein Aven is cleaved by Cathepsin D to unleash its anti-apoptotic potential. *Cell. Death Differ.* 19 (9), 1435–1445. Available at:. doi:10.1038/cdd.2012.17

Mills, C. C., Kolb, E. A., and Sampson, V. B. (2017). Recent advances of cell-cycle inhibitor therapies for pediatric cancer. *Cancer Res.* 77 (23), 6489–6498. Available at:. doi:10.1158/0008-5472.CAN-17-2066

Mizuno, H., Kitada, K., Nakai, K., and Sarai, A. (2009). PrognoScan: a new database for meta-analysis of the prognostic value of genes. *BMC Med. Genomics* 2, 18. Available at:. doi:10.1186/1755-8794-2-18

Ouzounova, M., Vuong, T., Ancey, P. B., Ferrand, M., Durand, G., Le-Calvez Kelm, F., et al. (2013). MicroRNA miR-30 family regulates non-attachment growth of breast cancer cells. *BMC Genomics* 14 (1), 139. Available at:. doi:10.1186/1471-2164-14-139

Petitprez, F., de Reyniès, A., Keung, E. Z., Chen, T. W. W., Sun, C. M., Calderaro, J., et al. (2020). B cells are associated with survival and immunotherapy response in sarcoma. *Nature* 577 (7791), 556–560. Available at:. doi:10.1038/s41586-019-1906-8

Qin, S., Xu, L., Yi, M., Yu, S., Wu, K., and Luo, S. (2019). Novel immune checkpoint targets: moving beyond PD-1 and CTLA-4. *Mol. Cancer* 18 (1), 155. Available at:. doi:10.1186/s12943-019-1091-2

Racle, J., et al. (2017). "Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data,"*eLife*, e26476. Editor A. Valencia, 6. Available at: . doi:10.7554/eLife.26476

Rapamycin hits the target (2023). Rapamycin hits the target | nature Reviews cancer. Available at: https://www.nature.com/articles/nrc2341 (Accessed: September 24, 2023).

Robert, C. (2020). A decade of immune-checkpoint inhibitors in cancer therapy. *Nat. Commun.* 11 (1), 3801. Available at:. doi:10.1038/s41467-020-17670-y

Rosen, R. D., and Sapra, A. (2022). TNM classification', in *StatPearls*. Treasure island (FL): statPearls publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK553187/ (Accessed: March 20, 2023).

Sarvaria, A., Madrigal, J. A., and Saudemont, A. (2017). B cell regulation in cancer and anti-tumor immunity. *Cell. Mol. Immunol.* 14 (8), 662–674. Available at:. doi:10.1038/cmi.2017.35

Skoulidis, F., and Heymach, J. V. (2019). Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nat. Rev. Cancer* 19 (9), 495–509. Available at:. doi:10.1038/s41568-019-0179-8

Sturm, G., Finotello, F., Petitprez, F., Zhang, J. D., Baumbach, J., Fridman, W. H., et al. (2019). Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics* 35 (14), i436–i445. Available at:. doi:10.1093/bioinformatics/btz363

van der Leun, A. M., Thommen, D. S., and Schumacher, T. N. (2020). CD8+ T cell states in human cancer: insights from single-cell analysis. *Nat. Rev. Cancer* 20 (4), 218–232. Available at:. doi:10.1038/s41568-019-0235-4

Wang, D., Qian, X., Du, Y. C. N., Sanchez-Solana, B., Chen, K., Kanigicherla, M., et al. (2023). cProSite: a web based interactive platform for online proteomics, phosphoproteomics, and genomics data analysis. *bioRxiv* 2023, 543932. Available at: . doi:10.1101/2023.06.10.543932

Wang, X., He, Q., Shen, H., Xia, A., Tian, W., Yu, W., et al. (2019). TOX promotes the exhaustion of antitumor CD8+ T cells by preventing PD1 degradation in hepatocellular carcinoma. *J. Hepatology* 71 (4), 731–741. Available at:. doi:10.1016/j.jhep.2019.05.015

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. Available at:. doi:10.1038/ncomms3612

Yu, Z., Du, M., and Lu, L. (2022). A novel 16-genes signature scoring system as prognostic model to evaluate survival risk in patients with glioblastoma. *Biomedicines* 10 (2), 317. Available at:. doi:10.3390/biomedicines10020317

Zirlik, K., and Duyster, J. (2018). Anti-angiogenics: current situation and future perspectives. *Oncol. Res. Treat.* 41 (4), 166–171. Available at:. doi:10.1159/000488087

Zou, Z., Tao, T., Li, H., and Zhu, X. (2020). mTOR signaling pathway and mTOR inhibitors in cancer: progress and challenges. *Cell. and Biosci.* 10 (1), 31. Available at:. doi:10.1186/s13578-020-00396-1

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 17 October 2023
DOI 10.3389/fmolb.2023.1248410

Check for updates

OPEN ACCESS

EDITED BY
Qingyu Luo,
Dana–Farber Cancer Institute,
United States

REVIEWED BY
Lin Cheng,
Peking University People's Hospital,
China
Shu Peng,
Huazhong University of Science and
Technology, China

*CORRESPONDENCE
Xiang Wang,
xiangw@vip.sina.com

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

RECEIVED 27 June 2023
ACCEPTED 02 October 2023
PUBLISHED 17 October 2023

CITATION
Feng K, Xing Z, Dai Q, Cheng H and Wang X (2023), Role of aggressive locoregional surgery in treatment strategies for ipsilateral supraclavicular lymph node metastasis of breast cancer: a real-world cohort study.
Front. Mol. Biosci. 10:1248410.
doi: 10.3389/fmolb.2023.1248410

COPYRIGHT
© 2023 Feng, Xing, Dai, Cheng and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Role of aggressive locoregional surgery in treatment strategies for ipsilateral supraclavicular lymph node metastasis of breast cancer: a real-world cohort study

Kexin Feng†, Zeyu Xing†, Qichen Dai, Han Cheng and Xiang Wang*

Department of Breast Surgical Oncology, National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Background: Breast cancer patients with synchronous ipsilateral supraclavicular lymph node metastases (ISLNM) have unfavorable prognoses. The role of supraclavicular lymph node dissection (SLND) as a surgical intervention in the treatment of this condition remains controversial. In this study, we aimed to evaluate the prognostic factors associated with breast cancer with ISLNM and to assess the potential impact of aggressive locoregional surgical management on patient outcomes.

Methods: We conducted a retrospective analysis of 250 breast cancer patients with ISLNM who were treated with curative intent at our institution between 2000 and 2020. The cohort was stratified into groups based on the extent of axillary surgery. The first group, comprising 185 patients, underwent level I/II axillary dissection. The second group, consisting of 65 patients, underwent aggressive locoregional surgery, including levels I/II/III (infraclavicular) dissection in 37 patients and levels I/II/III + SLND in 28 patients. Our study evaluated overall survival (OS) and disease-free survival (DFS) as primary endpoints, and locoregional recurrence-free survival (LRRFS) and distant metastasis-free survival (DMFS) as secondary endpoints.

Results: The median follow-up time among all patients was 5.92 years (1.05–15.36 years). The 5-year OS rate was 71.89%, while the DFS rate, LRRFS rate, and DMFS rates were 59.25%, 66.38%, and 64.98%, respectively. A significant difference in OS, DFS, LRRFS, and DMFS was observed between the second group and the first group ($p < 0.01$). No beneficial impact on recurrence, metastasis, or survival outcomes was observed in the levels I/II/III + SLND group compared to the levels I/II/III dissection group. Multivariate logistic regression analysis revealed that levels I/II/III ± SLND surgery and T stage were associated with OS ($p = 0.006$ and $p = 0.026$), while levels I/II/III ± SLND surgery, ER+/HER2-, and histologic grade were associated with DFS ($p = 0.032$, $p = 0.001$, $p = 0.032$).

Conclusion: Breast cancer with ISLNM may be considered a locoregional disease, requiring a combination of systemic and local therapies. Aggressive locoregional surgery has been shown to positively impact recurrence, metastasis, and survival

outcomes. This approach may provide improved management of the ISLNM for breast cancer patients.

KEYWORDS

breast cancer, supraclavicular lymph node, supraclavicular lymph node dissection (SLND), survival, prognostic factors

1 Introduction

The occurrence rate of ipsilateral supraclavicular lymph node metastases (ISLNM) in breast cancer patients who do not exhibit any distant metastases varies between 1% and 4% (Chen et al., 2006; Chen et al., 2006). The lymphatic dissemination of axillary breast cancer often adheres to a specific pattern: it initiates in the axillary nodes on the same side as the affected breast, advances to the infraclavicular nodes, and ultimately involves the supraclavicular nodes. It is noteworthy that a significant majority of cases with invasive lobular cancer of the breast with ISLNM are accompanied by metastases in the axillary lymph nodes, with a range of 94%–100%.The presence of ISLNM serves as a noteworthy prognostic indicator, indicating an adverse prognosis. A considerable proportion of these individuals demonstrates the occurrence of distant metastases after a year of identifying isolated tumour cells in regional lymph nodes. Furthermore, the survival rate after 5 years remains disappointingly low, ranging from 5% to 34% (Brito et al., 2001). In the past, the classification of ISLNM in breast cancer was considered as a form of distant metastasis. Consequently, the treatment strategy mostly emphasised palliative measures, whereas local surgical treatments received less attention. Nevertheless, the dynamic and progressive nature of diagnostic and therapeutic developments has introduced a positive outlook. Although the incidence of distant metastases is high in patients with ISLNM, there is a possibility of curative interventions for these individuals. In support of this shift in perspective, the American Joint Committee on Cancer, in its sixth edition TNM staging system (2002), revised the classification of breast cancer patients with isolated tumour cells in regional lymph nodes from stage M1 to clinical N3c (cN3c). The reclassification in question highlights the notion of ISLNM as a localised ailment rather than a disorder characterised by distant metastasis (Singletary et al., 2002). According to the NCCN guidelines (Gradishar et al., 2022), patients with ISLNM should undergo complete or partial breast resection and axillary level I/II lymph node dissection, preceded by preoperative chemotherapy/targeted therapy and followed by postoperative radiotherapy (RT) to the chest wall, supraclavicular, and/or internal mammary lymph nodes. However, it has been reported that the complete response (CR) rate of supraclavicular lymph nodes (SLN) following neoadjuvant chemotherapy (NAC) is only 50% (Zhu et al., 2019), and residual tumor load is a significant concern. Current radiation therapy in the supraclavicular and subclavian regions is limited by normal tissue tolerance and achieving curative agents' volume is difficult. Simply increasing the radiation dose does not improve patient prognosis but rather increases the occurrence of radiotherapy-related complications.

In addition to primary breast tumor surgery and axillary level I/II dissection, aggressive locoregional surgery, such as axillary level III (infraclavicular) and SLND, may offer a more precise and thorough therapeutic effect than radiotherapy. This approach theoretically reduces local tumor load and prevents the spread of tumor cells through lymph-vessels. Additionally, the surgical removal of the metastatic SLN can provide a more precise assessment of the SLN status following NAC. However, it is crucial to acknowledge that increased radiation doses may heighten the likelihood of problems following radiotherapy (Huang et al., 2007). The existing body of literature lacks a sufficient number of comparative trials that investigate the results of SLND surgery in comparison to neck RT, and the establishment of definite guidelines remains challenging. A comprehensive review of the existing literature reveals a prevalent lack of consensus regarding the optimal localised therapy approach. The objective of this research study was to investigate the possible advantages of intensive locoregional surgical interventions in improving the prognosis for patients with ISLNM and identifying individuals who may benefit from more aggressive therapies. In addition, we initiated an assessment of prognostic factors associated with survival rates in breast cancer patients affected by ISLNM, along with a thorough examination of research focused on SLND. The results of our study have the potential to enhance the existing NCCN guidelines on the management of locally advanced breast cancer by providing additional clinical perspectives.

2 Methods and materials

2.1 Patients

With approval from the institutional review board of the National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, a total of 250 patients with ISLNM and no evidence of distant metastatic disease were identified retrospectively. These patients received intended curative multidisciplinary therapy, which consisted of neoadjuvant systemic medication, surgical resection, and adjuvant radiotherapy (RT). The treatment was administered at our institution over the period from 2000 to 2020.

Complete physical examination and accessory tests were used for preoperative diagnosis, including ultrasonography (US), chest CT, breast magnetic resonance imaging (MRI), and/or positron emission tomography/computed tomography scans (PET/CT). Inclusion criteria were as follows: primary breast tumor confirmed by biopsy, ISLNM confirmed by biopsy or needle cytology, ISLNM defined as cervical lymph node metastasis levels IV and VB, no internal mammary lymph node metastasis or distant metastasis found, no previous history of other malignant tumors, and ability to accept comprehensive treatment, including chemotherapy, targeted therapy, and radiotherapy, without serious organ dysfunction. Patients were included regardless of tumor stage and pathotype.

Patients who had a prior history of breast cancer or other malignancies, as well as those experiencing significant problems in other organs, were excluded based on predetermined criteria. Participants were ineligible for inclusion if they presented with additional distant metastases during the initial treatment, which included surgery or prior chemotherapy. In order to consider the impact on the M stage, we choose to omit patients who exhibited indications of neck lymph node metastases in levels I, II, III, and VI.

2.2 Clinicopathological data

The clinical data of the patients were collected, including age at first diagnosis, clinical TNM stage, histology, receptor status, use of chemotherapy, pathological response of the tumor, axillary lymph node, and SLN after NAC, pathologic response, type of surgery and nodal dissection, radiation technique, and initial and boost radiation dose. This information was used to analyze the effectiveness of aggressive locoregional surgical treatment for improving the prognosis of patients with ISLNM.

2.3 Treatment

The treatments were administered in accordance with established national criteria and were customised to suit the unique circumstances of each individual patient. After receiving an initial breast cancer diagnosis, participants underwent conventional treatment and underwent routine follow-up, which included physical and imaging exams every three to 6 months following breast surgery. Patients with ISLNM were provided with personalised treatment strategies, which encompassed systemic therapeutic approaches such as chemotherapy, endocrine therapy, and targeted therapy specifically targeting HER2. Additionally, local treatment modalities such as neck RT and/or SLND surgery were employed.

2.4 Outcomes and follow up

In our investigation, we assessed recurrences at local, regional, and distant levels. Local recurrences refer to instances where the cancer reappears in either the same breast or the chest wall on the same side. On the other hand, recurrences that occur in the axillary node on the same side or the internal mammary sentinel lymph node (ISLN) are classified as regional recurrences. Tumours that reoccur outside the specified locoregional borders are classified as distant recurrences. The interval between the commencement of systemic therapy to the initial documented relapse was designated as disease-free survival (DFS). This statistic included occurrences of recurrence at the local, regional, or distant level, as well as instances of fatality. The metric known as locoregional recurrence-free survival (LRRFS) was defined as the duration until the initial occurrence of a local or regional recurrence, including recurrences inside the mediastinum or sternum. Overall survival (OS) was measured from the time of surgery to death from any cause or last follow-up. Distant metastasis-free survival (DMFS) was measured to the first occurrence of distant metastasis.

2.5 Statistical analysis

The comparison of clinicopathologic categorical factors was conducted using Pearson's chi-square test. The researchers utilised the Kaplan-Meier method to construct survival curves for factors related to time-to-event. Subsequently, the differences between the curves were evaluated by log-rank testing. A univariate Cox regression analysis was performed to investigate the relationship between clinicopathological characteristics and the likelihood of recurrence. The subsequent multivariate Cox proportional hazards regression analysis considered variables with a p-value of less than 0.15 to be relevant and included them. The statistical analyses were conducted using SPSS 22.0 (IBM Corp.) and R 4.1.2 (obtained from https://www.r-project.org/). A significance level of 0.05 was established as the threshold for statistical significance in the bilateral test.

3 Results

3.1 The clinicopathological characteristics of patients

This analysis incorporated data from a sequential sample of 250 patients. Out of the total sample, 185 patients (74.0%) received sole level I/II axillary dissection, 37 patients (14.8%) underwent dissection of levels I/II/III (infraclavicular), and 28 patients (11.2%) underwent dissection of levels I/II/III in conjunction with supraclavicular lymph node dissection (SLND). The provided information offers valuable insights into the surgical methodologies employed for the management of individuals diagnosed with isolated spinal metastatic lesions. Biopsy tissue test confirmation of the SLN was performed in 249 patients (99.6%). The reason for the 1 patient without a biopsy tissue test was that the result of the PET-CT scan was sufficient for diagnosis. The median age of the patients was 57 years, with a range of 24–75 years. In terms of clinical T stage, 47 patients (18.8%) were classified as T1, 116 patients (46.4%) as T2, 51 patients (20.4%) as T3, and 36 patients (14.4%) as T4. Among the 250 patients, 141 (56.4%) were estrogen receptor (ER) positive and 93 (37.2%) were human epidermal growth factor receptor 2 (HER2) positive. The clinicopathological characteristics of this cohort are summarized in Table 1.

3.2 Surgical approach and survival outcomes

The median follow-up time among all patients in this study was 5.92 years, with a range of 1.05–15.36 years. The 5-year OS rate was 71.89%, while the DFS rate, LRRFS rate, and DMFS rates were 59.25%, 66.38%, and 64.98%, respectively. The 10-year OS rate, DFS rate, LRRFS rate, and DMFS rates were 45.71%, 28.76%, 33.26%, and 43.57%, respectively. The median OS time was 9.57 years and the median DFS time was 6.81 years, while the median LRRFS and DMFS were 7.91 years and 8.84 years, respectively. Figure 1 shows the Kaplan-Meier curves of OS, DFS, LRRFS, and DMFS for breast cancer patients with ISLNM.

Of the 250 patients included in the analysis, 185 (74.0%) underwent level I/II axillary dissection only, while 65 (26.0%)

10.3389/fmolb.2023.1248410

TABLE 1 Patient and treatment characteristics.

Variable	All patients (N = 250) (n, %)
Age, years	—
<50	108 (43.2)
≥50	142 (56.8)
Pre-NAC largest SLN size[a]	—
<1 cm	116 (46.4)
≥1 cm	134 (53.6)
Clinical T-stage	—
T1	47 (18.8)
T2	116 (46.4)
T3	51 (20.4)
T4	36 (14.4)
Histology	—
Invasive ductal carcinoma	247 (98.8)
Others	3 (1.2)
Histologic grade	—
Low	60 (24.0)
Median	142 (56.8)
High	48 (19.2)
Receptor Status	
ER+/HER2-	96 (38.4)
ER+/HER2+	45 (18.0)
ER-/HER2+	48 (19.2)
ER-/HER2-	61 (24.4)
Type of primary breast surgery	—
Yes	246 (98.4)
No	4 (1.6)
Type of primary nodal surgery	—
Level I/II	185 (74.0)
Level I/II/III	37 (14.8)
Level I/II/III + SLND	28 (11.2)
Tumor chemotherapy response (ypT0)	—
Yes	63 (25.2)
No	183 (73.2)
NA	4 (1.6)
Nodal pathologic complete response (ypN0)	—
Yes	68 (27.2)
No	178 (71.2)
NA	4 (1.6)

(Continued on following page)

TABLE 1 (*Continued*) **Patient and treatment characteristics.**

Variable	All patients (N = 250) (n, %)
SLN radiographic response to NAC[a]	—
Yes	150 (60.0)
No	100 (40.0)
Involvement of infraclavicular lymph nodes after NAC	—
Yes	26 (10.4)
No	224 (89.6)
Post-NAC largest SLN size[a]	—
<1 cm	224 (89.6)
≥1 cm	26 (10.4)
SLN cumulative dose, Gy	—
<60	172 (68.8)
≥60	56 (22.4)
NA	22 (8.8)
Adjuvant targeted therapy	—
Yes	41 (16.4)
No	209 (83.6)
Adjuvant endocrine therapy	—
Yes	170 (68.0)
No	80 (32.0)

[a]Confirmed by BUS.
SLN, supraclavicular lymph nodes; NAC, neoadjuvant chemotherapy; SLND, supraclavicular lymph node dissection.

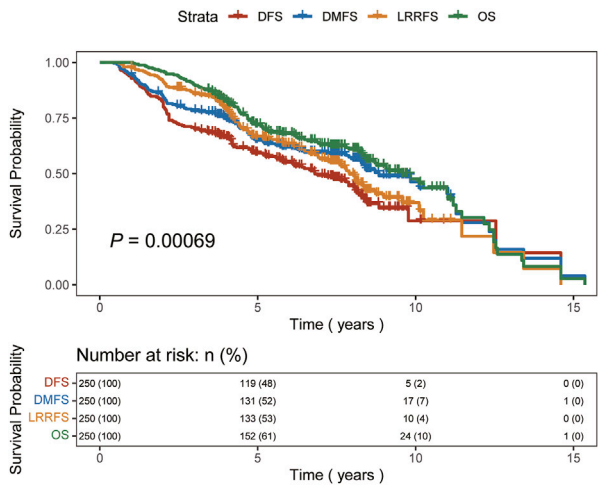


FIGURE 1
Kaplan-Meier curves of overall survival (OS), disease-free survival (DFS), locoregional recur-rence-free survival (LRRFS), and distant metastasis-free survival (DMFS) for breast cancer patients with ISLNM.

underwent aggressive locoregional surgical treatment (level I/II/III ± SLND), including 37 (14.8%) who underwent levels I/II/III (infraclavicular) dissection and 28 (11.2%) who underwent levels I/II/III with SLND. The clinicopathological characteristics of two subgroup are shown in Supplementary Tables S1, S2.

A significant difference in OS, DFS, and DMFS was observed between patients who underwent levels I/II axillary dissection only and those who underwent aggressive locoregional surgery (levels I/II/III lymph node dissection ± SLND) ($p < 0.01$ Figure 2). The results indicated that the implementation of a comprehensive surgical strategy targeting the local and regional areas had a beneficial impact on the rates of recurrence, metastasis, and overall survival.Nevertheless, upon comparing the patient groups, specifically those who underwent levels I/II/III dissection and those who underwent levels I/II/III combined with SLND, no significant differences were observed in terms of OS, DFS, LRRFS, or DMFS outcomes ($p = 0.63$, $p = 0.23$, $p = 0.31$, $p = 0.13$, respectively, Figure 3). This implies that the incorporation of SLND into stages I/II/III dissection may not provide any further benefits in terms of survival.

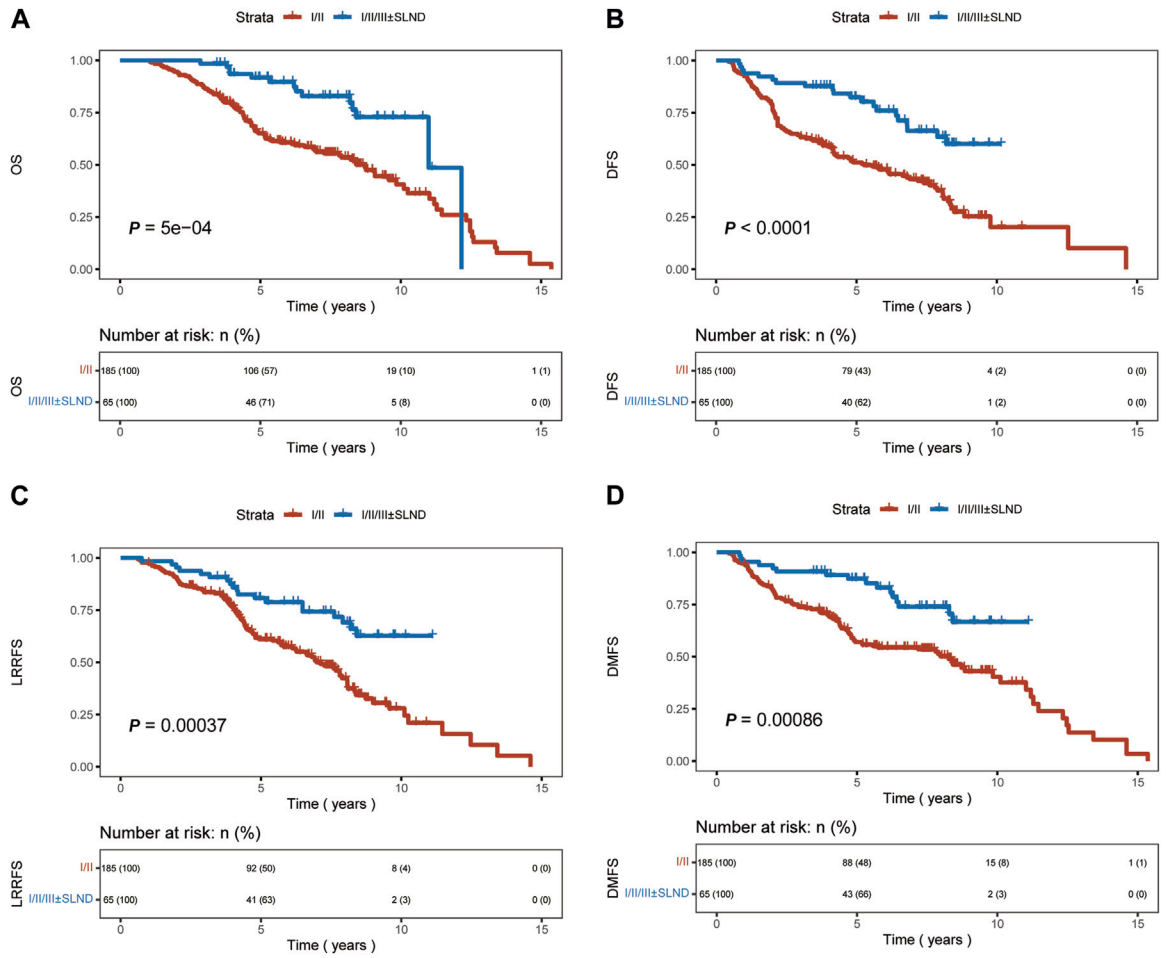


FIGURE 2
Kaplan-Meier curves of overall survival (OS), disease-free survival (DFS), locoregional recurrence-free survival (LRRFS), and distant metastasis-free survival (DMFS) for patients who underwent I/II axillary dissection only, and those with 1/11/lll ± SLND. **(A)** OS. **(B)** DFS. **(C)** LRRFS. **(D)** DMFS.

3.3 Prognostic factors

Our study's univariate analysis identified certain factors, including the nature of primary nodal surgery (Level I/II/III ± SLND), T stage (specifically T2 stage), and ER+/HER2-status, as being significantly related to OS, as detailed in Table 2 ($p < 0.05$). When delving into a multivariate logistic regression analysis that incorporated factors such as the primary nodal surgery type, breast surgery type, T stage, and receptor status ($p < 0.15$, Table 2), both the nodal surgery (Level I/II/III ± SLND) and T stage (T2) emerged as significantly linked to OS ($p = 0.006$ and $p = 0.026$, Table 2).

Regarding DFS, the univariate analysis pinpointed factors, namely, the primary nodal surgery (Level I/II/III ± SLND), size of the largest supraclavicular lymph node prior to NAC, post-NAC infraclavicular lymph node involvement, histologic grade (specifically Grade 3), and receptor status, as being significantly associated with DFS ($p < 0.05$, Table 2). Upon performing a multivariate logistic regression analysis, which incorporated variables such as primary nodal surgery type, pre-NAC supraclavicular lymph node size, receptor status, post-NAC infraclavicular lymph node involvement, and primary breast surgery type ($p < 0.15$, Table 2), it was discerned that nodal surgery (Level I/II/III ± SLND), ER+/HER2-status, and histologic grade (Grade 3) stood out as independent determinants influencing DFS ($p = 0.02$, Table 2). These results provide insight into the factors that may impact survival outcomes in patients with ISLNM.

3.4 Adjuvant treatment

Before surgery, chemotherapy was administered to all patients in our study, averaging six rounds (ranging from one to eight cycles). A total of 156 patients, accounting for 62.4% of the sample, received postoperative chemotherapy. The average number of cycles administered was two, with a range of 1–8 cycles. Consistently, regimens based on anthracyclines and/or taxanes were employed. Within the cohort of patients, a total of 93 individuals, accounting for 37.2% of the sample, were identified as having HER2-positive disease based on the diagnostic methods of immunohistochemistry or fluorescence *in situ* hybridization. Out of these individuals, 41 patients, equivalent to 44.1% of the HER2-positive subgroup,

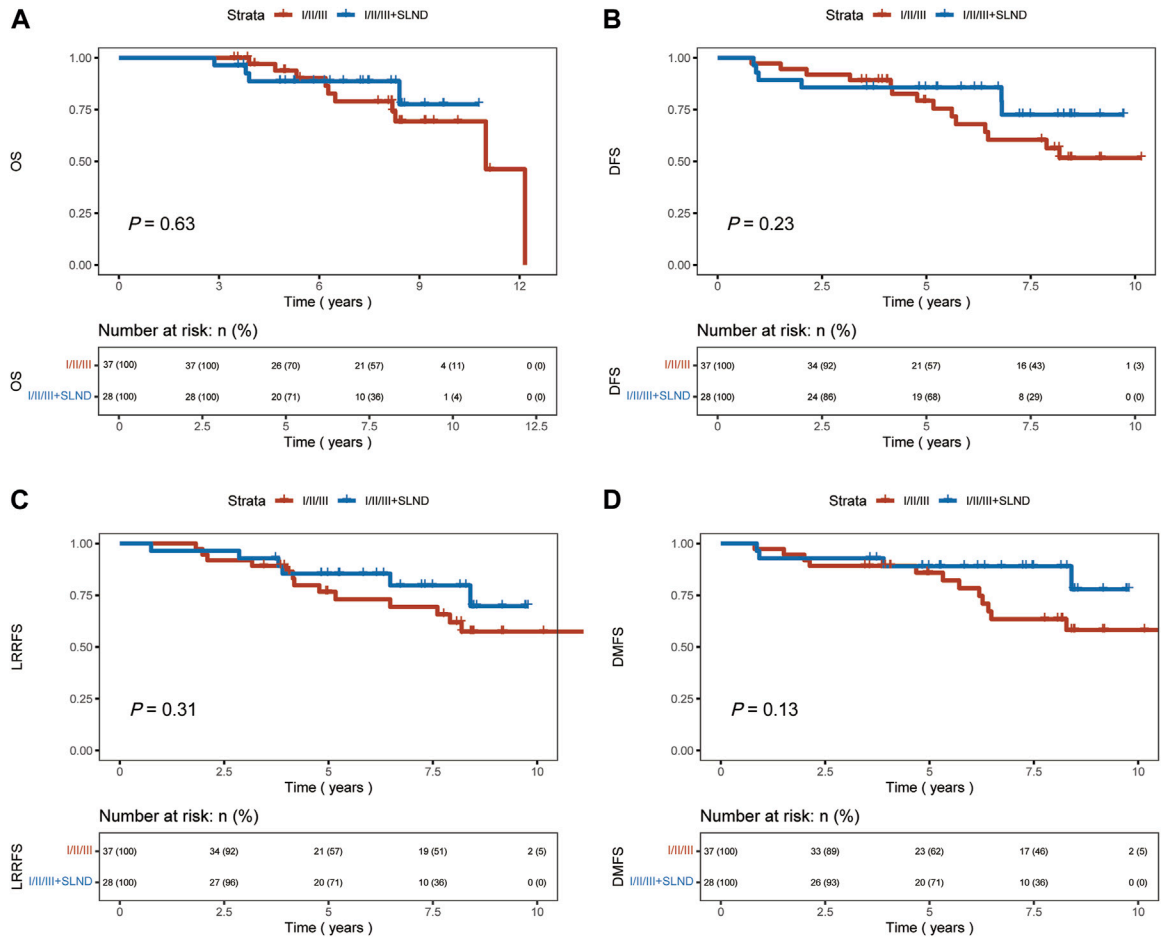


FIGURE 3
Kaplan-Meier curves of overall survival (OS), disease-free survival (DFS), locoregional recurrence-free survival (LRRFS), and distant metastasis-free survival (DMFS) for patients who underwent I/II/III dissection, and those with I/II/III + SLND. **(A)** OS. **(B)** DFS. **(C)** LRRFS. **(D)** DMFS.

received treatment specifically targeting the HER2 protein. No patients were prescribed preoperative endocrine treatment. However, a total of 170 patients, accounting for 68.0% of those diagnosed with ER/PR-positive breast cancer, received adjuvant endocrine therapy. Regarding the application of radiotherapy, a significant proportion of our study participants, specifically 91.2% (228 individuals), underwent this therapeutic intervention. This ensured that all patients got irradiation to the supraclavicular, infraclavicular, and axilla level II-III nodes, with the inclusion of the metastatic supraclavicular (SCV) node or its nodal bed boost, as deemed necessary. The primary radiation dose administered to the chest wall/breast and regional lymphatics was 50 Gy, with a range of 43.5–60 Gy, delivered over a course of 25 fractions, varying between 15 and 33 fractions, for 92.1% of the patients in this study. A small proportion, specifically 5.7%, was administered a radiation dosage of 43.5 Gy divided into 15 portions. The cumulative dose administered to metastatic SCV nodes or the nodal bed had an average of 60 Gy (with a range of 43.5–66 Gy) delivered over 28 fractions (encompassing 25–35) and 43.5 Gy (with a range of 43.5–52.2 Gy) delivered over 15 fractions (reaching 15–18). This distinction provides a clear explanation of the therapeutic techniques employed in our research.

3.5 Safety and toxicities

Among the 67 patients who underwent aggressive locoregional surgery, none experienced postoperative complications such as supraclavicular hemorrhage, edema, cervical chylous lymphocele, head and neck movement disorders, pleural effusion, or chylothorax. Postoperative over-drainage, defined as a 24-h volume of drainage greater than 200 mL, was noted in 48 patients (19.2%) and managed by compression and drainage without secondary surgery.

Radiotherapy side effects occurred in 63 patients, all of whom experienced grade 1–2 events of radiation dermatitis and 32 of whom also experienced grade 1–2 events of radiation esophagitis. No grade ≥ 3 acute or late toxicities were observed. These results suggest that the treatments used in this study were generally well-tolerated by patients.

4 Discussion

Breast cancer is the most common malignant tumor in females worldwide, with rapidly increasing incidence rates in China and other parts of Asia (Lei et al., 2021). However, compared to Western countries, where the detection rate of early breast cancer is over 60%,

TABLE 2 Univariate and multivariate analysis using Cox-regression model of the survival of patients with ISLNM.

Variables	Overall survival				Disease-free survival			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	P	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)
Age, years								
<50								
≥50	0.709	1.003 (0.986–1.021)			0.280	1.009 (0.993–1.024)		
Pre-NAC largest SLN size[a]								
<1 cm								
≥1 cm	0.298	1.217 (0.841–1.760)			**0.025**	1.490 (1.052–2.111)	0.237	1.243 (0.867–1.782)
Clinical T-stage						-		
T1								
T2	**0.045**	1.729 (1.012–2.954)	**0.026**	1.895 (1.080–3.324)	0.239	1.348 (0.820–2.216)	0.352	1.275 (0.765–2.124)
T3	0.693	1.137 (0.600–2.155)	0.508	1.250 (0.645–2.425)	0.360	1.306 (0.738–2.310)	0.688	1.126 (0.631–2.008)
T4	0.070	1.783 (0.954–3.331)	0.064	1.861 (0.964–3.593)	0.057	1.709 (0.983–3.151)	0.234	1.440 (0.789–2.627)
Histology								
Invasive ductal carcinoma								
Others	0.439	0.575 (0.141–2.338)			0.368	0.526 (0.130–2.130)		
Histologic grade								
Low								
Median	0.385	0.808 (0.499–1.308)			0.525	1.154 (0.742–1.793)	0.590	1.149 (0.693–1.906)
High	0.516	1.201 (0.691–2.086)			**0.026**	1.790 (1.072–2.989)	**0.032**	1.859 (1.056–3.273)
Receptor Status								
ER+/HER2-								
ER+/HER2+	**0.028**	0.593 (0.372–0.946)	0.072	0.647 (0.403–1.039)	**<0.001**	0.432 (0.282–0.661)	**0.001**	0.459 (0.298–0.709)
ER-/HER2+	0.409	0.8082 (0.476–1.353)	0.835	1.064 (0.595–1.900)	**0.047**	0.604 (0.367–0.994)	0.270	0.270 (0.428–1.268)
ER-/HER2-	0.201	0.708 (0.417–1.202)	0.728	0.893 (0.472–1.689)	**0.045**	0.604 (0.369–0.989)	0.353	0.353 (0.403–1.384)
Type of primary breast surgery								
Yes								
No	0.097	0.377 (0.119–1.193)	0.89	1.103 (0.271–4.521)	0.074	0.402 (0.148–1.092)	0.362	0.594 (0.194–1.819)
Type of primary nodal surgery								
Level I/II								
Level I/II/III ± SLND	**0.001**	0.382 (0.217–0.671)	**0.006**	0.331 (0.152–0.722)	**<0.001**	0.379 (0.235–0.610)	**0.007**	0.417 (0.221–0.785)

(Continued on following page)

TABLE 2 (*Continued*) Univariate and multivariate analysis using Cox-regression model of the survival of patients with ISLNM.

Variables	Overall survival				Disease-free survival			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	P	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)
Tumor chemotherapy response (ypT0)	0.81	1.05 (0.68–1.63)						
Yes								
No	0.913	0.976 (0.629–1.514)			0.988	0.997 (0.671–1.481)		
Nodal pathologic complete response (ypN0)								
Yes								
No	0.828	1.049 (0.682–1.613)			0.429	0.851 (0.570–1.269)		
SLN radiographic response to NAC[a]								
Yes								
No	0.876	0.489 (0.603–1.274)			0.769	0.949 (0.671–1.343)		
Post-NAC largest SLN size[a]								
<1 cm								
≥1 cm	0.533	1.204 (0.672–2.159)			0.841	1.509 (0.605–1.853)		
Involvement of infraclavicular lymph nodes after NAC								
Yes								
No	0.106	0.531 (0.246–1.145)	0.500	1.441 (0.498–4.167)	**0.023**	0.456 (0.232–0.898)	0.993	1.004 (0.412–2.446)
SLN cumulative dose, Gy								
<60								
≥60	0.299	1.259 (0.815–1.944)			0.563	1.127 (0.751–1.690)		
Adjuvant targeted therapy								
Yes								
No	0.100	0.649 (0.387–1.086)	0.098	0.579 (0.304–1.106)	0.228	0.742 (0.456–1.206)		
Adjuvant endocrine therapy								
Yes								
No	0.369	1.203 (0.804–1.800)			0.293	1.127 (0.840–1.782)		

[a]Confirmed by BUS.

SLN, supraclavicular lymph nodes; NAC, neoadjuvant chemotherapy; SLND, supraclavicular lymph node dissection.

The bold value indicates a *p*-value less than 0.05.

the detection rate in China is less than 30% (Duggan et al., 2021). As a result, the majority of breast cancer patients are diagnosed at the middle or late stages, with frequent metastasis to axillary and supraclavicular lymph nodes. Breast cancer with ISLNM is generally considered an indicator of poor prognosis, and the fifth edition of the AJCC TNM staging system defines it as M1 due to its poor outcome and the likelihood of developing distant metastasis within 1 year (Debois, 1997).

However, with improvements in diagnosis and multidisciplinary approaches, patient outcomes are improving (Brito et al., 2001) were the first to report that patients with breast cancer and ISLNM who received multimodal comprehensive treatment had a 5-year overall survival rate of 41.1%, significantly superior to that of patients with distant metastasis. Similar findings were subsequently reported by Chen et al. (2006), Huang et al. (2007), Fan et al. (2010), Park et al. (2011), Ogino et al. (2011), Dellapasqua et al. (2014), and Noh et al.

TABLE 3 Summary of studies on SLND for breast cancer patients with ISLNM.

Author, year	Study period	No.	No. SLND	Follow-up (years)	5-yr OS (%)	5-yr DFS/PFS/RFS (%)	*P* of SLND + RT VS. RT
Chang (2013)	2005.1–2005.10	29	13	3.92	46.2%	NA	NS
June 2015	1990–2010	111	73	NA	64.2%	56.2%	S
Diao (2022)	2004–2016	173	10	2.83	73%	50%	NS
Kim (2020)	2000–2014	78	35	4.65	68.6%	68.4%	NS
Ma (2020)	2009–2019	99	27	3.00	76.8% (3-yr)	41.4%	NS
Sun (2020)	2010–2019	108	84	6.25	67.8%	30.6%	NS
Ai (2020)	2004–2017	305	146	4.00	73.9%	54.8%	NS; S in stratified analyses
Ai (2020)	2018.1–2018.12	61	25	-	-	-	NS
Liu (2021)	2000–2016	142	104	NA	NA	NA	S for RFS NS for OS and DMFS

SLND, supraclavicular lymph node dissection; ISLNM, ipsilateral supraclavicular lymph node metastasis; OS, overall survival; DFS, disease-free survival; DMFS, distant metastasis-free survival; NS, no significant; S, significant; NA, not available.

(2015), which are comparable to those with stages IIIb/c disease but distinct from those with stage IV disease. As a result, the sixth edition of the AJCC-TNM breast cancer staging system reclassified breast cancer with ISLNM as stage IIIc instead of stage IV, as it is no longer considered to be distant metastasis (Singletary et al., 2002). Consequently, patients with ISLNM but no other evidence of distant metastases should receive therapeutic interventions aimed at achieving a curative outcome; i.e., combined-modality therapy consisting of NAC, targeted therapy, surgery, radiotherapy, and endocrine therapy.

In terms of surgery, there is still controversy regarding the treatment method for the local supraclavicular lymph node region. The latest NCCN guidelines recommend total or partial mastectomy with axillary dissection limited to axillary levels I and II for patients achieving clinical remission, with residual tumors in infraclavicular and supraclavicular regions usually managed through radiation (Gradishar et al., 2022). While the majority of studies indicate that locoregional control rates can exceed 80%, the 5-year OS and DFS rates typically remain at approximately 50%, ranging from 33.3% to 47% and 25%–34%, respectively (Brito et al., 2001; Huang et al., 2007; Park et al., 2011; Chang et al., 2013; Noh et al., 2015; Kim et al., 2020). This may be due to residual tumor leading to tumor development and growth because the dose of supraclavicular radiotherapy did not reach the radical dose, which is limited by surrounding vital organs and blood vessels such as the brachial plexus nerve. Moreover, complications from supraclavicular radiotherapy cannot be ignored. Radioactive nerve damage, especially brachial plexus injury induced by high-dose radiation therapy, may cause allodynia and movement disorder of the affected limb. Other common radiotherapy complications such as radiation dermatitis and radiation pneumonitis can reduce patients' quality of life. In a study by the Texas MD Anderson Cancer Center, 156 patients reached a cumulative SCV dose ≥60 Gy and 120 of them experienced radiation dermatitis grade ≥2 while 13 patients experienced hyperpigmentation grade ≥2. Incidences of such events increase with dose and irradiated volume while radiation dose limits may lead to insufficient doses in treating tumors and failure of local control. Thus, radical excision of tumors in infraclavicular and supraclavicular regions may be a more thorough and effective treatment. Due to the special anatomical location of the SLN and the lack of a unified standard for the extent of lymph node dissection, studies on aggressive locoregional surgery for ISLNM have been unsystematic and sporadic, with no clear recommendation in the NCCN guidelines. The primary question is whether more radical surgery can improve local control, reduce distant metastasis, and prolong survival outcomes for breast cancer patients with ISLNM.

A review of current literature between January 1975 and June 2020 using PubMed and Web of Science databases found 115 studies on breast cancer with ISLNM, but only 10 studies on the role of SLND. The largest study to date, by Ai et al. (2020), included 146 patients who received SLND with RT and 159 patients who received RT alone. They found that SLND with RT was not associated with superior survival overall, but in stratified analyses, patients with non-luminal A tumors and 4-9 positive axillary lymph nodes who underwent SLND with RT had superior OS and DFS compared to those who received RT alone. Jung et al. (2015) also found that the 5-year OS for 73 patients who received local aggressive treatment (SLND + RT) was superior to that of 38 patients who received non-aggressive treatment. However, other studies have reported opposite conclusions. Diao et al. (2022) found that SLND had a limited role, as evidenced by comparable LRRFS rates with and without neck dissection. This conclusion is supported by Kim et al., who reported no improvement in locoregional control or DFS with SLND. Similar trends were observed by Ma et al. (2020), Chang et al. (2013), and Sun et al. (2020). In summary, the available evidence on the role of SLND in the treatment of breast cancer with ISLNM is mixed and further research is needed to determine its effectiveness in improving patient outcomes. The optimal extent of SLND is a topic of debate. The supraclavicular fossa is defined as the IV region, upper part of the V region of the neck, and the entire supraclavicular region. Clinically, SLND involves removing lymph nodes containing adipose tissue in the triangle formed by the posterior border of the sternocleidomastoid muscle, the inferior edge of the omohyoid muscle, and the upper border of the clavicle. The lymph nodes of regions II and III are not reliably visualized as conventional lymphatic drainage from the breasts and are staged as M1 (Sesterhenn et al., 2006). After reviewing the published literature (Table 3), we made a forest plot to summarize the available data (Supplementary Figure S1) and conducted a retrospective study at our center. Our results show a marked increase in 5-year OS and DFS rates to 71.89% and 59.25%, respectively, when compared to historical outcomes. This improvement

in long-term survival may be attributed to the implementation of more aggressive locoregional management strategies, such as infraclavicular lymph node dissection and SLND.

In our study, we categorized axillary surgery into two groups: levels I/II axillary dissection only and aggressive locoregional surgery group including levels I/II/III ± SLND. The results indicate that patients who underwent aggressive locoregional surgery experienced significant benefits in OS, DFS, LRRFS, and DMFS ($p < 0.01$). Further analysis comparing the effects of two types of aggressive locoregional surgery on survival outcomes indicated that the addition of SLND to levels I/II/III surgery did not confer a significant advantage in terms of OS, DFS, LRRFS, or DMFS.

Moreover, the implementation of ultrasound-guided sentinel lymph node biopsy has facilitated precise staging and treatment planning for patients. Our institution achieved a notable proportion of patients with biopsy tissue-confirmed SLN involvement (99.6%). Reasons for excluding biopsy tissue test results typically included initial assessment at an external facility, commencement of systemic therapy prior to imaging or diagnosis of sentinel lymph node disease, presence of imaging adequate for diagnosis (usually involving large, positively identified nodes through positron emission tomography), or anatomically unfavorable location of the node necessitating biopsy tissue testing on a more accessible node.

A significant limitation of existing studies on breast cancer with ISLNM is their failure to specifically examine LRRFS. For example, Kim et al. (2020) conducted a study involving 78 patients with ipsilateral cervical lymph node involvement at presentation and reported a 5-year LRRFS rate of 68%, which is comparable to our findings. However, the low 5-year LRRFS rate may be attributed to the fact that only 83% of patients received radiation fields encompassing the ipsilateral cervical lymph nodes, and only 3% underwent SLND. Our study reports a 5-year LRRFS rate of 66.38%, which may have been influenced by the fact that only 65 (26.0%) patients underwent aggressive locoregional surgery. In order to determine any relationship between clinical features, treatment, and survival, univariate and multivariate analyses of prognostic factors were carried out. In both the univariate and multivariate analyses, nodal surgery as level I/II/III ± SLND surgery was significantly associated with OS and DFS. This result indicates that aggressive locoregional surgery may be a strong prognostic factor.

RT after surgery for breast cancer with ISLNM has provided acceptable in-field regional control rates in previous studies. In regards to radiotherapy dose, trials have established the role of adjuvant breast RT after a lumpectomy using a dose of 50 Gy in 25 fractions (Veronesi et al., 1981; Poortmans et al., 2015), and 50 Gy for elective regional nodal radiation (Poortmans et al., 2015). However, the optimal dose for gross disease and high-risk nodal regions in cN3c-stage breast cancer is unknown. The radiotherapy dose presented in the published literature was mainly 50 Gy, with the fields of radiotherapy chest wall generally including axilla, infraclavicular, and supraclavicular areas. Some clinicians have suggested a dose of 46–50 Gy in 23–25 fractions to the breast/chest wall and regional nodes without an SLN boost, while others have prescribed 50.4 Gy in 28 fractions with an SLN boost ≥54 Gy. Diao et al. (2022) investigated the optimal dose for patients with ISLNM as 156 patients (90%) received a cumulative SLN dose of ≥60 Gy. They found that a cumulative SLN dose of ≥60 Gy was associated with improved OS but not LRRFS or DFS. In our study, 56 patients (22.4%) had an SLN boost and the cumulative dose was ≥60 Gy, and RT was well-tolerated as no grade ≥3 acute or late toxicities were observed.

This study has certain limitations due to its retrospective design. Specifically, the retrospective nature of data collection limits the strength of statistical comparisons and the inferences that can be made. Additionally, the non-randomized retrospective nature of the study precludes random assignment of treatment types, which were instead selected individually based on the patient's condition. This selection of treatment may have been biased, limiting the interpretation of the survival analysis. The study interval was extensive, and the emergence of systemic treatment resulted in some clinical data being absent. While a preliminary evaluation suggested that the enhancement of treatment modalities had a negligible effect on the survival analysis, it may still introduce some degree of deviation. Further large-scale, randomized studies are necessary to evaluate the benefits of neck surgery in patients with ISLNM.

Despite these limitations, this study has several strengths. It is one of the largest examinations of individuals with N3c breast cancer, with a median follow-up duration (5.92 years) that exceeds that of previously reported literature. The cohort was divided into three distinct groups based on the extent of axillary surgery: level I/II axillary dissection, levels I/II/III (infraclavicular) dissection, and level I/II/III + SLND group. Additionally, published literature on this topic was reviewed and summarized in a forest plot. The study aimed to investigate the influence of the extent of surgery on survival outcomes. While more comprehensive stratification requires a larger sample size to accurately identify patients who may benefit from SLND, this study provides valuable insights and directions for future research.

5 Conclusion

Breast cancer with ISLNM may be classified as a locoregional disease that requires a combination of systemic and local therapies. The incorporation of more aggressive regional therapies, including level I/II/III ± SLND surgery, as part of a multimodal treatment strategy, may yield favorable outcomes in terms of recurrence, metastasis, and survival. This approach may give patients with ISLNM the best chance for a positive outcome.

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 contributed to the study conception and design. KF and ZX made the material preparation, data collection and analysis were performed. KF wrote the main manuscript and all authors commented on previous versions of the manuscript. QD and HC prepared Figures 1–3. XW reviewed and edited 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/fmolb.2023.1248410/full#supplementary-material

References

Ai, X., Wang, M., Li, J., Hu, Y., Hou, L., Zheng, X., et al. (2020). Supraclavicular lymph node dissection with radiotherapy versus radiotherapy alone for operable breast cancer with synchronous ipsilateral supraclavicular lymph node metastases: a real-world cohort study. *Gland. Surg.* 9 (2), 329–341. doi:10.21037/gs.2020.03.09

Brito, R. A., Valero, V., Buzdar, A. U., Booser, D. J., Ames, F., Strom, E., et al. (2001). Long-term results of combined-modality therapy for locally advanced breast cancer with ipsilateral supraclavicular metastases: the University of Texas M.D. Anderson Cancer Center experience. *J. Clin. Oncol.* 19 (3), 628–633. doi:10.1200/jco.2001.19.3.628

Chang, X. Z., Yin, J., Sun, J., Zhang, X. H., and Cao, X. C. (2013). A retrospective study of different local treatments in breast cancer patients with synchronous ipsilateral supraclavicular lymph node metastasis. *J. Cancer Res. Ther.* 9, S158–S161. doi:10.4103/0973-1482.122514

Chen, S. C., Chang, H. K., Lin, Y. C., Leung, W. M., Tsai, C. S., Cheung, Y. C., et al. (2006). Prognosis of breast cancer after supraclavicular lymph node metastasis: not a distant metastasis. *Ann. Surg. Oncol.* 13 (11), 1457–1465. doi:10.1245/s10434-006-9012-1

Debois, J. M. (1997). The significance of a supraclavicular node metastasis in patients with breast cancer. A literature review. *Strahlenther Onkol.* 173 (1), 1–12. doi:10.1007/bf03039187

Dellapasqua, S., Bagnardi, V., Balduzzi, A., Iorfida, M., Rotmensz, N., Santillo, B., et al. (2014). Outcomes of patients with breast cancer who present with ipsilateral supraclavicular or internal mammary lymph node metastases. *Clin. Breast Cancer* 14 (1), 53–60. doi:10.1016/j.clbc.2013.09.008

Diao, K., Andring, L. M., Barcenas, C. H., Singh, P., Carisa Le-Petross, H., Reed, V. K., et al. (2022). Contemporary outcomes after multimodality therapy in patients with breast cancer presenting with ipsilateral supraclavicular node involvement. *Int. J. Radiat. Oncol. Biol. Phys.* 112 (1), 66–74. doi:10.1016/j.ijrobp.2021.08.026

Duggan, C., Trapani, D., Ilbawi, A. M., Fidarova, E., Laversanne, M., Curigliano, G., et al. (2021). National health system characteristics, breast cancer stage at diagnosis, and breast cancer mortality: a population-based analysis. *Lancet Oncol.* 22 (11), 1632–1642. doi:10.1016/s1470-2045(21)00462-9

Fan, Y., Xu, B., Liao, Y., Yao, S., and Sun, Y. (2010). A retrospective study of metachronous and synchronous ipsilateral supraclavicular lymph node metastases in breast cancer patients. *Breast* 19 (5), 365–369. doi:10.1016/j.breast.2010.03.022

Gradishar, W. J., Moran, M. S., Abraham, J., Aft, R., Agnese, D., Allison, K. H., et al. (2022). Breast cancer, version 3.2022, NCCN clinical practice guidelines in oncology. *J. Natl. Compr. Canc Netw.* 20 (6), 691–722. doi:10.6004/jnccn.2022.0030

Huang, E. H., Strom, E. A., Valero, V., Fornage, B., Perkins, G. H., Oh, J. L., et al. (2007). Locoregional treatment outcomes for breast cancer patients with ipsilateral supraclavicular metastases at diagnosis. *Int. J. Radiat. Oncol. Biol. Phys.* 67 (2), 490–496. doi:10.1016/j.ijrobp.2006.08.040

Jung, J., Kim, S. S., Ahn, S. D., Lee, S. W., Ahn, S. H., Son, B. H., et al. (2015). Treatment outcome of breast cancer with pathologically proven synchronous ipsilateral supraclavicular lymph node metastases. *J. Breast Cancer* 18 (2), 167–172. doi:10.4048/jbc.2015.18.2.167

Kim, J. S., Kim, K., Shin, K. H., Kim, J. H., Ahn, S. D., Kim, S. S., et al. (2020). Cervical lymph node involvement above the supraclavicular fossa in breast cancer: comparison with stage IIIC (KROG 18-02). *J. Breast Cancer* 23 (2), 194–204. doi:10.4048/jbc.2020.23.e14

Lei, S., Zheng, R., Zhang, S., Wang, S., Chen, R., Sun, K., et al. (2021). Global patterns of breast cancer incidence and mortality: a population-based cancer registry data analysis from 2000 to 2020. *Cancer Commun. (Lond)* 41 (11), 1183–1194. doi:10.1002/cac2.12207

Ma, T., Mao, Y., and Wang, H. (2020). Combined therapy can improve the outcomes of breast cancer with isolated supraclavicular lymph node involvement. *Cancer Manag. Res.* 12, 11857–11869. doi:10.2147/cmar.S276996

Noh, J. M., Kim, K. H., Park, W., Suh, C. O., Huh, S. J., Choi, D. H., et al. (2015). Prognostic significance of nodal involvement region in clinical stage IIIc breast cancer patients who received primary systemic treatment, surgery, and radiotherapy. *Breast* 24 (5), 637–641. doi:10.1016/j.breast.2015.07.016

Ogino, T., Komoike, Y., Ishitobi, M., Motomura, K., Koyama, H., and Inaji, H. (2011). Breast cancer with ipsilateral supraclavicular metastases. *Breast J.* 17 (5), 555–557. doi:10.1111/j.1524-4741.2011.01141.x

Park, H. J., Shin, K. H., Cho, K. H., Park, I. H., Lee, K. S., Ro, J., et al. (2011). Outcomes of positron emission tomography-staged clinical N3 breast cancer treated with neoadjuvant chemotherapy, surgery, and radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* 81 (5), e689–e695. doi:10.1016/j.ijrobp.2010.11.061

Poortmans, P. M., Struikmans, H., and Bartelink, H. (2015). Regional nodal irradiation in early-stage breast cancer. *N. Engl. J. Med.* 373 (19), 1879–1880. doi:10.1056/NEJMc1510505

Sesterhenn, A. M., Albert, U. S., Barth, P. J., Wagner, U., and Werner, J. A. (2006). The status of neck node metastases in breast cancer-loco-regional or distant? *Breast* 15 (2), 181–186. doi:10.1016/j.breast.2005.05.006

Singletary, S. E., Allred, C., Ashley, P., Bassett, L. W., Berry, D., Bland, K. I., et al. (2002). Revision of the American Joint Committee on Cancer staging system for breast cancer. *J. Clin. Oncol.* 20 (17), 3628–3636. doi:10.1200/jco.2002.02.026

Sun, X. F., Wang, Y. J., Huang, T., Niu, L. J., Zhang, Q., and Liu, Z. Z. (2020). Comparison between surgery plus radiotherapy and radiotherapy alone in treating breast cancer patients with ipsilateral supraclavicular lymph node metastasis. *Gland. Surg.* 9 (5), 1513–1520. doi:10.21037/gs-20-691

Veronesi, U., Saccozzi, R., Del Vecchio, M., Banfi, A., Clemente, C., De Lena, M., et al. (1981). Comparing radical mastectomy with quadrantectomy, axillary dissection, and radiotherapy in patients with small cancers of the breast. *N. Engl. J. Med.* 305 (1), 6–11. doi:10.1056/nejm198107023050102

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 24 October 2023
DOI 10.3389/fonc.2023.1264259

Check for updates

OPEN ACCESS

EDITED BY
Li Chen,
Huazhong University of Science and Technology, China

REVIEWED BY
Lu Cao,
Shanghai Jiao Tong University, China
Liu Jing,
Shandong University, China
Xie Su,
Second Affiliated Hospital of Guangxi Medical University, China

*CORRESPONDENCE
Jingfei Ma
jma@mdanderson.org

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

RECEIVED 20 July 2023
ACCEPTED 09 October 2023
PUBLISHED 24 October 2023

CITATION
Panthi B, Mohamed RM, Adrada BE, Boge M, Candelaria RP, Chen H, Hunt KK, Huo L, Hwang K-P, Korkut A, Lane DL, Le-Petross HC, Leung JWT, Litton JK, Pashapoor S, Perez F, Son JB, Sun J, Thompson A, Tripathy D, Valero V, Wei P, White J, Xu Z, Yang W, Zhou Z, Yam C, Rauch GM and Ma J (2023) Longitudinal dynamic contrast-enhanced MRI radiomic models for early prediction of response to neoadjuvant systemic therapy in triple-negative breast cancer.
Front. Oncol. 13:1264259.
doi: 10.3389/fonc.2023.1264259

COPYRIGHT
© 2023 Panthi, Mohamed, Adrada, Boge, Candelaria, Chen, Hunt, Huo, Hwang, Korkut, Lane, Le-Petross, Leung, Litton, Pashapoor, Perez, Son, Sun, Thompson, Tripathy, Valero, Wei, White, Xu, Yang, Zhou, Yam, Rauch and Ma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Longitudinal dynamic contrast-enhanced MRI radiomic models for early prediction of response to neoadjuvant systemic therapy in triple-negative breast cancer

Bikash Panthi[1], Rania M. Mohamed[2], Beatriz E. Adrada[2], Medine Boge[2,3], Rosalind P. Candelaria[2], Huiqin Chen[4], Kelly K. Hunt[5], Lei Huo[6], Ken-Pin Hwang[1], Anil Korkut[7], Deanna L. Lane[2], Huong C. Le-Petross[2], Jessica W. T. Leung[2], Jennifer K. Litton[8], Sanaz Pashapoor[2], Frances Perez[2], Jong Bum Son[1], Jia Sun[4], Alastair Thompson[9], Debu Tripathy[8], Vicente Valero[8], Peng Wei[4], Jason White[8], Zhan Xu[1], Wei Yang[2], Zijian Zhou[1], Clinton Yam[8], Gaiane M. Rauch[2,10†] and Jingfei Ma[1*†]

[1]Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [2]Department of Breast Imaging, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [3]Koc University Hospital, Istanbul, Türkiye, [4]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [5]Department of Breast Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [6]Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [7]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [8]Department of Breast Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [9]Department of Surgery, Baylor College of Medicine, Houston, TX, United States, [10]Department of Abdominal Imaging, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

Early prediction of neoadjuvant systemic therapy (NAST) response for triple-negative breast cancer (TNBC) patients could help oncologists select individualized treatment and avoid toxic effects associated with ineffective therapy in patients unlikely to achieve pathologic complete response (pCR). The objective of this study is to evaluate the performance of radiomic features of the peritumoral and tumoral regions from dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) acquired at different time points of NAST for early treatment response prediction in TNBC. This study included 163 Stage I-III patients with TNBC undergoing NAST as part of a prospective clinical trial (NCT02276443). Peritumoral and tumoral regions of interest were segmented on DCE images at baseline (BL) and after two (C2) and four (C4) cycles of NAST. Ten first-order (FO) radiomic features and 300 gray-level-co-occurrence matrix (GLCM) features were calculated. Area under the receiver operating characteristic curve (AUC) and Wilcoxon rank sum test were used to determine the most predictive features. Multivariate logistic regression models were used for performance assessment. Pearson correlation was used to assess intrareader and interreader variability. Seventy-eight patients (48%) had pCR (52 training, 26 testing), and 85 (52%) had non-pCR (57 training, 28 testing). Forty-six radiomic

features had AUC at least 0.70, and 13 multivariate models had AUC at least 0.75 for training and testing sets. The Pearson correlation showed significant correlation between readers. In conclusion, Radiomic features from DCE-MRI are useful for differentiating pCR and non-pCR. Similarly, predictive radiomic models based on these features can improve early noninvasive treatment response prediction in TNBC patients undergoing NAST.

KEYWORDS

triple-negative breast cancer, dynamic contrast-enhanced MRI, neoadjuvant systemic therapy, radiomic analysis, pathologic complete response

1 Introduction

Triple-negative breast cancer (TNBC) accounts for approximately 10% to 20% of breast cancers but almost 30% to 40% of breast cancer–related deaths (1). TNBC is defined by a profile of negative immunohistochemical staining of the receptors for progesterone, estrogen, and human epidermal growth factor (HER2) and therefore not responsive to endocrine or HER2-targeted therapies (2). In addition, TNBC, when not responsive to chemotherapy, is generally associated with a poor prognosis with a high recurrence rate and a low long-term survival rate (3).

In patients with TNBC, neoadjuvant systemic therapy (NAST) with chemotherapy agents such as anthracyclines, taxanes, and cyclophosphamide, carboplatin plus Food and Drug Administration–approved immunotherapy agents such as pembrolizumab is usually administered before surgery to downstage the tumor (2, 4–8). Patients with a pathologic complete response (pCR) to NAST have favorable long-term overall survival and event-free survival, whereas patients without a pCR to NAST have higher recurrence and mortality rates (9, 10). However, only 50% to 60% of TNBC patients achieve a pCR to NAST. Thus, early prediction of NAST response is crucial to avoid exposure of predicted non-responders to ineffective NAST and unnecessary toxicity of neoadjuvant immunotherapy. Early prediction of NAST response can also help oncologists triage patients to a clinical trial and has the potential to better personalize therapy.

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) and the associated temporal enhancement curves can be used in pharmacokinetic modeling to determine the vascular properties of a tumor, such as the integrity and density of the tumor micro-vessels (11). Volumetric changes along with quantitative and semiquantitative kinetic parameters derived from DCE-MRI have been found to be applicable for breast cancer diagnosis, measurement of breast tumor size, classification of breast tumors, and detection of residual breast cancer after NAST (12–18).

In recent years, several studies have been conducted to assess the role of quantitative radiomic imaging features extracted from MRI images in predicting prognosis and treatment response in patients with breast cancer (19, 20). Wu et al. (21) studied quantitative radiomic features from DCE-MRI extracted before and after one cycle of NAST and demonstrated that these features could be used to develop a clinical biomarker for the prediction of the tumor response to NAST. Other investigators also studied radiomic features from DCE-MRI images for the early prediction of tumor response to NAST (12, 19, 21–24). In addition, radiomic features of breast MRI images have been widely investigated for the noninvasive characterization of tumors (25–28).

Studies suggest that peritumoral features assessed by MRI could provide vital information on the tumor microenvironment, cancer development, chemoresistance, and treatment response (29, 30). While previous studies have shown the potential of radiomic analysis in predicting response to NAST, most of these studies have focused on the tumoral region alone. Furthermore, few studies have explored the performance of radiomic models, specifically in patients with TNBC, and even fewer have evaluated the performance of models based on features extracted from DCE-MRI images obtained at multiple time points during NAST.

In this study, we systematically evaluated the performance of models based on radiomic features of both the peritumoral and tumoral regions from DCE-MRI images at baseline (BL), after two cycles (C2), and after four cycles (C4) of NAST for early prediction of response to NAST in TNBC.

2 Materials and methods

2.1 Patient population

In this institutional review board–approved study, we included 163 patients with biopsy-confirmed stage I-III TNBC enrolled in the prospective clinical trial "ARTEMIS: A Robust TNBC Evaluation FraMework to Improve Survival (NCT02276443)". Informed consent was obtained from all the patients before enrollment in this trial. The inclusion and exclusion criteria are shown in Figure 1. TNBC was defined from standard pathologic assays of biopsy specimens as negative for estrogen receptor and progesterone receptor (<10% of tumor staining) and negative for HER2 (immunohistochemistry score < 3, gene copy number not amplified) (31).

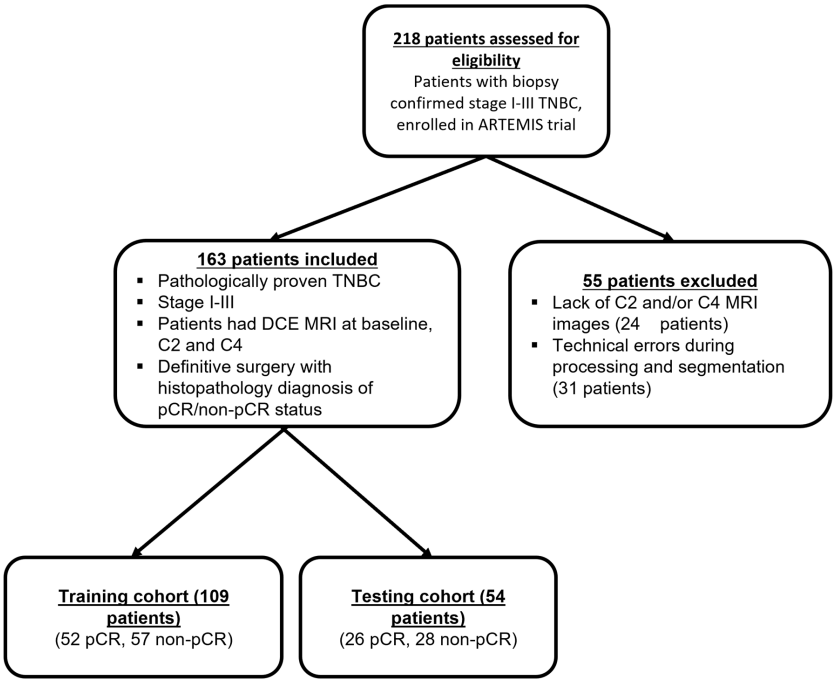


FIGURE 1
Inclusion and exclusion criteria for the study patients.

All patients in this study received NAST consisting of dose-dense doxorubicin and cyclophosphamide for four cycles, followed by paclitaxel every two weeks for four cycles or weekly for a total of 12 doses. All patients underwent DCE-MRI scans at baseline (BL), after 2 cycles(C2) and after 4 cycles(C4) of NAST, followed by surgery after the completion of NAST. Patient demographic data, clinical information, and pathologic findings were obtained from patients' medical records. Patients were classified as, having a pCR or a non-pCR according to the findings on the pathologic review of surgical specimens. pCR was defined as the absence of invasive carcinoma in the breast and axillary lymph nodes.

2.2 DCE-MRI acquisition

For all the patients, DCE-MRI images were acquired on a GE 3.0 T MR750w whole body scanner (Waukesha, WI) with a bilateral 8-channel phased array coil. The patients were imaged in a prone and feet-in-first position. The imaging protocol included a DCE-MRI series based on the differential subsampling with cartesian ordering (DISCO) sequence. Typical MRI scan parameters used for the DISCO acquisition were as follows: field of view = 34 × 34 cm, slice thickness = 3.0 mm, slice spacing = -1.5 mm, flip angle = 12°, repetition time = 7.6 ms, echo time 1/echo time 2 = 1.1/2.3 ms, total acquisition time = 7 minutes, acquisition matrix = 320 ×320, number of acquired slices = 60-115, temporal resolution = 8-15.5 s, receiver bandwidth = ± 166.7 kHz, and number of excitations = 0.69. During the DCE-MRI acquisition, each patient was injected with a single bolus of gadobutrol (Gadovist, Bayer Health Care) contrast agent (0.1 mL/kg at ~2 mL/s followed by a saline flush) after at least one mask phase was obtained.

2.3 Image processing and feature extraction

Manual tumor segmentations, followed by semiautomatic refinement of regions of interest (ROIs), were carried out on 2.5-minute early-phase DCE subtraction images by two fellowship-trained breast radiologists with eight years (MB) and five years (RMM) of experience, respectively. The tumoral region was defined as the region exhibiting contrast enhancement on the DCE images. All the segmentations on BL, C2, and C4 images were performed using an in-house image analysis software program (Image-I v2.0) coded in MATLAB (MathWorks Inc, Natick, MA, USA; RRID: SCR_001622) (31). Additionally, peritumoral regions were automatically generated by expanding the tumor ROIs outward with a fixed thickness of 10 mm (Figure 2) (32). A tumor bed was segmented for the cases with no visible tumor enhancement at C2 and/or C4.

A total of 310 radiomic features per imaging time point were extracted separately from the segmented tumor and peritumoral ROIs using an in-house source code based on MATLAB (Figure 3). Of the 310 features extracted, 10 were the first order (FO) histogram features (minimum; maximum; mean; standard deviation; 1st, 5th, 95th, and 99th percentiles; skewness; and kurtosis), and the remaining 300 features were gray-level co-occurrence matrix (GLCM) features generated as 60 rotation-invariant features obtained from five different gray levels. For each of the five gray levels (8, 16, 32, 64, and 256), the mean, range, and angular variance of the following 20 GLCM features were calculated to generate the 60 rotation-invariant features: autocorrelation, correlation, contrast, cluster prominence, cluster prominence, cluster shade, dissimilarity, energy, homogeneity, maximum probability, sum of squares/

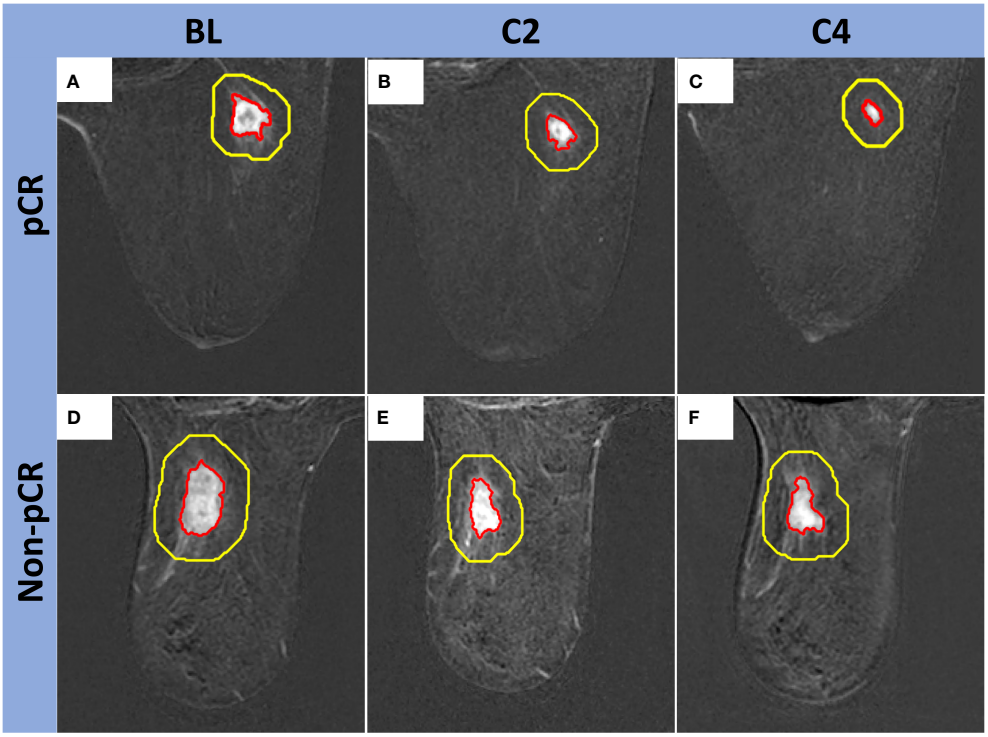


FIGURE 2
Examples of peritumoral regions automatically generated by the expansion of the tumor ROIs outward. **Top row,** Fifty-seven-year-old woman with a right breast TNBC (red contour) that measured 1.7 x 1.5 x 1.5 cm at BL **(A)**, 1.4 x 1.2 x 0.9 cm at C2 **(B)**, and 1.2 x 0.8 x 0.7 cm at C4 **(C)**. Final surgical pathology showed pCR. **Bottom row,** Seventy-eight-year-old woman with a left breast TNBC (red contour) that measured 2.6 x 2.0 x 1.8 cm at BL **(D)**, 2.8 x 1.5 x 1.4 cm at C2 **(E)**, and 2.2 x 1.7 x 1.3 cm at C4 **(F)**. Final surgical pathology showed non-pCR. Peritumoral segmentations are shown with yellow contours. The area between the yellow and red contours (thickness = 10 mm) was used for peritumoral features measurement.

variance, sum average, sum variance, sum entropy, difference variance, difference entropy, information measure of correlation 1, information measure of correlation 2, inverse difference normalized, and inverse difference moment normalized (33).

2.4 Statistical analysis

The absolute differences (ADs) and the relative differences (RDs) of radiomic features between the three imaging time points (C4 vs. BL [C4BL], C4 vs. C2 [C4C2], and C2 vs. baseline [C2BL]) were calculated. The calculated radiomic features and their differences from peritumoral and tumoral regions were compared between the patients with pCR and non-pCR using the Wilcoxon rank sum test and Fisher's exact test. The patient cohort was split into a training set (n=109, 67%) and a testing set (n=54, 33%) in a 2:1 ratio.

For the univariate analysis, area under the receiver operating characteristic curve (AUC) was measured for each radiomic feature from peritumoral and tumoral ROIs at all three imaging time points. Furthermore, AUCs were calculated for the ADs and RDs in the features between these time points. For multivariate analysis, logistic regression with elastic net regularization was performed for texture feature selection. The tuning parameter was optimized by using five-fold cross-validation based on the mean cross-validation AUC. Furthermore, independent testing was performed using the testing data set.

A second set of tumor segmentation was produced independently by a fellowship-trained breast radiologist (SP) with

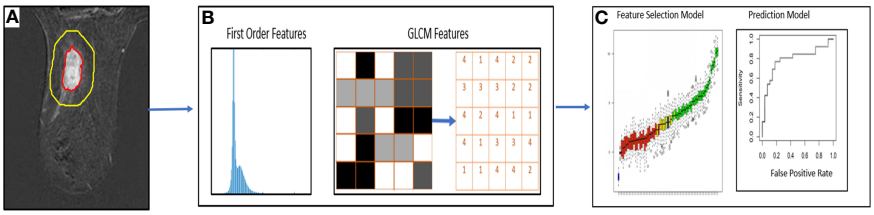


FIGURE 3
Workflow of DCE-MRI radiomic analysis for prediction of pCR. **(A)** Segmentation of regions of interest in DCE-MRI images **(B)** Extraction of first order radiomic features and GLCM features **(C)** Statistical models for feature selection and prediction of pCR.

12 years of experience. The same methodology used to extract features from the initial set of ROIs was applied to extract radiomic features from these new ROIs. The Pearson correlation coefficient was calculated to measure the linear relationship of each feature between two data sets. For the Pearson correlation coefficient, a value of 1 indicates a perfect positive correlation, 0 indicates no correlation, and -1 indicates a perfect negative correlation. Interreader differences were evaluated using the Wilcoxon signed-rank test. The ratio of interreader to intrareader variance was assessed to determine the degree of agreement between two sets of measurements. A lower ratio indicates greater agreement between the readers, suggesting that the measurements are reproducible and consistent.

Statistical analyses were conducted using R software (version 4.0.3, R Development Core Team, Vienna, Austria; RRID: SCR_001905). A p-value less than 0.05 was considered statistically significant.

3 Results

3.1 Patient characteristics

Of the 218 TNBC patients from the ARTEMIS trial who were assessed for eligibility, 55 were excluded from this study, leaving 163 TNBC patients in the final cohort (Figure 1). Patient characteristics are reported in Table 1. Seventy-eight patients (48%) had a pCR, and 85 (52%) had a non-pCR. There were no statistically significant differences in demographic and clinical characteristics between the patients who achieved a pCR and those who did not.

3.2 Univariate analysis

Forty-six radiomic features (21 extracted from the peritumoral region and 25 from the tumoral region) achieved statistical significance in predicting the pCR status of a patient and had AUC at least 0.70 for both the training and testing cohorts (Tables 2, 3). Two tumoral radiomic features (RD-C4BL_percentile 5 and RD-C4BL_percentile 1) had AUC greater than 0.80 for both the training and testing cohorts. The AUCs for the 21 significant features from the peritumoral region ranged from 0.70 to 0.82 for the training cohort and 0.70 to 0.77 for the testing cohort (Table 2). Similarly, the AUCs for the 25 significant features from the tumoral region ranged from 0.70 to 0.84 for the training cohort and 0.70 to 0.81 for the testing cohort (Table 3). None of the 46 features that were significant and had AUC at least 0.70 in both the training and testing cohorts were GLCM features. Additionally, none of the features extracted from BL and C2 had AUC at least 0.70 for both the training and testing cohorts.

3.3 Multivariate analysis

Multivariate analysis identified 13 radiomic models with AUC at least 0.75 for the training and testing cohorts (Table 4).

The multivariate radiomic models with the best AUCs in the testing cohort were based on peritumoral FO features acquired at BL, C2, and C4 along with their ADs and RDs (model 1: AUC in the training cohort, 0.95 [95% confidence interval (CI): 0.91-0.98], AUC in the testing cohort, 0.79 [95% CI: 0.65-0.92]) and a model based on features from both the peritumoral and tumoral regions (model 2: AUC in the training cohort, 0.96 [95% CI: 0.93-0.99], AUC in the testing cohort, 0.78 [95% CI: 0.65-0.91]). The receiver operative characteristic curves for model 1 and model 2 are shown in Figure 4.

The radiomic models based on peritumoral features at BL only, C2 only, a combination of BL and C2, as well as the models based on tumoral features at BL only, a combination of C2 and C4, had AUCs less than 0.70 for testing cohorts. However, the radiomic models based on peritumoral features at C4 only, a combination of BL and C4, a combination of C2 and C4, as well as tumoral features at C2 only, C4 only, a combination of BL and C2, and a combination of BL and C4 had AUCs greater than 0.70 for both training and testing

TABLE 1 Clinical and pathologic characteristics of patients with TNBC undergoing NAST included in the study.

Characteristic	Non-pCR (n = 85) (52%)	pCR (n = 78) (48%)	P value
Median age (range), years	50 (31–78)	48 (23-78)	
T category, n (%)			
T1	12 (14.1)	19 (24.4)	
T2	59 (69.4)	53 (67.9)	
T3	13 (15.3)	5 (6.4)	
T4	1 (1.2)	1 (1.3)	
N category, n (%)			0.850
N0	55 (64.7)	53 (68.0)	
N1	19 (22.4)	17 (21.8)	
N2	4 (4.7)	3 (3.8)	
N3	7 (8.2)	5 (6.4)	
Clinical stage, n (%)			0.690
I	11 (12.9)	11 (14.1)	
II	59 (69.4)	55 (70.5)	
III	15 (17.7)	12 (15.4)	
Histologic type, n (%)			0.741
Invasive ductal carcinoma	76 (89.4)	73 (93.6)	
Metaplastic	8 (9.4)	5 (6.4)	
Apocrine	1 (1.2)	0 (0)	
Type of surgery, n (%)			0.575
Breast-conserving surgery	47 (55.3)	49 (62.8)	
Total mastectomy	38 (44.7)	29 (37.2)	

10.3389/fonc.2023.1264259

TABLE 2 Significant features (AUC ≥ 0.70 for both training and testing sets) extracted from the peritumoral region as identified from univariate analysis.

Feature	Training (n = 109)		Testing (n = 54)	
	AUC	95% CI	AUC	95% CI
C4				
Peritumoral_DCE_C4_FO_Percentile.95	0.80	0.72-0.88	0.71	0.56-0.86
Peritumoral_DCE_C4_FO_Maximum	0.79	0.71-0.88	0.70	0.55-0.85
Peritumoral_DCE_C4_FO_Percentile.99	0.79	0.70-0.88	0.70	0.55-0.85
Peritumoral_DCE_C4_FO_Mean	0.78	0.70-0.87	0.73	0.59-0.87
Peritumoral_DCE_C4_FO_Percentile.5	0.70	0.60-0.80	0.73	0.60-0.87
AD C4BL				
Peritumoral_DCE_AD-C4BL_FO_Percentile.99	0.79	0.71-0.88	0.74	0.61-0.88
Peritumoral_DCE_AD-C4BL_FO_Maximum	0.78	0.69-0.86	0.72	0.58-0.87
Peritumoral_DCE_AD-C4BL_FO_Percentile.95	0.78	0.69-0.86	0.76	0.62-0.90
Peritumoral_DCE_AD-C4BL_FO_Mean	0.76	0.68-0.85	0.77	0.63-0.91
Peritumoral_DCE_AD-C4BL_FO_Percentile.5	0.70	0.61-0.80	0.70	0.56-0.85
RD C2BL				
Peritumoral_DCE_RD-C2BL_FO_Maximum	0.70	0.60-0.80	0.70	0.56-0.85
RD C4BL				
Peritumoral_DCE_RD-C4BL_FO_Percentile.95	0.82	0.74-0.90	0.76	0.63-0.90
Peritumoral_DCE_RD-C4BL_FO_Percentile.99	0.82	0.74-0.90	0.76	0.62-0.89
Peritumoral_DCE_RD-C4BL_FO_Maximum	0.82	0.74-0.90	0.75	0.61-0.89
Peritumoral_DCE_RD-C4BL_FO_Mean	0.81	0.73-0.90	0.77	0.64-0.90
Peritumoral_DCE_RD-C4BL_FO_Standard.Deviation	0.76	0.66-0.85	0.70	0.54-0.84
Peritumoral_DCE_RD-C4BL_FO_Percentile.5	0.75	0.66-0.84	0.70	0.55-0.85
RD C4C2				
Peritumoral_DCE_RD-C4C2_FO_Percentile.95	0.82	0.74-0.90	0.72	0.58-0.86
Peritumoral_DCE_RD-C4C2_FO_Percentile.99	0.79	0.70-0.88	0.72	0.58-0.86
Peritumoral_DCE_RD-C4C2_FO_Mean	0.79	0.70-0.88	0.72	0.58-0.86
Peritumoral_DCE_RD-C4C2_FO_Maximum	0.76	0.66-0.85	0.70	0.55-0.84

cohorts (Table 5). The radiomic model based on tumoral features from BL, C2 and C4 had an AUC of 0.99 [95% CI: 0.99-1.00] in the training cohort and an AUC of 0.73 [95% CI: 0.59-0.87] in the testing cohort. The radiomic models based on peritumoral features from BL, C2 and C4 with/without tumoral features from BL, C2, and C4 had AUCs greater than 0.75 for both training and testing cohort (Table 4).

3.4 Interreader variability analysis

The Pearson correlation analysis showed that both GLCM and FO features exhibited strong interreader correlation. Specifically, the correlation coefficient was greater than 0.8 for 83% (25/30) of original FO features, 60% (18/30) of AD FO features, and 67% (20/30) of RD FO features. For GLCM, the correlation coefficient was greater than 0.8 for 100% (300/300) of original and AD GLCM features and 90% (271/300) of RD GLCM features. For the ratio of interreader to intrareader variance for the FO features, the mean was 5.6×10^{-3} and the median was zero, indicating high agreement between the interreader datasets. The standard deviation was 0.02, indicating a slight variation in agreement between the readers across FO features. The range was relatively narrow, 0 to 0.08, suggesting that most features had a high level of agreement between the readers. There was no variability in the measurements between the readers for any of the GLCM features. The ratios were all zero, indicating a high level of agreement and consistency between the readers.

10.3389/fonc.2023.1264259

TABLE 3 Significant features (AUC ≥ 0.70 for both training and testing sets) extracted from the tumoral region as identified from univariate analysis.

Feature	Training (n = 109)		Testing (n = 54)	
	AUC	95% CI	AUC	95% CI
C4				
Tumor_DCE_C4_FO_Mean	0.81	0.73-0.89	0.72	0.57-0.86
Tumor_DCE_C4_FO_Percentile.5	0.80	0.72-0.88	0.71	0.57-0.86
Tumor_DCE_C4_FO_Percentile.1	0.79	0.70-0.88	0.74	0.60-0.88
Tumor_DCE_C4_FO_Minimum	0.74	0.64-0.83	0.77	0.63-0.90
AD C4BL				
Tumor_DCE_AD-C4BL_FO_Percentile.99	0.80	0.71-0.88	0.73	0.59-0.87
Tumor_DCE_AD-C4BL_FO_Percentile.95	0.79	0.70-0.87	0.76	0.62-0.90
Tumor_DCE_AD-C4BL_FO_Percentile.5	0.79	0.70-0.87	0.81	0.69-0.94
Tumor_DCE_AD-C4BL_FO_Mean	0.79	0.70-0.87	0.77	0.63-0.90
Tumor_DCE_AD-C4BL_FO_Percentile.1	0.78	0.70-0.87	0.81	0.69-0.93
Tumor_DCE_AD-C4BL_FO_Maximum	0.78	0.70-0.87	0.72	0.58-0.87
Tumor_DCE_AD-C4BL_FO_Minimum	0.71	0.61-0.81	0.76	0.62-0.90
AD C4C2				
Tumor_DCE_AD-C4C2_FO_Percentile.1	0.76	0.68-0.85	0.71	0.57-0.85
RD C2BL				
Tumor_DCE_RD-C2BL_FO_Maximum	0.70	0.60-0.80	0.71	0.57-0.85
RD C4BL				
Tumor_DCE_RD-C4BL_FO_Mean	0.84	0.77-0.92	0.78	0.64-0.91
Tumor_DCE_RD-C4BL_FO_Percentile.95	0.83	0.76-0.91	0.75	0.62-0.89
Tumor_DCE_RD-C4BL_FO_Maximum	0.83	0.76-0.91	0.75	0.61-0.89
Tumor_DCE_RD-C4BL_FO_Percentile.99	0.83	0.76-0.91	0.75	0.61-0.89
Tumor_DCE_RD-C4BL_FO_Percentile.5	0.83	0.75-0.90	0.80	0.68-0.93
Tumor_DCE_RD-C4BL_FO_Percentile.1	0.82	0.74-0.90	0.81	0.69-0.93
Tumor_DCE_RD-C4BL_FO_Minimum	0.79	0.70-0.87	0.79	0.65-0.92
RD C4C2				
Tumor_DCE_RD-C4C2_FO_Mean	0.84	0.77-0.92	0.71	0.57-0.85
Tumor_DCE_RD-C4C2_FO_Percentile.5	0.83	0.75-0.90	0.77	0.65-0.90
Tumor_DCE_RD-C4C2_FO_Percentile.1	0.80	0.72-0.89	0.76	0.63-0.89
Tumor_DCE_RD-C4C2_FO_Percentile.99	0.79	0.71-0.88	0.70	0.56-0.85
Tumor_DCE_RD-C4C2_FO_Minimum	0.70	0.59-0.80	0.76	0.62-0.89

4 Discussion

In this study, we used longitudinal DCE-MRI images obtained before the start of doxorubicin and cyclophosphamide-based NAST and after two and four cycles of NAST to assess radiomic features from the peritumoral and tumoral regions for early prediction of NAST response in TNBC patients. In the univariate analysis, we identified 46 radiomic features able to predict pCR with an AUC at least 0.70 for both the training and testing cohorts. Furthermore, in the multivariate analysis, we found that 13 multivariate radiomic models had AUC at least 0.75 for both the training and testing cohorts for early prediction of NAST response in TNBC.

Our results revealed that FO radiomic features from DCE-MRI were better predictors of treatment response than GLCM features. None of the 300 GLCM features had AUC at least 0.70 for both the training and testing cohorts; thus, their usefulness in predicting NAST response could not be established. The interreader variability analysis showed high reliability and reproducibility with a slightly

10.3389/fonc.2023.1264259

TABLE 4 Testing and training AUC for the 13 logistic regression models that achieved AUC at least 0.75 for both training and testing sets using radiomic features from BL, C2, and C4.

Radiomic model	Training (n = 109)		Testing (n = 54)	
	AUC	95% CI	AUC	95% CI
Peritumoral_DCE_AD-C4BL_FO	0.79	0.71-0.88	0.76	0.62-0.90
Peritumoral_DCE_RD-C4BL_FO	0.88	0.82-0.94	0.76	0.63-0.90
Peritumoral_DCE_BL_C2_C4_FO	0.95	0.92-0.99	0.76	0.62-0.90
Peritumoral_DCE_BL_C2_C4_AD-C2BL_RD-C2BL_AD-C4BL_RD-C4BL_AD-C4C2_RD-C4C2_FO	0.95	0.91-0.98	0.79	0.65-0.92
Tumoral_DCE_AD-C4BL_FO	0.82	0.74-0.90	0.76	0.63-0.90
Tumoral_DCE_RD-C4BL_FO	0.87	0.80-0.93	0.78	0.64-0.91
Tumoral_DCE_BL_C4_AD-C4BL_RD-C4BL_FO	0.88	0.81-0.94	0.75	0.61-0.90
Peritumoral_Tumoral_DCE_AD-C4BL_FO	0.79	0.71-0.88	0.76	0.62-0.89
Peritumoral_Tumoral_DCE_RD-C4BL_FO	0.89	0.84-0.95	0.77	0.64-0.91
Peritumoral_Tumoral_DCE_BL_C4_AD-C4BL_RD-C4BL_FO	0.91	0.86-0.96	0.75	0.61-0.90
Peritumoral_Tumoral_DCE_C2_C4_AD-C4C2_RD-C4C2_FO	0.93	0.88-0.98	0.75	0.61-0.89
Peritumoral_Tumoral_DCE_BL_C2_C4_FO	0.97	0.95-0.99	0.76	0.62-0.90
Peritumoral_Tumoral_DCE_BL_C2_C4_AD-C2BL_RD-C2BL_AD-C4BL_RD-C4BL_AD-C4C2_RD-C4C2_FO	0.96	0.93-0.99	0.78	0.65-0.91

better interreader agreement in GLCM features than in FO features. GLCM features capture texture information of the image, which is less subjective and more reproducible across readers than the FO features, which are mostly visual features. Furthermore, the features extracted from baseline DCE-MRI images and the models using features only from baseline DCE-MRI images showed poor performance in terms of AUC. Similar to this finding, Panthi et al. have previously reported that the tumor size measurements (longest dimension, tumor volume, enhanced tumor volume and functional tumor volume) extracted from DCE-MRI images at baseline showed poor performance. The tumor measurements extracted at C2, C4, and their relative differences (C2 vs BL and C4 vs BL) showed good correlation with the treatment response with a maximum AUC of 0.84 [95% CI: 0.76-0.92] for functional tumor volume at C4 (31). The possible explanation of these findings is that tumor biology can evolve over the course of treatment. The radiomic features measured after the initiation of NAST may capture these changes in tumor biology better than features

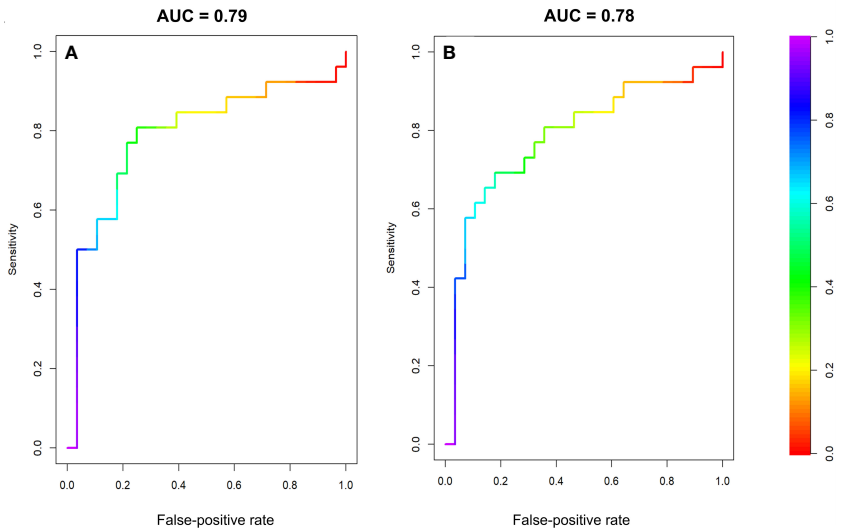


FIGURE 4
Receiver operator characteristic curves for the testing datasets of two multivariate models. **(A)** Model 1, using FO features from the peritumoral region at BL, C2, and C4 and the differences between these time points. **(B)** Model 2, using FO features from both the peritumoral and tumoral regions BL, C2, and C4 and the differences between these time points.

TABLE 5 Logistic regression models that achieved AUC > 0.70 for both training and testing sets using first order radiomic features from BL, C2, and C4.

Radiomic model	Training (n = 109)		Testing (n = 54)	
	AUC	95% CI	AUC	95% CI
Peritumoral_DCE_C4_FO	0.86	0.79-0.93	0.71	0.57-0.85
Peritumoral_DCE_BL_C4_FO	0.90	0.84-0.96	0.73	0.59-0.87
Peritumoral_DCE_C2_C4_FO	0.93	0.88-0.97	0.74	0.60-0.88
Tumoral_DCE_C2_FO	0.87	0.80-0.94	0.71	0.56-0.85
Tumoral_DCE_C4_FO	0.81	0.73-0.89	0.71	0.57-0.85
Tumoral_DCE_BL_C2_ FO	0.81	0.73-0.89	0.72	0.58-0.86
Tumoral_DCE_BL_C4_FO	0.89	0.83-0.95	0.74	0.61-0.88

measured at baseline, contributing to the better predictive performance of the features measured after NAST initiation.

In our study, almost 50% (6/13) of the radiomic models with AUC at least 0.75 for both the training and testing cohorts were based on a combination of peritumoral and tumoral features. The inclusion of the peritumoral features along with the tumoral features may have improved the performance of radiomic models by capturing the infiltrative tumor margins and information about the tumor microenvironment. Additionally, peritumoral features have the potential to aid in predicting treatment response, as tumor behavior heterogeneity likely reflects variation both of the tumor and surrounding environment.

MRI-based radiomic analysis has shown promise in predicting treatment response in breast cancer (12, 19, 22, 23, 34). Cain et al. used pretreatment tumor features from 151 patients with TNBC and HER2-enriched breast cancer in a multivariate machine learning model and showed an AUC of 0.707 for pCR prediction (35). In a study including 83 breast cancer patients, Pesapane et al. extracted 136 representative radiomic features of the tumoral region from pretreatment T1-weighted contrast-enhanced MRI images to predict the response to NAST. Their radiomic model had an AUC of 0.64 (95% CI, 0.51-0.75), which increased to 0.83 (95% CI, 0.73-0.92) after the investigators combined the radiomic model with clinical and biological features (36). Similarly, Fan et al. performed a radiomic analysis of pretreatment DCE-MRI images in 57 patients and demonstrated that combining features from the tumoral region and background parenchyma significantly improved the performance of their radiomic model (19). Li et al, in a study of 33 patients, compared AUCs for pretreatment images and images obtained after one cycle of NAST and observed better predictive AUCs after one cycle of treatment (37). In contrast to the aforementioned studies, we created multivariate radiomic models based on DCE-MRI images at three imaging time points to extract features from peritumoral as well as tumoral regions in a much larger cohort limited to patients with TNBC.

Braman et al. conducted a study including 117 breast cancer patients (with hormone-receptor-positive, HER2-negative, triple-negative, and HER2-positive disease) and showed that combined peritumoral and tumoral radiomic features obtained at baseline could be utilized to predict pCR to neoadjuvant chemotherapy (22). These authors reported an AUC of 0.74 for the testing cohort, which was slightly lower than the AUC of 0.79 for the testing cohort in our study, which had a larger patient pool, was limited to patients with TNBC, and was based on radiomic features obtained at multiple time points. Caballo et al. (38) used a four-dimensional machine learning radiomics approach to assess breast cancer response to NAST. They extracted 348 features from peritumoral and tumoral regions from DCE-MRI images to develop predictive radiomic models. These authors reported a multivariate model with an AUC of 0.71 for the data set of 251 patients (107 with luminal A subtype, 47 with luminal B subtype, 25 with HER2-enriched subtype, and 72 with TNBC). Furthermore, their study showed that the predictive performance improved when the radiomic models were tailored to specific subtypes of breast cancer. Caballo et al. found an AUC of 0.80 in 72 TNBC patients (38), which is comparable to the testing-group AUC of 0.79 in our study of 163 TNBC patients.

Fan et al. reported a study in which clinical features and radiomic features from the tumoral region before and after two cycles of treatment were acquired for 114 patients with primary breast cancer. These authors observed an AUC of 0.57 for the models based on pretreatment features and an AUC of 0.77 for the model based on posttreatment features. The combined model based on pretreatment features, posttreatment features, relative net feature change, Jacobian maps, and clinical features had an AUC of 0.81 (39). Our findings are consistent with their findings in that our findings showed improvement in the predictive performance with a combined model with features from BL, C2, C4, and their net feature changes.

Our study has several limitations. First, our study was conducted on a single scanner platform at a single institution, and therefore our findings will need to be validated on other scanner platforms and at different institutions. Second, although our study is the largest radiomic analysis reported to date for treatment response prediction in TNBC patients, additional validation may be necessary in a larger study with better statistical power and was only tested on patients with preoperative AC chemotherapy. Finally, the peritumoral regions were uniformly defined with a fixed dilation of 10 mm around the tumor ROIs for all patients, regardless of tumor size. Future studies

with different dilation radii based on individual tumor properties can offer a more comprehensive understanding of the role of the peritumoral region in response prediction.

5 Conclusions

In summary, our study shows that radiomic models based on peritumoral and tumoral features from longitudinal DCE-MRI images were able to serve as noninvasive biomarkers for early prediction of NAST response in patients with early-stage TNBC. Radiomic features extracted after four cycles of treatment and their change relative to baseline were better predictors of response than those from baseline only. Further, FO radiomic features showed better predictive performance than GLCM features.

Data availability statement

The datasets presented in this article are not readily available because we are not allowed to share patient data. Requests to access the datasets should be directed to jma@mdanderson.org.

Ethics statement

The studies involving humans were approved by The UT MD Anderson Cancer Center IRB. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

BP: Conceptualization, Formal Analysis, Methodology, Writing – original draft. RMM: Methodology, Writing – review & editing. BA: Data curation, Methodology, Writing – review & editing. MB: Formal Analysis, Methodology, Writing – review & editing. RC: Data curation, Methodology, Writing – review & editing. HC: Formal Analysis, Writing – review & editing. KKH: Writing – review & editing. LH: Writing – review & editing. K-PH: Writing – review & editing. AK: Writing – review & editing. DL: Writing – review & editing. HL-P: Writing – review & editing. JWTL: Writing – review & editing. JKL: Writing – review & editing. SP: Formal Analysis, Methodology, Writing – review & editing. FP: Writing – review & editing. JBS: Formal Analysis, Methodology, Writing – review & editing. JS: Formal Analysis, Writing – review & editing. AT: Writing – review & editing. DT: Writing – review & editing. VV: Writing – review & editing. PW: Writing – review & editing. JW: Writing – review & editing. ZX: Formal Analysis, Methodology, Writing – review & editing. WY: Writing – review & editing. ZZ: Formal Analysis, Methodology, Writing – review & editing. CY: Writing – review & editing. GR: Conceptualization, Formal Analysis, Methodology, Supervision, Writing – review & editing. JM: Formal Analysis, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Supported by the NIH/NCI under award number P30CA016672 and the resources from the Department of Biostatistics Resource Group were used.

Acknowledgments

We thank Stephanie Deming, Research Medical Library, MD Anderson Cancer Center, for editing the manuscript. This study was supported by the MD Anderson Moon Shots Program and the Robert D. Moreton Distinguished Chair Funds in Diagnostic Radiology.

Conflict of interest

KKH serves on the medical advisory boards for ArmadaHealth and AstraZeneca and receives research funding from Cairn Surgical, Eli Lilly and Company, and Lumicell. K-PH is currently receiving research funding from Siemens Healthineers and has received research funding from GE. JKL received grant or research support from Novartis, Medivation/Pfizer, Genentech, GSK, EMD-Serono, AstraZeneca, Medimmune, Zenith, and Merck; participated in the Speaker's Bureau for MedLearning, Physicians' Education Resource, Prime Oncology, Medscape, Clinical Care Options, and Medpage and receives royalty from UpToDate and Certis. DT declares research contracts with Pfizer, Novartis, and Ployphor and is a consultant of AstraZeneca, GlaxoSmithKline, OncoPep, Gilead, Novartis, Pfizer, Personalis, and Sermonix. WY receives royalties from Elsevier. GR receives research funding from GE Healthcare. JM is an inventor of United States patents licensed to Siemens Healthineers and GE Healthcare and is a consultant for C4 Imaging.

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

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

1. Wu Q, Siddharth S, Sharma D. Triple negative breast cancer: a mountain yet to be scaled despite the triumphs. *Cancers (Basel).* (2021) 13(15):3697. doi: 10.3390/cancers13153697

2. Pandy JGP, Balolong-Garcia JC, Cruz-Ordinario MVB, Que FVF. Triple negative breast cancer and platinum-based systemic treatment: a meta-analysis and systematic review. *BMC Cancer.* (2019) 19(1):1065. doi: 10.1186/s12885-019-6253-5

3. Liedtke C, Mazouni C, Hess KR, Andre F, Tordai A, Mejia JA, et al. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J Clin Oncol* (2008) 26(8):1275–81. doi: 10.1200/JCO.2007.14.4147

4. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. *N Engl J Med* (2010) 363(20):1938–48. doi: 10.1056/NEJMra1001389

5. Park JH, Ahn JH, Kim SB. How shall we treat early triple-negative breast cancer (TNBC): from the current standard to upcoming immuno-molecular strategies. *ESMO Open* (2018) 3(Suppl 1):e000357. doi: 10.1136/esmoopen-2018-000357

6. Sharma P, Connolly RM, Roussos Torres ET, Thompson A. Best foot forward: neoadjuvant systemic therapy as standard of care in triple-negative and HER2-positive breast cancer. *Am Soc Clin Oncol Educ Book.* (2020) 40:1–16. doi: 10.1200/EDBK_281381

7. Thompson AM, Moulder-Thompson SL. Neoadjuvant treatment of breast cancer. *Ann Oncol* (2012) 23 Suppl 10(Suppl 10):x231–6. doi: 10.1093/annonc/mds324

8. Schmid P, Cortes J, Pusztai L, McArthur H, Kümmel S, Bergh J, et al. Pembrolizumab for early triple-negative breast cancer. *N Engl J Med* (2020) 382 (9):810–21. doi: 10.1056/NEJMoa1910549

9. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. *Cancer* (2007) 109 (9):1721–8. doi: 10.1002/cncr.22618

10. Spring LM, Fell G, Arfe A, Sharma C, Greenup R, Reynolds KL, et al. Pathologic complete response after neoadjuvant chemotherapy and impact on breast cancer recurrence and survival: a comprehensive meta-analysis. *Clin Cancer Res* (2020) 26 (12):2838–48. doi: 10.1158/1078-0432.CCR-19-3492

11. Yankeelov TE, Gore JC. Dynamic contrast enhanced magnetic resonance imaging in oncology: theory, data acquisition, analysis, and examples. *Curr Med Imaging Rev* (2009) 3(2):91–107. doi: 10.2174/157340507780619179

12. Fusco R, Granata V, Maio F, Sansone M, Petrillo A. Textural radiomic features and time-intensity curve data analysis by dynamic contrast-enhanced MRI for early prediction of breast cancer therapy response: preliminary data. *Eur Radiol Exp* (2020) 4 (1):8. doi: 10.1186/s41747-019-0141-2

13. Peng S, Chen L, Tao J, Liu J, Zhu W, Liu H, et al. Radiomics analysis of multi-phase DCE-MRI in predicting tumor response to neoadjuvant therapy in breast cancer. *Diagnostics (Basel)* (2021) 11(11):2086. doi: 10.3390/diagnostics11112086

14. Peng Y, Cheng Z, Gong C, Zheng C, Zhang X, Wu Z, et al. Pretreatment DCE-MRI-based deep learning outperforms radiomics analysis in predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer. *Front Oncol* (2022) 12:846775. doi: 10.3389/fonc.2022.846775

15. Rosen EL, Blackwell KL, Baker JA, Soo MS, Bentley RC, Yu D, et al. Accuracy of MRI in the detection of residual breast cancer after neoadjuvant chemotherapy. *AJR Am J Roentgenol.* (2003) 181(5):1275–82. doi: 10.2214/ajr.181.5.1811275

16. Loo CE, Straver ME, Rodenhuis S, Muller SH, Wesseling J, Vrancken Peeters MJ, et al. Magnetic resonance imaging response monitoring of breast cancer during neoadjuvant chemotherapy: relevance of breast cancer subtype. *J Clin Oncol* (2011) 29(6):660–6. doi: 10.1200/JCO.2010.31.1258

17. Zhang CC, Yan Z, Giddabasappa A, Lappin PB, Painter CL, Zhang Q, et al. Comparison of dynamic contrast-enhanced MR, ultrasound and optical imaging modalities to evaluate the antiangiogenic effect of PF-03084014 and sunitinib. *Cancer Med* (2014) 3(3):462–71. doi: 10.1002/cam4.215

18. Zhou Z, Adrada BE, Candelaria RP, Elshafeey NA, Boge M, Mohamed RM, et al. Prediction of pathologic complete response to neoadjuvant systemic therapy in triple negative breast cancer using deep learning on multiparametric MRI. *Sci Rep* (2023) 13 (1):1171. doi: 10.1038/s41598-023-27518-2

19. Fan M, Wu G, Cheng H, Zhang J, Shao G, Li L. Radiomic analysis of DCE-MRI for prediction of response to neoadjuvant chemotherapy in breast cancer patients. *Eur J Radiol* (2017) 94:140–7. doi: 10.1016/j.ejrad.2017.06.019

20. Kim S, Kim MJ, Kim EK, Yoon JH, Park VY. MRI radiomic features: association with disease-free survival in patients with triple-negative breast cancer. *Sci Rep* (2020) 10(1):3750. doi: 10.1038/s41598-020-60822-9

21. Wu J, Gong G, Cui Y, Li R. Intratumor partitioning and texture analysis of dynamic contrast-enhanced (DCE)-MRI identifies relevant tumor subregions to predict pathological response of breast cancer to neoadjuvant chemotherapy. *J Magn Reson Imaging.* (2016) 44(5):1107–15. doi: 10.1002/jmri.25279

22. Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res* (2017) 19(1):57. doi: 10.1186/s13058-017-0846-1

23. Kim Y, Kim SH, Song BJ, Kang BJ, Yim KI, Lee A, et al. Early prediction of response to neoadjuvant chemotherapy using dynamic contrast-enhanced MRI and ultrasound in breast cancer. *Korean J Radiol* (2018) 19(4):682–91. doi: 10.3348/kjr.2018.19.4.682

24. Thibault G, Tudorica A, Afzal A, Chui SY, Naik A, Troxell ML, et al. DCE-MRI texture features for early prediction of breast cancer therapy response. *Tomography* (2017) 3(1):23–32. doi: 10.18383/j.tom.2016.00241

25. Guo W, Li H, Zhu Y, Lan L, Yang S, Drukker K, et al. Prediction of clinical phenotypes in invasive breast carcinomas from the integration of radiomics and genomics data. *J Med Imaging (Bellingham).* (2015) 2(4):041007. doi: 10.1117/1.JMI.2.4.041007

26. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest.* (2011) 121(7):2750–67. doi: 10.1172/JCI45014

27. Leithner D, Horvat JV, Marino MA, Bernard-Davila B, Jochelson MS, Ochoa-Albiztegui RE, et al. Radiomic signatures with contrast-enhanced magnetic resonance imaging for the assessment of breast cancer receptor status and molecular subtypes: initial results. *Breast Cancer Res* (2019) 21(1):106. doi: 10.1186/s13058-019-1187-z

28. Xie T, Zhao Q, Fu C, Bai Q, Zhou X, Li L, et al. Differentiation of triple-negative breast cancer from other subtypes through whole-tumor histogram analysis on multiparametric MR imaging. *Eur Radiol* (2019) 29(5):2535–44. doi: 10.1007/s00330-018-5804-5

29. Gillies RJ, Raghunand N, Karczmar GS, Bhujwalla ZM. MRI of the tumor microenvironment. *J Magn Reson Imaging.* (2002) 16(4):430–50. doi: 10.1002/jmri.10181

30. Hussain L, Huang P, Nguyen T, Lone KJ, Ali A, Khan MS, et al. Machine learning classification of texture features of MRI breast tumor and peri-tumor of combined pre- and early treatment predicts pathologic complete response. *BioMed Eng Online.* (2021) 20(1):63. doi: 10.1186/s12938-021-00899-z

31. Panthi B, Adrada BE, Candelaria RP, Guirguis MS, Yam C, Boge M, et al. Assessment of response to neoadjuvant systemic treatment in triple-negative breast cancer using functional tumor volumes from longitudinal dynamic contrast-enhanced MRI. *Cancers* (2023) 15(4):1025. doi: 10.3390/cancers15041025

32. Li C, Lu N, He Z, Tan Y, Liu Y, Chen Y, et al. A noninvasive tool based on magnetic resonance imaging radiomics for the preoperative prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer. *Ann Surg Oncol* (2022) 29(12):7685–93. doi: 10.1245/s10434-022-12034-w

33. Hwang K-P, Elshafeey NA, Kotrotsou A, Chen H, Son JB, Boge M, et al. A radiomics model based on synthetic MRI acquisition for predicting neoadjuvant systemic treatment response in triple-negative breast cancer. *Radiol: Imaging Cancer* (2023) 5(4):e230009. doi: 10.1148/rycan.230009

34. Choudhery S, Gomez-Cardona D, Favazza CP, Hoskin TL, Haddad TC, Goetz MP, et al. MRI radiomics for assessment of molecular subtype, pathological complete response, and residual cancer burden in breast cancer patients treated with neoadjuvant chemotherapy. *Acad Radiol* (2022) 29 Suppl 1(Suppl 1):S145–S54. doi: 10.1016/j.acra.2020.10.020

35. Cain EH, Saha A, Harowicz MR, Marks JR, Marcom PK, Mazurowski MA. Multivariate machine learning models for prediction of pathologic response to neoadjuvant therapy in breast cancer using MRI features: a study using an independent validation set. *Breast Cancer Res Treat* (2019) 173(2):455–63. doi: 10.1007/s10549-018-4990-9

36. Pesapane F, Rotili A, Botta F, Raimondi S, Bianchini L, Corso F, et al. Radiomics of MRI for the prediction of the pathological response to neoadjuvant chemotherapy in breast cancer patients: a single referral centre analysis. *Cancers (Basel).* (2021) 13 (17):4271. doi: 10.3390/cancers13174271

37. Li X, Abramson RG, Arlinghaus LR, Kang H, Chakravarthy AB, Abramson VG, et al. Multiparametric magnetic resonance imaging for predicting pathological response after the first cycle of neoadjuvant chemotherapy in breast cancer. *Invest Radiol* (2015) 50(4):195–204. doi: 10.1097/RLI.0000000000000100

38. Caballo M, Sanderink WBG, Han L, Gao Y, Athanasiou A, Mann RM. Four-dimensional machine learning radiomics for the pretreatment assessment of breast cancer pathologic complete response to neoadjuvant chemotherapy in dynamic contrast-enhanced MRI. *J Magn Reson Imaging.* (2023) 57(1):97–110. doi: 10.1002/jmri.28273

39. Fan M, Chen H, You C, Liu L, Gu Y, Peng W, et al. Radiomics of tumor heterogeneity in longitudinal dynamic contrast-enhanced magnetic resonance imaging for predicting response to neoadjuvant chemotherapy in breast cancer. *Front Mol Biosci* (2021) 8:622219. doi: 10.3389/fmolb.2021.622219

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 21 November 2023
DOI 10.3389/fmolb.2023.1325552

Check for updates

OPEN ACCESS

EDITED BY
Li Chen,
Huazhong University of Science and Technology, China

REVIEWED BY
Jiandong Diao,
Jilin University, China
Chun-jiao Wu,
Jilin Cancer Hospital, China

*CORRESPONDENCE
Xiang Wang,
xiangw630505@163.com

RECEIVED 21 October 2023
ACCEPTED 13 November 2023
PUBLISHED 21 November 2023

CITATION
Feng K, Ren F and Wang X (2023), Relationships among breast, gut, and oral microbiota across diverse pathological types of breast cancer, a Chinese cohort study.
Front. Mol. Biosci. 10:1325552.
doi: 10.3389/fmolb.2023.1325552

COPYRIGHT
© 2023 Feng, Ren and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Relationships among breast, gut, and oral microbiota across diverse pathological types of breast cancer, a Chinese cohort study

Kexin Feng, Fei Ren and Xiang Wang*

Department of Breast Surgical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Background: Recent research has unveiled the association between microbiota and the onset and progression of breast cancer (BC). This study investigates the microbiota in breast tissue, the gut, and the oral cavity in relation to different pathological types of breast diseases, aiming to unveil the microbiota-BC relationship and provide new perspectives for BC diagnosis and treatment.

Methods: The study encompassed a total of 98 breast cancer patients, with 52 diagnosed with Luminal A BC, 17 with Luminal B BC, 18 with HER2 BC, and 11 with TNBC. In addition, there were 46 patients with non-malignant breast diseases. The V3-V5 region of the 16S rRNA gene of breast tissue, feces, and the oral cavity was sequenced. Based on Amplicon Sequence Variants (ASV) representative sequences and abundance information, a series of statistical analyses were conducted including community diversity analysis, community composition analysis, species difference analysis, correlation analysis, and functional prediction analysis.

Results: Notable divergences in α-diversity and β-diversity were discerned in breast tissue between BC patients and non-malignant breast disease patients. The linear discriminant analysis effect size (LEfSe) and random forest examinations pinpoint Pasteurellaceae as a significant predictor in BC cohorts. Further exploration revealed significant microbial distribution divergences across distinct pathological types of BC, with notable variations in the relative abundance of microbial species such as Streptococcus, Serratia, and Pseudomonas, underscoring the diverse microbial diversity across BC subtypes and sample origins.

Conclusion: This venture sheds light on the complex microbiota milieu across varying body sites and pathological types of BC, emphasizing microbiota-BC connectivity. This articulation of a multisite microbiota-BC interrelation significantly advances a holistic grasp of BC pathogenesis.

KEYWORDS

breast cancer, fecal microbiome, non-malignant breast diseases, oral microbiome, Chinese

1 Introduction

Breast cancer (BC) reigns as the most prevalent cancer among women globally, with estimations in 2023 projecting nearly 300,000 women will be diagnosed and around 43,170 women will succumb to this ailment (Cancer Statistics, 2023), accentuating the critical necessity for in-depth etiological investigation to foster enhanced early detection and treatment strategies. The etiological foundations of BC, embodying a complex orchestration of genetic and environmental determinants, continue to be enigmatically nuanced. Hence, the imperatives of enhancing early diagnosis and orchestrating efficacious treatment strategies have emerged as pressing concerns.

The human microbiome, a refined and dynamic aggregate of microbial entities, exhibits its ubiquity across a plethora of anatomical niches, spanning from the epidermis to the oral and vaginal cavities. Emerging scientific paradigms robustly accentuate the profound impact of host microbiota in modulating the tumor microenvironmental ambit. Dysbiosis within this microbial aggregate can precipitate chronic inflammatory states, reconfigure immune responses, and potentially undermine genome fidelity, thereby instigating DNA aberrations and dysregulating metabolic pathways. Such perturbations furnish a fertile substrate for BC pathogenesis and advancement (Rossi et al., 2020; Sepich-Poore et al., 2021).

The intestinal tract harbors the most substantial contingent of bacterial flora within the human body (Yang et al., 2023). Pertinent seminal inquiries have elucidated disturbances in fecal microbiota diversity in individuals afflicted with BC, with a pronounced emphasis on bacterial taxa such as Clostridiaceae, Faecalibacteriaceae, Ruminococcaceae, Dorea, and Lachnospiraceae (Terrisse et al., 2021; Tzeng et al., 2021; Papakonstantinou et al., 2022). Furthermore, with its intricate complexity, the oral microbiome has been conjectured to play a contributory, albeit partial, role in BC ontogenesis (Nearing et al., 2023. Research suggests a potential link between periodontal disease and an elevated risk of breast cancer (Freudenheim et al., 2016). A meta-analysis of 11 studies demonstrated a significant elevation in breast cancer risk associated with periodontal disease, with a relative risk of 1.22 (Chung et al., 2016). Common pathogenic factors like microorganisms and inflammation, shared between periodontal disease and breast cancer, may impact the onset and progression of breast cancer (Zhang et al., 2023). Existing literature suggests a possible association between oral microbiota and breast cancer. Particularly, studies reveal variations in oral microbial communities between breast cancer patients and healthy women (Thu et al., 2023). Additionally, research identifies a potential association between oral microbiota, particularly those related to menopause and menstrual status, and breast cancer risk. Recent scholarly ventures have delved into the micro-ecology of breast tissue. Sepich-Poore et al.'s (Sepich-Poore et al., 2021) pioneering exploration meticulously cataloged the tumor microbiota across a spectrum of malignancies, unveiling a particularly diverse bacterial milieu in BC. Augmenting this discourse, Fu et al. (Fu et al., 2022) unearthed the hitherto unexplored presence of "intracellular bacteria" within BC tissues, a phenomenon with profound ramifications for metastatic inclinations. This revelation embodies a seismic shift in our ontological comprehension of tumor metastasis. Moreover, the nuanced interplay between microbial niches, as epitomized by the presence of *Fusobacterium* nucleatum, an oral anaerobe with established oncogenic tendencies, in breast cancer, further accentuates the complexity of these interactions (Parhi et al., 2020; Little et al., 2023). More specific research indicates an interaction between the microbial communities of the gut and breast, potentially impacting breast health and breast cancer development. For instance, gut microbes may be transmitted to the breast, potentially altering the breast microbial community and thereby affecting the development of breast diseases (Zhang et al., 2021).

In the realm of composite research on microbiota across diverse body regions, presently, there exists solely one study investigating the oral and fecal microbiota within the Ghanaian populace (Wu et al., 2022), and another delving into the microbiota of breast tissue, oral cavity, and urine (Wang et al., 2017). All remaining studies focus on single-site microbial analyses. The two aforementioned studies exclusively target breast cancer patients and healthy control cohorts. As of now, no research has concurrently explored the oral, gut, and breast microbiota of breast cancer patients, nor ventured into examining microbiota variations across different body regions for diverse pathological types of breast cancer. Aiming to bridge this knowledge gap, our study leverages the precision of 16S sequencing to evaluate fecal, saliva, and breast tissue specimens across a broad spectrum of BC pathologies. This comprehensive analysis aims to elucidate the complex bacterial interactions within these specific anatomical regions and explore their potential implications for breast cancer pathogenesis. Consequently, the notion of an oral-gut-breast axis has been proposed, wherein dysregulated oral bacteria infiltrate the gut, trigger adverse events in the resident breast microbiome, and contribute to breast diseases. The insights gleaned from this inquiry hold the potential to fundamentally reshape existing paradigms, thereby propelling the development of tailored therapeutic approaches in the realm of breast cancer.

2 Materials and methods

2.1 Patient enrollment and tissue collection

The investigation was carried out at the National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences, and Peking Union Medical College during the timeframe from January 2022 to March 2022. The confirmation of breast cancer diagnoses was achieved through ultrasonography, radiography, or breast MRI, and further substantiated by fine needle aspiration or core needle biopsy of breast tissue. Individuals were omitted from the study if they were below 18 years old, had a history of other malignancies, suffered from oral afflictions, had been treated with antibiotics or probiotics within the preceding 2 months, or lacked complete data. The exclusion criteria for control subjects mirrored those for patients, with additional exclusions for gastrointestinal maladies, history of malignancies, chronic non-communicable diseases, or incomplete data.

2.2 Sample collection and storage

Salivary specimens were obtained utilizing the Salivettes® sampling apparatus (Sarstedt, Nümbrecht, Germany) promptly

upon participants' arousal (between 7 and 8 a.m.), with 5–10 mL being collected. Fecal samples were self-collected pre-operatively using stool sampling kits. Fresh-frozen breast tissues were acquired in adherence to standard biorepository protocols from individuals undergoing surgical intervention for breast cancer. A specimen container, housing 5 mL of sterile saline or water, was unveiled in the surgical suite during breast surgery to mitigate potential microbial contamination from the environment. Concurrently with tissue specimens, these environmental controls were processed. All specimens were preserved at −80°C within a 4-h window post-receipt, pending conveyance to the laboratory for processing and analysis.

2.3 16S rRNA gene sequencing

2.3.1 Extraction of genomic DNA

Genomic DNA was extracted from the specimens employing the CTAB/SDS methodology. The concentration and purity of DNA were ascertained by electrophoresis on a 1% agarose gel. DNA was diluted to a concentration of 1 ng/μL using sterile water, the specific dilution factor being contingent on the initial concentration.

2.3.2 Amplicon generation

Ultrapure water was utilized in the DNA extraction procedure to preclude false-positive PCR outcomes. The 16S rRNA genes from varied loci, namely, 16S V4, 16S V3, 16S V3-V4, and 16S V4-V5, were amplified employing designated primers (e.g., 16S V4: 515F-806R et al.) inclusive of barcodes. The PCR reactions comprised 15 μL of Phusion® High-Fidelity PCR Master Mix, 0.2 μM of both forward and reverse primers, and approximately 10 ng of template DNA. The thermal cycling regimen entailed an inaugural denaturation step at 95°C for 4 min, succeeded by 30 cycles, each encompassing denaturation at 95°C for 10 s, annealing at 50°C for 30 s, and elongation at 72°C for 30 s, culminating with a 72°C hold for 4 min. Ultrapure water was utilized in the DNA extraction procedure to preclude false-positive PCR outcomes. The 16S rRNA genes from varied loci, namely, 16S V4, 16S V3, 16S V3-V4, and 16S V4-V5, were amplified employing designated primers (e.g., 16S V4: 515F-806R et al.) inclusive of barcodes. The PCR reactions comprised 15 μL of Phusion® High-Fidelity PCR Master Mix, 0.2 μM of both forward and reverse primers, and approximately 10 ng of template DNA. The thermal cycling regimen entailed an inaugural denaturation step at 95°C for 4 min, succeeded by 30 cycles, each encompassing denaturation at 95°C for 10 s, annealing at 50°C for 30 s, and elongation at 72°C for 30 s, culminating with a 72°C hold for 4 min.

2.3.3 PCR product quantification and qualification

An equal volume of 1× loading buffer was amalgamated with the PCR outputs, followed by electrophoresis on a 2% agarose gel for detection. PCR products were pooled in equimolar proportions, thereafter undergoing purification via the Qiagen Gel Extraction Kit.

2.3.4 Library preparation and sequencing

Sequencing libraries were generated employing the TruSeq® DNA PCR-Free Sample Preparation Kit, in congruence with the manufacturer's stipulations. Index codes were integrated into the libraries. Library quality was evaluated using the Qubit@ 2.0 Fluorometer and the Agilent Bioanalyzer 2100 apparatus. The library was sequenced using Illumina NovaSeq technology, yielding 250 base pair paired-end reads. Initially, quality control and filtering, based on sequencing quality, are performed on the paired-end reads. Concurrently, the overlapping relationships between reads are utilized for assembly, yielding optimized data post-assembly. Subsequently, the DADA2 sequencing denoising method is utilized to obtain Amplicon Sequence Variant (ASV) representative sequences and abundance information. Utilizing the ASV representative sequences and abundance information, a series of statistical or visual analyses are conducted, including taxonomic, community diversity, community composition, species difference, correlation, and functional prediction analyses.

2.4 Clinicopathologic parameters

Baseline clinical and pathological data were documented, encompassing patients' age, pre-existing basal diabetes, hypertension and diabetes history, body mass index (BMI), along with pathological and surgical details. TNM staging was evaluated according to the guidelines set forth in the 8th edition of the American Joint Committee on Cancer (Giuliano et al., 2018). Based on the results from immunohistochemistry tests, the statuses of estrogen receptor (ER) and progesterone receptor (PgR) were established. A positive hormone receptor (HR) status was identified if either ER or PgR tested positive in immunohistochemistry, while a negative status was recorded if both were negative. The classification of Human Epidermal Growth Factor Receptor 2 (HER2) was determined as negative when immunohistochemistry showed negative or 1+ results, and positive for 3+ results; for results of 2+, HER2 positivity was ascertained through fluorescence *in situ* hybridization findings. The calculation of Body Mass Index (BMI) was performed as the ratio of body weight (in kilograms) to the square of height (in meters). The surgical interventions in breast tissue were distinguished as either lumpectomy or mastectomy, contingent upon the extent of surgery.

3 Statistical analysis

Bioinformatics and statistical analysis were performed on the derived sequence data. The DADA2 method (Callahan et al., 2016), contained within the QIIME2 software (Bolyen et al., 2019), was employed for denoising and the acquisition of ASVs. Normalized ASV abundance tables were utilized as the foundation for subsequent analyses. Alpha diversity (α-diversity) primarily assesses community diversity within a specific ecological environment or sample, utilizing indices like Chao1, Shannon, and Simpson to gauge species richness and diversity. Beta diversity (β-diversity) analysis, conducted with QIIME2 software, evaluates the similarity in species diversity across different samples. Beta diversity analysis primarily utilizes four algorithms: Binary Jaccard, Bray Curtis, Weighted Unifrac, and Unweighted Unifrac, to compute the distance between samples and derive β values. These four algorithms are divided into two main classes: weighted (Bray-

TABLE 1 The clinical characteristics of patients with breast cancer and non-malignant breast diseases.

Characteristics	Breast cancer (N = 98)	Non-malignant breast diseases (N = 46)	*p*-Value
Age, years Median (IQR)	49 (18.5)	54 (15.25)	0.163
BMI—kg/m^2 Median (IQR)	24.46 (5.01)	24.8 (5.47)	0.460
Smoking status			0.109
Never smoker	95 (96.9%)	40 (87.0%)	
Former smoker	3 (3.1%)	6 (13.0%)	
Current smoker	0	0	
Alcohol consumption			0.711
Never drink	65 (66.3%)	29 (63.0%)	
	33 (33.7%)	17 (37.0%)	
1 standard drink per day	0	0	
Diabetes			0.212
Yes	48 (49.0%)	28 (60.9%)	
No	50 (51.0%)	18 (39.1%)	
Oral contraceptives use past			0.854
Yes	36 (36.7%)	18 (39.1%)	
No	62 (63.3%)	28 (60.9%)	
Number of live births			0.475
0	33 (33.7%)	12 (26.1%)	
1–2	49 (50%)	28 (60.9%)	
≥3	16 (16.3%)	6 (12.5%)	

Curtis and Weighted Unifrac) and unweighted (Jaccard and Unweighted Unifrac). Unweighted methods primarily compare the presence or absence of species; a smaller β diversity value between two communities indicates higher species similarity. Weighted methods consider both the presence or absence, and the abundance of species. Linear discriminant analysis Effect Size (LEfSe) identifies species features that best explain inter-group differences among two or more groups of samples under varying biological conditions or environments, and evaluates the extent of these features' impact on the differences. PICRUSt2 estimates functional composition of microbial communities using marker genes. KEGG is a database integrating genomic, chemical, and systemic functional information for interpreting biological systems. Following this, random forest analysis was conducted, and receiver operating characteristic (ROC) analysis was performed using the pROC package based on the results of the random forest analysis. All statistical analyses were performed using the R program (Version 4.2.2), with a *p*-value of <0.05 denoting statistical significance.

4 Results

4.1 Baseline characteristics of the participants

The study encompassed a total of 98 breast cancer patients, with 52 diagnosed with Luminal A BC, 17 with Luminal B BC, 18 with HER2 BC, and 11 with TNBC. In addition, there were 46 patients with non-malignant breast diseases. The baseline characteristics of the cohort are in Table 1. For some of these patients, fecal and saliva samples were paired. An illustrative delineation of the study cohort is presented in Figure 1A.

4.2 Diversity of microbial abundance among breast cancer and healthy controls

Initially, comparative scrutiny of microbial compositions across three sample types from individuals with BC and those with non-malignant breast ailments was performed.

Employing α-diversity and β-diversity as pivotal descriptors, we gained a more nuanced understanding of the microbiota's overall assembly and dispersion in relation to BC susceptibility. Table 2 illustrates the comparison of microbial diversity indices in breast tissue, fecal, and saliva samples between breast cancer patients and patients with non-malignant breast diseases. In breast tissue samples, we observed significant differences in α-diversity indices between breast cancer patients and non-malignant cases for the following metrics: Chao1 (p = 0.019, Figure 1B), ACE (p = 0.019), Shannon (p = 0.015), and Simpson (p = 0.021). The indicators of α-diversity of microbial communities in breast tissue, fecal and saliva samples are in Supplementary Figure S1. Additionally, β-diversity indices, namely, Unweighted UniFrac (p = 0.032) and Weighted UniFrac (p = 0.019), also demonstrated significant disparities between breast cancer patients and non-malignant cases. In feces and saliva samples, no significant differences were observed in the breast tissue, feces, or saliva samples between breast cancer patients and patients with non-malignant breast diseases. The indicators of β-diversity of microbial communities in breast tissue, fecal and saliva samples are in Supplementary Figure S1. The heatmaps of species composition (at the genus level) of the three samples is in Supplementary Figure S2.

Since only the α-diversity and β-diversity of breast tissue samples showed significant differences, we conducted further analysis on the breast tissue samples using LEfSe and Random Forest, along with functional prediction analysis.

10.3389/fmolb.2023.1325552

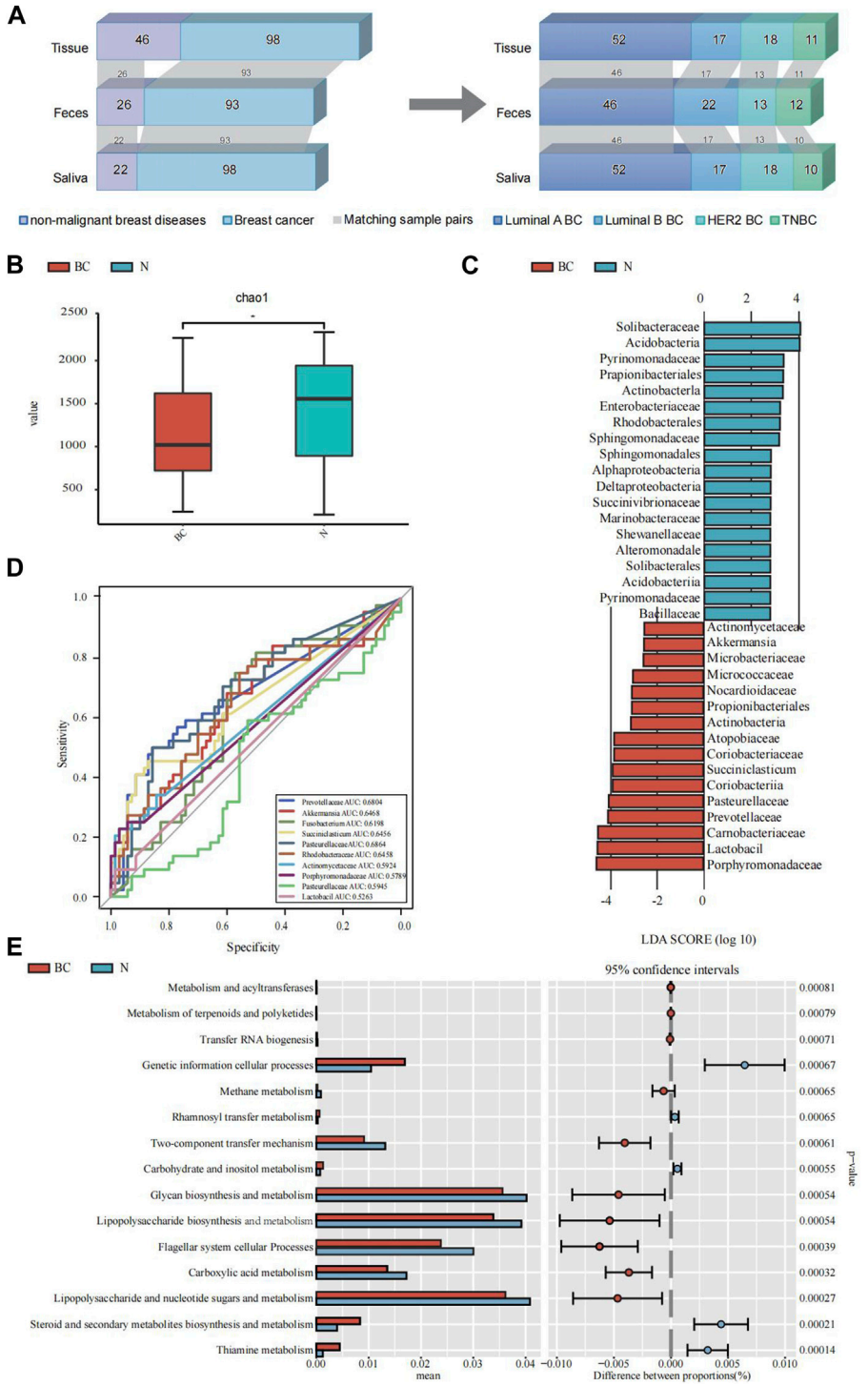


FIGURE 1
An illustrative delineation of the study cohort and the microbial composition and difference between breast cancer (BC) and the non-malignant groups (N). **(A)** Overview of the study population. Grey bands between bar plots represent samples of matching body regions within individuals. **(B)** Alpha diversity index (Chao1 indices) of for the breast tissue samples. **(C)** The common characteristic bacteria of the BC and N groups found in LEfSe of breast tissue samples. **(D)** The top 10 genera were tested by ROC analysis of BC in breast tissue samples. **(E)** The major KEGG pathways between BC and the N groups with the 16S sequencing data of breast tissue samples. Differential shotgun metagenomic sequence-based KEGG pathways in microbiota between the two groups detected by Diamond software. The top 15 items are listed along with the appropriate 95% confidence intervals and adjusted-p values.

We employed the LEfSe analysis method to compare the microbial composition across different groups. Through Linear Discriminant Analysis (LDA), we estimated the effect size of abundance differences for each species (LDA score >3, Bonferroni-adjusted $p < 0.05$). The results indicated that Pasteurellaceae, Prevotellaceae, Carnobacteriaceae, Lactobacil,

TABLE 2 Comparison of microbial diversity indices in tissue, feces, and saliva samples between breast cancer patients and patients with non-malignant breast diseases.

Microbial diversity index	Breast cancer cases VS. non-malignant cases		
	Tissue	Feces	Saliva
Alpha diversity index			
Chao1	0.019	0.490	0.630
ACE	0.019	0.440	0.650
Shannon	0.015	0.560	0.770
Simpson	0.021	0.630	0.920
Beta-diversity index			
Bray-Curtis	0.166	0.147	0.240
Jaccard	0.084	0.096	0.069
Unweighted UniFrac	0.032	0.055	0.153
Weighted UniFrac	0.019	0.092	0.225

Porphyromonadaceae, and Actinomycetaceae exhibited significant abundance differences in the breast cancer group. A subsequent random forest examination of the top 10 genera within the BC cohorts pinpointed Pasteurellaceae as having a prime predictive acumen with an area under the curve (AUC) of 68.64%, followed by Prevotellaceae (68.04%), Akkermansia (64.68%), Rhodobacteraceae (64.58%), Succiniclasticum (64.56%) and *Fusobacterium* (61.98%) (Figure 1D). The PICRUSt2 analysis for KEGG pathway envisaged a marked decline in multiple metabolic activities in the BC cohort vis-a-vis the non-malignant cohort, like the carbohydrate and inositol metabolism, lipopolysaccharide and nucleotide sugars and metabolism, and steroid and secondary metabolites biosynthesis and metabolism (Figure 1E).

4.3 Diversity of microbial abundance among four subtypes of breast cancer

Further exploration ensued on microbiological variances across three paired samples from BC-afflicted individuals showcasing diverse pathological types, including Luminal A, Luminal B, HER2, and TNBC. The Shannon diversity index of breast tissues, fecal and saliva samples, which considers both species richness and evenness, indicate a high level of diversity within four subtypes of breast cancer (Figures 2A–C). To visualize the dissimilarities in microbial community composition between samples, a non-metric multidimensional scaling (NMDS) analysis was performed based on Bray-Curtis dissimilarity. The analysis revealed a stress metric value of 0.24 and 0.06 for fecal and saliva microbiota and 0.18 for breast tumor microbiota, indicating the accuracy of the clustering results (Figures 2D–F). In analyzing microbial compositional differences across breast cancer subtypes, Venn Diagram identified unique microbial species in various samples (Figures 2G–I). Significant divergences in microbial distribution among distinct pathological types within all three samples were evident, elucidating the distinctive microbial diversity across BC subtypes and sample origins.

Microbial species composition and relative abundance were investigated across various pathological types within breast tissue, fecal, and saliva samples. In breast tissue samples (Figure 3A), microbial structure similarity was observed between Luminal A and Luminal B, with Firmicutes and Bacteroidetes predominating. However, the HER2 type exhibited an increased abundance of Proteobacteria, while the TNBC type showed a slight increase in Actinobacteria. In fecal samples (Figure 3B), Luminal A displayed higher abundances of Verrucomicrobia and Cyanobacteria compared to other pathological types. Luminal B exhibited increased levels of Fusobacteria, while Triple negative showed a relative increase in Actinobacteria. In saliva samples (Figure 3C), both Luminal A and Luminal B showed a relative abundance of Bacteroidetes and Firmicutes, whereas HER2 and Triple negative types exhibited higher levels of Proteobacteria.

In the microbial abundance differential analysis, distinct variations were observed across the pathological types. Within breast tissue samples (Figure 3D), significant variations in *Streptococcus* and *Lactobacillus* abundance were noted across all pathological types, with HER2 and Triple negative types exhibiting higher levels of *Pseudomonas*. In fecal samples (Figure 3E), Luminal B exhibited a notably higher abundance of *Bacteroides* compared to Luminal A and HER2, while TNBC showed the lowest abundance of Faecalibacterium. Additionally, Luminal B and HER2 types showed an increased abundance of *Escherichia*. In saliva samples (Figure 3F), *Streptococcus* was highly abundant across all pathological types, though slightly less in TNBC; *Neisseria* was the most abundant in Luminal A and least in Triple negative. Conversely, HER2 type exhibited a higher abundance of Porphyromonas.

The Circos plot reveals the association strength between four pathological types of breast cancer and major microbes in three samples. The present study aimed to examine the relationships between microbial genera and breast cancer pathology kinds using three separate biological materials, namely, breast tissue, fecal, and saliva. Our findings consistently demonstrated both consistent and diverse patterns. The presence of Proteobacteria demonstrated a constant and significant relationship with the Luminal B subtype of breast cancer in all examined samples.

From the breast tissue samples (Figure 4A), there's a pronounced affiliation between Luminal A and microbial communities such as Proteobacteria and Bacteroidetes. Notably, the TNBC subtype showcases a remarkable association with Actinobacteria, hinting at the potential role of

10.3389/fmolb.2023.1325552

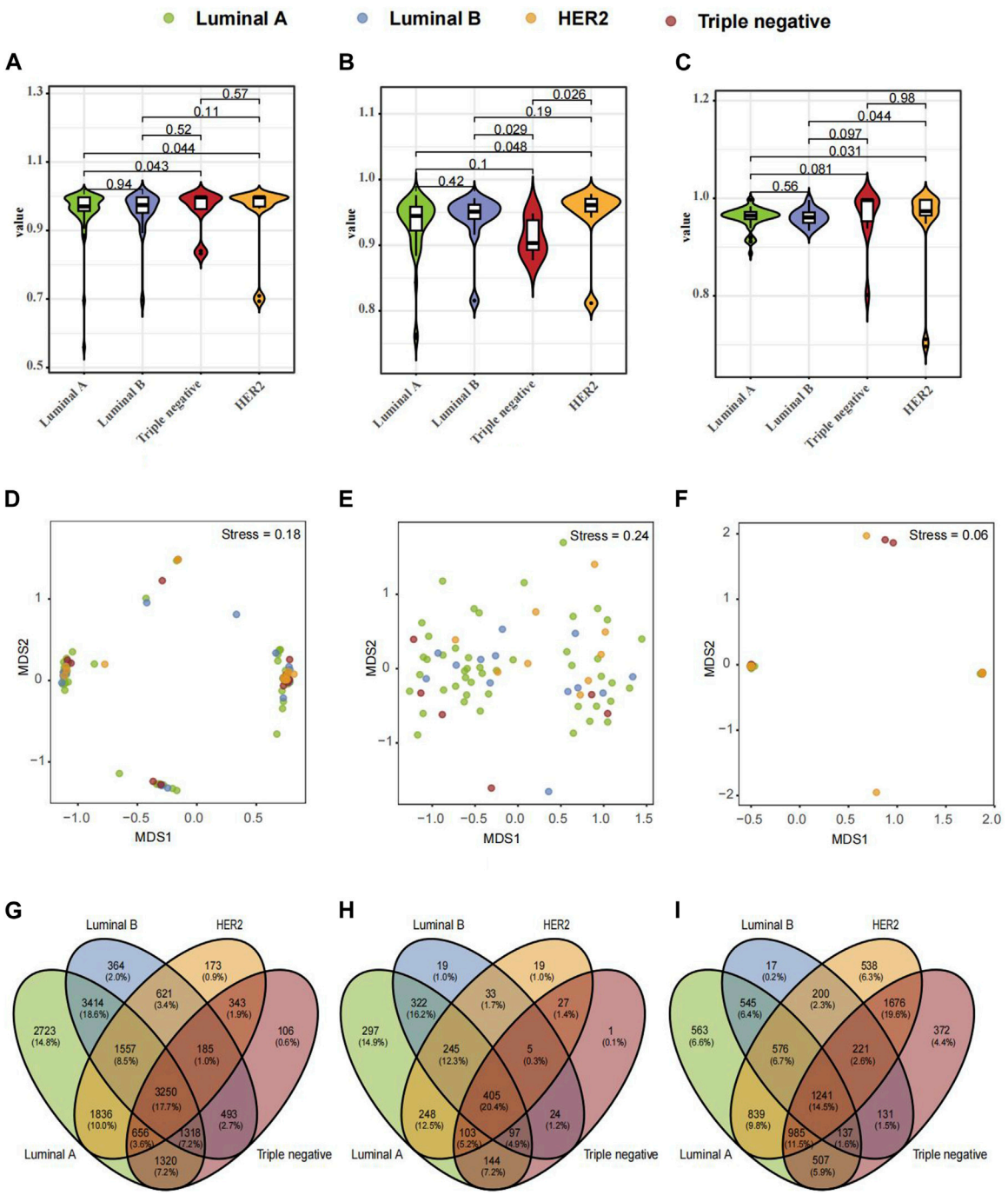


FIGURE 2
Composition and Differential Microbial Presence in Breast Tissue, Fecal, and Saliva Samples across Four Breast Cancer (BC) Types. **(A)** Simpson's α-diversity index for breast tissue samples. **(B)** Simpson's α-diversity index for fecal samples. **(C)** Simpson's α-diversity index for saliva samples. **(D)** NMDS-based β-diversity for breast tissue samples. **(E)** NMDS-based β-diversity for fecal samples. **(F)** NMDS-based β-diversity for saliva samples. **(G)** Venn diagram of microbial taxa in breast tissue. **(H)** Venn diagram of microbial taxa in fecal samples. **(I)** Venn diagram of microbial taxa in saliva samples.

specific microbes within breast tissue in relation to certain cancer subtypes. In the fecal samples (Figure 4B), we observe a differentiated microbial landscape. The bond between HER2 and Bacteroidetes emerges prominently, while Luminal B intertwines closely with Actinobacteria and Proteobacteria. This underscores the possibility that the fecal microbiota-breast cancer relationship possesses its own unique dynamics, distinct from that in breast tissue. Within the saliva samples (Figure 4C), Luminal B's ties with both Proteobacteria and Bacteroidetes shine through, complemented by the evident relationship between Triple-negative and Actinobacteria. This alludes to the theory that saliva's microbial matrix might offer specific

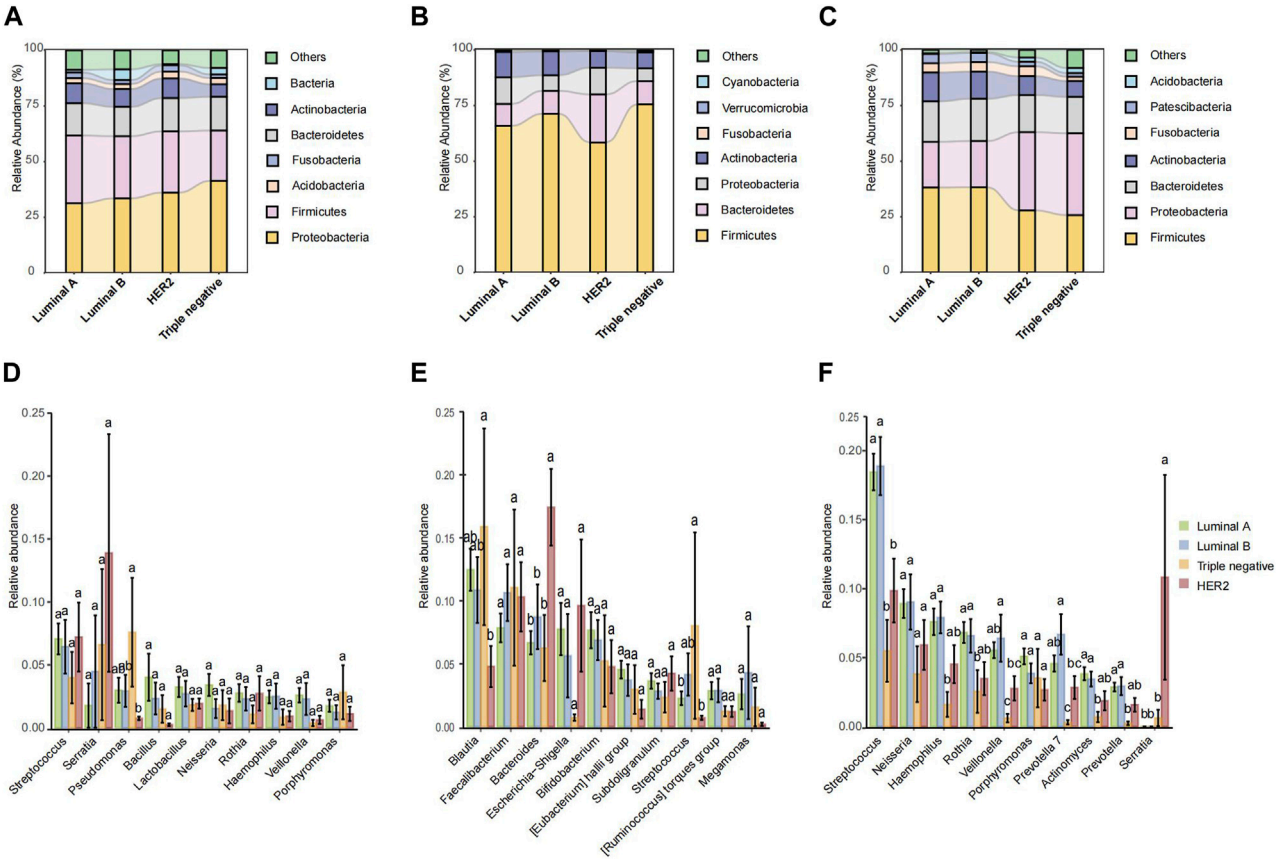


FIGURE 3
The microbiota relative abundances of all groups. **(A–C)** Stacked bar plot of mean proportions of breast, fecal and oral derived taxonomic composition of four types of breast cancer at genus levels. **(D–F)** Statistically differential genera of breast, fecal and saliva microbiota were evaluated with box plots. Different small letters in the bar chart represent statistical differences among the four groups.

insights into certain breast cancer subtypes. In analyzing KEGG functional prediction bubble charts across varied pathological breast cancer microbiomes, breast tissue samples chiefly exhibit lower abundance in most metabolic pathways, particularly within Luminal A and Luminal B types, most prominently in the pathways of "Biosynthesis of ansamycins" and "Alanine, aspartate, and glutamate metabolism" (Figure 4D). Fecal samples demonstrate a significant variation in the "Synthesis and degradation of ketone bodies" pathway, particularly associated with the HER2 type (Figure 4E). Saliva samples indicate a higher abundance of the "Biosynthesis of ansamycins" pathway in Triple-negative breast cancer (Figure 4F). Collectively, this data suggests potential variations in specific metabolic pathways within the microbiome across different breast cancer types, offering a potential avenue for further biological exploration and insights into breast cancer pathogenesis.

5 Discussion

The relationship between host microbiota and breast cancer is complex and multi-layered. Recent studies have greatly enhanced our understanding of dysbiosis as a factor contributing to breast cancer development. Studies show that unique shifts in the composition and function of the mammary and gut microbiomes in breast cancer patients may act as early biomarkers for tumor development (Alpuim Costa et al., 2021). Research indicates that tumor microbiota complexity varies according to breast cancer subtypes, stages, and racial and ethnic groups, suggesting a detailed interplay between microbial composition and genetic and environmental factors. The tumor microenvironment may harbor microbial populations, yet the relationship is bidirectional. Bacteria can promote tumor progression; conversely, tumor progression can lead to bacterial dysbiosis, making it challenging to discern the interactions' causality and direction (Gilbert et al., 2018). Certain microbes are associated with genetic alterations in host cells that may influence cancer progression *Enterococcus faecalis* and *Staphylococcus* hominis exhibit anti-cancer properties by notably reducing cancer cell growth without harming normal cells (Hassan et al., 2016). *Lactobacillus* species can downregulate genes linked to aggressive breast tumor pathways, indicating a potential shift toward microbial-based cancer therapies (Riaz Rajoka et al., 2019). Conversely, microbes such as *Fusobacterium* nucleatum may promote tumor growth and metastasis, (Li et al., 2023; Van der Merwe et al., 2021) reflecting the microbiome's dual role in breast cancer

10.3389/fmolb.2023.1325552

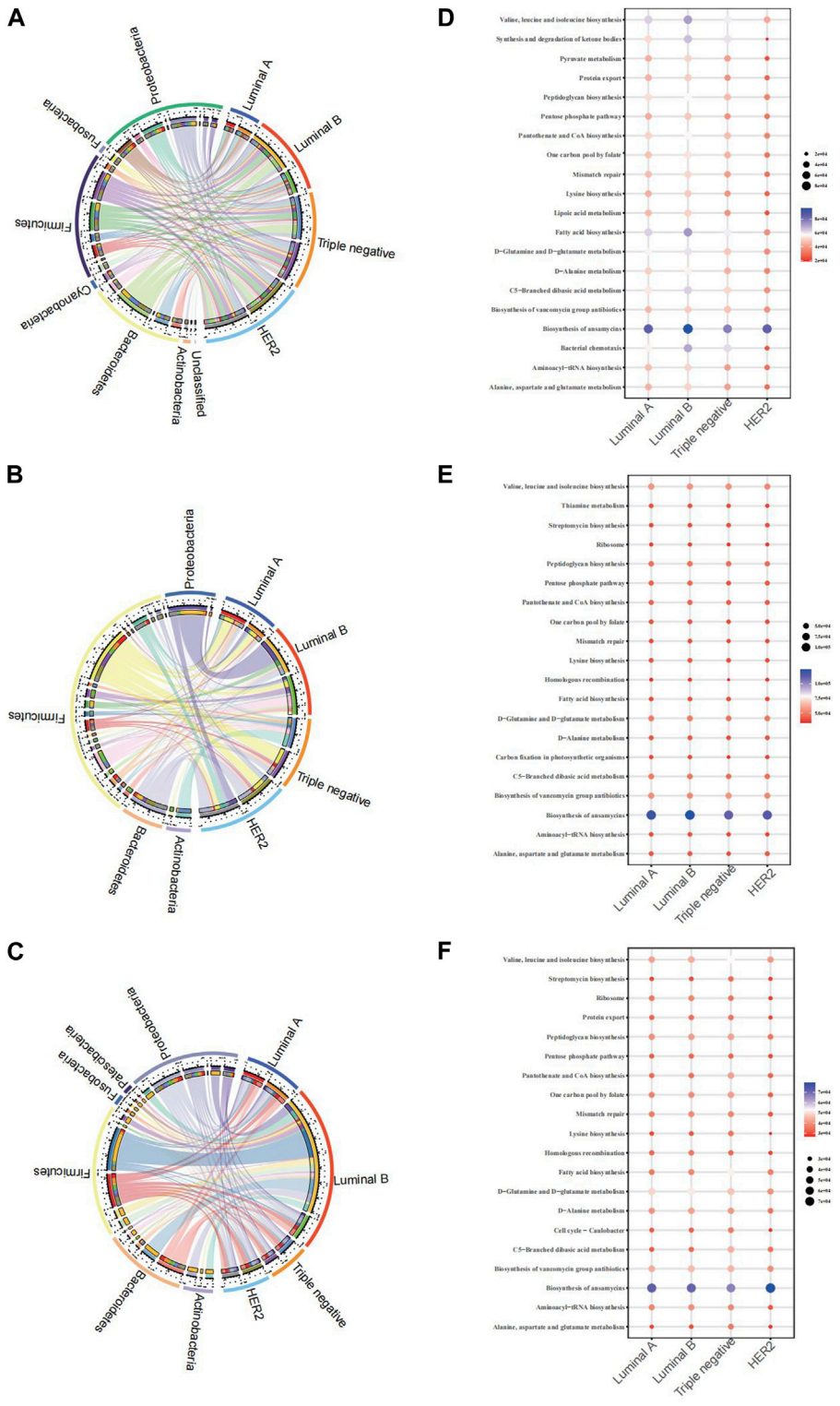


FIGURE 4
Chord diagram of the microbiota and BC pathological types of breast cancer and microbiome function prediction according to the KEGG pathway database **(A)** Chord diagram of breast tissue **(B)** Chord diagram of fecal samples **(C)** Chord diagram of saliva samples **(D)** KEGG for breast tissue **(E)** KEGG for fecal samples **(F)** saliva samples.

pathology. Urbaniak et al.'s groundbreaking research has revealed a unique breast tissue microbiome in cancer patients and shown that certain microbes can cause DNA damage, pointing to a direct microbial role in cancer development (Urbaniak et al., 2016). These findings emphasize the microbiome's role in breast cancer development and propose

microbiome modification as a treatment strategy. The potential for oral microbiota's involvement in breast health is also being investigated, as seen in a Ghanaian case-control study. A case-control study conducted in Ghana explored the association between oral microbiota and breast cancer, as well as benign breast diseases, establishing a connection between the microbiota of the oral cavity and that of the fecal matter. This investigation highlights the potential systemic interplay between different microbial communities within the body and their collective impact on breast health (Wu et al., 2022). Key studies have identified microbial families associated with BC, emphasizing the role of microbiota in BC susceptibility and progression. German et al. (German et al., 2023) analyzed the microbiome of breast tissue from 403 women without cancer and 76 with breast cancer. They discovered a potential association between the presence of Lactobacillaceae, Acetobacterraceae, Xanthomonadaceae, and Ralstonia and breast cancer development. Additionally, analysis of transcriptome data from normal breast tissue revealed that a high abundance of *Acetobacter* aceti, *Lactobacillus* vini, *Lactobacillus* paracasei, and Xanthonomas sp. correlates with enriched metabolism and immune-related genes. Bacteroidetes, Firmicutes, Proteobacteria, and Actinobacteria were also identified by Klann et al. (Klann et al., 2020). Liu et al. (Liu et al., 2023) performed a 16S RNA analysis on 70 FFPE samples, identifying Bacteroidetes, Firmicutes, and Proteobacteria as the primary differential microbes. Additionally, in ER+/HER2-, ER+/HER2+, and ER-/HER2+ tumors, the genera Prevotella_9, *Bacteroides*, and Alloprevotella were the most active, whereas *Lactobacillus* showed heightened activity in triple-negative samples. Our study also validated the presence of differential microbes across these distinct pathological types of breast cancer.

The current investigation meticulously navigates through the symbiotic nexus between host microbiota, sprawling across diverse anatomical realms, and BC pathogenesis. By conducting an extensive appraisal of microbial compositions within fecal, saliva, and breast tissue specimens, spanning various BC pathological spectrums, the study elucidates the potential microbial underpinnings in BC onset and progression. Through a robust methodological framework, this research pioneers in unraveling the microbial intricacies, propelling a significant stride towards a profound understanding of BC's microbial etiology.

The discernible variances in microbial diversity between individuals afflicted with BC and those with non-malignant breast conditions underscore the pivotal role of microbiota in BC pathogenesis. The significant disparities in alpha and beta diversity indices within breast tissue samples between the two cohorts accentuate the plausible influence of microbial consortia on BC susceptibility. Also, the analysis conducted using PICRUSt2 for KEGG pathway prediction indicated disparities in carbohydrate and inositol metabolism, consistent with the findings reported by German et al., with the presence of Ralstonia is linked to the dysregulation of genes in the carbohydrate metabolism pathway (German et al., 2023). Transitioning into the exploration of microbial compositions across various BC subtypes, a meticulous examination unveils a distinct microbial blueprint associated with each pathological subtype, hinting at the microbial predilections towards specific BC subtypes. For instance, a higher abundance of Proteobacteria in HER2 and Actinobacteria in Triple Negative cases reveal the intricate microbial-tumor crosstalk, which potentially modulates tumor behavior and thereby, the clinical outcomes. The review article by Rizzatti et al. emphasizes that Proteobacteria play a role not only in intestinal diseases but also may be related to extra-intestinal diseases, suggesting they might be a microbial signature of some diseases, including breast cancer (Rizzatti et al., 2017).

Expanding the scope to different anatomical niches, the study delineates the distinct microbial landscapes across the breast tissue, fecal, and saliva specimens. This underlines the potential of employing a multi-niche microbial analysis approach in deciphering the complex microbial dynamics in BC pathogenesis. For instance, the pronounced affiliation between the Luminal A subtype and Proteobacteria and Bacteroidetes within breast tissue, contrasted by different microbial associations within fecal and saliva samples, elucidates the necessity of a multi-dimensional microbial analysis for a nuanced understanding of BC-microbiota interactions.

This study, exploring the link between microbial communities and breast cancer, holds certain limitations. The small sample size and single-center approach may hinder generalizability, necessitating larger, multi-center studies in the future. The focus on specific microbial taxa might not provide a comprehensive microbial landscape, calling for deeper microbial analysis using advanced technologies. The absence of long-term follow-up data leaves the impact of microbial variations on clinical outcomes unexplored. Unaccounted potential confounders like dietary habits could influence the findings. The study also does not explore microbial intervention effects on treatment efficacy, nor delves into the molecular mechanisms involved, suggesting avenues for future in-depth investigations.

Looking forward, the findings furnish a vital foundation for future explorations aimed at elucidating the precise microbial interactions in BC ontogenesis. The potential of microbial profiling in early BC detection and monitoring, along with its prognostic and therapeutic implications, merits further in-depth investigations. Moreover, the substantial microbial divergences across different BC subtypes advocate for a more personalized microbial analysis approach in BC management. Future studies should pivot to extensive longitudinal research, engaging larger and more varied populations, and harnessing omics and translational methodologies to fully elucidate the microbiome's role in breast cancer (BC). This research has the potential to shape innovative predictive and preventive strategies, as well as personalized treatment protocols that may include microbial modulation. The dynamic relationship between microbial communities and the immune system, alongside the capacity of certain microbes to either induce genetic instability or offer therapeutic advantages, presents promising avenues for investigation. Embracing a precision medicine framework could transform BC management by integrating microbial, genomic, and environmental data to customize prevention and treatment for each patient's unique profile.

6 Conclusion

The study highlights the intricate connection between microbial structures, their functional roles, and the diversity observed among

different subtypes of breast cancer. Further investigation is required to fully comprehend the possible contributions of these microbial organisms to the dynamics of breast cancer.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://www.ncbi.nlm.nih.gov/sra/PRJNA1036382.

Ethics statement

The studies involving humans were approved by National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, National GCP Center for Anticancer Drugs, The Independent Ethics Committee 20/468-2664. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from primarily isolated as part of your previous study for which ethical approval was obtained. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

KF: Formal Analysis, Methodology, Writing–original draft, Conceptualization, Data curation, Investigation. FR: Formal Analysis, Validation, Visualization, Writing–original draft. XW: Supervision, Writing–review and editing.

Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article..

Conflict of interest

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

Publisher's note

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/fmolb.2023.1325552/full#supplementary-material

References

Alpuim Costa, D., Nobre, J. G., Batista, M. V., Ribeiro, C., Calle, C., Cortes, A., et al. (2021). Human microbiota and breast cancer-is there any relevant link? a literature review and new horizons toward personalised medicine. *Front. Microbiol.* 12, 584332. doi:10.3389/fmicb.2021.584332

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using qiime 2. *Nat. Biotechnol.* 37 (8), 852–857. doi:10.1038/s41587-019-0209-9

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., and Holmes, S. P. (2016). Dada2: high-resolution sample inference from Illumina Amplicon data. *Nat. Methods* 13 (7), 581–583. doi:10.1038/nmeth.3869

Cancer Statistics (2023). Cancer. Available at: https://cancerstatisticscenter.cancer.org/#!/.

Chung, S. D., Tsai, M. C., Huang, C. C., Kao, L. T., and Chen, C. H. (2016). A population-based study on the associations between chronic periodontitis and the risk of cancer. *Int. J. Clin. Oncol.* 21 (2), 219–223. doi:10.1007/s10147-015-0884-6

Freudenheim, J. L., Genco, R. J., LaMonte, M. J., Millen, A. E., Hovey, K. M., Mai, X., et al. (2016). Periodontal disease and breast cancer: prospective cohort study of postmenopausal women. *Cancer Epidemiol. Biomarkers Prev.* 25 (1), 43–50. doi:10.1158/1055-9965.Epi-15-0750

Fu, A., Yao, B., Dong, T., Chen, Y., Yao, J., Liu, Y., et al. (2022). Tumor-resident intracellular microbiota promotes metastatic colonization in breast cancer. *Cell* 185 (8), 1356–1372.e26. doi:10.1016/j.cell.2022.02.027

German, R., Marino, N., Hemmerich, C., Podicheti, R., Rusch, D. B., Stiemsma, L. T., et al. (2023). Exploring breast tissue microbial composition and the association with breast cancer risk factors. *Breast Cancer Res.* 25 (1), 82. doi:10.1186/s13058-023-01677-6

Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., and Knight, R. (2018). Current understanding of the human microbiome. *Nat. Med.* 24 (4), 392–400. doi:10.1038/nm.4517

Giuliano, A. E., Edge, S. B., and Hortobagyi, G. N. (2018). Eighth edition of the ajcc cancer staging manual: breast cancer. *Ann. Surg. Oncol.* 25 (7), 1783–1785. doi:10.1245/s10434-018-6486-6

Hassan, Z., Mustafa, S., Rahim, R. A., and Isa, N. M. (2016). Anti-breast cancer effects of live, heat-killed and cytoplasmic fractions of Enterococcus faecalis and Staphylococcus hominis isolated from human breast milk. *Vitro Cell Dev. Biol. Anim.* 52 (3), 337–348. doi:10.1007/s11626-015-9978-8

Klann, E., Williamson, J. M., Tagliamonte, M. S., Ukhanova, M., Asirvatham, J. R., Chim, H., et al. (2020). Microbiota composition in bilateral healthy breast tissue and breast tumors. *Cancer Causes Control* 31 (11), 1027–1038. doi:10.1007/s10552-020-01338-5

Li, G., Sun, Y., Huang, Y., Lian, J., Wu, S., Luo, D., et al. (2023). Fusobacterium nucleatum-derived small extracellular vesicles facilitate tumor growth and metastasis via Tlr4 in breast cancer. *BMC Cancer* 23 (1), 473. doi:10.1186/s12885-023-10844-z

Little, A., Tangney, M., Tunney, M. M., and Buckley, N. E. (2023). Fusobacterium nucleatum: a novel immune modulator in breast cancer? *Expert Rev. Mol. Med.* 25, e15. doi:10.1017/erm.2023.9

Liu, E., Zhang, F., Xu, T., Ye, L., Ma, S. S. Q., and Ji, Z. S. (2023). Relationship between tumor microbiota transcriptional activity and gene expression in breast cancer. *BMC Cancer* 23 (1), 252. doi:10.1186/s12885-023-10726-4

Nearing, J. T., DeClercq, V., and Langille, M. G. I. (2023). Investigating the oral microbiome in retrospective and prospective cases of prostate, colon, and breast cancer. *NPJ Biofilms Microbiomes* 9 (1), 23. doi:10.1038/s41522-023-00391-7

Papakonstantinou, A., Nuciforo, P., Borrell, M., Zamora, E., Pimentel, I., Saura, C., et al. (2022). The conundrum of breast cancer and microbiome - a comprehensive review of the current evidence. *Cancer Treat. Rev.* 111, 102470. doi:10.1016/j.ctrv.2022.102470

Parhi, L., Alon-Maimon, T., Sol, A., Nejman, D., Shhadeh, A., Fainsod-Levi, T., et al. (2020). Breast cancer colonization by Fusobacterium nucleatum accelerates tumor

growth and metastatic progression. *Nat. Commun.* 11 (1), 3259. doi:10.1038/s41467-020-16967-2

Riaz Rajoka, M. S., Zhao, H., Mehwish, H. M., Li, N., Lu, Y., Lian, Z., et al. (2019). Anti-tumor potential of cell free culture supernatant of Lactobacillus rhamnosus strains isolated from human breast milk. *Food Res. Int.* 123, 286–297. doi:10.1016/j.foodres.2019.05.002

Rizzatti, G., Lopetuso, L. R., Gibiino, G., Binda, C., and Gasbarrini, A. (2017). Proteobacteria: a common factor in human diseases. *Biomed. Res. Int.* 2017, 9351507. doi:10.1155/2017/9351507

Rossi, T., Vergara, D., Fanini, F., Maffia, M., Bravaccini, S., and Pirini, F. (2020). Microbiota-derived metabolites in tumor progression and metastasis. *Int. J. Mol. Sci.* 21 (16), 5786. doi:10.3390/ijms21165786

Sepich-Poore, G. D., Zitvogel, L., Straussman, R., Hasty, J., Wargo, J. A., and Knight, R. (2021). The microbiome and human cancer. *Science* 371 (6536), eabc4552. doi:10.1126/science.abc4552

Terrisse, S., Derosa, L., Iebba, V., Ghiringhelli, F., Vaz-Luis, I., Kroemer, G., et al. (2021). Intestinal microbiota influences clinical outcome and side effects of early breast cancer treatment. *Cell Death Differ.* 28 (9), 2778–2796. doi:10.1038/s41418-021-00784-1

Thu, M. S., Chotirosniramit, K., Nopsopon, T., Hirankarn, N., and Pongpirul, K. (2023). Human gut, breast, and oral microbiome in breast cancer: a systematic review and meta-analysis. *Front. Oncol.* 13, 1144021. doi:10.3389/fonc.2023.1144021

Tzeng, A., Sangwan, N., Jia, M., Liu, C. C., Keslar, K. S., Downs-Kelly, E., et al. (2021). Human breast microbiome correlates with prognostic features and immunological signatures in breast cancer. *Genome Med.* 13 (1), 60. doi:10.1186/s13073-021-00874-2

Urbaniak, C., Gloor, G. B., Brackstone, M., Scott, L., Tangney, M., and Reid, G. (2016). The microbiota of breast tissue and its association with breast cancer. *Appl. Environ. Microbiol.* 82 (16), 5039–5048. doi:10.1128/aem.01235-16

Van der Merwe, M., Van Niekerk, G., Botha, A., and Engelbrecht, A. M. (2021). The onco-immunological implications of Fusobacterium nucleatum in breast cancer. *Immunol. Lett.* 232, 60–66. doi:10.1016/j.imlet.2021.02.007

Wang, H., Altemus, J., Niazi, F., Green, H., Calhoun, B. C., Sturgis, C., et al. (2017). Breast tissue, oral and urinary microbiomes in breast cancer. *Oncotarget* 8 (50), 88122–88138. doi:10.18632/oncotarget.21490

Wu, Z., Byrd, D. A., Wan, Y., Ansong, D., Clegg-Lamptey, J. N., Wiafe-Addai, B., et al. (2022). The oral microbiome and breast cancer and nonmalignant breast disease, and its relationship with the fecal microbiome in the Ghana breast health study. *Int. J. Cancer* 151 (8), 1248–1260. doi:10.1002/ijc.34145

Yang, Q., Wang, B., Zheng, Q., Li, H., Meng, X., Zhou, F., et al. (2023). A review of gut microbiota-derived metabolites in tumor progression and cancer therapy. *Adv. Sci. (Weinh)* 10 (15), e2207366. doi:10.1002/advs.202207366

Zhang, J., Xia, Y., and Sun, J. (2021). Breast and gut microbiome in health and cancer. *Genes Dis.* 8 (5), 581–589. doi:10.1016/j.gendis.2020.08.002

Zhang, Y., Ren, X., Hu, T., Cheng, R., and Bhowmick, N. A. (2023). The relationship between periodontal disease and breast cancer: from basic mechanism to clinical management and prevention. *Oral Health Prev. Dent.* 21 (1), 49–60. doi:10.3290/j.ohpd.b3904343

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 21 November 2023
DOI 10.3389/fmolb.2023.1327893

Check for updates

OPEN ACCESS

EDITED BY
Li Chen,
Huazhong University of Science and Technology, China

REVIEWED BY
Long Bai,
Sun Yat-sen University Cancer Center (SYSUCC), China
Sharvesh Raj Seeruttun,
Sun Yat-sen University Cancer Center (SYSUCC), China

*CORRESPONDENCE
Xiang Wang,
xiangw@vip.sina.com

RECEIVED 25 October 2023
ACCEPTED 10 November 2023
PUBLISHED 21 November 2023

CITATION
Feng K, Ren F and Wang X (2023), Association between oral microbiome and seven types of cancers in East Asian population: a two-sample Mendelian randomization analysis.
Front. Mol. Biosci. 10:1327893.
doi: 10.3389/fmolb.2023.1327893

COPYRIGHT
© 2023 Feng, Ren and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Association between oral microbiome and seven types of cancers in East Asian population: a two-sample Mendelian randomization analysis

Kexin Feng, Fei Ren and Xiang Wang*

Department of Breast Surgical Oncology, National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Background: The oral microbiome has been intricately linked to various pathological conditions, notably cancer, though clear causal links remain elusive. This study aimed to investigate the potential causal relationships between the oral microbiome and seven major cancers: breast, lung, pancreatic, colorectal, gastric, ovarian, and prostate cancers, leveraging Mendelian randomization (MR).

Methods: A two-sample MR analysis was conducted using genome-wide association study (GWAS) data specific to oral microbiota in individuals of East Asian descent. Single nucleotide polymorphisms (SNPs) independent of confounders served as instrumental variables (IVs) to deduce causality. MR methodologies such as the inverse variance weighted (IVW) method, weighted median (WM) method, and Mendelian randomization-Egger (MR-Egger) method were employed. The study utilized datasets encapsulating a multitude of cancer cases and controls, focusing on Asian populations.

Results: Our analysis revealed intricate associations between specific bacterial genera of the oral microbiome and diverse cancers. Notably, *Fusobacterium* showed mixed associations with various cancers, while genera like Prevotella and *Streptococcus* exhibited nuanced roles across malignancies. The genus Aggregatibacter demonstrated a multifaceted influence, positively correlating with some cancers while inhibiting others.

Conclusion: Our findings underscore the profound implications of the oral microbiome in systemic malignancies, suggesting potential modulatory roles in cancer etiology. These insights, though preliminary, accentuate the need for deeper exploration and could pave the way for novel therapeutic strategies.

KEYWORDS

Mendelian randomization, cancers, oral microbiome, case-control study, 16S

1 Introduction

Globally, cancer incidence and mortality rates are escalating alarmingly, with projections indicating about a 50% surge in the global oncologic burden over the next two decades (Sung et al., 2021). As delineated in the Cancer Statistics 2022 report (Siegel et al., 2022), prostate cancer (PCa), lung cancer (LC), and colorectal cancer (CC) collectively account for an estimated 48% of all new cases among males. In females, cancers of the breast (BC), lung (LC), colon, and rectum comprise 52% of cancer diagnoses. Particularly in China, ovarian cancer (OC) has become the second leading cause of death among gynecological malignancies.

The intricate landscape of the human oral microbiome is a dynamic consortium of microorganisms that play pivotal roles in maintaining oral health and systemic wellness. Over recent decades, advances in metagenomic sequencing have unveiled the profound complexity and diversity of these microbial communities. While the oral microbiome has been historically associated with oral-specific diseases such as periodontitis and dental caries, emerging evidence has spotlighted its potential influence on systemic conditions, notably cancers. Traditionally linked to oral-specific diseases like periodontitis and dental caries, emerging evidence now highlights the potential influence of the oral microbiome on systemic conditions, notably cancers. The complex interactions between oral pathogens and host immunity, coupled with the metabolites produced, have been implicated in tumorigenesis, suggesting a plausible link between oral microbial dysbiosis and cancer progression (Teles et al., 2020). Previous research has demonstrated links between the oral microbiome and several types of cancers, including colorectal cancer (Warren et al., 2013), lung cancer (Vogtmann et al., 2022), and pancreatic cancer (Fan et al., 2018; Herremans et al., 2022). This connection, although nascent, poses profound implications for cancer diagnostics, therapeutics, and prevention strategies. In this context, understanding the intricate interplay between the oral microbiome and carcinogenic pathways emerges as a frontier in oncological research. However, traditional observational studies were impeded by inherent limitations, such as local confounding, and the potential for reverse causality.

Mendelian randomization (MR), based on the principles of Gregor Mendel's laws of heredity, has become a useful approach in the field. By exploiting the randomness inherent in genetic classification during meiosis, MR transforms genetic variation into a powerful tool for inferring causality, avoiding traditional confounding and reducing the bias inherent in observational studies. The advent of widespread genome-wide association studies (GWAS) has amplified the potential of MR to dissect the complexities of genetic causation, enabling researchers to distinguish between pure correlation and true causation. Furthermore, as the line between genetics and epidemiology continues to blur, MR acts as a beacon, guiding the scientific community to more robust and reliable causal inferences (Bowden and Holmes, 2019; Burgess et al., 2019).

In delving into the research realm of the interrelations between the oral microbiome and various cancers, the shortcomings of existing studies begin to emerge. Initially, compared to the extensive research on gut microbiome, studies on the oral microbiome are relatively scarce. While the gut microbiome has been extensively studied, the exploration of the oral microbiome remains in its nascent stages. This research bias not only limits our understanding of the relationship between the oral microbiome and cancer, but also hinders the discovery of potential preventive and therapeutic methods. Secondly, most existing studies tend to focus on a single type of cancer or 2-3 types of gastrointestinal cancers, lacking a comprehensive understanding of the microbial interactions across different cancer types. This limitation reduces our ability to understand how the oral microbiome functions across different cancer types, thereby restricting the potential to translate research findings into practical applications. Additionally, traditional observational studies usually rely on microbiome sequencing technologies, which, due to technical and sampling limitations, result in significant heterogeneity in research outcomes. For instance, different sequencing platforms and data processing workflows may yield varying results, while the collection, handling, and storage of samples may also affect the composition and diversity analysis of the microbiom. These technical and methodological constraints diminish the repeatability and comparability of research findings, posing challenges for researchers to obtain more accurate, consistent, and interpretable data.

The primary objective of our inquiry is to explore the putative causal relationship between the oral microbiome and a spectrum of predominant malignancies, including breast, lung, pancreatic, colorectal, gastric (GC), prostate, and ovarian cancers. By constructing a theoretical framework underscoring the role of oral microbial communities in oncogenesis, our scholarly endeavors aim to catalyze the development of innovative therapeutic strategies. In the emerging field of microbiome-cancer interconnections, our study serves as a forefront endeavor exploring the diverse associations between the oral microbiome and a broad range of prevalent cancers. Unlike previous studies that often focused on individual cancer types or specific microbial species, our research adopts a comprehensive approach, examining the interplay between the diverse oral microbial communities and various cancer types. This holistic perspective is crucial for unveiling the widespread implications of oral microbial dysbiosis across various oncologic landscapes. Moreover, our study pioneers the use of MR to delineate the causal relationships between the oral microbiome and cancer, overcoming the inherent limitations of traditional observational studies. This innovative methodology not only enhances the robustness of our findings but also advances the scientific discourse towards more precise causal inferences in the microbiome-cancer nexus.

2 Materials and methods

2.1 Study design and data sources

In this study, the oral microbiota represented the exposure variable, whereas seven different types of cancers were the outcomes. We have taken into consideration the specific genetic and environmental factors present in Asian populations, which may influence the relationship between oral microbiota and cancer. Previous research has demonstrated that the composition of oral microbiota in Asian populations may differ from other ethnic groups, potentially impacting the onset and progression of cancer. Therefore, both our exposure and outcome data are derived from GWAS studies conducted on East Asian populations.

TABLE 1 Summary of the GWAS included in this Mendelian randomization study.

Exposures/outcomes	Consortium	Ethnicity	Sample sizes	N. SNPs	Year
Oral microbiome	CNGBdb	East Asian	2948	Tongue N = 8426	2021
			Tongue N = 2017	Saliva N = 8009	
			Saliva N = 1914		
Breast cancer	NBDC Human Database	East Asian	N = 95283	N = 8919992	2020
Lung cancer	NBDC Human Database	East Asian	N = 212453	N = 8885805	2020
Pancreatic cancer	NBDC Human Database	East Asian	N = 196187	N = 8885075	2020
Colorectal cancer	NBDC Human Database	East Asian	N = 202807	N = 8885369	2020
Gastric cancer	NBDC Human Database	East Asian	N = 202308	N = 8885324	2020
Prostate cancer	NBDC Human Database	East Asian	N = 109347	N = 8878753	2020
Ovarian cancer	NBDC Human Database	East Asian	N = 90451	N = 8876088	2020

GWAS, genome-wide association studies; SNPs, single nucleotide polymorphisms; IVs, instrumental variables; CNGBdb, China National GeneBank DataBase; NBDC, national bioscience database centre.

A two-sample MR analysis was conducted using summary statistics derived from a previously published GWAS that investigated the oral microbiota of individuals of East Asian descent. This GWAS, being the first of its kind on a large scale within an East Asian population, targeted 2017 tongue dorsum samples and 1915 salivary samples, employing high-depth whole-genome sequencing. The dataset utilized in this study comprised 309 tongue dorsum microbiomes (N = 2,017) and 285 salivary microbiomes (N = 1,915) (Liu X. et al., 2021). The samples underwent rigorous inclusion criteria, which involved ensuring a variant calling rate of at least 98%, a mean sequencing depth of over 20×, absence of population stratification in principle component analysis (PCA), and the removal of related individuals based on pairwise identity by descent estimates. In addition, the study used stringent criteria, including a minimum mean depth of 8×, Hardy-Weinberg equilibrium (HWE) values over 10–5, and a genotype calling rate higher than 98% for the analyzed variations. Following stringent quality control protocols, a comprehensive cohort of 2,984 participants was assembled, consisting of 2,017 individuals with tongue dorsum samples and 1,915 persons with salivary samples. Subsequently, a dataset comprising about 10 million variants, encompassing both common and low-frequency variants with a minor allele frequency (MAF) of at least 0.5%, was maintained for further analysis. For a more comprehensive understanding of the methods employed in this study, including sample collection, sequencing techniques, creation of microbiome traits, and observational and genotyping studies, readers are referred to the work of Liu X. et al. (2021).

The data included in this investigation for the sevencancers originated from a comprehensive GWAS performed on an Asian population. This is a large-scale GWAS conducted on an East Asian population for 27 diseases (Ishigaki et al., 2020). The study included a total of 5552 cases and 89731 controls of BC; 6563 cases and 195745 controls of GC, 442 cases and 195745 cotrols of PC, 4050 cases and 208403 controls of LC; 720 cases and 89731 controls of OC, 5408 cases and 103903 controls of PCa; 7062 cases and 195745 controls of CC (Ishigaki et al., 2020). The summary of the GWAS included in this MR study was in Table 1.

2.2 Selection of genetic instrumental variables

In MR, single nucleotide polymorphisms (SNPs) are harnessed as instrumental variables (IVs) to attenuate reverse causation and a litany of inaccuracies endemic to observational epidemiological analyses. For the purpose of to validate the initial assumption, SNPs were initially chosen based on a genome-wide significance threshold of $p < 5 \times 10^{-8}$. Nevertheless, given the limited number of SNPs that exhibited an association with the oral microbiome at this particular threshold, a less stringent criterion of $p < 5 \times 10^{-6}$ was employed. In order to account for potential linkage disequilibrium (LD) among the chosen SNPs, a clumping technique was performed. This procedure involved utilizing a window size of 10,000 kilobases and setting a threshold of an R2 value less than 0.001, as outlined by Abecasis et al. (2010) and Purcell et al. (2007). In addition, in order to maintain allele consistency, the exposure and outcome datasets underwent a process of harmonization. This involved removing SNPs with non-concordant alleles and SNPs with intermediate allele frequencies, hence reducing ambiguity. The SNPs that were selected with great care were afterwards employed as the definitive genetic IVs for the subsequent MR analysis. Furthermore, the F statistics were computed for each SNP both individually and cumulatively. The calculation was performed using the formula: F = R2 * (N–2)/(1–R2), where R2 denotes the proportion of the variance in the exposure variable that is accounted for by each IV. In the study conducted by Burgess et al. (2018) and Burgess et al. (2011), IVs with F statistics below ten were deemed to be poor instruments and were consequently removed from the MR analysis.

2.3 MR analysis

Various statistical methodologies were utilized to examine the causal relationship between breast cancer and the oral microbiome. These methodologies encompassed the inverse variance weighted (IVW) method (Burgess and Thompson, 2015), the Simple mode, Weighted mode, weighted median (WM) method (Bowden et al.,

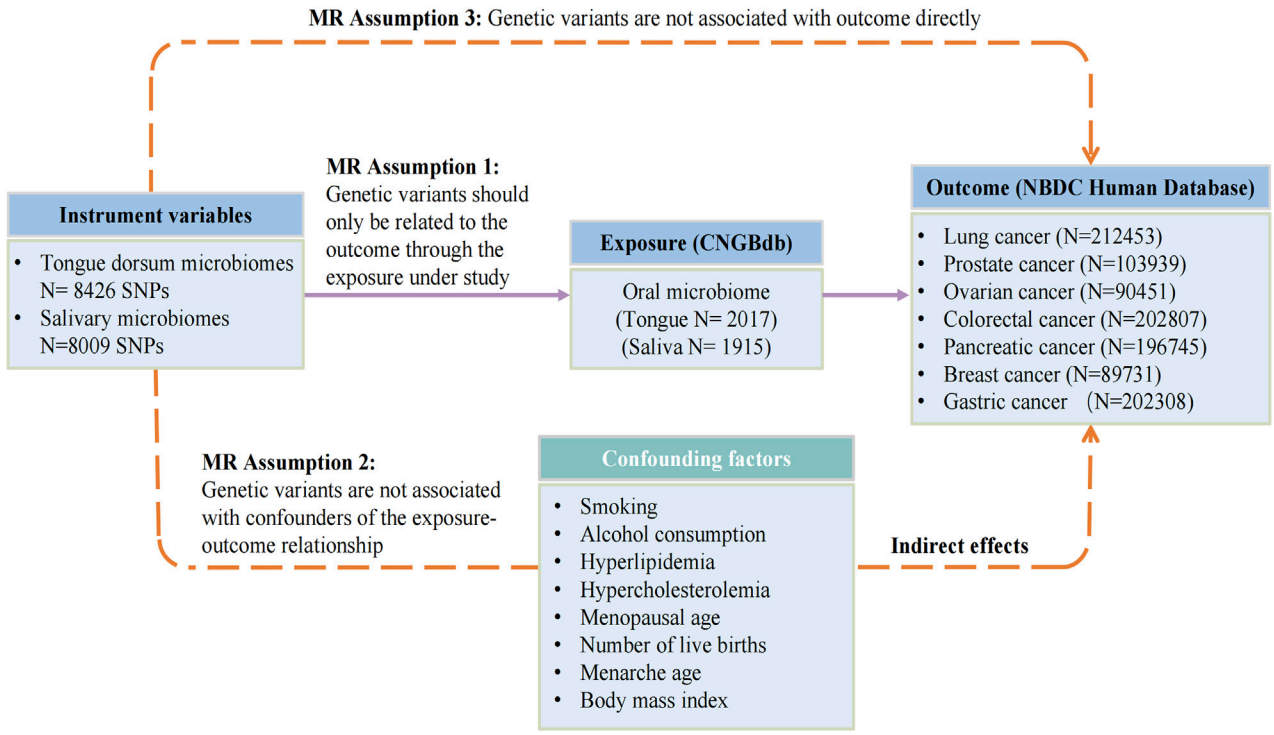


FIGURE 1
The flowchart of Mendelian randomization analysis. IVW, inverse variance weighted; MR, Mendelian randomization; MVMR, multivariable Mendelian randomization; SNPs, single nucleotide polymorphisms; IVs, instrumental variables; CNGBdb, China National GeneBank DataBase; BBJ, BioBank Japan Project.

2016), and the Mendelian randomization-Egger (MR-Egger) method (Burgess and Thompson, 2017). The IVW method is a widely utilized technique, most effective when all IVs adhere to the fundamental assumptions of MR: absence of horizontal pleiotropy and unbiased estimations. Alternatively, the WM method, by computing the median of all instrumental variable effect estimates, emerges as a commendable choice when some instrumental variables do not satisfy MR assumptions, such as exhibiting horizontal pleiotropy. The MR-Egger method, besides estimating causal effects, detects and corrects for horizontal pleiotropy, proving valuable when horizontal pleiotropy is suspected (Burgess and Thompson, 2015). Associations between variables were deemed significant if the resulting p-value of IVW method was less than 0.05 with the estimate direction of other four MR methods were consistent with IVW.

2.4 Sensitivity analysis

Multiple tests were employed, encompassing the heterogeneity test, pleiotropy test, and leave-one-out sensitivity test. The Cochrane's Q test was employed to evaluate the comprehensive pleiotropy in the IVW MR findings as p-value <0.05 implying the presence of heterogeneity. The determination of the average horizontal pleiotropy of the IVs in MR-Egger regression was based on the intercept term and the evaluation of funnel plot asymmetry (Hemani et al., 2018). The existence of heterogeneity was considered significant if the significance level was below $p <$ 0.05. Furthermore, the MR-PRESSO approach was employed to assess the existence of pleiotropy and address the issue of horizontal pleiotropy through the identification and exclusion of probable outliers. Following this, we conducted leave-one-out assessment to determine if significant alterations in the causal effects were observed both before and after the elimination of outliers (Verbanck et al., 2018).

The study provides estimates of effect sizes or odds ratios (ORs) together with their respective 95% confidence intervals (CIs). The statistical tests employed in this study were conducted using a two-sided approach. The studies were conducted using the open-source statistical software R (version: 4.2.2). The analyses were mostly conducted using the TwoSampleMR program (version: 0.5.6) (Verbanck et al., 2018) and MR-PRESSO (version 1.0).

3 Results

3.1 Instrumental variables selection

The flowchart depicted in Figure 1 provides an overview of the MR analysis procedure. Following the removal of SNPs influenced by linkage disequilibrium and palindrome structure, a combined count of 8009 SNPs linked to salivary microbiomes and 8426 SNPs linked to tongue microbiomes were retained for subsequent analysis, utilizing a suggestive significance threshold of $p < 5.0 \times 10^{-6}$. The SNPs included in this study represent species. The F-statistics of the instrumental variables varied between 20.01 and 32.44, all of which

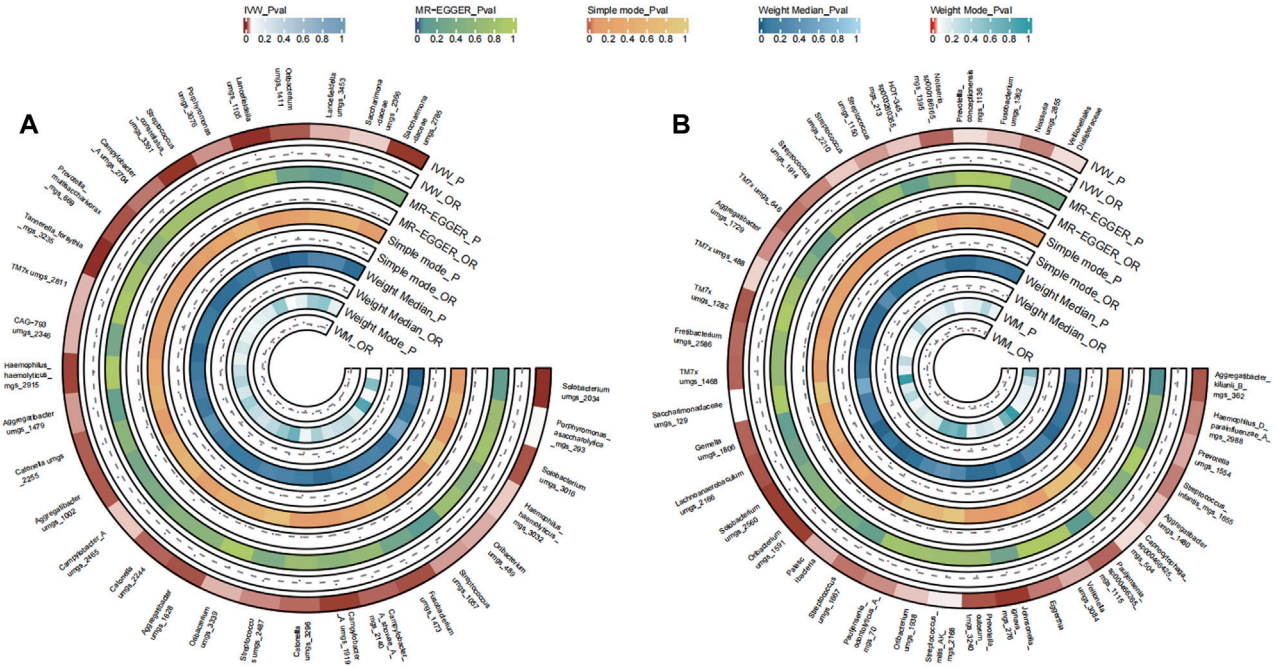


FIGURE 2
Circular Heatmap of Mendelian randomization results of oral microflora at species level and breast cancer **(A)**, tongue, **(B)** saliva.

were statistically significant and exceeded the threshold of 10. These results suggest that there is no indication of weak instrument bias. The statistical analysis conducted by Cochran's Q test revealed no statistically significant heterogeneity ($p > 0.05$). The P- values of MR-PRESSO and the MR-Egger regression were all over 0.05, indicating no horizontal pleiotropy among the chosen SNPs. The leave-one-out sensitivity analysis revealed that individual SNPs did not exhibit a dominant influence on the overall evaluation. Detailed data regarding the MR results of specific tongue microflora at genus level related to seven cancers can be found in the Supplementary Table S1, and results of saliva microflora were in Supplementary Table S2.

3.2 Causal effects of oral microbiota on the development of seven cancer types

3.2.1 Breast cancer

A total of 31 bacterial species in the tongue (16 genera and 12 family) and 36 bacterial species in saliva (19 genera and 16 family) had statistically significant relationships with BC. Seven genera were shared by both the tongue and saliva, including Aggregatibacter, *Fusobacterium*, *Streptococcus* and Saccharimonadaceae TM7x (OR >1), Prevotella, Oribacterium, and Solobacterium (OR <1). The causal effects of tongue and saliva bacterial species on breast cancer were shown in Figures 2A, B.

3.2.2 Lung cancer

A total of 32 bacterial species in the tongue (11 genus and 11 family) and 37 bacterial species in saliva (18 genus and 24 family) had statistically significant relationships with LC. For LC, six genera manifested in both tongue and saliva samples. Among them, Aggregatibacter and Gemella were positively correlated (OR >1), whereas *Fusobacterium*, *Streptococcus*, *Campylobacter* A, and Saccharimonadaceae TM7x were negatively associated with disease progression (OR <1). The causal effects of tongue and saliva bacterial species on lung cancer were shown in Figures 3A, B.

3.2.3 Pancreatic cancer

A total of 48 bacterial species in the tongue (26 genus and 18 family) and 51 bacterial species in saliva (29 genus and 21 family) had statistically significant relationships with PC. For PC, a total of ten genera were present across both sample sources. *Fusobacterium* and Veillonellaceae F0422 emerged as potential risk factors (OR >1), while genera such as Prevotella, Oribacterium, Aggregatibacter, Solobacterium, Pauljensenia, *Streptococcus*, Gemella, Porphyromonas, Saccharimonadaceae TM7x, and Lancefieldella might confer protective benefits (OR <1). The causal effects of tongue and saliva bacterial species on pancreatic cancer were shown in Figures 4A, B.

3.2.4 Colorectal cancer

A total of 39 bacterial species in the tongue (21 genera and 17 family) and 49 bacterial species in saliva (20 genera and 27 family) had statistically significant relationships with CC. Thirteen genera were shared by both the tongue and saliva, including Pauljensenia, *Fusobacterium*, Catonella, *Campylobacter*_A, *Haemophilus*, Granulicatella, Saccharimonadaceae TM7x (OR >1), Prevotella, Solobacterium, *Streptococcus*, Gemella, Lachnoanaerobaculum, Lancefieldella (OR <1). The causal effects of tongue and saliva bacterial species on colorectal cancer were shown in Figures 5A, B.

10.3389/fmolb.2023.1327893

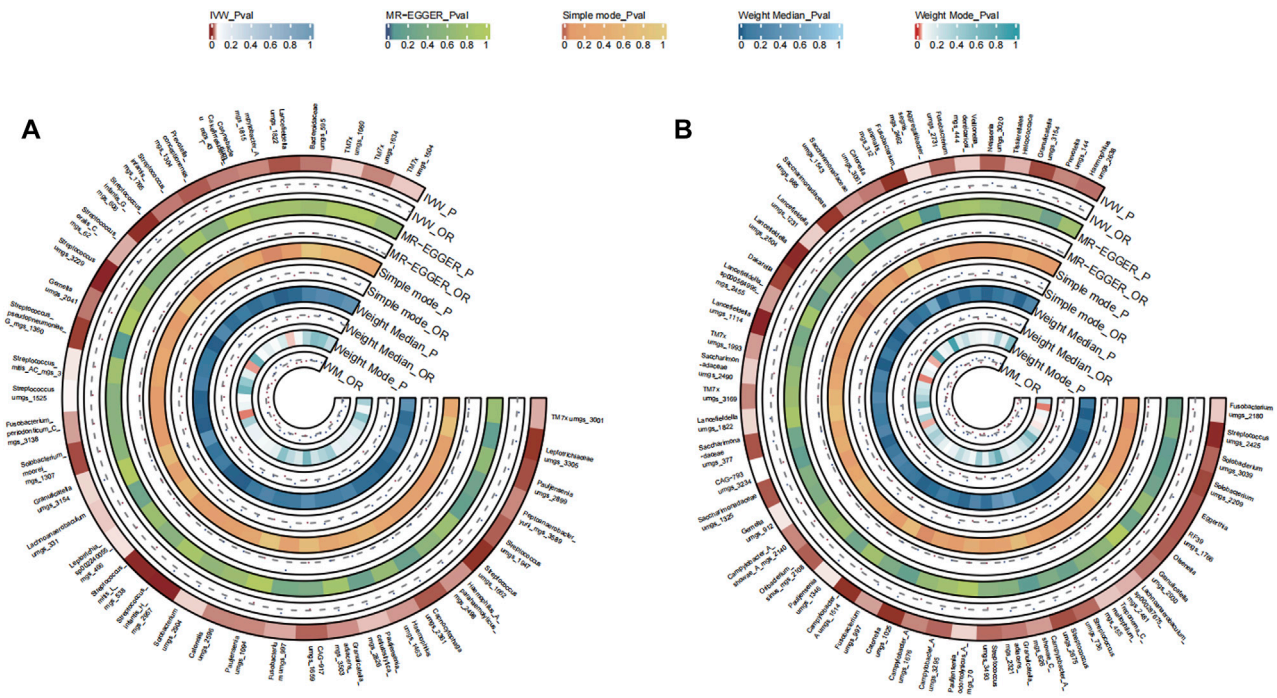


FIGURE 3
Circular Heatmap of Mendelian randomization results of oral microflora at species level and lung cancer **(A)**, tongue, **(B)** saliva.

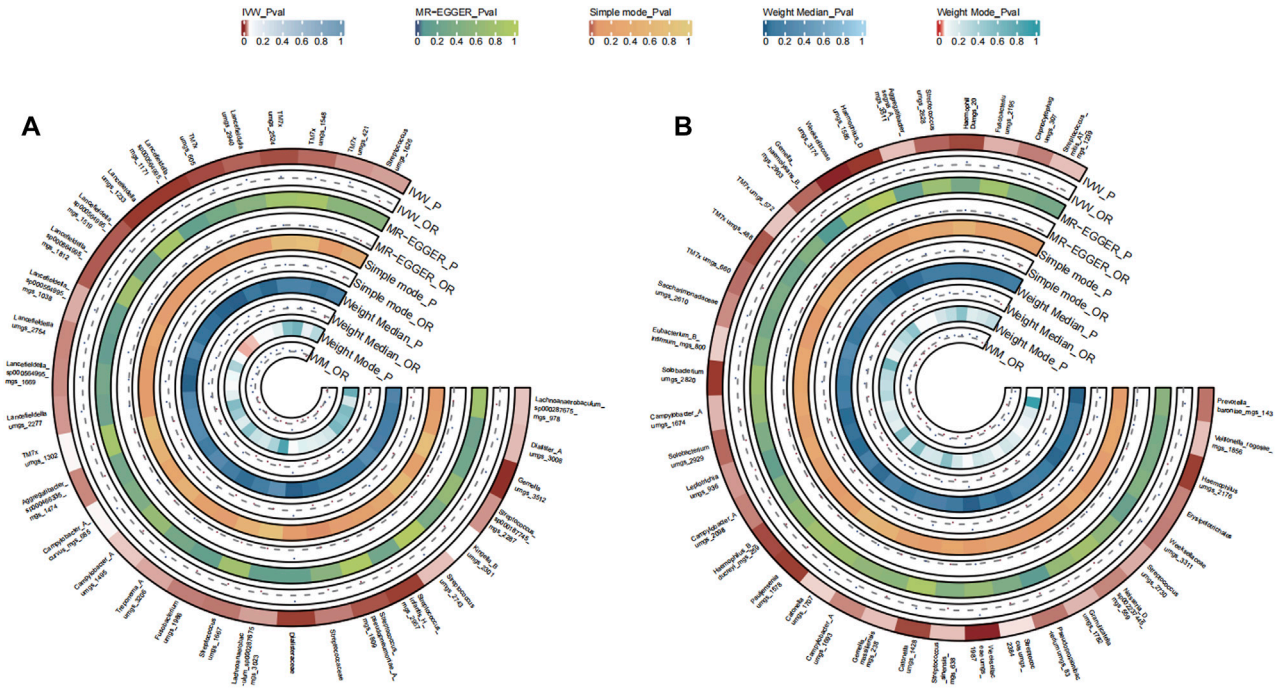


FIGURE 4
Circular Heatmap of Mendelian randomization results of oral microflora at species level and pancreatic cancer **(A)**, tongue, **(B)** saliva.

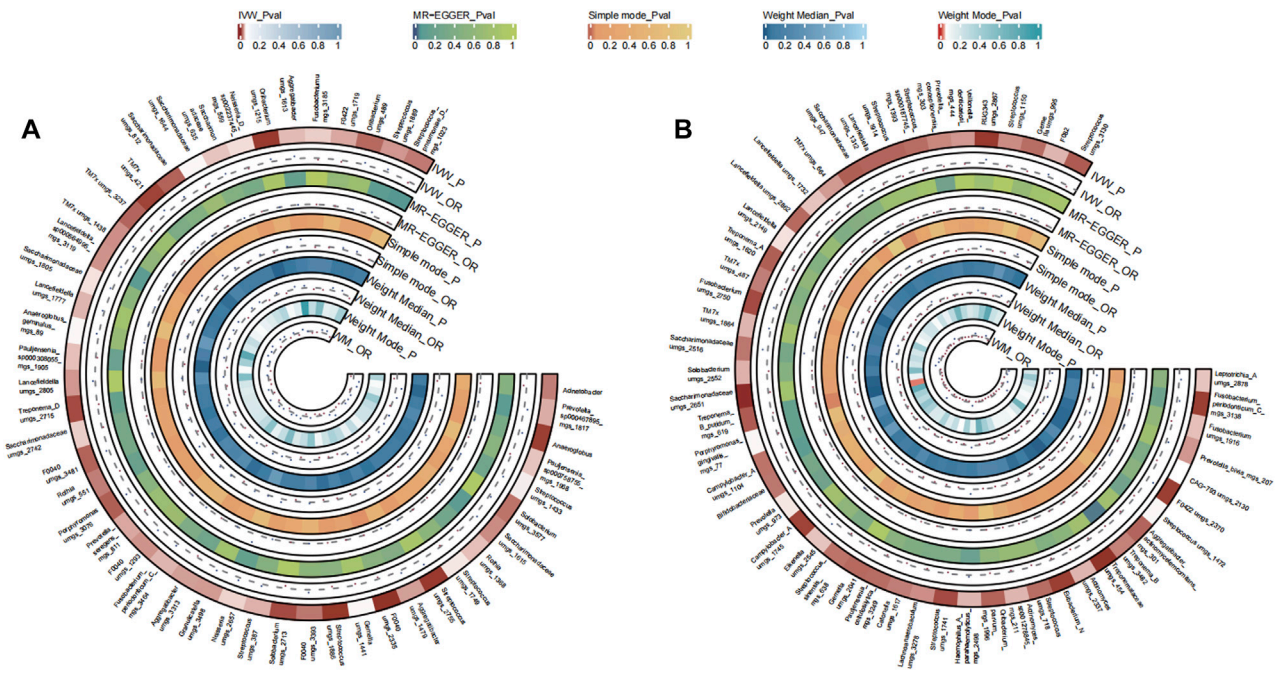


FIGURE 5
Circular Heatmap of Mendelian randomization results of oral microflora at species level and colorectal cancer **(A)**, tongue, **(B)** saliva.

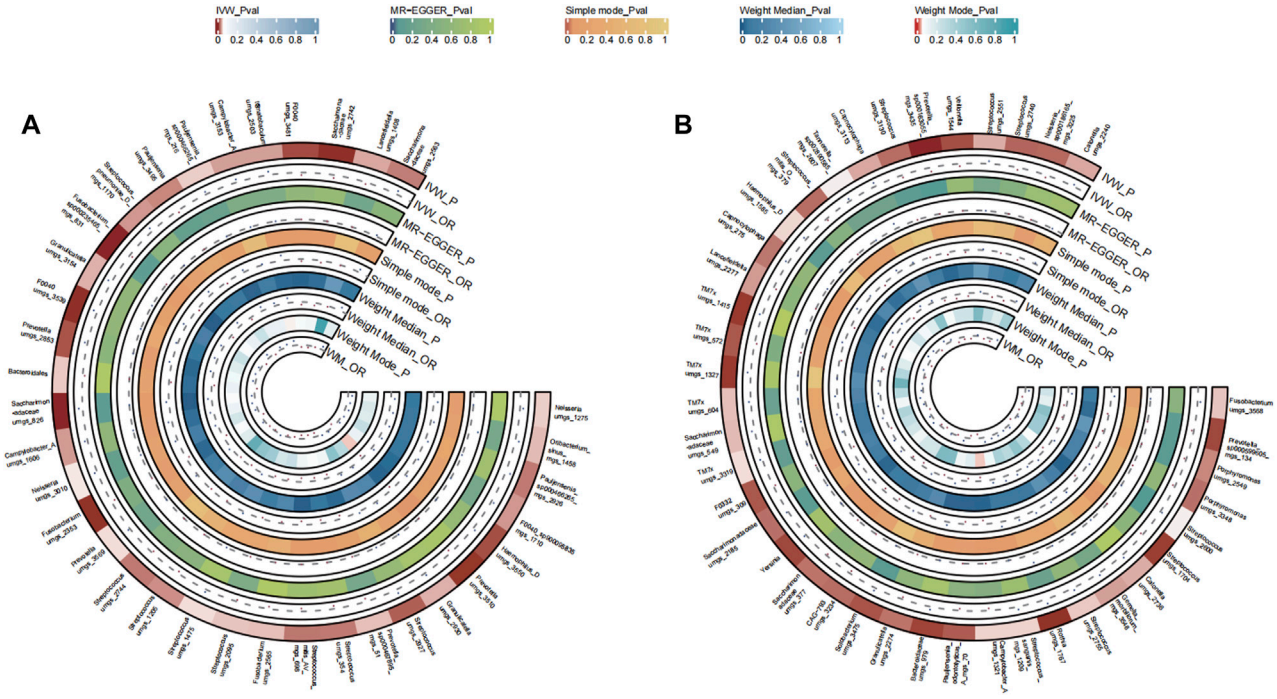


FIGURE 6
Circular Heatmap of Mendelian randomization results of oral microflora at species level and gastric cancer **(A)**, tongue, **(B)** saliva.

3.2.5 Gastric cancer

A total of 35 bacterial species in the tongue (12 genera and 15 family) and 40 bacterial species in saliva (25 genera and 20 family) had statistically significant relationships with GC. Seven genera were identified both in the tongue and saliva samples associated with gastric cancer. Notably, *Neisseria*, *Fusobacterium*, *Haemophilus* D,

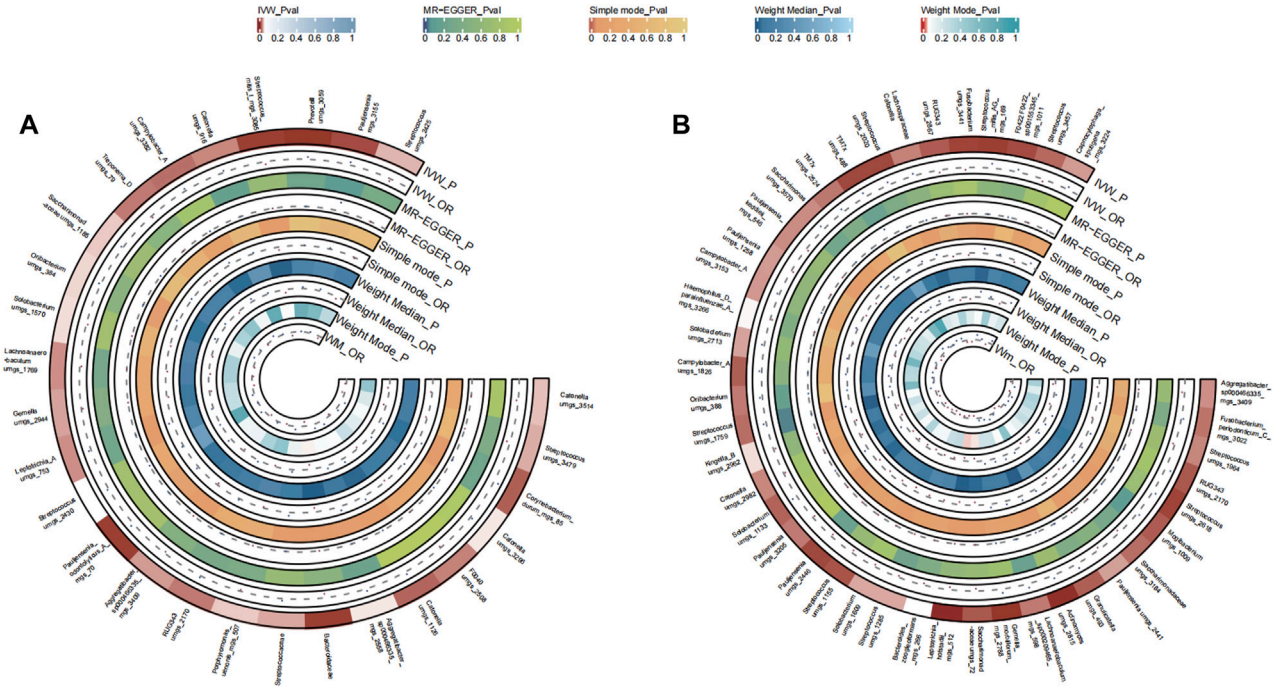


FIGURE 7
Circular Heatmap of Mendelian randomization results of oral microflora at species level and prostate cancer **(A)**, tongue, **(B)** saliva.

and Granulicatella were observed to potentially promote the disease (OR >1), while Prevotella, Pauljensenia, *Streptococcus*, *Campylobacter* A, and Lancefieldella might exhibit protective roles (OR <1). The causal effects of tongue and saliva bacterial species on gastric cancer were shown in Figures 6A, B.

3.2.6 Prostate cancer

A total of 27 bacterial species in the tongue (17 genus and 14 family) and 42 bacterial species in saliva (24 genus and 17 family) had statistically significant relationships with PCa. In prostate cancer samples, nine genera were identified from both the tongue and saliva. Strikingly, Oribacterium, Pauljensenia, *Campylobacter* A, Catonella, Lachnoanaerobaculum, and RUG343 were implicated in possibly elevating the disease risk (OR >1). In contrast, Aggregatibacter, Solobacterium, *Streptococcus*, and Gemella were potentially inhibitory (OR <1). The causal effects of tongue and saliva bacterial species on prostate cancer were shown in Figures 7A, B.

3.2.7 Ovarian cancer

A total of 39 bacterial species in the tongue (20 genus and 16 family) and 33 bacterial species in saliva (22 genus and 15 family) had statistically significant relationships with OC. In the context of OC, seven genera coexisted in both sample types. Interestingly, *Streptococcus*, *Campylobacter* A, Granulicatella, and Saccharimonadaceae TM7x might enhance disease risk (OR >1), in contrast to Solobacterium, *Fusobacterium*, Gemella, and Saccharimonadaceae umgs 1558 which appeared to be protective (OR <1). The causal effects of tongue and saliva bacterial species on ovarian cancer were shown in Figures 8A, B.

3.3 Comprehensive relationship between specific oral microbiota genera and diverse cancers

Within the diverse ecosystem of oral microbiota, certain bacterial genera have arisen as potential modulators of various cancer types, as discerned from an integrated analysis of both saliva and tongue datasets. The genus Aggregatibacter show cases a multifaceted influence, with a promotive role in breast and lung cancers, an inhibitory stance towards pancreatic and prostate cancers, a nuanced effect on colorectal, gastric and ovarian cancers. *Haemophilus* D reveals an intriguing profile, acting as a promoter for breast, stomach, and lung cancers, while seemingly offering protection against prostate cancer. *Fusobacterium* stands out with its diverse associations, promoting breast, colorectal and pancreatic cancers, while adopting a more complex mixed stance on prostate cancers. Interestingly, it exhibits inhibitory effects against ovarian and lung cancers. The genus Prevotella leans towards an inhibitory role in breast, stomach, and colorectal cancers, but its association with pancreatic cancer is mixed, and it distinctly promotes lung cancer. *Streptococcus*, a ubiquitous member of the oral microbiome, presents a mosaic of effects. While it demonstrates mixed associations with a suite of cancers, including breast, colorectal, stomach, prostate, and ovarian, its influence on pancreatic cancer is predominantly inhibitory, and for lung cancer, it remains diverse. This intricate dance of associations continues with Capnocytophaga, which suppresses breast and ovarian cancers, but promotes stomach, lung, and notably, prostate cancers. Porphyromonas, on the other hand, consistently inhibits stomach and pancreatic cancers. The canvas of associations

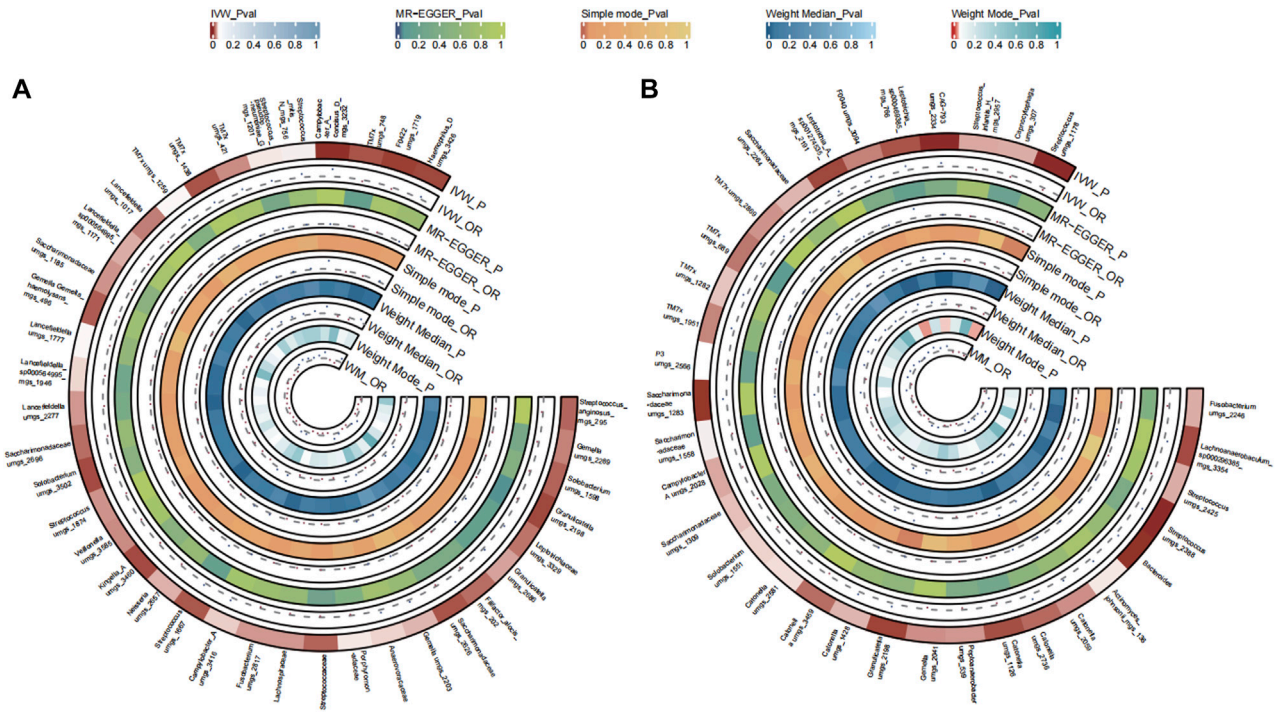


FIGURE 8
Circular Heatmap of Mendelian randomization results of oral microflora at species level and ovarian cancer **(A)**, tongue, **(B)** saliva.

broadens with Solobacterium promoting breast and stomach cancers, inhibiting pancreatic and ovarian, while showing mixed tendencies for colorectal, prostate, and lung cancers. Pauljensenia mostly exerts inhibitory effects, especially against stomach, pancreatic, prostate, and lung cancers, but its influence on breast and colorectal cancers remains mixed. Other genera, including Veillonella, Granulicatella, *Neisseria*, *Haemophilus*, Leptotrichia A, Veillonellaceae F0422, RUG343, and Leptotrichia, further enrich this complex tapestry of interactions, each with their distinct patterns of promotion, inhibition, or mixed effects on various cancers. Summary of multiple cancers in relation to oral microbiota genera was presented in Table 2. The comprehensive relationships between specific oral microflora at genus level and diverse cancers were shown in Figure 9.

4 Discussion

The oral microbiome has the second highest amount of diversity, with a diverse range of bacterial genera and families, which was second only to that of gut (Caselli et al., 2020). The relationship between the oral microbiota and various forms of cancers has been the focus of numerous studies, but the exact mechanisms and implications remain partially understood. Our study, using MR, provides insights into this relationship (Allen and Sears, 2019).

Numerous studies have delved into the relationship between oral microbiota and breast cancer, uncovering various facets of their interaction. A study conducted in Ghana probed the association between the oral microbiome and breast cancer, identifying a potential linkage akin to that observed between the fecal microbiome and breast cancer risk (Wu et al., 2022). Particularly, a linkage seems to exist between breast cancer and the oral microbiota. Women with periodontal disease, triggered by specific bacteria like the red complex (Porphyromonas, Tannerella, and *Treponema*) and the orange complex (*Fusobacterium*, Prevotella, *Peptostreptococcus*, *Streptococcus*, Eubacterium, and *Campylobacter*), are observed to have a heightened risk of breast cancer (Thompson et al., 2017; Wang et al., 2017; Huang et al., 2018). The presence of oral pathogens has also been discovered in breast cancer tumor tissues, notably *Fusobacterium*. The suggested mechanism involves these bacteria entering breast tissues via the bloodstream, potentially driving inflammation and tumorigenic processes (Parhi et al., 2020; Little et al., 2023). In addition, the ability of *Fusobacterium* to colonize malignant breast tumors by adhering to GalNAc receptors points towards a possible mechanism. Such colonization could alter the tumor microenvironment, modulating host cell signaling, immune responses, or even producing metabolites that promote cancers. *Streptococcus* species are abundant in the oral cavity (Yu et al., 2022). In women without breast tumors, researchers observed elevated levels of Lactococcus and *Streptococcus*, suggesting these bacteria may possess anti-cancer properties. Another study found that in healthy patients, the relative abundance of *Streptococcus* was higher, while there was a negative correlation between the stage of breast cancer and bacterial load in tumor tissue. Through MR analysis, we confirmed the association of Aggregatibacter, *Fusobacterium*, *Streptococcus*, and Saccharimonadaceae, along with Prevotella, Oribacterium, and Solobacterium, with breast cancer.

10.3389/fmolb.2023.1327893

TABLE 2 Summary of multiple cancers in relation to bacteria (genera) present in both tongue and saliva.

Outcomes	Mechanism	Quantity	Bacterial (genus)
Breast cancer	promote	5	Aggregatibacter, *Campylobacter*, *Fusobacterium*, *Streptococcus*, TM7x
	inhibit	1	Prevotella
	mix	7	Aggregatibacter, *Fusobacterium*, Oribacterium, Saccharimonadaceae, *Streptococcus*, Solobacterium, *Haemophilus*
Colorectal cancer	promote	1	*Haemophilus*
	inhibit	1	Gemella
	mix	9	*Campylobacter*, Catonella, *Fusobacterium*, Granulicatella, Lancefieldella, Pauljensenia, Solobacterium, TM7x, *Streptococcus*
Lung cancer	promote	0	None
	inhibit	0	None
	mix	5	Aggregatibacter, *Campylobacter*, *Fusobacterium*, *Streptococcus*, TM7x
Gastric cancer	promote	1	*Haemophilus*
	inhibit	2	Lancefieldella, Prevotella
	mix	6	*Campylobacter*, *Fusobacterium*, Granulicatella, Pauljensenia, Saccharimonadaceae, *Streptococcus*
Pancreatic cancer	promote	0	None
	inhibit	3	Gemella, Pauljensenia, *Treponema*
	mix	8	F0422, *Fusobacterium*, Lancefieldella, Prevotella, Saccharimonadaceae, Solobacterium, *Streptococcus*, TM7x
Ovarian cancer	promote	1	*Fusobacterium*
	inhibit	1	Solobacterium
	mix	6	*Campylobacter*, Gemella, Granulicatella, Saccharimonadaceae, *Streptococcus*, TM7x
Prostate cancer	promote	2	*Campylobacter*, Oribacterium
	inhibit	0	None
	mix	8	Aggregatibacter, Catonella, Leptotrichia, Pauljensenia, RUG343, Saccharimonadaceae, Solobacterium, *Streptococcus*

Lung cancer, another major global health concern, may find its etiological roots intertwined with oral health. In three US cohort studies, Vogtmann et al. found certain types of oral flora, like *Streptococcus*, has a favourable association with the risk of developing lung cancer. *Streptococcus* abundance was associated with 1.14 times the risk of lung cancer (95% confidence interval = 1.06–1.22) (Vogtmann et al., 2022). Sun et al. (2023) and Zhang et al. (2019) also reported the levels of *Streptococcus* in saliva samples collected from patients with LC are considerably elevated compared to those in control samples. A study by Yan et al. demonstrated Capnocytophaga and Veillonella were significantly higher in the saliva from lung cancer patients (Yan et al., 2015). A study sought to prospectively investigate the association of the oral microbiome with lung cancer risk, involving 156 incident lung cancer cases and 156 individually matched controls (Shi et al., 2021). Capnocytophaga was associated with a decreased risk of lung cancer with ORs and 95% CIs of 0.53 (0.31–0.92). These studies suggested that the oral microbiome may present new avenues for lung cancer prevention. In our study, we found a positive correlation between Aggregatibacter and Gemella with LC, while *Fusobacterium*, *Streptococcus*, *Campylobacter* A, and Saccharimonadaceae were negatively associated with LC progression. These microbial communities have also been validated in previous literature.

Pancreatic cancer, a highly aggressive digestive system malignancy, has ascended to the third leading cause of cancer-related deaths with an increasing incidence rate, imposing significant global health and economic burdens. Characterized by late symptom manifestation and limited therapeutic options, its 5-year survival rate is below 9%, earning it the moniker king of cancers. In a landmark study utilizing the Cancer Prevention Study II (CPSII) and Prostate, Lung, Colorectal and Ovarian (PLCO) prospective databases, Fan et al. found Porphyromonas gingivalis and Aggregatibacter, were associated with higher risk of pancreatic cancer (adjusted OR for presence vs. absence = 1.60 and 95% CI 1.15 to 2.22; OR = 2.20 and 95% CI 1.16 to 4.18, respectively) and Phylum Fusobacteria was associated with decreased PC risk (Fan et al., 2018). Herremans et al.'s review summarized the correlation between oral microbiota and pancreatic cancer (Herremans et al., 2022). Oral bacteria associated with an increased risk of pancreatic cancer include Porphyromonas (Michaud et al., 2013), Aggregatibacter, Enterobacteriaceae, Lachnospiraceae G7 (35), Bacteroidaceae (Vogtmann et al., 2020), Staphylococcaceae (Vogtmann et al., 2020), Gemella adiacens (Farrell et al., 2012), and Firmicutes (Al-Zyoud et al., 2019; Wei et al., 2020). Conversely,

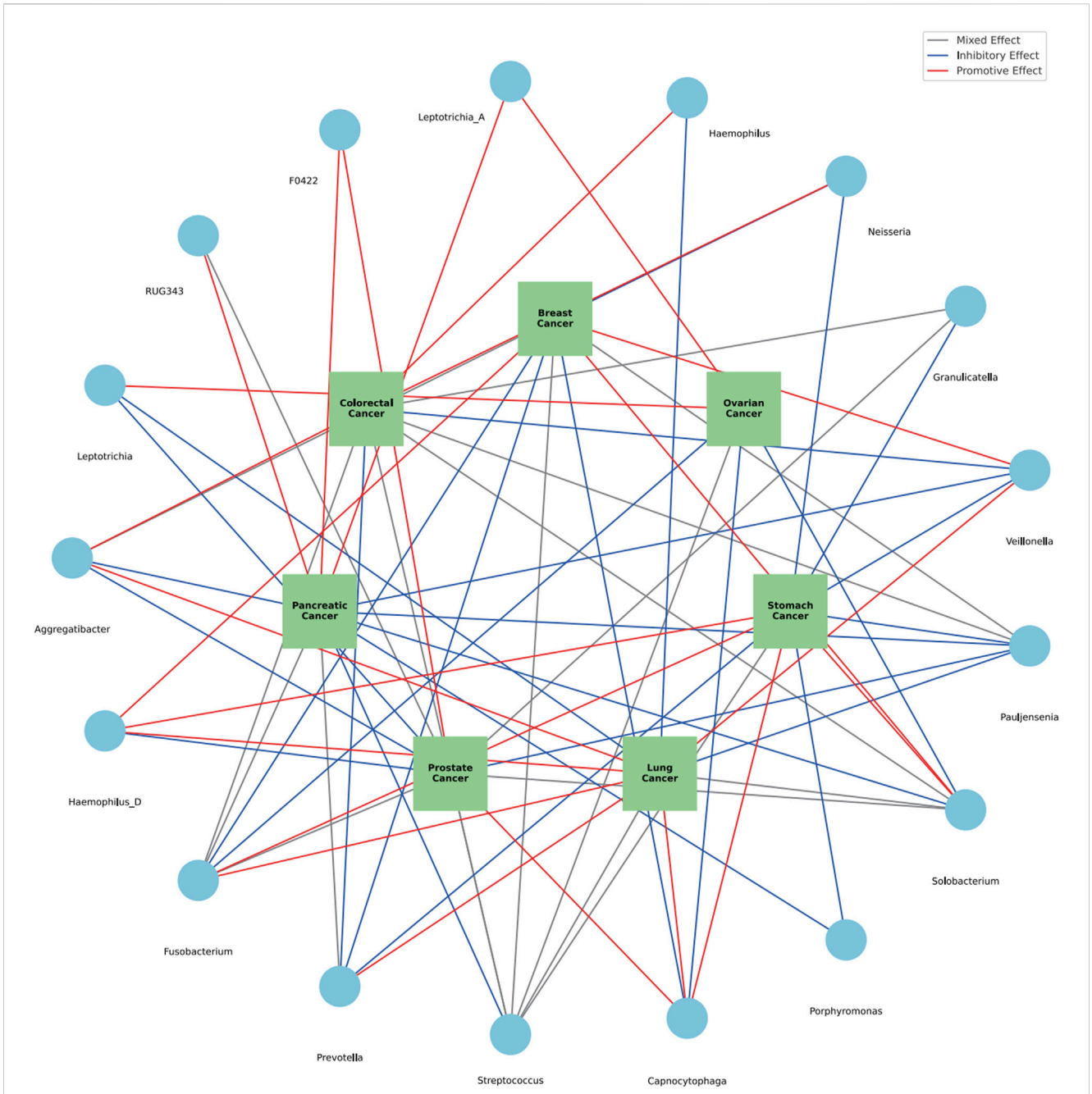


FIGURE 9
The comprehensive relationships between specific oral microflora at genus level and diverse cancers.

oral bacteria associated with a decreased risk of pancreatic cancer include *Fusobacterium* (Mitsuhashi et al., 2015; Wei et al., 2020) and Leptotrichia (Torres et al., 2015; Fan et al., 2018), *Haemophilus* (Vogtmann et al., 2020), *Streptococcus* mitis (Wei et al., 2020), *Neisseria* (Wei et al., 2020), and Proteobacteria (Al-Zyoud et al., 2019). In our study, by MR analysis, the association of *Fusobacterium*, Aggregatibacter, Solobacterium, *Streptococcus*, Gemella, Porphyromonas, Saccharimonadaceae, and Lancefieldella with pancreatic cancer was demonstrated, aligning with previous research. We also found Veillonellaceae, Prevotella, Oribacterium were associated with pancreatic cancer. Veillonellaceae F0422 belongs to the phyla Firmicutes, a common oral microbiome that may be related to gastrointestinal cancers (Asili et al., 2023). Research has indicated a correlation between Prevotella bacteria presence and symptoms of pancreatic cancer. Notably, a higher abundance of Prevotella ($p = 0.008$) was observed in patients reporting jaundice (Wei et al., 2020).

Colorectal cancer, with its burgeoning global incidence, shown associations with gut microbiome. Disruptions in the intestinal flora can stimulate an excessive immune response, leading to the release of inflammatory cytokines such as TNF-α, IL-16, and IL-1β (Chen F. et al., 2022; Kong et al., 2023). Specific periodontopathogens, like

Fusobacterium nucleatum, have been linked to the development of CC and hold potential as biomarkers for cancer diagnosis (Negrut et al., 2023). Alterations in the oral microbiome, including varying levels of specific taxa like *Streptococcus* and Prevotella, have been noted in individuals with CC. Furthermore, a higher abundance of certain oral-like bacterial networks in colonic tissue correlated with CC, whereas an increased abundance of Lachnospiraceae was inversely associated with the colonization of colonic tissue by these oral-like bacterial networks, hinting at a potential protective effect against CC (Negrut et al., 2023). Evidence indicates that the oral microbiome may colonize the gut, resulting in the dysregulation of gut microbes. This colonization fosters an intestinal inflammatory and immunosuppressive microenvironment, which could potentially facilitate tumorigenesis and progression of CC (Mo et al., 2022).

The oral microbiota is a crucial factor in the onset and progression of gastric cancer in humans, associated with inflammation of the gastric mucosa, interactions in the upper gastrointestinal tract, and the infection and transmission of *Helicobacter pylori* (Stasiewicz and Karpiński, 2022). The oral microbiota of gastric cancer patients exhibits significant differences compared to healthy individuals, with an increased abundance of certain oral bacteria such as Porphyromonas gingivalis, *Fusobacterium*, and Streptococci, potentially linked to the onset and advancement of gastric cancer (Şurlin et al., 2020). Like CC, the abundance of certain oral bacteria can migrate to the stomach, altering the gastric microbial milieuoral, producing harmful metabolites or carcinogens in the stomach, and elevating cancer risks. In study of Shu et al., the abundance of *Fusobacterium*, Prevotella, *Neisseria* were significantly changed, compared to healthy controls (Zhang et al., 2022). Chronic prostatitis (CP) and benign prostatic hyperplasia (BPH) are chronic inflammation in the prostate and the etiology is linked to the disorders of oral microbiome (Boland et al., 2013).

The migration of bacteria and chronic inflammation also apply for the correlation between oral microbiota and prostate cancer. Research suggests a potential association between oral microbiota and the development of prostate diseases, notably prostate cancer (Fang et al., 2021). A common chronic inflammatory condition, periodontal disease, shares risk factors with prostate diseases. Inflammation is deemed a significant factor in the progression of prostate diseases. Periodontal disease could potentially lead to increased expression of pro-inflammatory cytokines, thereby impacting prostate health. Additionally, periodontal treatment can significantly alleviate symptoms of prostatic inflammation. Furthermore, oral pathogens, such as Porphyromonas gingivalis and *Fusobacterium* nucleatum, might migrate to the prostate via the bloodstream, inducing analogous pathogenic effects (Wu et al., 2019). These bacterial aggregates may lead to the development of chronic inflammation and may cause damage and abnormal proliferation of prostate cells, or even cancer, thereby increasing the risk of prostate cancer (Fang et al., 2021).

OC is leading oncological cause of death among women. Unlike the above cancers, there is no sufficient evidence substantiating causal relationships between oral bacteria and OC. While the literature on associations of ovarian cancers with oral microbiota is nascent, emerging evidence suggests potential risk elevations tied to alterations in bacterial profiles in numerous compartments, including vaginal, cervicovaginal (Nené et al., 2019), upper genital tract (Zhou et al., 2019), peritoneal (Miao et al., 2020), serum (Kim et al., 2020), and fecal (Mori et al., 2019) compartments in patients with OC.

In the present study, we obtained several important findings. First of all, *Streptococcus* is the sole bacterial genus associated with all seven cancer types. *Streptococcus* is predominantly found in the oral cavity, pharynx, and nasal passages. Its association with multiple cancers is established, underscored by studies indicating shifts in its abundance at respective cancerous sites. Specifically, increased levels were discerned in the mammary tissues of breast cancer patients, bronchoalveolar lavage fluids of lung cancer patients, gastric mucosa of those with stomach cancer, ovarian tissues in ovarian cancer patients, prostate tissues in prostate cancer subjects, and within the microbial communities of the pancreas (Xie et al., 2022). Then, excluding prostate cancer, *Fusobacterium* exhibits associations with six other cancer types. *Fusobacterium* is an oral bacterium, recent investigations have revealed the its intricate involvement in colorectal, breast and pancreatic cancers pathogenesis (Bullman et al., 2017; Parhi et al., 2020; Alon-Maimon et al., 2022; Ou et al., 2022; Udayasuryan et al., 2022). In our MR analysis, *Fusobacterium* is a favorable factor for breast, colorectal and pancreatic cancers, consistent with published studies. This bacterium's capability to adhere to both healthy and neoplastic cells hinge on specific molecular recognition, subsequently activating β-catenin-centric transcriptional pathways, potentiating carcinogenesis. *Fusobacterium* can migrate to the intestines, reshaping the microbial landscape (Yu et al., 2017; Chen S. et al., 2022). This translocation plays a pivotal role in establishing a tumor-immunosuppressive environment, augmenting cancer cell spread by triggering the host's innate immune mechanisms (Dai et al., 2018; Zhao et al., 2022). Noteworthy is the concurrent identification of *Fusobacterium* in oral and colorectal cancer specimens from patients, highlighting the profound interplay between oral microbiota and colorectal malignancy. Furthermore, Prevotella, recognized as the second most abundant genus in the human oral microbiome, has inhibitory role in breast, stomach, and colorectal cancers in our MR analysis. In the published literature (Niccolai et al., 2020; Huh et al., 2022), the influence of Prevotella on cancer progression appears to be multifaceted, contingent upon the cancer type, its stage, the host's immunological profile, and different body sites. Emerging evidence delineates its potential anticancer properties, manifested through mechanisms such as the amplification of immune responses, attenuation of inflammation, or induction of cellular apoptosis. Conversely, other studies implicate Prevotella in exacerbating cancer progression, possibly by bolstering cellular proliferation, invasiveness, or metastatic activities. Such dichotomies emphasize the need for nuanced interpretations of Prevotella's role in oncogenesis and progression. Such disparities underscore the premise that identical bacterial taxa might manifest contrasting effects, contingent upon the environmental conditions they inhabit. Finally, in both

colorectal and gastric cancers, *Haemophilus* stands out as the sole bacterial genus exerting a purely promotive effect. *Haemophilus* emerges as a commonly found inhabitant, predominantly residing within the oral cavity. Intriguingly, a heightened abundance of this bacterial genus has been observed in individuals diagnosed with colorectal cancer. This proliferation might be attributed to its potential role in inducing inflammatory responses and aberrations in the immune system, thereby fostering the progression of colorectal cancer. Complementing this, certain studies have illuminated the significant distinction in the absolute abundance of *Haemophilus* between gastric cancer patients and their healthy counterparts ($p \leq 0.05$) (Liu D. et al., 2021; Zhang et al., 2022). Such findings dovetail with our research, underscoring the contributory role of *Haemophilus* in the advancement of both gastric and colorectal cancers.

In essence, a singular bacterium may manifest dual roles in oncogenesis, utilizing distinct mechanisms to alternately inhibit or advance cancer progression. The association between the oral microbiome and cancer has significant implications for public health and clinical practice. Screening for alterations in the oral microbiome may become part of routine cancer risk assessments. Patients with a dysbiotic oral microbiome may benefit from early intervention strategies, such as microbiome modulation through diet, prebiotics, probiotics, or even microbiota transplants. Furthermore, our findings suggest that manipulation of the oral microbiome may serve as a novel therapeutic avenue. If a certain bacterial genus is found to promote carcinogenesis, strategies to suppress this microbial population may be pursued. Conversely, if certain bacteria are found to exert protective effects, these may be encouraged through targeted therapies.

This study possesses several notable strengths. Firstly, it utilizes the most up-to-date GWAS data pertaining to the oral microbiome and employs MR as a methodological approach to establish causal connections. Furthermore, our comprehensive analysis reveals the unexplored and multifaceted causal relationship between oral microbiota and diverse cancers. The complex network of relationships between oral microbiota and cancer development highlights the importance of a comprehensive understanding of this intricate system. As we continue to explore this area, these findings provide new insights for future research and potential treatment approaches. This is the first Mendelian Randomization analysis on the relationship between oral microbiota and various cancers in an East Asian population. Additionally, it was precisely identified down to the species level.

However, it is important to note that this study has several limitations that need to be addressed. The potential for horizontal pleiotropy may impact the selection of instrumental variables in MR studies. The oral microbiome can be influenced by a variety of factors, including genetic inheritance, lifestyle choices, dietary changes, and environmental factors. Instrumental variables may only account for a small portion of the observed variability, and further research is needed to fully understand the complex changes in the oral microbiota. Additionally, our MR analysis focused on populations of Asian ancestry, and our findings may not be generalizable to populations of European ancestry.

5 Conclusion

Utilizing recent GWAS datasets and Mendelian randomization, the findings of this study support the plausibility of a causal relationship between oral microbiota and seven cancers. These links suggest diverse roles in influencing cancer evolution, utilizing distinct mechanisms to alternately inhibit or advance cancer progression. Particular bacteria, notably *Streptococcus* and *Fusobacterium*, displaying significant correlations with various cancers. The study offers a foundational step towards understanding the profound implications of the oral microbiome in systemic malignancies.

Data availability statement

Publicly available datasets were analyzed in this study. The data of oral microbime can be found here: https://db.cngb.org/search/project/CNP0001664. The data of seven types of cancers can be found here: https://gwas.mrcieu.ac.uk/.

Author contributions

KF: Conceptualization, Data curation, Investigation, Writing–original draft. FR: Formal Analysis, Methodology, Visualization, Writing–original draft. XW: Supervision, 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.

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/fmolb.2023.1327893/full#supplementary-material

References

Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., Gibbs, R. A., et al. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467 (7319), 1061–1073. Epub 2010/10/29. doi:10.1038/nature09534

Allen, J., and Sears, C. L. (2019). Impact of the gut microbiome on the genome and epigenome of colon epithelial cells: contributions to colorectal cancer development. *Genome Med.* 11 (1), 11. Epub 2019/02/26. doi:10.1186/s13073-019-0621-2

Alon-Maimon, T., Mandelboim, O., and Bachrach, G. (2022). Fusobacterium nucleatum and cancer. *Periodontol* 89 (1), 166–180. Epub 2022/03/05. doi:10.1111/prd.12426

Al-Zyoud, W., Hajjo, R., Abu-Siniyeh, A., and Hajjaj, S. (2019). Salivary microbiome and cigarette smoking: a first of its kind investigation in Jordan. *Int. J. Environ. Res. Public Health* 17 (1), 256. Epub 20191230. doi:10.3390/ijerph17010256

Asili, P., Mirahmad, M., Rezaei, P., Mahdavi, M., Larijani, B., and Tavangar, S. M. (2023). The association of oral microbiome dysbiosis with gastrointestinal cancers and its diagnostic efficacy. *J. Gastrointest. Cancer*. Epub 2023/01/05. doi:10.1007/s12029-022-00901-4

Boland, M. R., Hripcsak, G., Albers, D. J., Wei, Y., Wilcox, A. B., Wei, J., et al. (2013). Discovering medical conditions associated with periodontitis using linked electronic health records. *J. Clin. Periodontol.* 40 (5), 474–482. Epub 2013/03/19. doi:10.1111/jcpe.12086

Bowden, J., Davey Smith, G., Haycock, P. C., and Burgess, S. (2016). Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* 40 (4), 304–314. Epub 2016/04/12. doi:10.1002/gepi.21965

Bowden, J., and Holmes, M. V. (2019). Meta-analysis and mendelian randomization: a review. *Res. Synth. Methods* 10 (4), 486–496. Epub 2019/03/13. doi:10.1002/jrsm.1346

Bullman, S., Pedamallu, C. S., Sicinska, E., Clancy, T. E., Zhang, X., Cai, D., et al. (2017). Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. *Science* 358 (6369), 1443–1448. Epub 2017/11/25. doi:10.1126/science.aal5240

Burgess, S., Davey Smith, G., Davies, N. M., Dudbridge, F., Gill, D., Glymour, M. M., et al. (2019). Guidelines for performing mendelian randomization investigations. *Wellcome Open Res.* 4, 186. Epub 2020/08/11. doi:10.12688/wellcomeopenres.15555.2

Burgess, S., Foley, C. N., and Zuber, V. (2018). Inferring causal relationships between risk factors and outcomes from genome-wide association study data. *Annu. Rev. Genomics Hum. Genet.* 19, 303–327. Epub 2018/05/02. doi:10.1146/annurev-genom-083117-021731

Burgess, S., Thompson, S. G., and CRP CHD Genetics Collaboration (2011). Avoiding bias from weak instruments in mendelian randomization studies. *Int. J. Epidemiol.* 40 (3), 755–764. Epub 2011/03/19. doi:10.1093/ije/dyr036

Burgess, S., and Thompson, S. G. (2015). Multivariable mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol.* 181 (4), 251–260. Epub 20150127. doi:10.1093/aje/kwu283

Burgess, S., and Thompson, S. G. (2017). Interpreting findings from mendelian randomization using the mr-egger method. *Eur. J. Epidemiol.* 32 (5), 377–389. Epub 2017/05/21. doi:10.1007/s10654-017-0255-x

Caselli, E., Fabbri, C., D'Accolti, M., Soffritti, I., Bassi, C., Mazzacane, S., et al. (2020). Defining the oral microbiome by whole-genome sequencing and resistome analysis: the complexity of the healthy picture. *BMC Microbiol.* 20 (1), 120. Epub 2020/05/20. doi:10.1186/s12866-020-01801-y

Chen, F., Dai, X., Zhou, C. C., Li, K. X., Zhang, Y. J., Lou, X. Y., et al. (2022a). Integrated analysis of the faecal metagenome and serum metabolome reveals the role of gut microbiome-associated metabolites in the detection of colorectal cancer and adenoma. *Gut* 71 (7), 1315–1325. Epub 2021/09/01. doi:10.1136/gutjnl-2020-323476

Chen, S., Zhang, L., Li, M., Zhang, Y., Sun, M., Wang, L., et al. (2022b). Fusobacterium nucleatum reduces mettl3-mediated M(6)a modification and contributes to colorectal cancer metastasis. *Nat. Commun.* 13 (1), 1248. Epub 2022/03/12. doi:10.1038/s41467-022-28913-5

Dai, Z., Coker, O. O., Nakatsu, G., Wu, W. K. K., Zhao, L., Chen, Z., et al. (2018). Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. *Microbiome* 6 (1), 70. Epub 2018/04/13. doi:10.1186/s40168-018-0451-2

Fan, X., Alekseyenko, A. V., Wu, J., Peters, B. A., Jacobs, E. J., Gapstur, S. M., et al. (2018). Human oral microbiome and prospective risk for pancreatic cancer: a population-based nested case-control study. *Gut* 67 (1), 120–127. Epub 2016/11/02. doi:10.1136/gutjnl-2016-312580

Fang, C., Wu, L., Zhu, C., Xie, W. Z., Hu, H., and Zeng, X. T. (2021). A potential therapeutic strategy for prostatic disease by targeting the oral microbiome. *Med. Res. Rev.* 41 (3), 1812–1834. Epub 20201230. doi:10.1002/med.21778

Farrell, J. J., Zhang, L., Zhou, H., Chia, D., Elashoff, D., Akin, D., et al. (2012). Variations of oral microbiota are associated with pancreatic diseases including pancreatic cancer. *Gut* 61 (4), 582–588. Epub 20111012. doi:10.1136/gutjnl-2011-300784

Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., et al. (2018). The mr-base platform supports systematic causal inference across the human phenome. *Elife* 7, e34408. Epub 2018/05/31. doi:10.7554/eLife.34408

Herremans, K. M., Riner, A. N., Cameron, M. E., McKinley, K. L., Triplett, E. W., Hughes, S. J., et al. (2022). The oral microbiome, pancreatic cancer and human diversity in the age of precision medicine. *Microbiome* 10 (1), 93. Epub 20220615. doi:10.1186/s40168-022-01262-7

Huang, Y. F., Chen, Y. J., Fan, T. C., Chang, N. C., Chen, Y. J., Midha, M. K., et al. (2018). Analysis of microbial sequences in plasma cell-free DNA for early-onset breast cancer patients and healthy females. *BMC Med. Genomics* 11 (Suppl. 1), 16. Epub 20180213. doi:10.1186/s12920-018-0329-y

Huh, J. W., Kim, M. J., Kim, J., Lee, H. G., Ryoo, S. B., Ku, J. L., et al. (2022). Enterotypical Prevotella and three novel bacterial biomarkers in preoperative stool predict the clinical outcome of colorectal cancer. *Microbiome* 10 (1), 203. Epub 2022/11/30. doi:10.1186/s40168-022-01388-8

Ishigaki, K., Akiyama, M., Kanai, M., Takahashi, A., Kawakami, E., Sugishita, H., et al. (2020). Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat. Genet.* 52 (7), 669–679. Epub 20200608. doi:10.1038/s41588-020-0640-3

Kim, S. I., Kang, N., Leem, S., Yang, J., Jo, H., Lee, M., et al. (2020). Metagenomic analysis of serum microbe-derived extracellular vesicles and diagnostic models to differentiate ovarian cancer and benign ovarian tumor. *Cancers (Basel)* 12 (5), 1309. Epub 2020/05/28. doi:10.3390/cancers12051309

Kong, C., Liang, L., Liu, G., Du, L., Yang, Y., Liu, J., et al. (2023). Integrated metagenomic and metabolomic analysis reveals distinct gut-microbiome-derived phenotypes in early-onset colorectal cancer. *Gut* 72 (6), 1129–1142. Epub 2022/08/12. doi:10.1136/gutjnl-2022-327156

Little, A., Tangney, M., Tunney, M. M., and Buckley, N. E. (2023). Fusobacterium nucleatum: a novel immune modulator in breast cancer? *Expert Rev. Mol. Med.* 25, e15. Epub 20230403. doi:10.1017/erm.2023.9

Liu, D., Chen, S., Gou, Y., Yu, W., Zhou, H., Zhang, R., et al. (2021b). Gastrointestinal microbiota changes in patients with gastric precancerous lesions. *Front. Cell Infect. Microbiol.* 11, 749207. Epub 2021/12/28. doi:10.3389/fcimb.2021.749207

Liu, X., Tong, X., Zhu, J., Tian, L., Jie, Z., Zou, Y., et al. (2021a). Metagenome-genome-wide association studies reveal human genetic impact on the oral microbiome. *Cell Discov.* 7 (1), 117. Epub 2021/12/08. doi:10.1038/s41421-021-00356-0

Miao, R., Badger, T. C., Groesch, K., Diaz-Sylvester, P. L., Wilson, T., Ghareeb, A., et al. (2020). Assessment of peritoneal microbial features and tumor marker levels as potential diagnostic tools for ovarian cancer. *PLoS One* 15 (1), e0227707. Epub 2020/01/10. doi:10.1371/journal.pone.0227707

Michaud, D. S., Izard, J., Wilhelm-Benartzi, C. S., You, D. H., Grote, V. A., Tjønneland, A., et al. (2013). Plasma antibodies to oral bacteria and risk of pancreatic cancer in a large European prospective cohort study. *Gut* 62 (12), 1764–1770. Epub 20120918. doi:10.1136/gutjnl-2012-303006

Mitsuhashi, K., Nosho, K., Sukawa, Y., Matsunaga, Y., Ito, M., Kurihara, H., et al. (2015). Association of Fusobacterium species in pancreatic cancer tissues with molecular features and prognosis. *Oncotarget* 6 (9), 7209–7220. doi:10.18632/oncotarget.3109

Mo, S., Ru, H., Huang, M., Cheng, L., Mo, X., and Yan, L. (2022). Oral-intestinal microbiota in colorectal cancer: inflammation and immunosuppression. *J. Inflamm. Res.* 15, 747–759. Epub 20220204. doi:10.2147/jir.S344321

Mori, G., Orena, B. S., Cultrera, I., Barbieri, G., Albertini, A. M., Ranzani, G. N., et al. (2019). Gut microbiota analysis in postoperative lynch syndrome patients. *Front. Microbiol.* 10, 1746. Epub 2019/08/17. doi:10.3389/fmicb.2019.01746

Negrut, R. L., Cote, A., and Maghiar, A. M. (2023). Exploring the potential of oral microbiome biomarkers for colorectal cancer diagnosis and prognosis: a systematic review. *Microorganisms* 11 (6), 1586. Epub 20230615. doi:10.3390/microorganisms11061586

Nené, N. R., Reisel, D., Leimbach, A., Franchi, D., Jones, A., Evans, I., et al. (2019). Association between the cervicovaginal microbiome, Brca1 mutation status, and risk of ovarian cancer: a case-control study. *Lancet Oncol.* 20 (8), 1171–1182. Epub 2019/07/14. doi:10.1016/s1470-2045(19)30340-7

Niccolai, E., Russo, E., Baldi, S., Ricci, F., Nannini, G., Pedone, M., et al. (2020). Significant and conflicting correlation of il-9 with Prevotella and Bacteroides in human colorectal cancer. *Front. Immunol.* 11, 573158. Epub 2021/01/26. doi:10.3389/fimmu.2020.573158

Ou, S., Wang, H., Tao, Y., Luo, K., Ye, J., Ran, S., et al. (2022). Fusobacterium nucleatum and colorectal cancer: from phenomenon to mechanism. *Front. Cell Infect. Microbiol.* 12, 1020583. Epub 2022/12/17. doi:10.3389/fcimb.2022.1020583

Parhi, L., Alon-Maimon, T., Sol, A., Nejman, D., Shhadeh, A., Fainsod-Levi, T., et al. (2020). Breast cancer colonization by Fusobacterium nucleatum accelerates tumor growth and metastatic progression. *Nat. Commun.* 11 (1), 3259. Epub 20200626. doi:10.1038/s41467-020-16967-2

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). Plink: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81 (3), 559–575. Epub 2007/08/19. doi:10.1086/519795

Shi, J., Yang, Y., Xie, H., Wang, X., Wu, J., Long, J., et al. (2021). Association of oral microbiota with lung cancer risk in a low-income population in the southeastern USA. *Cancer Causes Control* 32 (12), 1423–1432. Epub 20210825. doi:10.1007/s10552-021-01490-6

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2022). Cancer statistics, 2022. *CA Cancer J. Clin.* 72 (1), 7–33. Epub 2022/01/13. doi:10.3322/caac.21708

Stasiewicz, M., and Karpiński, T. M. (2022). The oral microbiota and its role in carcinogenesis. *Semin. Cancer Biol.* 86 (Pt 3), 633–642. Epub 20211104. doi:10.1016/j.semcancer.2021.11.002

Sun, Y., Liu, Y., Li, J., Tan, Y., An, T., Zhuo, M., et al. (2023). Characterization of lung and oral microbiomes in lung cancer patients using culturomics and 16s rrna gene sequencing. *Microbiol. Spectr.* 11 (3), e0031423. Epub 2023/04/24. doi:10.1128/spectrum.00314-23

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. Epub 2021/02/05. doi:10.3322/caac.21660

Şurlin, P., Nicolae, F. M., Şurlin, V. M., Ş, P., Ungureanu, B. S., Didilescu, A. C., et al. (2020). Could periodontal disease through periopathogen Fusobacterium nucleatum Be an aggravating factor for gastric cancer? *J. Clin. Med.* 9 (12), 3885. Epub 20201129. doi:10.3390/jcm9123885

Teles, F. R. F., Alawi, F., Castilho, R. M., and Wang, Y. (2020). Association or causation? Exploring the oral microbiome and cancer links. *J. Dent. Res.* 99 (13), 1411–1424. Epub 2020/08/20. doi:10.1177/0022034520945242

Thompson, K. J., Ingle, J. N., Tang, X., Chia, N., Jeraldo, P. R., Walther-Antonio, M. R., et al. (2017). A comprehensive analysis of breast cancer microbiota and host gene expression. *PLoS One* 12 (11), e0188873. Epub 20171130. doi:10.1371/journal.pone.0188873

Torres, P. J., Fletcher, E. M., Gibbons, S. M., Bouvet, M., Doran, K. S., and Kelley, S. T. (2015). Characterization of the salivary microbiome in patients with pancreatic cancer. *PeerJ* 3, e1373. Epub 20151105. doi:10.7717/peerj.1373

Udayasuryan, B., Ahmad, R. N., Nguyen, T. T. D., Umaña, A., Monét Roberts, L., Sobol, P., et al. (2022). Fusobacterium nucleatum induces proliferation and migration in pancreatic cancer cells through host autocrine and paracrine signaling. *Sci. Signal* 15 (756), eabn4948. Epub 2022/10/19. doi:10.1126/scisignal.abn4948

Verbanck, M., Chen, C. Y., Neale, B., and Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat. Genet.* 50 (5), 693–698. Epub 2018/04/25. doi:10.1038/s41588-018-0099-7

Vogtmann, E., Han, Y., Caporaso, J. G., Bokulich, N., Mohamadkhani, A., Moayyedkazemi, A., et al. (2020). Oral microbial community composition is associated with pancreatic cancer: a case-control study in Iran. *Cancer Med.* 9 (2), 797–806. Epub 20191121. doi:10.1002/cam4.2660

Vogtmann, E., Hua, X., Yu, G., Purandare, V., Hullings, A. G., Shao, D., et al. (2022). The oral microbiome and lung cancer risk: an analysis of 3 prospective cohort studies. *J. Natl. Cancer Inst.* 114 (11), 1501–1510. doi:10.1093/jnci/djac149

Wang, H., Altemus, J., Niazi, F., Green, H., Calhoun, B. C., Sturgis, C., et al. (2017). Breast tissue, oral and urinary microbiomes in breast cancer. *Oncotarget* 8 (50), 88122–88138. Epub 20170814. doi:10.18632/oncotarget.21490

Warren, R. L., Freeman, D. J., Pleasance, S., Watson, P., Moore, R. A., Cochrane, K., et al. (2013). Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome* 1 (1), 16. Epub 2014/01/24. doi:10.1186/2049-2618-1-16

Wei, A. L., Li, M., Li, G. Q., Wang, X., Hu, W. M., Li, Z. L., et al. (2020). Oral microbiome and pancreatic cancer. *World J. Gastroenterol.* 26 (48), 7679–7692. doi:10.3748/wjg.v26.i48.7679

Wu, L., Li, B. H., Wang, Y. Y., Wang, C. Y., Zi, H., Weng, H., et al. (2019). Periodontal disease and risk of benign prostate hyperplasia: a cross-sectional study. *Mil. Med. Res.* 6 (1), 34. Epub 2019/11/14. doi:10.1186/s40779-019-0223-8

Wu, Z., Byrd, D. A., Wan, Y., Ansong, D., Clegg-Lamptey, J. N., Wiafe-Addai, B., et al. (2022). The oral microbiome and breast cancer and nonmalignant breast disease, and its relationship with the fecal microbiome in the Ghana breast health study. *Int. J. Cancer* 151 (8), 1248–1260. Epub 20220630. doi:10.1002/ijc.34145

Xie, Y., Xie, F., Zhou, X., Zhang, L., Yang, B., Huang, J., et al. (2022). Microbiota in tumors: from understanding to application. *Adv. Sci. (Weinh)* 9 (21), e2200470. Epub 2022/05/24. doi:10.1002/advs.202200470

Yan, X., Yang, M., Liu, J., Gao, R., Hu, J., Li, J., et al. (2015). Discovery and validation of potential bacterial biomarkers for lung cancer. *Am. J. Cancer Res.* 5 (10), 3111–3122. Epub 20150915.

Yu, L., Maishi, N., Akahori, E., Hasebe, A., Takeda, R., Matsuda, A. Y., et al. (2022). The oral bacterium Streptococcus mutans promotes tumor metastasis by inducing vascular inflammation. *Cancer Sci.* 113 (11), 3980–3994. Epub 20220912. doi:10.1111/cas.15538

Yu, T., Guo, F., Yu, Y., Sun, T., Ma, D., Han, J., et al. (2017). Fusobacterium nucleatum promotes chemoresistance to colorectal cancer by modulating autophagy. *Cell* 170 (3), 548–563. Epub 2017/07/29. doi:10.1016/j.cell.2017.07.008

Zhang, C., Hu, A., Li, J., Zhang, F., Zhong, P., Li, Y., et al. (2022). Combined non-invasive prediction and new biomarkers of oral and fecal microbiota in patients with gastric and colorectal cancer. *Front. Cell Infect. Microbiol.* 12, 830684. Epub 2022/06/07. doi:10.3389/fcimb.2022.830684

Zhang, W., Luo, J., Dong, X., Zhao, S., Hao, Y., Peng, C., et al. (2019). Salivary microbial dysbiosis is associated with systemic inflammatory markers and predicted oral metabolites in non-small cell lung cancer patients. *J. Cancer* 10 (7), 1651–1662. Epub 2019/06/18. doi:10.7150/jca.28077

Zhao, L., Zhang, X., Zhou, Y., Fu, K., Lau, H. C., Chun, T. W., et al. (2022). Parvimonas micra promotes colorectal tumorigenesis and is associated with prognosis of colorectal cancer patients. *Oncogene* 41 (36), 4200–4210. Epub 2022/07/27. doi:10.1038/s41388-022-02395-7

Zhou, B., Sun, C., Huang, J., Xia, M., Guo, E., Li, N., et al. (2019). The biodiversity composition of microbiome in ovarian carcinoma patients. *Sci. Rep.* 9 (1), 1691. Epub 2019/02/10. doi:10.1038/s41598-018-38031-2

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 21 November 2023
DOI 10.3389/fmolb.2023.1264553

Check for updates

OPEN ACCESS

EDITED BY
Li Chen,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

REVIEWED BY
Xuantong Zhou,
Peking University, China
Shu Peng,
Huazhong University of Science and Technology, China

*CORRESPONDENCE
Feng Xu,
xufengxh19@163.com

RECEIVED 21 July 2023
ACCEPTED 31 August 2023
PUBLISHED 21 November 2023

CITATION
Chen X, Zhao Q, Wang H, Qiu K, Deng X and Xu F (2023), Period2 is associated with immune cell infiltration and is a potential diagnostic and prognostic marker for hepatocellular carcinoma. *Front. Mol. Biosci.* 10:1264553. doi: 10.3389/fmolb.2023.1264553

COPYRIGHT
© 2023 Chen, Zhao, Wang, Qiu, Deng and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Period2 is associated with immune cell infiltration and is a potential diagnostic and prognostic marker for hepatocellular carcinoma

Xiaolin Chen[1], Qiaosu Zhao[1], Haibiao Wang[2], Kaijie Qiu[2], Xi Deng[3] and Feng Xu[1]*

[1]Department of Gastroenterology, The Affiliated Lihuili Hospital of Ningbo University, Ningbo, China, [2]Department of Hepatobiliary Surgery, The Affiliated Lihuili Hospital of Ningbo University, Ningbo, China, [3]Department of Histopathology, Ningbo Clinical Pathology Diagnostic Center, Ningbo, China

Background: Hepatocellular carcinoma (HCC) health challenge worldwide. Many studies showed that circadian rhythms play a critical role in tumor development. This study aimed to investigate the role of the circadian gene period2 (PER2) in HCC development and explore the possible mechanisms involved.

Methods: From fresh HCC tissues and paired paracancerous tissues, we measured PER2 mRNA and protein expression levels and calculated the correlations between PER2 expression and clinicopathological parameters in patients with HCC. We used transcriptome data from The Cancer Genome Atlas to mine the PER2 gene, including single gene difference analysis, single gene co-expression analysis, gene set enrichment analysis, immune infiltration analysis, and methylation analysis to explore its role and mechanism in HCC occurrence and development.

Results: PER2 expression levels were significantly lower in HCC tissues than in the paired paracancerous tissues. PER2 expression in HCC significantly correlated with neural invasion, Child-Pugh classification, and China liver cancer staging stage in HCC patients. The differentially expressed genes associated with PER2 were significantly enriched in mitochondrial oxidative phosphorylation, transcriptional translation, amino acid metabolism, and other related pathways. PER2 expression levels significantly correlated with immune cell infiltration and immune checkpoint genes and positively correlated with TP53 expression in HCC tissues. The DNA methylation status in eight CpG islands of the PER2 gene was associated with HCC outcomes.

Conclusion: PER2 is a potential diagnostic and prognostic biomarker and a promising therapeutic target in HCC.

KEYWORDS

Period2, hepatocellular carcinoma, prognostic biomarker, tumor microenvironment, immune cell infiltration, Methylation

1 Introduction

Liver cancer is one of the most common malignant digestive system tumors worldwide. The World Health Organization GLOBOCAN's Global Cancer Statistics 2020 report stated that there were 19.3 million new cancer cases worldwide in 2020, of which liver cancer accounted for 4.7% of new cases (ranking sixth) and 9.9 million cancer deaths worldwide in 2020. Liver cancer accounted for 8.3% of deaths, ranking third (Sung et al., 2021). Hepatocellular carcinoma (HCC) is the predominant form of primary liver cancer,

accounting for 75%–85% of primary liver cancers (Sung et al., 2021). For patients with HCC, despite progress in early diagnosis, chemotherapy, immunotherapy, and surgical treatment, outcomes remain unsatisfactory, with a 5-year survival rate of about 14.1% (Allemani et al., 2018). There is an urgent need to identify therapeutic targets and treatment options to improve survival (Hartke et al., 2017; El Jabbour et al., 2019; Xie et al., 2020).

Period2 (PER2) gene is essential to circadian rhythm maintenance in mammals (Shafi and Knudsen, 2019; Jiang et al., 2021). Disruption of circadian homeostasis is an independent risk factor for cancer (Shafi and Knudsen, 2019). Recent studies identified significant changes in PER2 expression in tumors such as chronic lymphocytic leukemia, kidney cancer, head and neck squamous cell carcinoma, and colorectal cancer (Momma et al., 2017; Xiong et al., 2018; Qiu et al., 2019; Wang et al., 2020). Low expression of PER2 protein in HCC tissues correlated with tumor diameter, portal invasion, TNM staging, and outcomes in HCC patients (Li et al., 2018).

The specific mechanism by which the circadian gene PER2 is involved in developing HCC is unknown. Therefore, we compared PER2 mRNA and protein expression in human HCC and paracancerous tissues and explored the relationship between PER2 expression and clinicopathological tumor characteristics. Bioinformatics analysis of the PER2 gene was performed using public databases to explore its potential functions and mechanisms of action in HCC development.

2 Materials and methods

2.1 Patients and databases

We collected HCC tissues and their paired adjacent tissues from 80 HCC patients who underwent radical HCC surgery at Ningbo Medical Center Lihuili Hospital from December 2021 to August 2022. Inclusion criteria were preoperative clinical diagnosis or pathological diagnosis of primary HCC and complete clinical data. Exclusion criteria were concurrent carcinomas *in situ* and absence of paracancerous tissues. All liver cancer tissues and their paired paracancerous tissues were collected during surgery. The specimens were partly snap-frozen in liquid nitrogen, stored at −80°C, and partly fixed in formalin and embedded in paraffin to create tissue wax blocks. The Ethics Committee of Ningbo Medical Center Lihuili Hospital approved the study (number: KY2021PJ231); patients provided informed written consent.

We downloaded the RNA sequencing (RNAseq) data in the Fragments Per Kilobase per Million (FPKM) format from the HCC Project from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/). The RNAseq data in the FPKM format was converted into the transcripts per million reads format and log2-transformed. The Entrez IDs were converted to gene symbols using the "org.Hs.eg.db" (v3.10.0) R package.

2.2 Expression of PER2 mRNA and protein in HCC tissues and paired paracancerous tissues

Total RNA from tissues was extracted using a TransZol UP Plus RNA Kit (TransGen Biotech, Beijing, China), and the extracted total RNA (OD 260 nm/OD 280 nm > 1.8) was reverse transcribed into cDNA using a TransScript All-in-One First-Strand cDNA Synthesis Supermix for QPCR kit (TransGen Biotech, Beijing, China). Real-time fluorescence quantification was performed on an ABI 7500 PCR instrument with PerfectStart® Green qPCR SuperMix (TransGen Biotech, Beijing, China). The following primers were PER2 forward, 5′-GACATGAGACCAACGAAAACTGC-3'; PER2 reverse, 5′-AGG CTAAAGGTATCTGGACTCTG-3'; GAPDH forward, 5′- GCACCG TCAAGGCTGAGAAC-3′, and GAPDH reverse, 5′-TGGTGAAGA CGCCAGTGGA-3'. The $2^{-\Delta\Delta Ct}$ method was used to compare the mRNA levels of liver cancer samples and their paired adjacent tissues.

Total protein was extracted as protein lysates, then protein quantification and sample preparation were performed. Protein samples were separated using 7.5% Express Cast polyacrylamide gel electrophoresis (NCM Biotech, Suzhou, China) and transferred to polyvinylidene fluoride membranes (Millipore Company, United States). The membranes were blocked with quick blocking buffer (NCM Biotech, Suzhou, China) at room temperature for 10 min, then incubated overnight at 4°C with a primary antibody working solution (PER2 Mouse monoclonal antibody, 67513-1-Ig, 1: 5000, Proteintech Company, United States; β-tubulin rabbit monoclonal antibody, A12289, 1:1000, ABclonal, China). The next day, the membranes were washed three times with Tris-buffered saline-Tween was incubated with a secondary antibody working solution (HRP-conjugated Affinipure Goat Anti-Mouse IgG (H + L), SA00001-1, 1:2000, Proteintech Company, United States; HRP-conjugated Affinipure Goat Anti-Rabbit IgG (H + L), SA00001-2, 1:2000, Proteintech Company, United States) for 2 h at room temperature, then the membranes were washed three times with Tris-buffered saline-Tween. The chemiluminescence method was used to expose the membranes in Image Quant LAS 500 instrument. Finally, the gray value of protein bands was analyzed using ImageJ software, and protein expression levels in liver cancer samples and their paired adjacent tissues were compared.

The paraffin sections of specimens were stained with PER2 protein by Ningbo Pathological Center. Immunohistochemical staining score: In 400 times visual field, randomly select five visual fields, each counting 200 cells, and score the positive rate and intensity of staining. The PER2 staining positive rate score was obtained according to the proportion of positively stained cells. The score of the staining range was 0 (0%), 1 (1%–25%), 2 (26%–50%), 3 (51%–75%), and 4 (76%–100%). PER2 staining intensity score: negative 0, light yellow 1, light brown 2 and tan 3. The product of the positive staining rate score and staining intensity score was used to judge the positivity grade. Samples were classified as 0 negative (−), 1–4 weak positive (+), 5–8 moderate positive (++), or 9–12 strong positive (++). The average of five visual fields was recorded.

2.3 Correlation analysis between PER2 expression level and clinicopathological features of patients with HCC

According to the results of immunohistochemical staining of PER2 protein in liver cancer tissues, liver cancer tissues were divided into a low expression group where PER2 was negative (−) or weakly

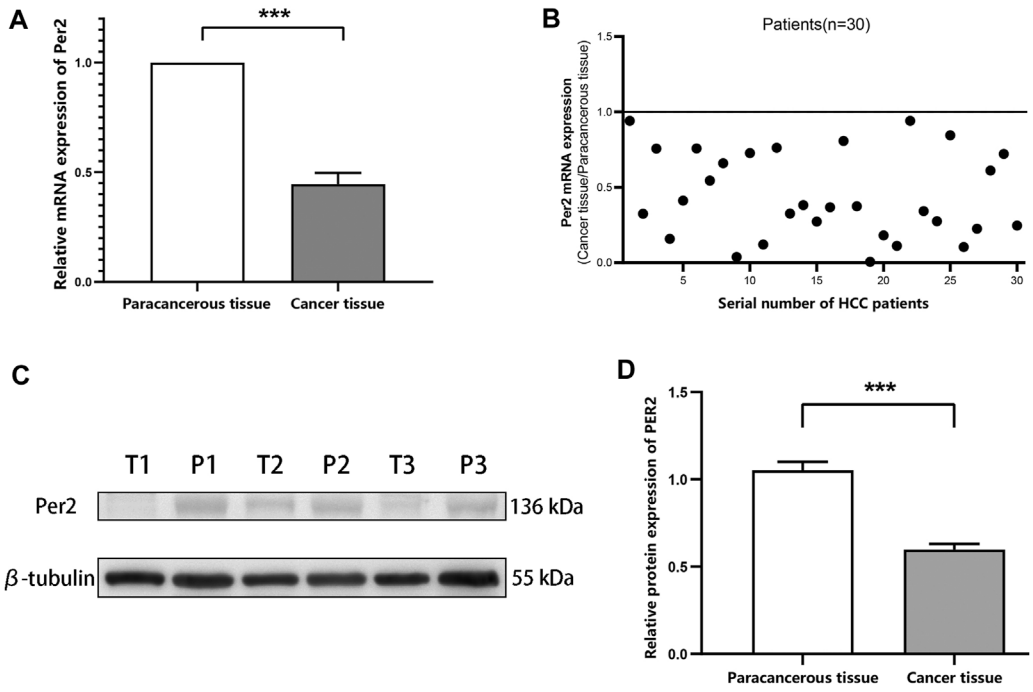


FIGURE 1
PER2 mRNA and protein expression levels in HCC tissues and paired paracancerous tissues. **(A,B)** The mRNA expression levels of PER2 were determined using qRT-PCR. **(C,D)** The protein expression levels of PER2 were determined using Western blotting. T, tumor tissue; P, paracancerous tissue; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

positive (+) and a high expression group where PER2 was moderately positive (++) or strongly positive (++). We recorded clinical data corresponding to liver cancer tissues, including gender, age, operation mode, smoking and drinking habits, body mass index (BMI), serum α-fetoprotein (AFP) level, whether they are infected with hepatitis B virus, tumor number, tumor maximum diameter, histological grade, microvascular invasion risk grading, Child-Pugh grading, China liver cancer staging (CNLC) to calculate correlations between PER2 expression and clinicopathological parameters and oncology behavior in patients with HCC.

2.4 Functional enrichment analysis of PER2-related differentially expressed genes (DEGs) in HCC

HCC patients were divided into high PER2 expression and low PER2 expression groups. DEGs between the two groups were analyzed using the "DESeq2" (V1.26.0) R package (Love et al., 2014), and the threshold parameters of differential analysis were set to |log2(FC)| >1.5 and p. adj <0.05. We used the "ggplot 2"(V3.3.3) R package to visualize DEGs in volcano and heat maps.

The "ClusterProfiler" (v3.14.3) R package (Yu et al., 2012) was used for the functional annotation and Gene Set Enrichment Analysis (GSEA) of the DEGs. The curated reference genesets from the MsigDB file (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp): c2. cp.v7.2. symbols.gmt were selected for GSEA (Subramanian et al., 2005). The analysis results were significantly enriched with |NES| > 1, p. adj <0.05, and we used q-value <0.25 as thresholds.

2.5 Analysis of immune cell infiltration of PER2 in HCC

The ssGSEA algorithm in the "GSVA" (v1.34.0) R package (Hänzelmann et al., 2013) evaluated the tumor infiltration of 24 immune cell types in HCC tissues. These 24 immune cell markers were plasmacytoid dendritic cells (pDC), cytotoxic cells, dendritic cells (DC), T cells, B cells, CD56 bright natural killer cells (NDC), CD56 bright natural killer cells (NK CD56bright cells), gamma delta T cells (Tgd), immature dendritic cells (iDC), macrophages, effector memory T cells (Tem), regulatory T cells (TReg), type 1 helper cells (Th1 cells), neutrophils, CD56 dim natural killer cells (NK CD56dim cells), mast cells, T follicular helper cells (TFH), activated dendritic cells (aDC), type 2 helper cells (Th2 cells), CD8 T cells, natural killer cells (NK cells), eosinophils, Type 17 helper cells (Th17 cells), Central memory T cells (Tcm), and helper T cells (T helper cells) (Bindea et al., 2013). The immune infiltration in HCC tissues was assessed using the StromalScore, ImmuneScore, and ESTIMATEScore algorithms in the "estimate" (v1.0.13) R package. The relationship between PER2 expression, immune cell infiltration status, and immune cell markers was determined using Spearman correlation analysis.

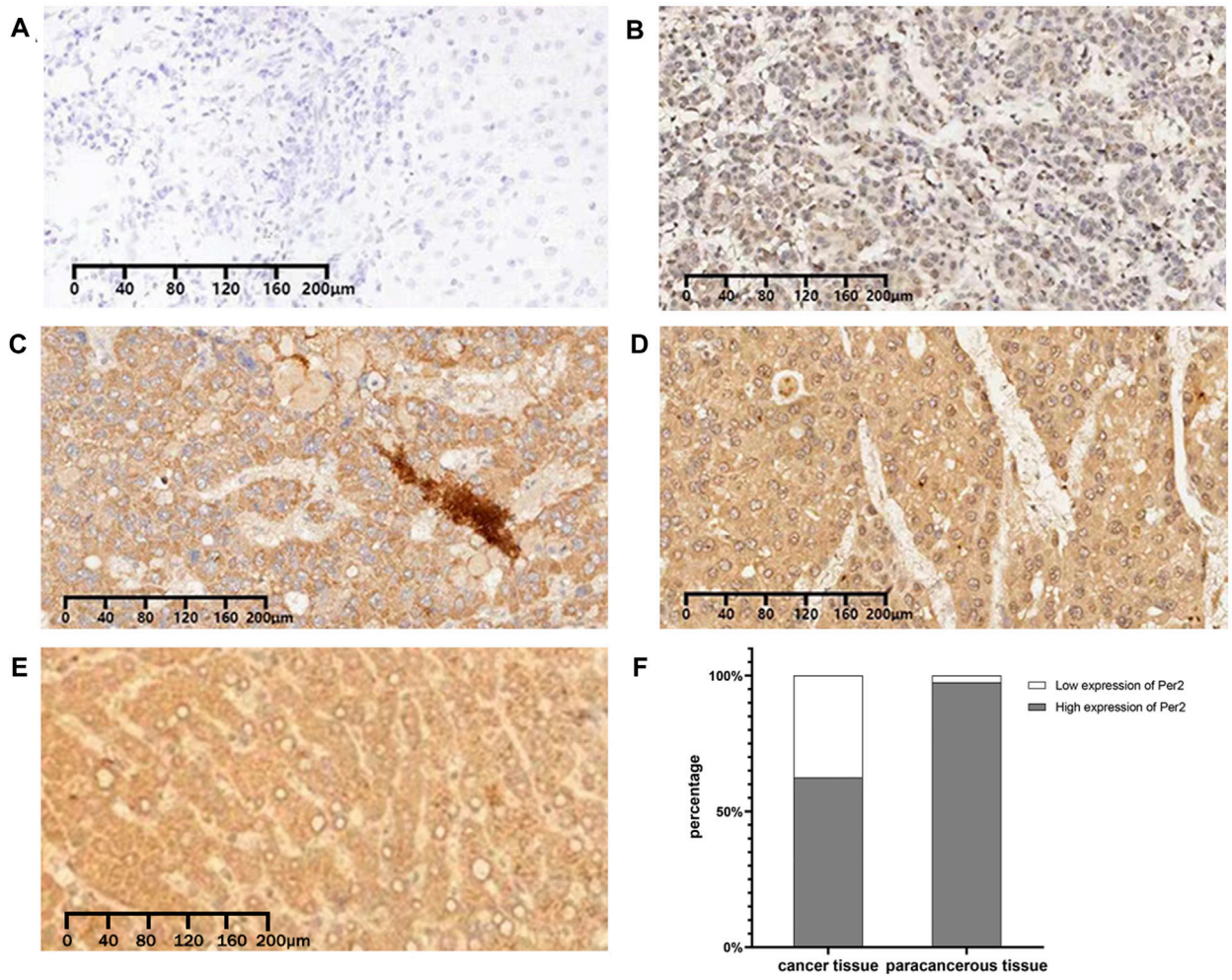


FIGURE 2
Immunohistochemical expression of PER2 protein in HCC tissues and paracancerous tissues. **(A–D)** The expression of PER2 was negative **(A)**, weakly positive **(B)**, moderately positive **(C)**, and strongly positive **(D)** in HCC. **(E)** The expression of PER2 was strongly positive in paracancerous tissues. **(F)** Stacked column diagram of PER2 protein expression in hepatocellular carcinoma and paracancerous tissues. IHC staining, ×100.

TABLE 1 Result of immunohistochemical staining of PER2 protein in HCC and paracancerous tissues.

Group	N	PER2 expression		χ^2	*p*-value
		Low	High		
Cancer tissues	80	30	50	30.625	<0.05
Paracancerous tissues	80	2	78		

2.6 Correlation analysis of PER2 expression level with other core circadian rhythm genes, immune checkpoint genes, and TP53 in HCC

Spearman correlation analysis determined the relationship between PER2 expression level and core circadian rhythm genes, immune checkpoint genes, and TP53. The core circadian rhythm gene (Jiang et al., 2021) and immune checkpoint gene (Hu et al., 2021) were obtained from a literature review. The correlation was considered significant, with $p < 0.05$ as the threshold. We used the "ggplot 2"(V3.3.3) R package to visualize the heat maps and scatter plots analysis results.

2.7 Analysis of DNA methylation status in the CpG islands of PER2

The DNA methylation status in the CpG sites of the PER2 gene was analyzed in TCGA using the MetSurv database (https://biit.cs.ut.ee/methsurv/). The prognostic value of the CpG methylation status of PER2 was evaluated in the HCC samples.

10.3389/fmolb.2023.1264553

TABLE 2 Clinicopathological characteristics of HCC patients with high- and low-PER2 expression levels.

Characteristics	Number	PER2 expression levels		*p*
		Low	High	
Total number of patients	80	30	50	
Gender				0.306
Male	70	28	42	
Female	10	2	8	
Age				0.649
<50	14	6	8	
≥50	66	24	42	
Operation				0.645
Hepatectomy	75	29	46	
Liver transplantation	5	1	4	
Drinking				0.712
No	54	21	33	
Yes	26	9	17	
Smoking				0.902
No	54	20	34	
Yes	26	10	16	
BMI				0.229
<18 kg/m^2	5	0	5	
18~24 kg/m^2	39	15	24	
≥24 kg/m^2	36	15	21	
AFP				0.424
<200 μg/L	60	21	39	
≥200 μg/L	20	9	11	
HBV				0.083
No	16	3	13	
Yes	64	27	37	
Tumor number				0.144
≤3	72	25	47	
>3	8	5	3	
Vascular tumor thrombus				0.628
No	76	28	48	
Yes	4	2	2	
Maximum tumor diameter				0.649
<5cm	66	24	42	
≥5cm	14	6	8	
Histological grade				0.124
Poor differentiation	30	15	15	

(Continued on following page)

Table 2 (*Continued*) Clinicopathological characteristics of HCC patients with high- and low-PER2 expression levels.

Characteristics	Number	PER2 expression levels		*p*
		Low	High	
Moderate differentiation	41	11	30	
High differentiation	9	4	5	
Satellite lesions				0.706
(-)	73	28	45	
(+)	7	2	5	
MVI risk grade				0.383
M0	45	15	30	
M1 + M2	35	15	20	
Nerve invasion				0.017
(-)	76	26	50	
(+)	4	4	0	
Hepatic capsule invasion				1.000
(-)	70	26	44	
(+)	10	4	6	
Involve peripheral organs				0.553
(-)	77	28	49	
(+)	3	2	1	
Child-Pugh grade				0.004
A	72	23	49	
B	8	7	1	
CNLC stage				0.004
I	62	18	44	
II+III+IV	18	12	6	

3 Results

3.1 The mRNA and protein expression levels of PER2 were significantly lower in HCC tissues than in their paired paracancerous tissues

To measure mRNA and protein expression levels of the circadian gene PER2 in HCC tissues and paired paracancerous tissues, we randomly selected liver cancer tissues and paired paracancerous tissues from 30 liver cancer patients from our liver cancer sample library and measured PER2 mRNA expression levels using qRT-PCR (Figures 1A, B), randomly selected liver cancer tissues and paired paracancerous tissues from 54 liver cancer patients. We measured PER2 protein expression levels using Western blotting (Figures 1C, D). The mRNA and protein expression levels of PER2 in HCC tissues were significantly lower than in the paired paracancerous tissues ($p < 0.001$).

We measured the expression of PER2 protein in 80 pairs of HCC tissues and paired paracancerous tissues using immunohistochemical staining. The Ningbo Pathology Center performed paraffin tissue sections of the specimens and immunohistochemical staining scoring. A negative (−) or weakly positive (+) staining result indicated low PER2 expression; a moderately positive (++) or strongly positive (++++) staining result indicated high PER2 expression. In liver cancer tissues, high expression of PER2 protein accounted for 62.5%, and low expression of PER2 protein accounted for 37.5%. In paracancerous tissues, high expression of PER2 protein accounted for 97.5%, and low expression of PER2 protein accounted for 2.5% (Figure 2). There was significantly less expression of PER2 protein in liver cancer tissues than in paracancerous tissues ($p < 0.05$) (Table 1).

3.2 Correlation between the expression of the PER2 gene and clinicopathological parameters and oncological behavior of patients with HCC

To analyze the relationship between the circadian rhythm gene PER2 expression level in liver cancer tissues and the clinicopathologic

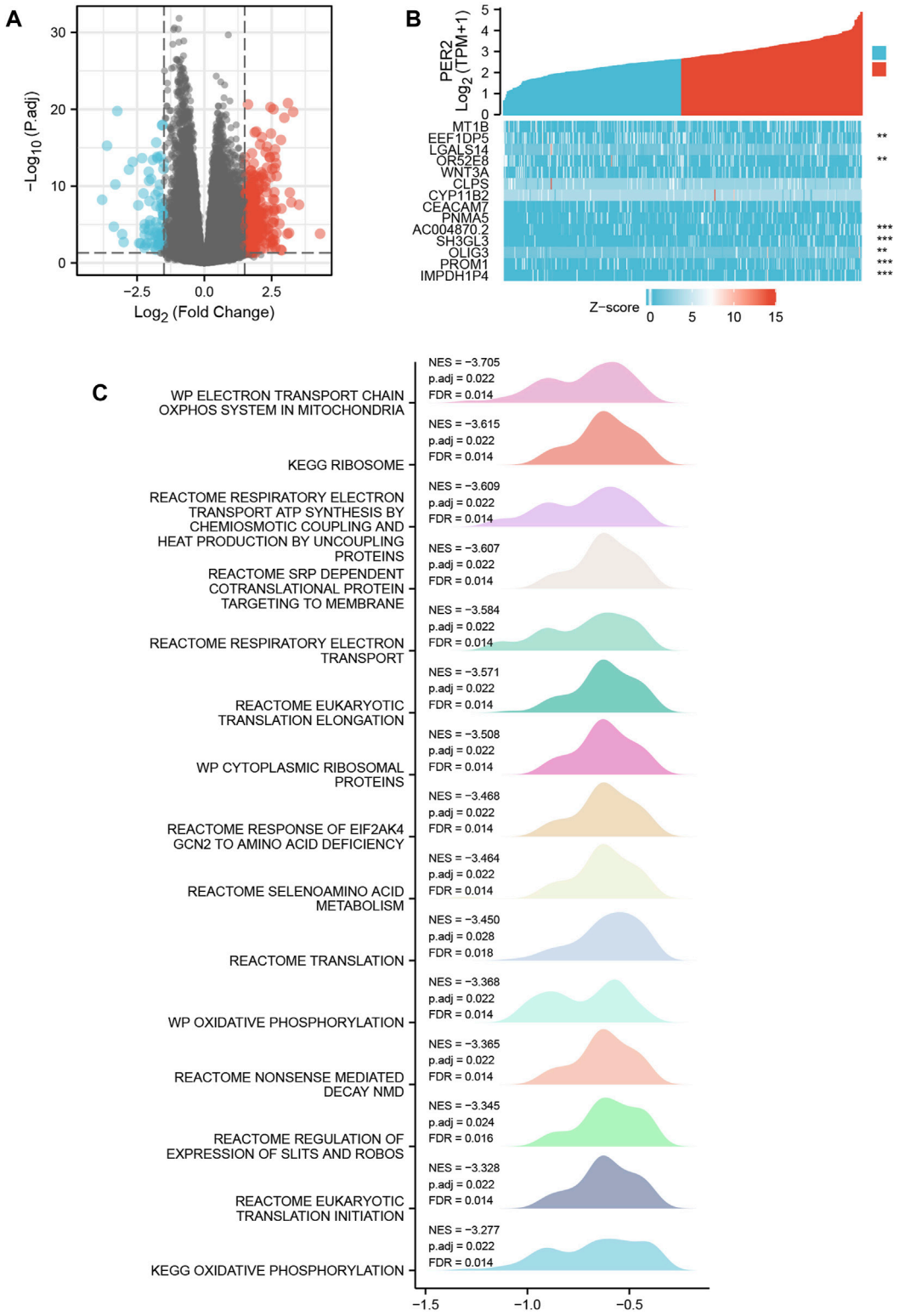


FIGURE 3
PER2-related differentially expressed genes (DEGs) and their functional enrichment analysis in HCC. **(A)** The volcano plot and **(B)** the correlation heat map of DEGs with high and low expression of PER2 in HCC. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. **(C)** Gene Set Enrichment Analysis of PER2 expression-related DEGs in HCC.

and oncological characteristics, we collected clinical data from 80 patients with liver cancer corresponding to liver cancer tissues. We recorded sex, age, operation mode, smoking and drinking habits, BMI, serum AFP level, whether they were infected with hepatitis B virus, tumor number, maximum tumor diameter, histological grading, microvascular invasion risk grading, Child-Pugh grading, and CNLC staging. Samples were

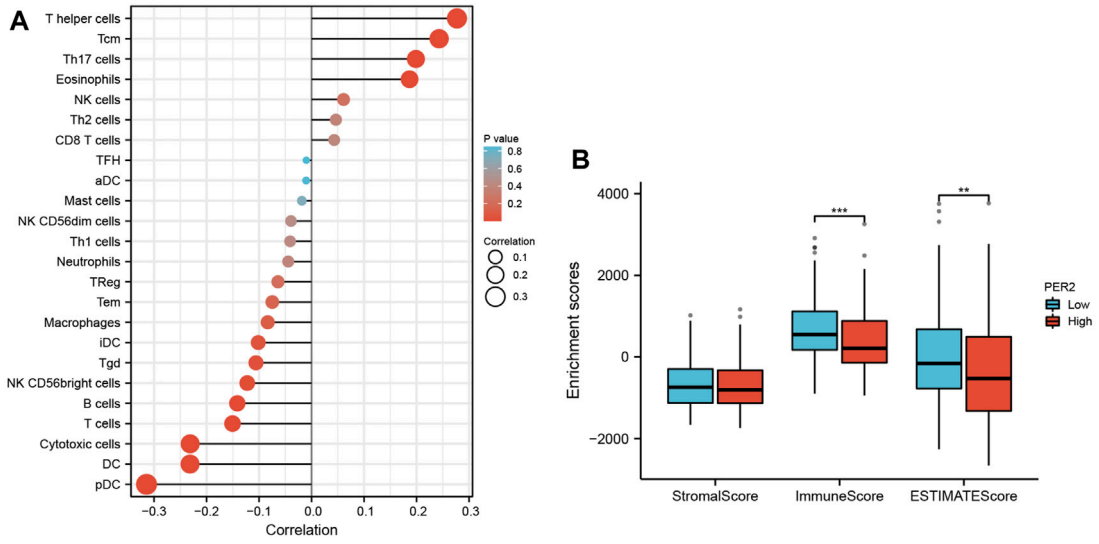


FIGURE 4
The correlation analysis between immune cell infiltration and PER2 expression in HCC. **(A)** Spearman correlation analysis results between infiltration levels of 24 immune cell types and PER2 expression levels in the HCC tissues. **(B)** Spearman correlation analysis results between the immune infiltration score and PER2 expression levels in the HCC tissues. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

divided into high and low PER2 expression groups, with 50 people in the high-expression group and 30 in the low-expression group (Table 2). PER2 expression significantly correlated with nerve invasion ($p = 0.017$), Child-Pugh grading ($p = 0.004$), and CNLC staging ($p = 0.004$).

3.3 DEGs with high and low PER2 expression in HCC

Based on the transcriptome data of HCC in TCGA, 424 HCC patients were divided into two groups according to the median expression value of the PER2 gene, and the differential expression between the two groups was analyzed. Taking |log2(FC)| >1.5 and p. adj <0.05 as threshold parameters, DEGs related to PER2 expression were obtained. Compared with the low expression group, there were 365 DEGs in the high PER2 expression group, of which 295 were upregulated, and 70 were downregulated (Figure 3A). Fourteen crucial DEGs were presented as a co-expression heat map (Figure 3B).

3.4 Functional enrichment analysis of PER2-related DEGs in HCC

To study the biological function and signal pathways related to the PER2 gene in HCC, we used the transcriptome data from TCGA to group samples according to the median expression value of PER2 and obtained the DEGs related to PER2. We used the "Cluster Profiler" R package and the GSEA of PER2-related DEGs in HCC patients; the possible functions or pathways involved in PER2 gene expression were inferred. The predefined gene set used in GSEA came from the c2.cp.v7.2.symbols.gmt gene set in the MSigDB database. The PER2-related DEGs were enriched in mitochondrial oxidative phosphorylation, transcription and translation, amino acid metabolism, and other related pathways (|NES| > 1, p. adj <0.05, q value < 0.2) (Figure 3C).

3.5 The expression of PER2 in HCC is related to immune cell infiltration

Transcriptome data from HCC in TCGA were used to evaluate the infiltration status of 24 immune cell types using the ssGSEA algorithm in the "GSVA" R package. The correlation between PER2 expression and immune cell infiltration was calculated using Spearman correlation analysis. The expression of PER2 mRNA was correlated with many immune cells, including plasma cell-like pDC, cytotoxic cells, DC, T cells, B cells, NK CD56 bright cells, Tgd, iDC, eosinophils, Th17 cells, Tcm, and helper cells (Figure 4A; Table 3). We also used the "ESTIMATE" Score package to score the immune infiltration of HCC tissues. The results showed that the low expression of PER2 was significantly correlated with the immune infiltration score ($p < 0.001$) and with the immune and matrix comprehensive score ($p < 0.01$) (see Figure 4B). PER2 expression in HCC is related to immune cell infiltration.

3.6 PER2 expression in HCC is related to the expression of many core circadian rhythm genes, immune checkpoint genes, and TP53

A literature review revealed that there are a least 15 core circadian rhythm genes: PER1, PER2, PER3, CLOCK, CRY1, CRY2, ARNTL/BMAL1, TIMLESS/TIM, RORA, RORB, RORC, NPAS2, NR1D1, NR1D2, and CSNK1E/CKIε (Jiang et al., 2021). Using the transcriptome data of HCC in TCGA, the correlation between the

10.3389/fmolb.2023.1264553

TABLE 3 Correlation analysis results between infiltration levels of 24 immune cell types and PER2 expression levels in HCC tissues.

Immune cell	Correlation	*p*
pDC	−0.309	<0.001
Cytotoxic cells	−0.237	<0.001
DC	−0.232	<0.001
T cells	−0.158	0.002
B cells	−0.151	0.004
NK CD56bright cells	−0.132	0.011
Tgd	−0.112	0.030
iDC	−0.111	0.031
Macrophages	−0.093	0.074
Tem	−0.078	0.13
TReg	−0.074	0.154
Th1 cells	−0.051	0.326
Neutrophils	−0.049	0.348
NK CD56dim cells	−0.046	0.376
Mast cells	−0.028	0.588
TFH	−0.02	0.697
aDC	−0.016	0.752
Th2 cells	0.044	0.396
CD8 T cells	0.046	0.373
NK cells	0.061	0.238
Eosinophils	0.177	<0.001
Th17 cells	0.204	<0.001
Tcm	0.253	<0.001
T helper cells	0.271	<0.001

PER2 gene and other core circadian rhythm genes was analyzed and expressed as a thermogram (Figure 5C). PER2 expression correlated with many core circadian rhythm genes ($p < 0.05$), including PER3, CLOCK, and NR1D2 (correlation >0.5) (Table 4).

Given the broad application of immune checkpoint inhibitors in anti-tumor treatment, we studied the relationship between PER2 expression and immunosuppression checkpoints. A literature review revealed 79 immune checkpoint genes (Hu et al., 2021). We analyzed the transcriptome data of HCC in TCGA to calculate the correlation between PER2 and immune checkpoint gene expression. PER2 positively correlated with 32 immune checkpoint genes and negatively correlated with one immune checkpoint gene ($p < 0.05$). These correlated immune checkpoint genes were presented as heat maps (Figure 5A).

TP53 is a tumor suppressor gene with low expression in normal cells and high expression in malignant tumors (Marei et al., 2021). Using correlation analysis of transcriptome data of HCC in TCGA, we found that PER2 expression positively correlated with TP53 expression (Figure 5B).

3.7 Correlation between methylation status of the PER2 gene and outcomes in HCC patients

The DNA methylation level in the PER2 gene and the prognostic value of each CpG locus in the PER2 gene were analyzed using the MetSurv tool. The thermogram results revealed the methylation levels of 21 CpG sites in HCC (Figure 6A). The methylation levels of eight CpG sites (cg03004097, cg04169774, cg20070418, cg22879834, cg21315421, cg24831107, cg12308675 and cg06259818) were significantly correlated HCC outcomes ($p < 0.05$) (Table 5). Patients with hypermethylation of cg03004097 had better survival than patients with hypomethylation ($p < 0.01$, hazard ratio [HR] = 0.579), while patients with hypermethylation of cg06259818 had significantly worse survival outcomes than those with hypomethylation ($p < 0.01$, HR = 1.982) (Figures 6B, C). These findings suggest that the methylation status of the PER2 gene is related to HCC outcomes.

4 Discussion

Downregulation of PER2 expression is related to poor outcomes in several tumors, including chronic lymphocytic leukemia, renal cancer, head, and neck squamous cell carcinoma, colorectal cancer, and others (Momma et al., 2017; Xiong et al., 2018; Qiu et al., 2019; Wang et al., 2020). Previous studies showed that PER2 mRNA and protein in HCC tissue were lower than in paired adjacent tissues (Yang et al., 2014; Li et al., 2017; Li et al., 2018). These findings are consistent with the results of the present study. Based on the clinical and pathological data analysis of patients with liver cancer, PER2 protein expression in liver cancer tissue was significantly related to nerve invasion, Child-Pugh grading, and CNLC staging. These findings suggest that the circadian rhythm gene PER2 is critical in HCC.

Circadian rhythm regulation involves a large and complex interactive regulation network that maintains an organism's steady state (Momma et al., 2017). Chronic disturbance of circadian rhythm is related to metabolic diseases, and is closely related to increased cancer risk (Crespo et al., 2021). Some core circadian rhythm genomes have become feedback loops of gene transcription and translation (Shafi and Knudsen, 2019). Basic helix-loop-helix heterodimeric transcription factors (CLOCK/BMAL1 or BMAL1/NPAS2) transcribe-translate the feedback loop in a negative feedback way, thus regulating the expression of CRY1, CRY2, PER1, PER2, and PER3 (Shafi and Knudsen, 2019). It is worth noting that circadian rhythm genes play a crucial role in biological processes such as apoptosis, cell aging, DNA damage repair, and metastasis by rhythmically regulating gene expression and gene activity in the whole genome (Shafi and Knudsen, 2019; Crespo et al., 2021; Huang et al., 2023). More and more pieces of evidence show the importance of circadian rhythm genes in the diagnosis, treatment, and cancer outcomes (Angelousi et al., 2019; Qiu et al., 2019; Zhang et al., 2020; Sahar et al., 2022). This study found a correlation between PER2 expression and several core circadian rhythm genes, confirming the complex regulation mechanism network. If the rhythm expression of the

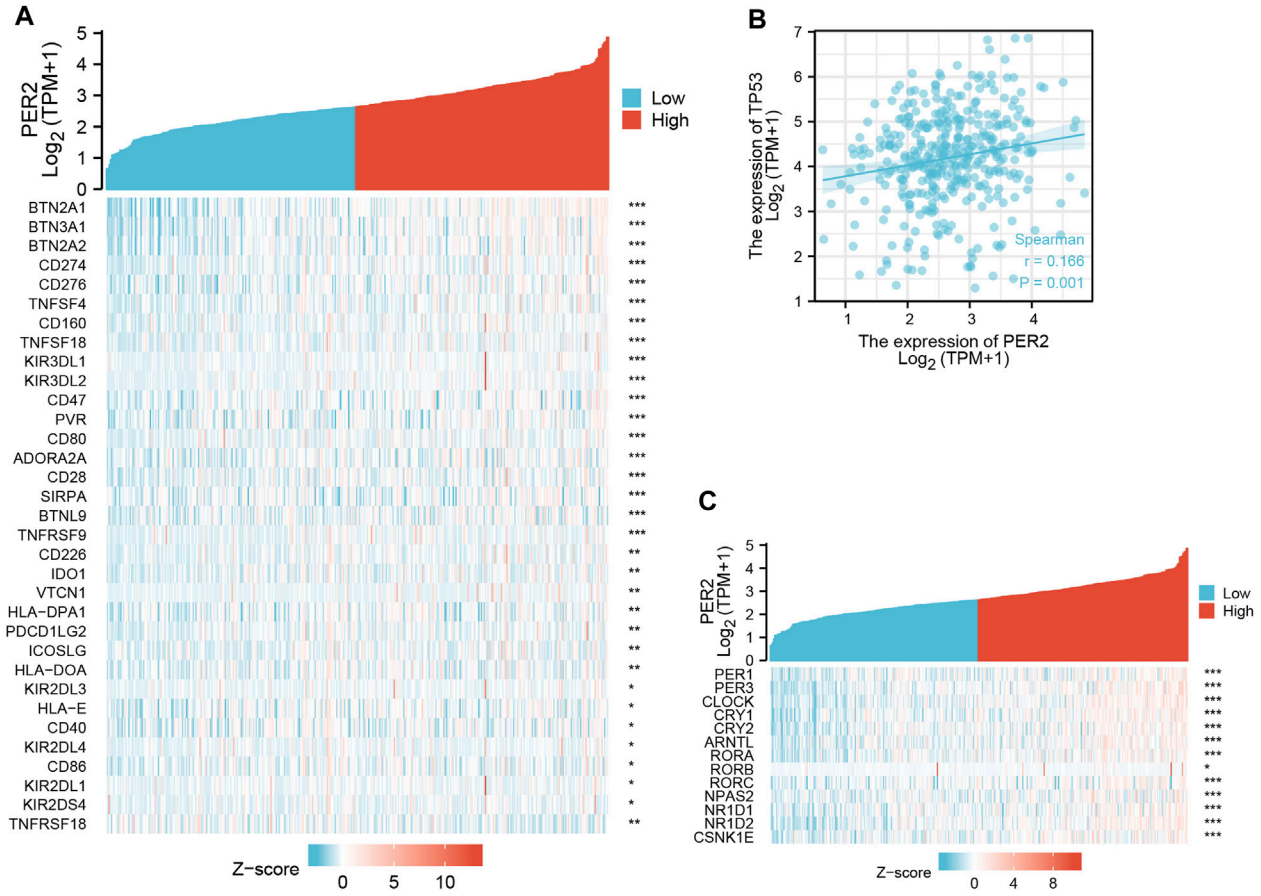


FIGURE 5
The correlation analysis between PER2 expression and **(A)** immune checkpoint genes, **(B)** TP53, and **(C)** other core circadian rhythm genes in HCC. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

PER2 gene oscillates and cannot be adjusted by the circadian rhythm regulation mechanism, it causes circadian rhythm disorders, which lead to many diseases, including liver cancer. Several studies found that the circadian rhythm plays an essential role in the occurrence and development of tumors, and "time therapy" has been recognized by many researchers and physicians (Zhou et al., 2021). Nevertheless, the mechanism of circadian rhythm genes, including PER2, in tumorigenesis and development is exploratory.

This study used bioinformatics analysis to determine the potential mechanisms of PER2 in liver cancer. We obtained a set of DEGs related to PER2 expression through single gene difference analysis and then analyzed their functions and pathways to understand their relationship to PER2. The DEGs related to PER2 expression were enriched in mitochondrial oxidative phosphorylation, transcription and translation, amino acid metabolism, and other related pathways. PER2 and other core circadian rhythm genes form a transcription-translation feedback loop; abnormal transcription and translation cause tumor proliferation, metastasis, and invasion. Experimental studies confirmed that the dysfunction of mitochondrial oxidative phosphorylation promotes the proliferation, metastasis, and invasion of primary liver cancer (Lee et al., 2021; Zhang et al., 2023). Branched-chain amino acid catabolism can break glutamine metabolism, maintaining HCC progression (Yang et al., 2022). These studies suggested that PER2 may be involved in the proliferation, metastasis, and invasion of HCC through mitochondrial oxidative phosphorylation, transcription and translation, and amino acid metabolism; nevertheless, we need further research and experiments to elucidate the mechanisms fully.

The role of immune cell infiltration in cancer development and progress has attracted substantial attention in recent years (Dai et al., 2022). This study found that PER2 expression in HCC showed a strongly negative correlation with pDC. There is pDC accumulation in blood, tumor tissue, and ascites in HCC patients, and high-density tumor infiltration of pDC is related to poor outcomes (Pang et al., 2021). Pang et al. found that IFNα secreted by pDC induced postoperative recurrence of HCC by promoting the recruitment of marrow-derived inhibitory cells (Pang et al., 2022). PER2 may play an essential role in regulating the tumor immune microenvironment. Nevertheless, the mechanism by which PER2 affects the immune microenvironment and tumor progression in HCC remains unclear, and further research is needed to clarify the biological effects of PER2 in HCC.

TABLE 4 The correlation analysis between PER2 and other core circadian rhythm genes expression.

Core circadian rhythm genes	Correlation	*p*
CLOCK	0.566	<0.001
PER3	0.558	<0.001
NR1D2	0.543	<0.001
CRY1	0.466	<0.001
CRY2	0.452	<0.001
RORA	0.421	<0.001
ARNTL	0.398	<0.001
RORC	0.39	<0.001
NR1D1	0.343	<0.001
CSNK1E	0.331	<0.001
PER1	0.315	<0.001
NPAS2	0.241	<0.001
RORB	0.13	0.012

In recent years, the application of immune checkpoint inhibitors has become common and has also significantly improved HCC outcomes (Wang Z. et al., 2023). TP53 mutates in about half of human malignant tumors, including breast, colon, lung, liver, prostate, bladder, and skin malignant tumors (Marei et al., 2021; Wang H. et al., 2023). TP53 mutations inhibit anti-tumor immunity and reduce the efficacy of cancer immunotherapy (Wang H. et al., 2023; Marvalim et al., 2023). Therefore, we evaluated the relationship between PER2 expression level, immune checkpoint genes, and the TP53 gene. PER2 expression positively correlated with immune checkpoint genes and TP53 expression in HCC tissues. Interestingly, studies showed that PER2 binding prevents MDM2-mediated ubiquitination of tumor suppressor p53, regulating the p53 stability (Gotoh et al., 2016). These findings suggest that PER2 is involved in liver cancer and may be a potential target to improve immunotherapy, contribute to the research and development of new drugs, and provide a strategy and research direction for immunotherapy for HCC.

DNA methylation is an epigenetic mechanism. Previous studies showed that the change of DNA methylation pattern is critical in HCC and has potential value for diagnosis and outcome prediction (Long et al., 2019; Qu et al., 2020). Ma et al. found that hypermethylation of five CpG loci of the EXO1 gene was related to poor overall survival of HCC patients (Ma et al., 2022), and Yuan et al. found that hypermethylation of two CpG loci of the DDX1 gene was related to poor outcomes (Yuan et al., 2022). However, no study has shown the relationship between the methylation level of the circadian rhythm gene PER2 and HCC outcomes. Therefore, we conducted methylation analysis and evaluated the relationship between the methylation level of each CpG site of the PER2 gene and HCC outcomes. We identified eight CpG loci, among which the hypermethylation of six CpG loci, cg20070418, cg22879834, cg21315421, cg24831107, cg12308675, and cg06259818 were associated with poor outcomes. These findings suggest that the methylation status of PER2 is related to HCC outcomes. Nevertheless, we need *in vivo* and *in vitro* evidence to confirm the relationship between PER2 methylation and HCC outcomes.

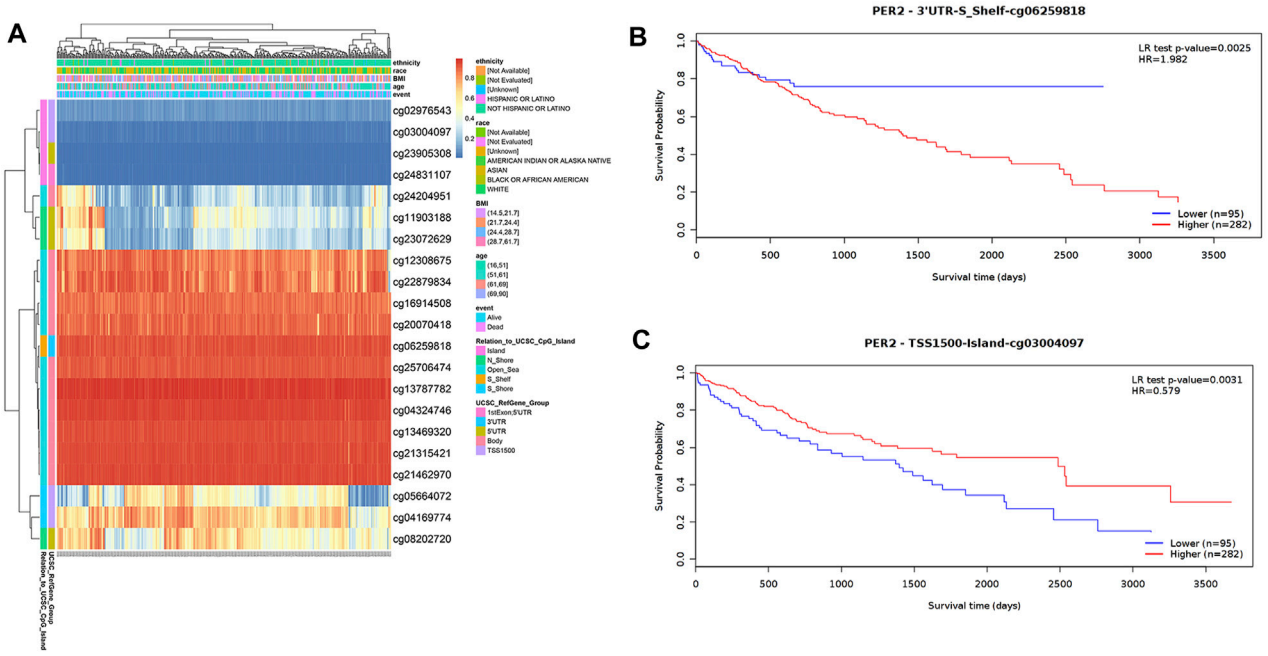


FIGURE 6
Analysis of DNA methylation status in the CpG islands of the PER2. **(A)** The visualization of 21 CpG sites on PER2 with methylation level in HCC. **(B,C)** The Kaplan-Meier survival curves of the DNA methylation of PER2 at cg06259818 **(B)** and cg03004097 **(C)** probes.

TABLE 5 Effects of methylation levels in the CpG sites of the PER2 gene on HCC outcomes.

CpG island	HR	*p*
TSS1500—Island—cg03004097	0.579	0.0024
TSS1500—S_Shore—cg04169774	0.703	0.044
TSS1500 - S_Shore - cg05664072	0.741	0.089
Body—Open_Sea—cg13787782	0.769	0.18
Body—Open_Sea—cg21462970	0.842	0.33
5′UTR—N_Shore—cg23072629	0.842	0.33
Body—Open_Sea—cg16914508	0.87	0.43
5′UTR—N_Shore—cg11903188	0.878	0.46
5′UTR—N_Shore—cg08202720	0.881	0.51
Body—Open_Sea—cg25706474	0.882	0.52
TSS1500—Island—cg02976543	1.134	0.48
Body—Open_Sea—cg13469320	1.229	0.24
Body—Open_Sea—cg24204951	1.29	0.23
5′UTR—Island—cg23905308	1.291	0.15
Body—Open_Sea—cg20070418	1.438	0.049
Body—Open_Sea—cg22879834	1.481	0.025
Body—Open_Sea—cg04324746	1.49	0.079
Body—Open_Sea—cg21315421	1.529	0.018
1stExon; 5′UTR—Island—cg24831107	1.562	0.038
Body—Open_Sea—cg12308675	1.645	0.024
3′UTR—S_Shelf—cg06259818	1.982	0.005

Our study has some limitations. Most of our results were based on RNA sequencing data of HCC tissues from the TCGA database. However, we could not verify the analysis results of bioinformatics by basic experiments. Therefore, further *in vivo* and *in vitro* experiments are necessary to investigate the mechanisms of PER2 in HCC.

5 Conclusion

The circadian rhythm gene PER2 was expressed at low levels in HCC, and its expression level was correlated with nerve invasion, Child-Pugh grading, CNLC staging, immune cell infiltration, and immune checkpoint genes. PER2 expression was positively correlated with TP53, while its methylation status was related to outcomes. DEGs related to PER2 are enriched in mitochondrial oxidative phosphorylation, transcription and translation, amino acid metabolism, and other pathways. Therefore, PER2 may be a new prognostic and therapeutic marker for HCC. However, further research is needed to validate our findings.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/; https://biit.cs.ut.ee/methsurv/.

Ethics statement

The studies involving humans were approved by the Ethics Committee of Ningbo Medical Center Lihuili Hospital (number: KY2021PJ231). 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

XC: Formal analysis, Investigation, Writing–original draft. QZ: Conceptualization, Writing–review and editing. HW: Resources. KQ: Resources. XD: Investigation. FX: Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and of this article. This work was supported by the Ningbo Natural Fund Project (grant no. 2021J295, 2022J261), Ningbo Medical Center Lihuili Hospital, “Huili Fund” (2022ZD005) and Ninghai County Social Development Science and Technology Project (202301).

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/fmolb.2023.1264553/full#supplementary-material

References

Allemani, C., Matsuda, T., Di Carlo, V., Harewood, R., Matz, M., Nikšić, M., et al. (2018). Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* 391 (10125), 1023–1075. doi:10.1016/s0140-6736(17)33326-3

Angelousi, A., Kassi, E., Ansari-Nasiri, N., Randeva, H., Kaltsas, G., and Chrousos, G. (2019). Clock genes and cancer development in particular in endocrine tissues. *Endocr. Relat. Cancer* 26 (6), R305–r317. doi:10.1530/erc-19-0094

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

Crespo, M., Leiva, M., and Sabio, G. (2021). Circadian clock and liver cancer. *Cancers (Basel)* 13 (14), 3631. doi:10.3390/cancers13143631

Dai, K., Liu, C., Guan, G., Cai, J., and Wu, L. (2022). Identification of immune infiltration-related genes as prognostic indicators for hepatocellular carcinoma. *BMC Cancer* 22 (1), 496. doi:10.1186/s12885-022-09587-0

El Jabbour, T., Lagana, S. M., and Lee, H. (2019). Update on hepatocellular carcinoma: pathologists' review. *World J. Gastroenterol.* 25 (14), 1653–1665. doi:10.3748/wjg.v25.i14.1653

Gotoh, T., Kim, J. K., Liu, J., Vila-Caballer, M., Stauffer, P. E., Tyson, J. J., et al. (2016). Model-driven experimental approach reveals the complex regulatory distribution of p53 by the circadian factor Period 2. *Proc. Natl. Acad. Sci. U. S. A.* 113 (47), 13516–13521. doi:10.1073/pnas.1607984113

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

Hartke, J., Johnson, M., and Ghabril, M. (2017). The diagnosis and treatment of hepatocellular carcinoma. *Semin. Diagn Pathol.* 34 (2), 153–159. doi:10.1053/j.semdp.2016.12.011

Hu, F. F., Liu, C. J., Liu, L. L., Zhang, Q., and Guo, A. Y. (2021). Expression profile of immune checkpoint genes and their roles in predicting immunotherapy response. *Brief. Bioinform* 22 (3), bbaa176. doi:10.1093/bib/bbaa176

Huang, C., Zhang, C., Cao, Y., Li, J., and Bi, F. (2023). Major roles of the circadian clock in cancer. *Cancer Biol. Med.* 20 (1), 1–24. doi:10.20892/j.issn.2095-3941.2022.0474

Jiang, Y., Shen, X., Fasae, M. B., Zhi, F., Chai, L., Ou, Y., et al. (2021). The expression and function of circadian rhythm genes in hepatocellular carcinoma. *Oxid. Med. Cell Longev.* 2021, 4044606. doi:10.1155/2021/4044606

Lee, H. Y., Nga, H. T., Tian, J., and Yi, H. S. (2021). Mitochondrial metabolic signatures in hepatocellular carcinoma. *Cells* 10 (8), 1901. doi:10.3390/cells10081901

Li, B., Wang, S., Yan, Z., Zhang, J., Yang, T., Zhang, H., et al. (2018). Prognostic value of Per2 in hepatocellular carcinoma and its effect on the growth and apoptosis of hepatocellular carcinoma cells. *Basic and Clin. Med.* 38 (4), 464–469. doi:10.16352/j.issn.1001-6325.2018.04.006

Li, H., Lu, Y. F., Chen, H., and Liu, J. (2017). Dysregulation of metallothionein and circadian genes in human hepatocellular carcinoma. *Chronobiol Int.* 34 (2), 192–202. doi:10.1080/07420528.2016.1256300

Long, J., Chen, P., Lin, J., Bai, Y., Yang, X., Bian, J., et al. (2019). DNA methylation-driven genes for constructing diagnostic, prognostic, and recurrence models for hepatocellular carcinoma. *Theranostics* 9 (24), 7251–7267. doi:10.7150/thno.31155

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 (12), 550. doi:10.1186/s13059-014-0550-8

Ma, J., Jin, J., Lu, H., Zhang, J., Li, Y., and Cai, X. (2022). Exonuclease 1 is a potential diagnostic and prognostic biomarker in hepatocellular carcinoma. *Front. Mol. Biosci.* 9, 889414. doi:10.3389/fmolb.2022.889414

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. doi:10.1186/s12935-021-02396-8

Marvalim, C., Datta, A., and Lee, S. C. (2023). Role of p53 in breast cancer progression: an insight into p53 targeted therapy. *Theranostics* 13 (4), 1421–1442. doi:10.7150/thno.81847

Momma, T., Okayama, H., Saitou, M., Sugeno, H., Yoshimoto, N., Takebayashi, Y., et al. (2017). Expression of circadian clock genes in human colorectal adenoma and carcinoma. *Oncol. Lett.* 14 (5), 5319–5325. doi:10.3892/ol.2017.6876

Pang, L., Ng, K. T., Liu, J., Yeung, W. O., Zhu, J., Chiu, T. S., et al. (2021). Plasmacytoid dendritic cells recruited by HIF-1α/eADO/ADORA1 signaling induce immunosuppression in hepatocellular carcinoma. *Cancer Lett.* 522, 80–92. doi:10.1016/j.canlet.2021.09.022

Pang, L., Yeung, O. W. H., Ng, K. T. P., Liu, H., Zhu, J., Liu, J., et al. (2022). Postoperative plasmacytoid dendritic cells secrete IFNα to promote recruitment of myeloid-derived suppressor cells and drive hepatocellular carcinoma recurrence. *Cancer Res.* 82 (22), 4206–4218. doi:10.1158/0008-5472.Can-22-1199

Qiu, M. J., Liu, L. P., Jin, S., Fang, X. F., He, X. X., Xiong, Z. F., et al. (2019). Research on circadian clock genes in common abdominal malignant tumors. *Chronobiol Int.* 36 (7), 906–918. doi:10.1080/07420528.2018.1477792

Qu, N., Qin, S., Zhang, X., Bo, X., Liu, Z., Tan, C., et al. (2020). Multiple m(6)A RNA methylation modulators promote the malignant progression of hepatocellular carcinoma and affect its clinical prognosis. *BMC Cancer* 20 (1), 165. doi:10.1186/s12885-020-6638-5

Sahar, N. E., Qadir, J., Riaz, S. K., Bagabir, S. A., Muneer, Z., Sheikh, A. K., et al. (2022). Dysregulated expression of suppressor loop of circadian rhythm genes in colorectal cancer pathogenesis. *Minerva Med.* 113 (3), 497–505. doi:10.23736/s0026-4806.22.07981-2

Shafi, A. A., and Knudsen, K. E. (2019). Cancer and the circadian clock. *Cancer Res.* 79 (15), 3806–3814. doi:10.1158/0008-5472.Can-19-0566

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 (43), 15545–15550. doi:10.1073/pnas.0506580102

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

Wang, H., Guo, M., Wei, H., and Chen, Y. (2023a). Targeting p53 pathways: mechanisms, structures, and advances in therapy. *Signal Transduct. Target Ther.* 8 (1), 92. doi:10.1038/s41392-023-01347-1

Wang, N., Mi, M., Wei, X., and Sun, C. (2020). Circadian clock gene Period2 suppresses human chronic myeloid leukemia cell proliferation. *Exp. Ther. Med.* 20 (6), 147. doi:10.3892/etm.2020.9276

Wang, Z., Wang, Y., Gao, P., and Ding, J. (2023b). Immune checkpoint inhibitor resistance in hepatocellular carcinoma. *Cancer Lett.* 555, 216038. doi:10.1016/j.canlet.2022.216038

Xiong, H., Yang, Y., Yang, K., Zhao, D., Tang, H., and Ran, X. (2018). Loss of the clock gene PER2 is associated with cancer development and altered expression of important tumor-related genes in oral cancer. *Int. J. Oncol.* 52 (1), 279–287. doi:10.3892/ijo.2017.4180

Xie, D. Y., Ren, Z. G., Zhou, J., Fan, J., and Gao, Q. (2020). 2019 Chinese clinical guidelines for the management of hepatocellular carcinoma: updates and insights. *Hepatobiliary Surg. Nutr.* 9 (4), 452–463. doi:10.21037/hbsn-20-480

Yang, D., Liu, H., Cai, Y., Lu, K., Zhong, X., Xing, S., et al. (2022). Branched-chain amino acid catabolism breaks glutamine addiction to sustain hepatocellular carcinoma progression. *Cell Rep.* 41 (8), 111691. doi:10.1016/j.celrep.2022.111691

Yang, S. L., Yu, C., Jiang, J. X., Liu, L. P., Fang, X., and Wu, C. (2014). Hepatitis B virus X protein disrupts the balance of the expression of circadian rhythm genes in hepatocellular carcinoma. *Oncol. Lett.* 8 (6), 2715–2720. doi:10.3892/ol.2014.2570

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

Yuan, M., Xu, J., Cao, S., and Sun, S. (2022). DDX1 is a prognostic biomarker and correlates with immune infiltrations in hepatocellular carcinoma. *BMC Immunol.* 23 (1), 59. doi:10.1186/s12865-022-00533-0

Zhang, J., Lv, H., Ji, M., Wang, Z., and Wu, W. (2020). Low circadian clock genes expression in cancers: A meta-analysis of its association with clinicopathological features and prognosis. *PLoS One* 15 (5), e0233508. doi:10.1371/journal.pone.0233508

Zhang, Y., Li, W., Bian, Y., Li, Y., and Cong, L. (2023). Multifaceted roles of aerobic glycolysis and oxidative phosphorylation in hepatocellular carcinoma. *PeerJ* 11, e14797. doi:10.7717/peerj.14797

Zhou, J., Wang, J., Zhang, X., and Tang, Q. (2021). New insights into cancer chronotherapies. *Front. Pharmacol.* 12, 741295. doi:10.3389/fphar.2021.741295

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 29 November 2023
DOI 10.3389/fmolb.2023.1298077

Check for updates

OPEN ACCESS

EDITED BY
Li Chen,
Huazhong University of Science and Technology, China

REVIEWED BY
Boxuan Zhou,
Third Affiliated Hospital of Sun Yat-sen University, China
Yancheng Song,
Qingdao University, China

*CORRESPONDENCE
Jinze Li,
1576328807@qq.com
Daoming Liang,
daomingliangkm@sina.com
Wence Zhou,
zhouwc129@163.com

†These authors share first authorship

RECEIVED 21 September 2023
ACCEPTED 15 November 2023
PUBLISHED 29 November 2023

CITATION
Chen H, Xu N, Xu J, Zhang C, Li X, Xu H, Zhu W, Li J, Liang D and Zhou W (2023), A risk signature based on endoplasmic reticulum stress-associated genes predicts prognosis and immunity in pancreatic cancer.
Front. Mol. Biosci. 10:1298077.
doi: 10.3389/fmolb.2023.1298077

COPYRIGHT
© 2023 Chen, Xu, Xu, Zhang, Li, Xu, Zhu, Li, Liang and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A risk signature based on endoplasmic reticulum stress-associated genes predicts prognosis and immunity in pancreatic cancer

Haofei Chen[1,2†], Ning Xu[3†], Jia Xu[4†], Cheng Zhang[3†], Xin Li[2], Hao Xu[5], Weixiong Zhu[1], Jinze Li[6*], Daoming Liang[3*] and Wence Zhou[1,2*]

[1]The Second Clinical Medical School, Lanzhou University, Lanzhou, China, [2]Department of General Surgery, Lanzhou University Second Hospital, Lanzhou, China, [3]The Second Affiliated Hospital of Kunming Medical University, Kunming, China, [4]Wuhan Blood Center, Wuhan, China, [5]Department of General Surgery, The First Hospital of Lanzhou University, Lanzhou, China, [6]Department of Gastrointestinal Surgery, The Third People's Hospital of Hubei Province, Wuhan, China

Introduction: The involvement of endoplasmic reticulum (ER) stress in cancer biology is increasingly recognized, yet its role in pancreatic cancer (PC) remains unclear. This study aims to elucidate the impact of ER stress on prognosis and biological characteristics in PC patients.

Methods: A bioinformatic analysis was conducted using RNA-seq data and clinicopathological information from PC patients in the TCGA and ICGC databases. The ER stress-associated gene sets were extracted from MSigDB. ER stress-associated genes closely linked with overall survival (OS) of PC patients were identified via log-rank test and univariate Cox analysis, and further narrowed by LASSO method. A risk signature associated with ER stress was formulated using multivariate Cox regression and assessed through Kaplan-Meier curves, receiver operating characteristic (ROC) analyses, and Harrell's concordance index. External validation was performed with the ICGC cohort. The single-sample gene-set enrichment analysis (ssGSEA) algorithm appraised the immune cell infiltration landscape.

Results: Worse OS in PC patients with high-risk signature score was observed. Multivariate analysis underscored our ER stress-associated signature as a valuable and independent predictor of prognosis. Importantly, these results based on TCGA were further validated in ICGC dataset. In addition, our risk signature was closely associated with homeostasis, protein secretion, and immune regulation in PC patients. In particular, PC microenvironment in the high-risk cluster exhibited a more immunosuppressive status. At last, we established a nomogram model by incorporating the risk signature and clinicopathological parameters, which behaves better in predicting prognosis of PC patients.

Discussion: This comprehensive molecular analysis presents a new predictive model for the prognosis of PC patients, highlighting ER stress as a potential therapeutic target. Besides, the findings indicate that ER stress can have effect modulating PC immune responses.

KEYWORDS

ER stress, immune microenvironment, prognosis, bioinformatics, pancreatic cancer

1 Introduction

Pancreatic cancer (PC) is a relatively frequent malignant disease in digestive system, which mostly occurs as pancreatic adenocarcinoma. According to data from the GLOBOCAN, there exists an estimated 459,000 newly diagnosed cases all over the world with an upward trend (Globocan, 2020). Adding to the woes, the mortality rate of PC is almost equal to the morbidity rate, putting it the seventh cause in terms of cancer-associated mortality (Cai et al., 2021). PC has been recognized as a silent disease that hardly presents any clinical manifestations at the early stage. Most PC patients are at an advanced or metastatic stage and lost the opportunity of surgical treatment, when firstly diagnosed (Mizrahi et al., 2020). The advent of intraoperative revascularization and chemotherapy have led to some success in PC treatment (Conroy et al., 2018; Neoptolemos et al., 2018; Sohal et al., 2020). However, the sobering reality is that the overall 5-year survival rate of PC patients is still below 10% (Zeng et al., 2018; Siegel et al., 2020). Worse still, the clinical outcome of PC patients differs largely from a few months to several years. Traditional clinical indicators such as the AJCC stage have poor performance in predicting the survival time for PC, and it is hard for doctors to establish effective therapeutic strategies for PC patients. Thus, it is imperative and urgent to develop novel tools of predicting prognosis so as to provide effective and individualized therapies for PC patients.

Endoplasmic reticulum (ER) is the largest organelle characterized by a high dynamic sac or tube membrane structure that emerges from the nucleus (Schwarz and Blower, 2016). ER plays a vital role in proteostasis due to the exquisite activities of ER-assisted folding (Wiseman and Balch, 2005), ER-associated degradation (Allison, 2021), and COPII export pathway (Zanetti et al., 2011). Abnormal accumulation of misfolded or unfolded protein and alteration of Ca^{2+} concentration will provoke ER stress in eukaryotic cells, subsequently triggering unfolded protein reaction (UPR) that is a homeostatic mechanism mainly mediated by three transmembrane proteins deemed as sensors of ER stress: protein kinase RNA-like ER kinase (PERK), inositol-required enzyme 1 (IRE1α) and activating transcription factor 6 (ATF6) (Hetz et al., 2020). A variety of perturbations in the tumour microenvironment (TME), such as hypoxia, deprivation, acidosis, and accumulation of reactive oxygen species (ROS), which is aggravated by oncogenic events in cancer cells, have been confirmed to drive persistent ER stress. In turn, signal transduction through ER sensors can not only induce UPR but also modulate UPR-independent transcriptional and metabolic pathways, resultantly contributing to tumour initiation, progression, immune evasion, and chemoradiotherapy resistance (Chen and Cubillos-Ruiz, 2021). Prior research has elucidated that ER stress exhibits a significant correlation with the survival outcomes in patients with osteosarcoma (Shi et al., 2022), colon cancer (Liu et al., 2018), renal carcinoma (Zhang et al., 2022), glioma (Zhang et al., 2021). In particular, recent studies show that ER stress can promote stemness and chemoresistance of PC by interfering fatty acid metabolism of tumour cells (Tadros et al., 2017; Dauer et al., 2019). Additionally, hypoxia-induced ER stress has been demonstrated to facilitate PC metastasis by inducing release of vascular endothelial growth factor A (Kong et al., 2015).

A body of work has indicated that ER stress can have effect modulating tumoural immune responses. The study by Pommier et al. unveils that ER stress can induce PC cell dormancy and downregulate the expression of major histocompatibility complex (MHC) class I on PC cells to escape from immune killing, consequently leading to tumour distant metastasis (Pommier et al., 2018). In a subsequent investigation on melanoma, it was discovered that the MHC class I polypeptide-related sequence A/B (MICA/B) - the cognate ligands of major NK activating receptor NK group 2 member D (NKG2D), is negatively regulated by ER stress-associated IRE1/XBP1 pathway in human melanoma cell lines (Obiedat et al., 2019). Beyond modulating the expression of immune-related molecules in cancer cells, the tumour cell-intrinsic ER stress can also reprogram the function of immune cells within the surrounding milieu, which facilitates immune evasion. The findings of Song and colleagues discerned that cell-free ascites from ovarian cancer patients instigate ER stress and triggers IRE1α/XBP1 signaling in activated T cells from healthy women; tumour-bearing mice model further showed that the IRE1α/XBP1 signaling in T cells impairs its mitochondrial respiration by regulating glutamine transporter, ultimately attenuating their anti-neoplastic efficacy (Song et al., 2018). This suggested that ER-stressed cancer cells can release additional factors to remodel the tumour milieu. Another compelling evidence (Yuan et al., 2022) showed that ER-stressed human head and neck squamous cell carcinoma cell line HN4 evades immune surveillance by releasing exosome PD-L1 and upregulating the expression of PD-L1 in macrophages to skew macrophages towards a pro-tumoural phenotype. Cubillos-Ruiz et al. revealed that persistent activation of XBP1 in cancer-associated dendritic cells hindered its capacity to instigate anti-tumour T cells, subsequently catalyzing the progression of ovarian malignancies. (Cubillos-Ruiz et al., 2015). Of note, severe ER stress tip the tumour cells towards releasing damage-associated molecular patterns (DAMPs), triggering immunogenic cell death (ICD) (Rufo et al., 2017). A most recent study revealed that B16 melanoma cells release externalized calreticulin (ExoCRT) during severe ER stress, which is subsequently recognized by the NKp46 receptor on NK cell membrane. And the binding of NKp46–CRT curtails B16 melanoma growth and lung metastasis through amplifying NK killing (Sen Santara et al., 2023). Such revelations underscore that ER stress wields multifaceted impacts on tumours, contingent on its intensity and duration in a context-dependent manner. To

holistically understand the intricacies of ER stress in PC, we endeavored to devise a prognostic prediction model for PC patients, underpinned by ER stress-related genes, through an exhaustive exploration of large-scale genomic databases.

In current work, we first identified 46 ER stress-associated genes closely linked with OS of 164 PC patients from TCGA by log-rank test and univariate Cox regression, and further narrowed 12 optimal prognostic genes (SERP1, BCL2L1, RNF139, TSPYL2, ERO1A, BFAR, MAP3K5, ARFGAP1, ATP6V0D1, NCCRP1, PLA2G4B, USP 19) to build a risk signature via LASSO method. Next, these 164 PC patients were classified as two groups based on the risk signature score, and better OS in the individuals with low-risk score was observed by the Kaplan-Meier (K-M) survival analysis method. Meanwhile, the receiver operating characteristic (ROC) curves and multivariate hazard regression demonstrated that our ER stress-associated risk signature was a robust and independent prognostic biomarker. Furthermore, we established a nomogram model by integrating our risk signature and several clinicopathological parameters, which shows promise in predicting prognosis of PC patients. In addition, the risk signature based on ER stress-associated genes was tightly correlated with homeostasis, protein secretion, and immune regulation in PC. Particularly, PC microenvironment in high-risk group exhibited a more immunosuppressive status. In summary, this study suggested that ER stress is of much prognostic significance and may serve as promising therapeutic targets in PC.

2 Materials and methods

2.1 Datasets and samples

The RNA sequencing metrics of PC samples and normal pancreatic tissues and key clinicopathological characteristics were acquired from TCGA, which is considered as a training cohort. Similarly, we procured the transcriptome profile and clinical data from PACA-CA of the ICGC data portal as an external test cohort. All patients enrolled in this study met the following criteria: i availability of matched gene expression data; ii complete clinical information; iii survival time greater than 30 days. We opted to exclude individuals who, post neoadjuvant therapy, succumbed to extraneous factors such as non-cancerous ailments, other malignancies, or surgical complications. Donors' transcriptome matrix from Genotype-Tissue Expression Project (GTEx) was used for comparing the expression landscape between PC and normal samples. As for the transcriptome profile, probe IDs were mapped to ENSEMBL IDs and gene symbols using the *Annoprobe* package (Xiang and Zeng, 2019) in R, and genes were filtered based on their expression in more than half of the samples. The counts were just used in the differentiated analysis and were log2-transformed in other cases. The genes involved in ER stress were extracted from MSigDB v7.4 in the two GO pathways (Huang et al., 2021). A list containing Pan-cancer Immune Metagenes of 28 immune cells was obtained from a widely acknowledged research (Charoentong et al., 2017) to reveal the PC immune microenvironment landscape.

2.2 Identification candidate genes with prognostic potential

For determining the ER stress-related genes that own prognostic significance, the univariate Cox regression and log-rank tests were carried out (Terry and Patricia, 2001; Therneau, 2021) based on the TCGA-PAAD dataset. A p-value threshold of 0.05 was instituted as the criterion for statistical significance. Then the candidate genes associated with PC patients' survival were then selected based on the p-values of the above two test results and included in the following analyses.

2.3 Establishment and confirmation of the ER stress-associated risk signature

To assess the link of ER stress with OS of PC patients, we tried to build a risk signature according to the ER stress-associated genes. LASSO penalty algorithm avoiding the overfitting concern (Friedman et al., 2010) was employed to further filter prognostic genes associated with ER stress. In the LASSO regression, the optimal tuning parameter was identified by harnessing the minimal lamda value and the random seed is set to 15091102. Subsequently, multivariate Cox analysis was conducted to yield the ER stress-associated risk signature as previously reported (Song et al., 2022). Then, we scored each PC patient based on the risk signature and calculated the optimal cutoff point of the risk scores. PC patients were grouped low- and high-risk clusters. The OS time of the two clusters was compared by K-M analysis. In addition, the ROC curve was adopted for assessing the prognostic accuracy of our risk signature.

2.4 Development and assessment of a predictive nomogram

To better predict the prognosis of PC patients, we established a nomogram by integrating the ER stress-associated signature and several key clinicopathological parameters which have been reported elsewhere as risk factors in PC (Mizrahi et al., 2020; Cai et al., 2021).

2.5 Functional enrichment analyses

Differentially expressed genes (DEGs) were initially delineated utilizing the DESeq2 algorithm as (Love et al., 2014), with a criterion of $|\text{LogFC}| \geq 1$ coupled with an adjusted p-value of < 0.05 to signify statistical relevance. Following this, these identified DEGs underwent comprehensive KEGG enrichment and GO analytic evaluations, applying an adjusted p-value benchmark of < 0.05 to ascertain statistical noteworthiness. Furthermore, the paramount enriched KEGG pathways and GO terms were delineated using the GOplot visualization tool.

2.6 Evaluation of immune cell infiltration landscape

Utilizing the ssGSEA algorithm, we quantified the abundance of TME immune cells based on gene sets delineated by Charoentong

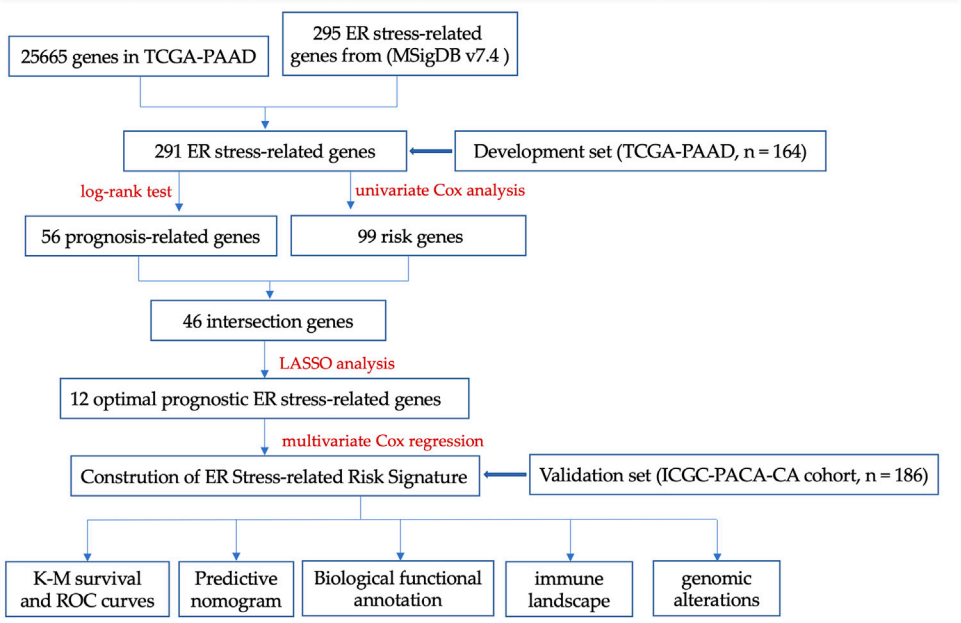


FIGURE 1
The key design and procedures of this research.

et al. (2017). In detail, 782 genes were regarded as the markers for 28 kinds of immune cells to detect the abundance of each kind of immune cell in PC samples. Additionally, the stromal and immune scores in the two different clusters were calculated via the estimate function (Kosuke et al., 2016), estimating tumour purity. Moreover, using the *complot* (Wei and Simko, 2021), we analyzed the correlation between the abundance of these 28 kinds of immune cells and risk scores, with correlation coefficients derived via the Pearson technique.

2.7 Relationship between the risk signature and genomic alterations

In consideration of the vital role of tumour somatic mutations in response to the TME and therapy-induced associated stresses that occur during evolution, we also investigate the association of risk signature with the mutation landscape. Somatic mutation data of PC patients in mutation annotation format, which has been processed under mutect software, was acquired at Genomic Data Commons (GDC) Data Portal. A genomic analysis of PC patients was conducted using the maftools. Tumour mutational burden was further estimated using tmb function to analyze the TMB in different groups.

2.8 Prediction of chemotherapy sensitivity for PC patients

To further elucidate the nexus between the ER stress-associated risk signature and therapeutic efficacy, we harnessed the Drug Sensitivity in Cancer (GDSC) database, a comprehensive resource boasting transcriptomic profiles of 805 cell lines and accompanying IC50 values for a staggering 198 bioactive compounds. Using *oncoPredict* R package to align PC transcriptome data in TCGA, we calculated the predicted IC50 value for each PC patient for several chemotherapeutic drugs commonly used in PC treatment. Subsequent comparative analysis between high and low-risk cohorts was facilitated using the Wilcoxon test.

2.9 Statistical analysis

The continuous data in two clusters were compared by conducting Wilcoxon test. As mentioned above, combined application of log-rank test and univariate Cox analysis were to screen genes linked with the PC patients' survival. The LASSO penalty algorithm further narrowed prognosis genes associated with ER stress. Independent prognostic factors were further assessed by multivariate analysis, supplemented with Pearson's correlation to elucidate the linear relationship between risk scores and immune cell abundance.

3 Results

3.1 Identification of ER stress-associated genes

The key design and procedures of this research were depicted in Figure 1. We finally included 164 PC patients and acquired 295 unique ER stress-associated genes (Supplementary Table S1). Univariate Cox analysis identified 99 risk genes affecting survival (Supplementary Table S2), while log-rank test determined 56 prognostic genes (Supplementary Table S3) within these

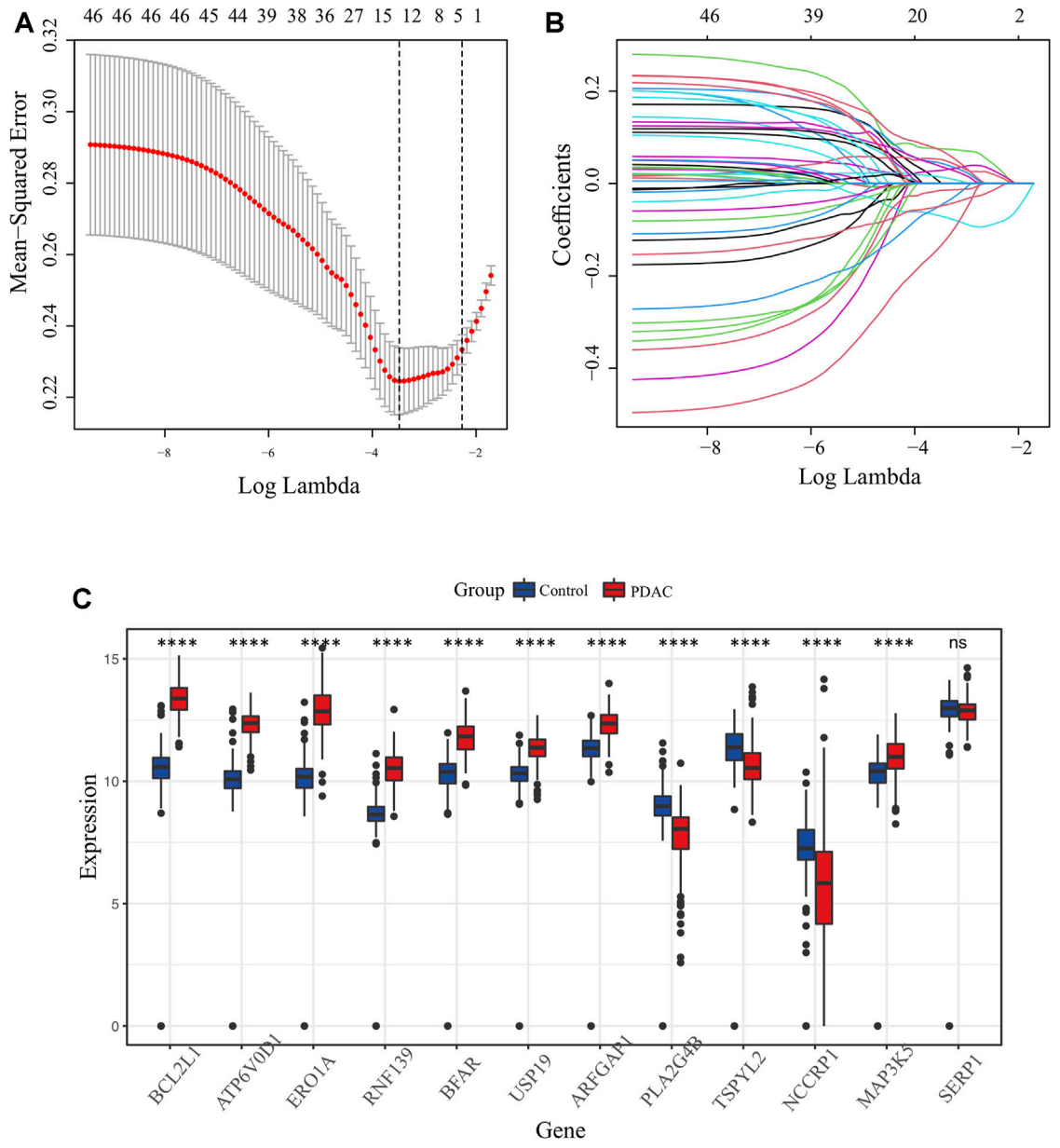


FIGURE 2
Identification of candidate ER stress-associated genes with prognostic potential. **(A)** Selection of lambda parameters for the LASSO model in TCGA-PAAD dataset. **(B)** Lambda-dependent coefficients for the 46 intersecting ER stress-associated genes. **(C)** Comparative expression profiling of the 12 prime ER stress-associated genes between pancreatic cancer and normal pancreatic tissues.

295 genes. Furthermore, these 46 intersection genes were considered to have significant relevance to the OS of PC patients and were introduced to LASSO analysis to screen ultimate genes with the best predictive performance. Finally, 12 optimal prognostic ER stress-associated genes with non-zero regression coefficients were determined to construct the risk signature (Figures 2A, B), namely stress-associated endoplasmic reticulum protein 1 (SERP1), B-cell lymphoma 2 like 1 (BCL2L1), ring finger protein 139 (RNF139), testis-specific protein Y-like 2 (TSPYL2), endoplasmic reticulum oxidoreductase 1-α (ERO1A), bifunctional apoptosis regulator (BFAR), Mitogen-Activated Protein 3 Kinase 5 (MAP3K5), ADP-ribosylation factor GTPase activating protein 1 (ARFGAP1), vacuolar proton ATPase subunit d 1 (ATP6V0D1), non-specific cytotoxic cell receptor protein 1 (NCCRP1), phospholipase A2, group IVB (PLA2G4B), ubiquitin specific peptidase 19 (USP 19). The K-M survival plot of every gene for PC patients was shown as Supplementary Figure S1, where the group is based on the best cut-off point for each gene. The mRNA profiles of these 12 genes were also compared between PC tissues and normal pancreatic tissues, and we found that eight of them (BCL2L1, ATP6V0D1, ERO1A, RNF139, BFAR, USP19, ARFGAP1, MAP3K5) were significantly upregulated, three (PLA2G4B, TSPYL2, NCCRP1) were downregulated, and there was no difference in SERP1 expression (Figure 2C).

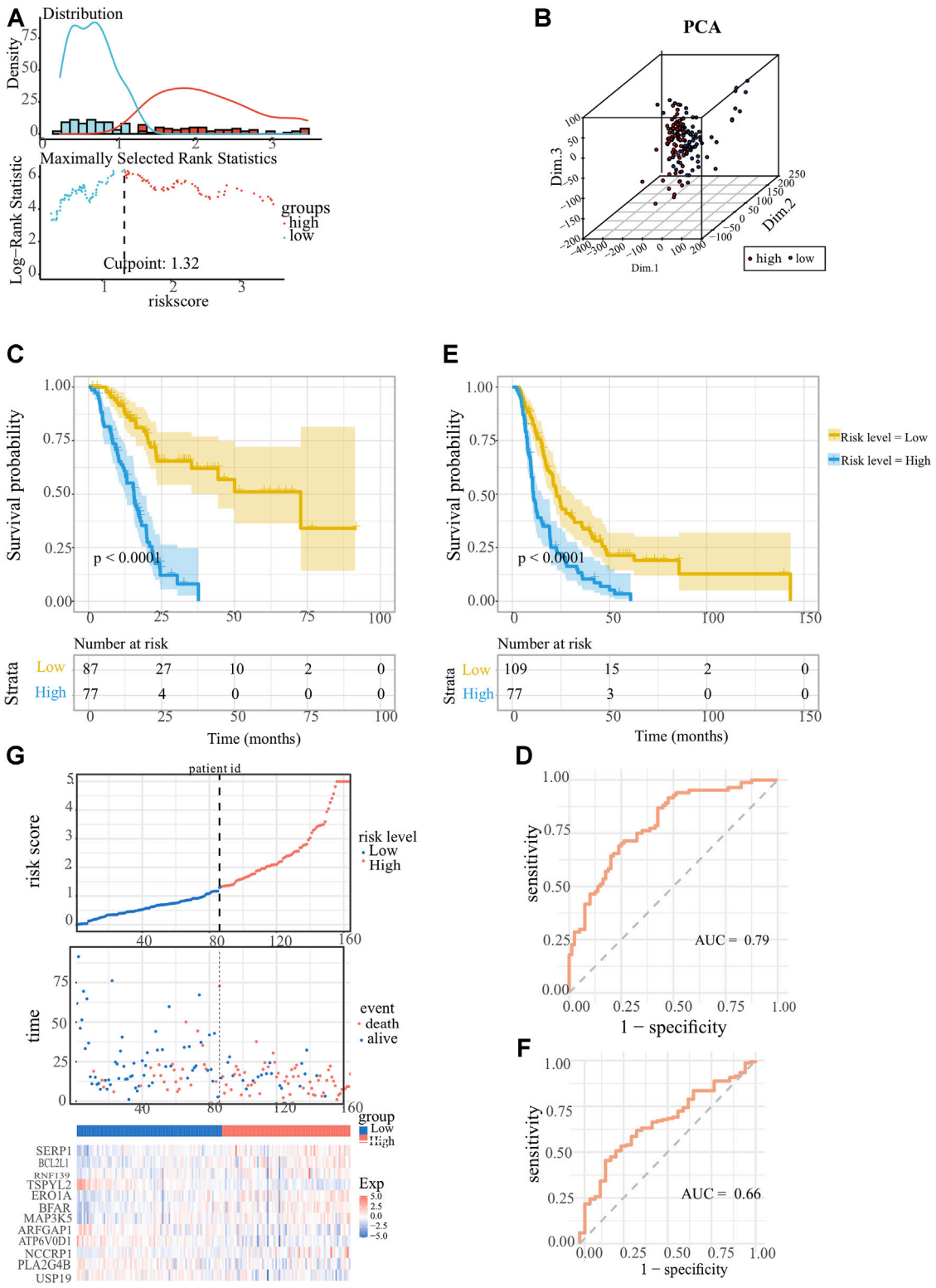


FIGURE 3
Establishment and confirmation of the ER stress-associated risk signature for pancreatic cancer. **(A)** Optimal cutoff ascertained via survival R package. **(B)** Genomic-wide expression-based Principal Component Analysis (PCA). **(C, E)** Kaplan-Meier survival curves showed indicating adverse overall survival (OS) in PC patients with lower risk indices, as observed in TCGA-PAAD and ICGC-PACA-CA datasets. **(D, F)** Receiver operating characteristic (ROC) curves Receiver Operating Characteristic (ROC) analyses confirming the robustness of the predictive model for pancreatic carcinoma patient survival within the TCGA-PAAD and ICGC-PACA-CA datasets. **(G)** Comprehensive display encompassing risk score distribution, survival status scatter plots, and a heatmap representation of the 12 optimal ER stress-correlated genes.

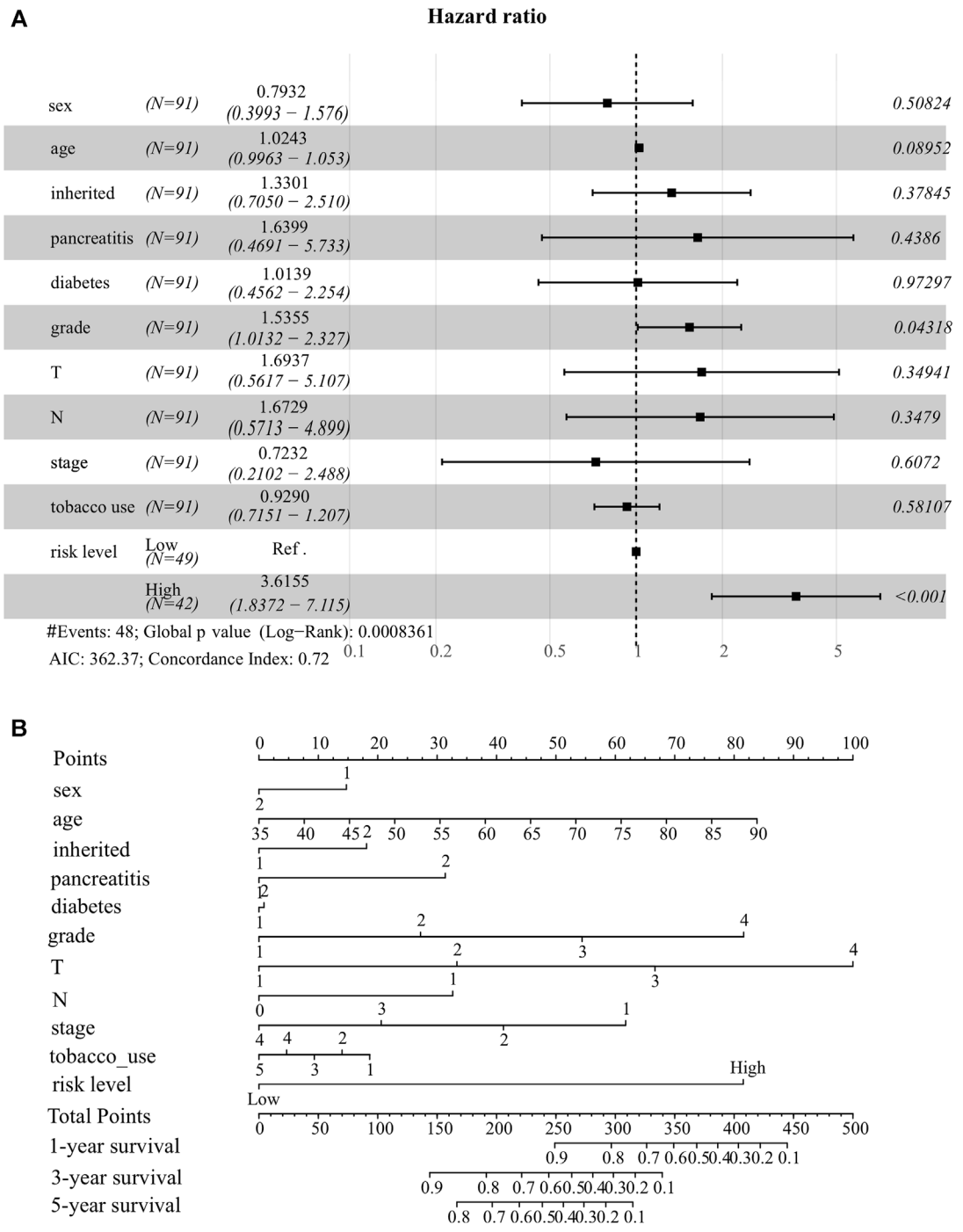


FIGURE 4
Establishment of nomogram model integrating ER stress-associated risk signature with clinicopathological parameters for prognostic predictions for pancreatic cancer. **(A)** Forest plot showing risk level as an independent prognostic biomarker for pancreatic cancer by multivariate analysis. **(B)** A prognostic nomogram amalgamating the ER stress-associated risk signature and multiple clinical determinants, predicting 1-, 2-, and 3-year overall survival likelihoods in pancreatic cancer patients.

3.2 ER stress-associated risk signature was established, and the low-risk score PC patients owned longer OS time

The ER stress-associated risk signature was established using the above 12 markers, and the risk score equation was: riskscore = 0.813468268 × [(0.496637586 × *SERP1*) + (0.589598808 × *BCL2L1*) + (0.406009527 × *RNF139*) + (−0.521549002 × *TSPYL2*) + (−0.027972152 × *ERO1A*) + (0.323475966 × *BFAR*) + (0.476688839 × *MAP3K5*) + (0.025209694 × *ARFGAP1*) + (-0.92171365 × *ATP6V0D1*) + (0.152614999 × *NCCRP1*) + (0.005952034 × *LA2G4B*) + (-1.026914644 × *USP19*)]. We scored

each PC patient with the above signature, and classified 164 PC patients into low-risk group (n = 87) and high-risk group (n = 77) according to the optimal cutoff point (1.322269) of the risk scores (Figure 3A). As Figure 3B showed, the profiles of gene expression in low- and high-risk clusters represented discrepant distribution patterns according to the principal components analysis (PCA), which implied that our ER stress-associated signature was correlated with a whole genome-wide difference. The K-M survival analysis is depicted in Figure 3C. It's obvious that low-risk score PC patients owned longer OS time ($p < 0.0001$) than those patients with high-risk scores. The area under the ROC curve was 0.79, confirming the prognostic accuracy of our risk signature (Figure 3D). We did the same in the ICGC-PACA-CA cohort to validate the prognostic significance of our ER stress-associated risk signature. Survival analysis also confirmed a statistically significant survival advantage for low-risk score PC patients despite a slight decrease in the AUC value (Figures 3E, F). Moreover, a linked graph of a smooth curve, scatterplot, and heatmap intuitively indicated that as the risk score increases, the survival status of patients with PC from TCGA is more likely to be death and shorter survival time (Figure 3G). Furthermore, the differential expression landscape of these 12 risk factors between the two subgroups was also observed (Supplementary Figure S2).

3.3 Establishment of nomogram model integrating ER stress-related risk signature and clinical parameters for predicting prognosis of PC patients

To better predict the survival probability of PC patients, we further built a nomogram based on our ER stress-associated risk signature and several readily available clinical parameters that recognized elsewhere as PC risk factors. Considering the number of patients with complete clinical information in the TCGA-PAAD cohort, our analysis did not include the M stage. Consequently, 91 patients provided complete information contributing to the predicting model. The forest plot showed that risk level and histological grade were independent prognostic indicators (Figure 4A). In particular, the risk level has the most apparent statistical significance (HR = 3.613, $p < 0.001$). Finally, we generated a nomogram model to predict the survival probability for each PC patient with a robust performance (Concordance Index = 0.72), as shown in Figure 4B. Total points can be calculated according to these parameters, and the 1-year, 3-year, and 5-year survival probability can be further obtained based on this total point.

3.4 GO and KEGG analyses of the risk signature reveal alteration of the homeostasis and immune signaling pathways

To investigate the underlying molecular mechanisms through which ER stress influences the prognosis of pancreatic cancer, we conducted a detailed functional enrichment analysis. Firstly, we identified DEGs between low- and high-risk clusters. A total of 1929 genes were differently expressed, containing 324 high-expressed and 1605 low-expressed genes. The enrichment of biological processes suggested that the ER stress-associated risk signature altered "membrane potential", "chemical synaptic transmission", " trans-synaptic signaling", "protein secretion" and "B-cell receptor signaling pathway" (Figure 5A). Interestingly, the genes enriched in biological processes terms of the "B cell receptor signaling pathway" were almost downregulated, indicating that high-risk patients may be "colder" regarding the immune microenvironment. The frontmost cellular components were "synaptic membrane," "presynapse," "neuronal cell body," "transport complex," and "ion channel complex" (Figure 5B). Molecular functions, including "passive transmembrane transporter activity," "gated channel activity," and "hormone activity," were enriched (Figure 5C). More details of GO enrichment analyses can be found in Supplementary Table S4. The top 10 down and upregulated KEGG pathways are shown in bar graphs (Figure 5D). As we can see, the upregulated KEGG pathways mainly involve the exocrine function of the pancreas. Meanwhile, enriched downregulated pathways include "dopaminergic synapse," "calcium ion signaling," and "insulin secretion." Thus, the KEGG and GO analysis results confirmed that the risk signature correlated with homeostasis and emphasized the role of ER status in during PC progression.

3.5 Relationship between ER stress-associated risk signature and immune microenvironment

The landscape of the immune microenvironment of the two risk clusters were analyzed from the two aspects of immune cell abundance and immune scores. Four kinds of cells, including CD56 bright NK cell, Th2 cell, and Th17 cell, were infiltrated in the high-risk cluster; whereas other 14 types of cells were enriched in l enriched in the low-risk cluster, such as activated $CD8^+$ T cells, NK cells, eosinophils, monocytes, activated DCs, as well as macrophages and MDSCs (Figure 6A). In addition, we calculated each PC patient's stromal and immune scores at different risk levels. Unsurprisingly, the high-risk group got lower scores (Figure 6B). To further explore whether the ER stress-related signature affects PC immune microenvironment, a correlation heatmap was constructed to reflect the correlations between these differentially infiltrated immune cells and risk scores (Figure 6C). All these results demonstrated that the high-risk cluster has a more suppressive immune microenvironment that interacts with the risk signature. Considering the close link between our risk signature and the immune microenvironment and the therapeutic potential of immunotherapy in selected tumours, we also conducted a correlation analysis to identify PC patients who could benefit from immunotherapeutic modalities. Several immunotherapy targets, such as PDCD1, CTLA4, LMTK3, and LAG3, were also expressed in PC tissues. Interestingly, PDCD1 expression level had an inverse correlation with the risk score in PC patients with histopathological grade 2 or AJCC stage II (Figures 6D, E).

3.6 Association of ER stress-associated risk with genomic alterations

Since somatic mutations are closely related to tumour initiation and progression, an integrated analysis was performed to establish

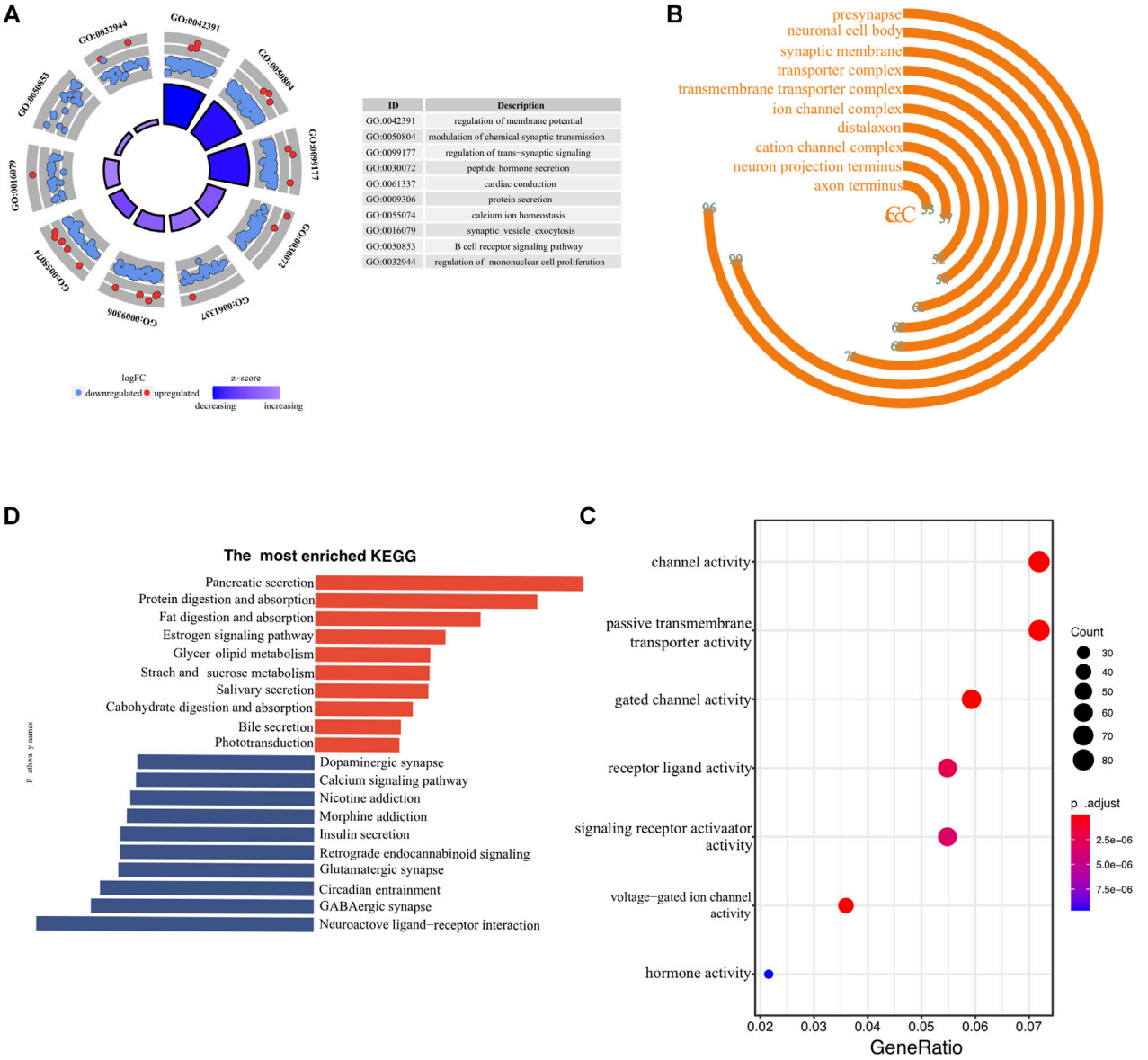


FIGURE 5
Biological functional annotation of differentially expressed genes between high- and low-risk samples. The front-most enriched biological processes **(A)**, cellular components **(B)**, and Molecular functions **(C)**. **(D)** The most prominently upregulated and downregulated KEGG pathways in the TCGA-PAAD cohort.

an association of the status of genomic alterations with our ER stress-associated signature in PC. A total of 9,957 genes were mutated among 178 PC samples obtained from the GDC website, of which *KRAS*, *TP53*, *SMAD4*, and *CDKN2A*-the four main drive genes, were the most frequently mutated. The summary of the mutation landscape shows that missense mutation, nonsense mutation, frameshift deletion or insertions, and splice site mutation are the leading fourth variant types (Figure 7A). A waterfall plot depicted the top 10 mutated genes under ER stress signature-based clusters and other clinicopathological groups, and we found that 77% of PC patients experienced *KRAS* mutation (Figure 7B). As for the 12 ER stress-associated genes for constructing the risk model, only *BFAR*, *MAP3K5*, and *ARFGAP1* have missense mutations or frameshift insertions, and *USP19* has multiple mutations (multi-hits). The mutation ratio is less than 1%. In addition, the TMB was also calculated to reveal a possible mechanism for the different immune microenvironments in the two risk groups. However, our result showed no difference in TMB between the two clusters (Figure 7C).

3.7 Correlation between ER stress-associated risk signature and therapeutic response

To better connect our ER stress-related risk signature to clinical practice, we employed oncoPredict tool to calculate the IC50 value of several commonly used chemotherapeutic agents for each PC

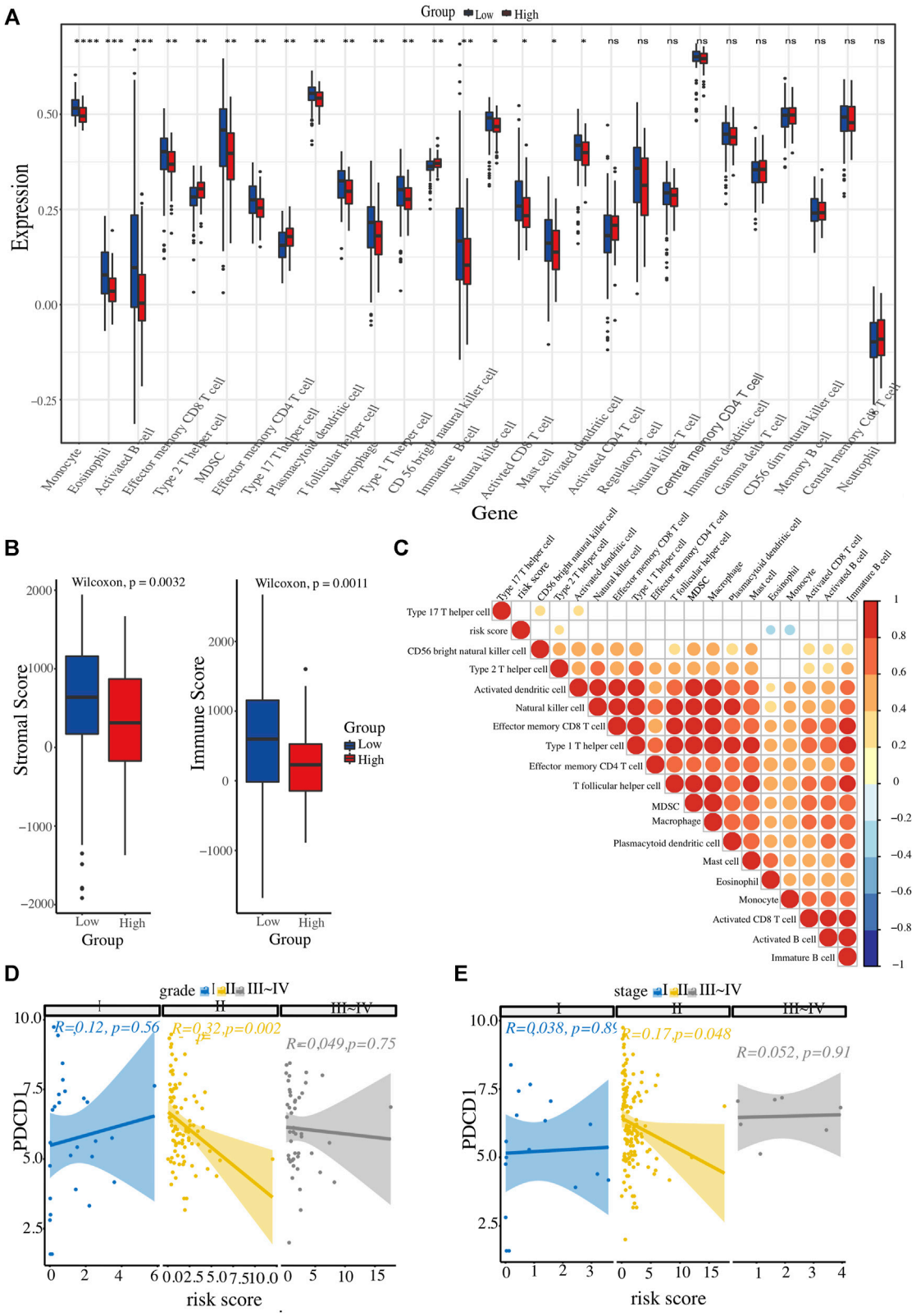


FIGURE 6
Evaluation of immune cell infiltration landscape. **(A)** Boxplot illustrating differential infiltrations of 28 immune cell subtypes between high- and low-risk groups. **(B)** Comparative evaluation of immune scores and stromal integrity between the two clusters. **(C)** Heatmap representation of risk score associations with 18 immune cell types displaying variable infiltration degrees. **(D, E)** The relationship between ER stress-associated risk score and immune checkpoint (PDCD1), stratified by histopathological grade and AJCC TNM stage, respectively.

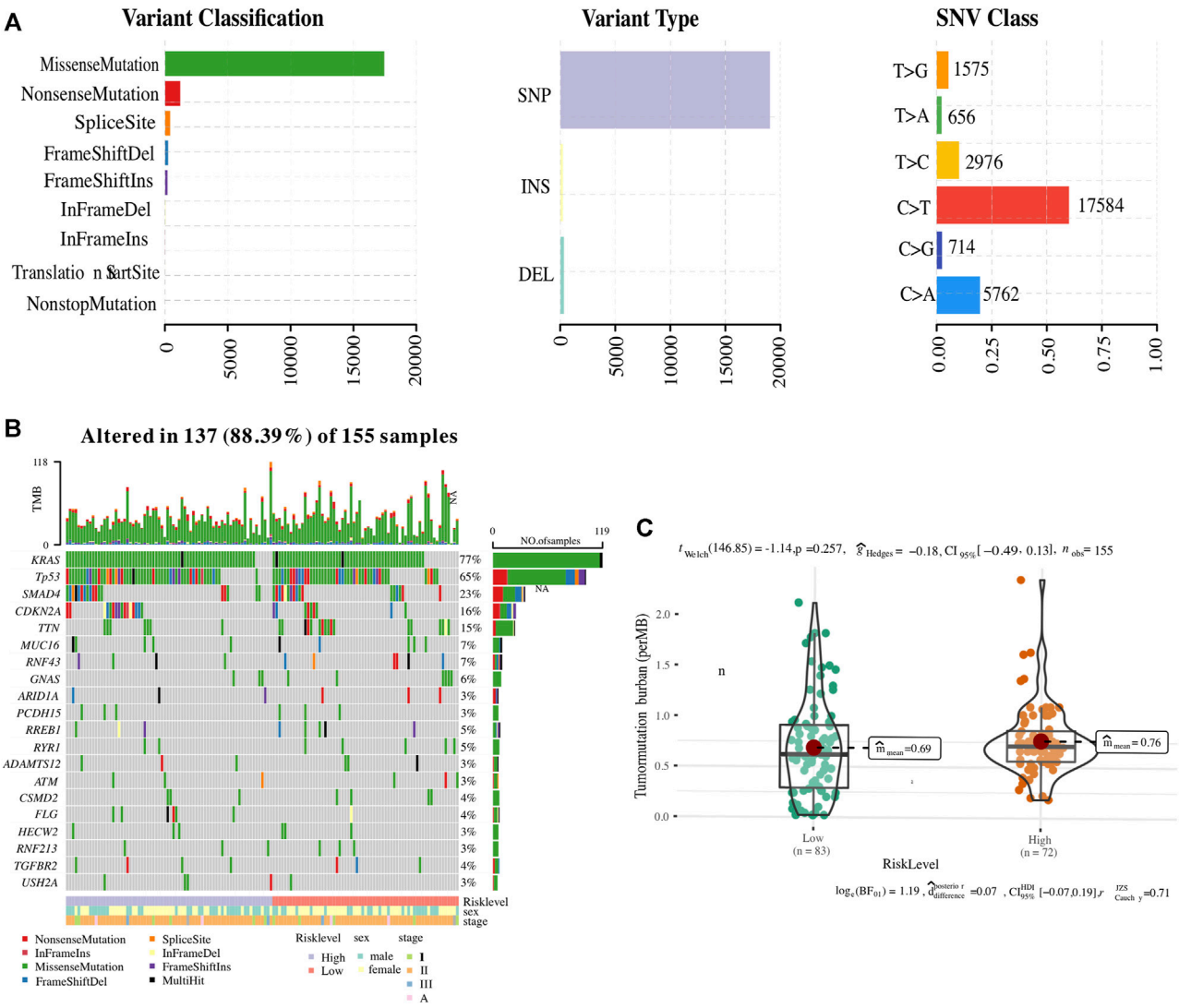


FIGURE 7
Association of ER stress-related risk signature with genomic alterations. **(A)** The summary of the mutation landscape of pancreatic cancer. **(B)** Waterfall representation highlighting the top 10 mutated genes. **(C)** Comparative tumour mutational burden (TMB) across the two ER stress-associated risk stratifications.

patients. The results showed that the predicted IC50 for agents such as "5-Fluorouracil", "Oxaliplatin", "Irinotecan", "Paclitaxel", and "Cisplatin" skews higher for the high-risk demographic., indicating that low-risk score patients may be more sensitive to chemotherapy (Figures 8A–F).

4 Discussion

The traditional clinical indicators are poor in predicting the survival time of PC, so it is imperative to develop new tools to predict survival and furnish effective individualized treatment for PC patients. Increasing evidence shows that ER stress plays a complex role in PC progression. Hence, we here performed a bioinformatic analysis based on TCGA database to study role of ER stress in prognosis and biological characteristics in PC patients. In this study, we initially established a twelve-gene risk signature in terms of ER stress according to the gene expression counts and corresponding survival time of 164 PC patients from TCGA. The twelve ER stress-related genes enrolled in the risk model could be divided into protective factors (*TSPYL2*, *ARFGAP1*, *ATP6V0D1*, *PLA2G4B*, *USP19*) and pernicious factors (*SERP1*, *BCL2L1*, *RNF139*, *ERO1A*, *BFAR*, *MAP3K5*, *NCCRP1*). The predictive potency of the ER stress-associated risk signature was further validated by another independent PC dataset from ICGC. Survival curve and forest plot demonstrated that our ER stress-associated signature was a valuable and independent prognostic biomarker, according to the ROC curves (AUC was 0.79) and the hazard ratio (HR was 3.613, $p < 0.001$, CI was 0.72). Besides, we constructed a credible nomogram integrated the risk signature and several clinicopathological parameters predicted the OS of individual PC patients. Function annotations elucidate the biological role of ER stress in PC, and the results indicated that the ER stress-related signature has an effect in shaping TME immune cell infiltration.

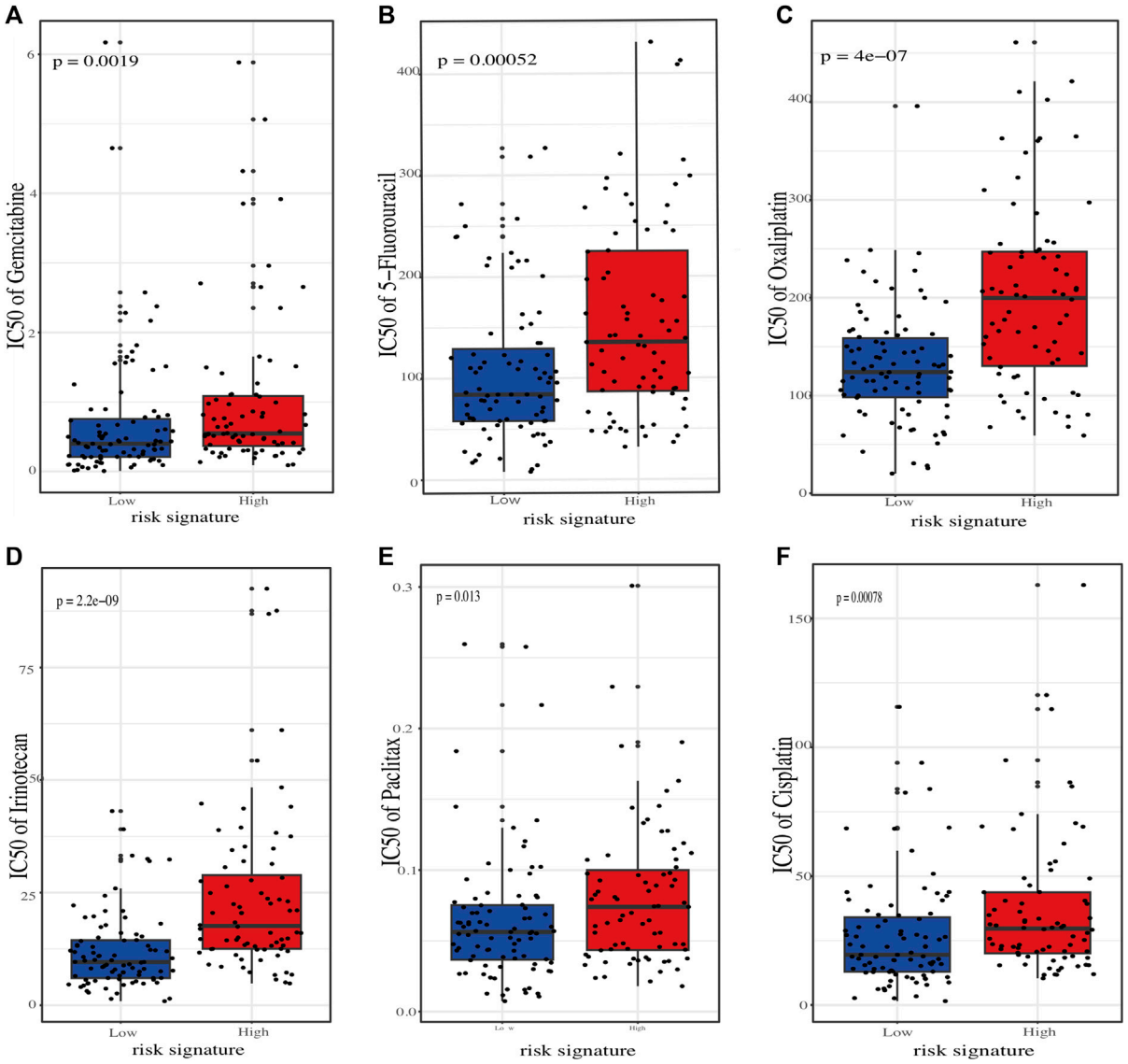


FIGURE 8
Estimation of chemotherapeutic sensitivity in stratified risk groups. Comparision of the predicted half-maximal inhibitory concentration (IC50) values for six chemotherapeutic agents: gemcitabine **(A)**, 5-fluorouracil **(B)**, oxaliplatin **(C)**, irinotecan **(D)**, paclitaxel **(E)**, and cisplatin **(F)**.

SERP1 has been uncovered to be overexpressed under stress, which helps to stabilize membrane proteins and functions as a translocon on the ER membrane (Yamaguchi et al., 1999). Evidence shows that SERP1 may act as an oncogene and could be developed as a valuable prognostic biomarker. For instance, a recent research uncovered that SERP1 can facilitate osteosarcoma progression through modulating circ_0085539/miR-526b-5p signalling axis (Liu et al., 2020). Additionally, SERP1 has been reported to maintain the viability of PC cells promote the survival of PC cells by down-regulating apoptosis-related protein SRPRB and activating the NF-κB signalling pathway (Ma et al., 2017). Consistently, our study demonstrated that SERP1 was upregulated in PC and its high expression predicted dismal patient survival prognosis. BCL2L1, also called Bcl-extra (Bcl-x), belongs to the B-cell lymphoma 2 apoptosis family. BCL2L1 translates two antagonistic variants including Bcl-X_S and Bcl-X_L due to alternative splicing. The short isoform Bcl-X_S exhibits pro-apoptosis biological function, whereas the long isoform Bcl-X_L contributes to survival of cells. Numerous studies have confirmed that the disequilibrium of BCL2L1 splice is related to the progression of multiple malignancies (Dou et al., 2021). In addition to promoting tumour cell survival, Bcl-X_L can also boost metastasis and EMT of pancreatic neuroendocrine tumour cells via epigenetically enhancing TGFβ signalling (Choi et al., 2016). Furthermore, recent evidence shows in melanoma that Bcl-X_L can positively regulate the expressions of IL-8 and CCL5, which not only recruit macrophages into tumour sites, but also polarize them into M2 phenotype (Lucianò et al., 2023). In PC, Bcl-XL is overexpressed, and correlated with worse prognosis of patients (Friess et al., 1998). Particularly, Bcl-X_L has been identified as a

crucial mediator of acquiring radioresistance and gemcitabine resistance (Lee et al., 1999; Thummuri et al., 2022). Similarly, we herein found that BCL2L1 was significantly increased in PC tissues and its high expression predicted unfavourable prognosis of patients. In general, our data and the literature evidence suggest that Bcl-X_L may become a valuable prognostic biomarker and serve as a therapeutic target for PC patients.

RNF139, also known as TRC8, encodes a multi-membrane spanning protein resides in ER, and it has been revealed to act as a tumour suppressor. A study reported that RNF139 was significantly downregulated in tongue cancer tissue and experimentally that silencing RNF139 enhanced the viability and aggressiveness of tongue cancer cells (Wang et al., 2017). Intriguingly, our data shows that RNF139 is dramatically upregulated in PC tissues, and low-expressed RNF139 PC patients have shorter survival time, indicating a potential role of RNF139 in propagating PC progression. Hence, the roles of RNF139 in cancer may depend upon the tumour types, and cell-based and animal experiments are warranted to determine the specific roles of RNF139 in PC. ERO1A is an oxidoreductase located in ER lumen. Recently, accumulating data uncovers that ERO1A functions as an oncogene in many types of cancers. According to Yan and colleague, the expression of ERO1A is increased in cholangiocarcinoma and its overexpression predicts poor prognosis of patients (Yan et al., 2019). Zilli et al. (2021) revealed that upregulation of ERO1A significantly promotes primary breast tumour growth and lung metastatic colonization by enhancing HIF1α-VEGFA mediated angiogenesis. Zhang et al. corroborated that overexpression of ERO1A can contribute to tumour progression by accelerating hydrogen peroxide-associated epithelial-mesenchymal transition, and it may be used as a predictor for dismal prognosis in cervical cancer (Zhang et al., 2019). In PC, ERO1A has been demonstrated to boost tumour growth via augmenting aerobic glycolysis of tumour cells (Zhang et al., 2020). In agreement with these evidence, we sighted that mRNA of ERO1A is highly in PC and high ERO1A expression is accompanying worse prognosis. Altogether, ERO1A may be a useful prognostic biomarker and promising therapeutic target for PC patients.

BFAR has been recognized as an inhibitor of apoptosis with an alary helix DNA-binding structure and is tightly linked with tumour progression. Cheng et al. reported that BFAR is upregulated in glioma tissues and glioma patients with lower BFAR expression have worse long-term survival prognosis (Cheng et al., 2020). In gastric cancer, BFAR has been proven to rebound to the tumour metastasis via activating the PI3K/AKT/mTOR signalling axis (Chen et al., 2023). Similar to the previous studies, we observed that BFAR is highly expressed in PC and its increasing expression predicts worse outcomes of PC patients. Thus, BFAR may also act as an oncogene in PC. MAP3K5 has been known as an upstream protein of mitogen-activated protein kinase cascade signalling pathway and plays a vital tumour-suppressive role in many types of cancers. For instance, Cheng et al. showed that overexpression of MAP3K5 promotes PC cell apoptosis via activating the p38/MAPK axis (Cheng et al., 2014). Consistent with this, our data showed that MAP3K5 is upregulated in PC tissues and its high expression predicts poor prognosis of PC patients. NCCRP1 is a type III transmembrane protein containing an antigen recognition motif, also named as FBXO50. Evidence shows that NCCRP1 plays an essential role in tumour development, and whether it acts as a tumour suppressor gene or oncogene depend upon tumour types. A most recent study suggested that NCCRP1 was highly expressed in triple-negative breast cancers (TNBCs) and its overexpression could markedly enhance cell proliferation (Zhou et al., 2022). In gastric cancer, higher NCCRP1 expression is associated with shorter recurrence-free and overall survival time, as it promotes proliferation, movement, and incursion of gastric cancers (Miwa et al., 2017b). Inversely, another study (Miwa et al., 2017a) found that esophageal squamous cell carcinoma patients with lower NCCRP1 expression tended to have an increased risk of disease recurrence and dismal survival prognosis versus those with higher NCCRP1 expression, suggesting a putative tumour-suppressive role of NCCRP1. Notably, when it comes to PC, the roles of NCCRP1 seems rather complex. In our study, we found that NCCRP1 is downregulated in PC tissues compared to normal pancreatic tissues, but the high NCCRP1 expression patients have little survival time. Our paradoxical results may be due to the fact that NCCRP1 in PC tissues is not expressed mainly in tumour cells, but in stroma cells such as immune cells and fibroblasts. More research is thus required to further decipher the biological roles of NCCRP1 in PC and its clinical prognostic value.

USP19 is a kind of deubiquitinating protease. Similar to NCCRP1, USP19 is aberrantly expressed in multiple cancers and may act as a tumour suppressor gene or oncogene, which relies upon the tumour type (Rossi and Rossi, 2022). For example, USP19 was found to be upregulated in TNBC tissues and promote the invasion and metastasis of TNBC (Rossi et al., 2021). Besides, a recent study reported that upregulation of USP19 contributed to tumourigenicity of colorectal cancer cells by stabilizing Survivin protein (Chandrasekaran et al., 2022). On contrast, diminished USP19 expression in ovarian cancer correlates with an adverse prognosis (Kang et al., 2021). Adding to this complexity, (Hu et al., 2020), reported in clear cell renal cell carcinoma that overexpression of USP19 markedly curtails tumour cell proliferation and migration via ERK pathway inactivation. Aligning with these findings, our research discerned that elevated USP19 expression portends a favorable survival outcome in PC, and it is of clinical significance to investigate its specific biological functions and the relevant mechanisms in the progression of PC. TSPYL2 is the nucleosome component of chromatin remodelling (Cardano et al., 2023). Recent work by Pan et al. (2022) recently confirmed that TSPYL2 is downregulated in PC and overexpression of TSPYL2 can effectively overcome gemcitabine resistance in PC cells. Likewise, in this study we also found that TSPYL2 is significantly downregulated in PC and its low expression predicts shorter survival time of PC patients. Within the cohort of ER stress-associated risk signature genes, the roles of PLA2G4B, ARFGAP1, and ATP6V0D1 in human malignancies remain largely unexplored. Our research determined that PLA2G4B exhibits decreased expression in PC tissues, correlating lower expression levels with a more guarded prognosis. In contrast, both ARFGAP1 and ATP6V0D1 manifested heightened expression in PC tissues, resonating with an improved prognostic outlook for PC patients. Future endeavors should be directed towards elucidating the precise biological and clinical implications of PLA2G4B, ARFGAP1, and ATP6V0D1 in the context of PC.

The tumour immune microenvironment is intertwined with PC growth, metastasis, and therapeutic resistance (Bear et al., 2020; Ho et al., 2020; Sherman and Beatty, 2023). Increasing evidence reveals that ER stress can confer immunomodulatory capacity on cancer cells (Cubillos-Ruiz et al., 2017; Di Conza et al., 2023), prompting us to evaluate the correlation of our ER stress-associated signature with the PC immune microenvironment. The functional enrichment analysis showed that ER stress status in PC may have a close relationship to a multiple of immune signalling pathways including T cell proliferation, B cell proliferation and activation, and interleukin-4 production. A body of findings in oncological research indicate that ER stress can have effects on the trafficking of immune cell to the tumour milieu. For instance, Song et al. showed XBP1 is upregulated in T cells isolated from ovarian cancer patient samples in contrast with healthy donor-derived T cells, and the upregulation of XBP1 is correlated with decreased intratumoural T cell infiltration (Song et al., 2018). Work by Harnoss and colleagues demonstrated that in TNBC xenograft, knockout of IRE1α can reduce MDSC infiltration by downregulating the transcription of several genes involved in recruitment, such as PTGS2, CXCL8, and CXCR4 (Harnoss et al., 2020). Histological analyses revealed that in human hepatocellular carcinoma (HCC) specimens, the expression levels of ER stress markers such as BiP, ATF6, PERK, and IRE1α were closely associated with an elevated infiltration of $CD68^{+}PDL1^{+}$ macrophages. The elevating PD-L1 expression on macrophages is primarily achieved by exosomes containing miR-23a-3p released from HCC cells subjected to ER stress mediated PTEN/PI3K/AKT pathway, which in turn impacts T cell functions and survival (Liu et al., 2019). In our study, we observed that low-risk cluster has higher accumulation of T cells and NK cells, comparing with high-risk cluster. It has been demonstrated that low infiltration and dysfunction of NK cells in PC tissues are tightly associated with tumour recurrence and survival of patients (Hoshikawa et al., 2018; Xie et al., 2019). As such, our ER stress risk signature could serve as a valuable tool for evaluating NK cell immune status and predicting the efficacy of NK cell-based immunotherapy in PC. Since the cytotoxic and cytokine-producing effector molecules functions of NK cells are governed by a delicate balance between activating and inhibitory receptors, we sought correlations between NK cell ligands and our risk signature, albeit with inconclusive outcomes. Comparative analyses revealed that our high-risk and low-risk cohorts exhibited no discernible disparities in the mRNA expression levels of the MICA/B, cognate ligands of major NK activating receptor NKG2D (Cadoux et al., 2021), and the inhibitory receptors ligands HLA-E (Liu et al., 2023) and CEACAM1 (Park et al., 2020). Interestingly, relative to healthy pancreatic tissue, all the aforementioned molecules exhibited augmented expression in PC. This observation may be attributed to the impairment of IRE1α arm in PC, as it was reported that the MICA/B expression was suppressed by the IRE1α/XBP1 arm of the ER stress response in melanoma cells (Obiedat et al., 2019). Eosinophils, primitive cells of the innate immune system primarily studied in allergic diseases, have recently garnered attention for their role in promoting antitumour immune responses (Grisaru-Tal et al., 2022; O'Flaherty et al., 2017; Blomberg et al., 2023) demonstrated that eosinophils can enhance NK cell chemotaxis and activation via secretion of CCL5, CXCL10 and IL-12. Moreover, colorectal cancer-infiltrated eosinophils have been shown to promote IFN-γ and TNF-α production by $CD8^{+}$ T cells via the GM-CSF-IRF5 signalling axis, thereby augmenting antitumour immune responses (Arnold et al., 2020). Intriguingly, our findings revealed a predominant concentration of tumour-infiltrating eosinophils in the low-risk cluster compared to its high-risk counterpart, potentially elucidating the abundant occurrence of NK and T cells within the low-risk group. It's hypothesized that ER stress might compromise the antitumour efficacy of NK and T cells by obstructing eosinophil recruitment and activation. Both macrophages and MDSCs, pivotal myeloid suppressive entities, can debilitate NK and T cell functionality, facilitating tumour immune evasion (Sui et al., 2022; Cassetta and Pollard, 2023). Our observations indicate an inverse relationship between MDSCs and ER stress scores, presenting an apparent paradox. Given the pronounced presence of NK and T cells within low-risk strata, this data intimates potential regulatory feedback exerted by MDSCs that modulates NK- and T-cell-mediated antitumour immune responses. Thus, ER stress's influence on the TME seems intricate, making targeted interventions demanding. Therefore, the effect of ER stress on the TME may be multifaceted and targeting ER stress remains challenging.

In summary, we established a 12-gene ER stress-associated risk signature, holding promise as a prognostic tool for PC patients. Importantly, this prognostic risk signature is positively correlated with an immunosuppressive state in PC, indicating its potential of guiding clinical decision making in the immunotherapy of PC. The analyses of functional enrichment and quantification of the immune infiltration landscape yielded essential findings for initiating future research on the mechanisms underlying ER stress-related genes and tumour immunity in PC. Nonetheless, it's crucial to highlight that our investigation relied solely on transcriptomic data from surgical biopsies, so further clinical trials and experimental validations are required to confirm our findings.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: The Cancer Genome Atlas (https://xenabrowser.net/), International Cancer Genome Consortium (https://icgc.org), and Molecular Signature Database v7.4 (https://www.gsea-msigdb.org/). The full original source data used in this study has been downloaded from the above public datasets and uploaded to jianguoyun along with codes. Detailed information can be found here: "https://www.jianguoyun.com/p/DYpnVlIQ4qX6CxjQ3ZoFIAA."

Ethics statement

Ethical approval was not required for the studies involving humans because the data were extracted from the public database. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from The Cancer

Genome Atlas Program (TCGA). Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

HC: Conceptualization, Data curation, Formal Analysis, Methodology, Visualization, Writing–original draft. NX: Conceptualization, Investigation, Methodology, Writing–original draft. JX: Conceptualization, Investigation, Methodology, Writing–original draft. CZ: Conceptualization, Investigation, Methodology, Writing–original draft. XL: Conceptualization, Data curation, Investigation, Validation, Writing–review and editing. HX: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Writing–original draft. WZ: Formal Analysis, Investigation, Writing–original draft. JL: Conceptualization, Methodology, Writing–review and editing. DL: Conceptualization, Methodology, Writing–review and editing. WZ: Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Supervision, 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 National Natural Science Foundation of China (grant number 82260555); Medical Innovation and Development Project of Lanzhou University (grant number lzuyxcx-2022-177) and Major Science and Technology Projects of Gansu Province (grant number 22ZD6FA021-4).

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/fmolb.2023.1298077/full#supplementary-material

References

Allison, S. (2021). ER-associated degradation of nephrin. *Nat. Rev. Nephrol.* 17 (5), 297. doi:10.1038/s41581-021-00421-7

Arnold, I. C., Artola-Boran, M., Gurtner, A., Bertram, K., Bauer, M., Frangez, Z., et al. (2020). The GM-CSF-IRF5 signaling axis in eosinophils promotes antitumor immunity through activation of type 1 T cell responses. *J. Exp. Med.* 217 (12), e20190706. doi:10.1084/jem.20190706

Bear, A. S., Vonderheide, R. H., and O'Hara, M. H. (2020). Challenges and opportunities for pancreatic cancer immunotherapy. *Cancer Cell* 38 (6), 788–802. doi:10.1016/j.ccell.2020.08.004

Blomberg, O. S., Spagnuolo, L., Garner, H., Voorwerk, L., Isaeva, O. I., van Dyk, E., et al. (2023). IL-5-producing CD4(+) T cells and eosinophils cooperate to enhance response to immune checkpoint blockade in breast cancer. *Cancer Cell* 41 (1), 106–123.e10. doi:10.1016/j.ccell.2022.11.014

Cadoux, M., Caruso, S., Pham, S., Gougelet, A., Pophillat, C., Riou, R., et al. (2021). Expression of NKG2D ligands is downregulated by β-catenin signalling and associates with HCC aggressiveness. *J. Hepatol.* 74 (6), 1386–1397. doi:10.1016/j.jhep.2021.01.017

Cai, J., Chen, H., Lu, M., Zhang, Y., Lu, B., You, L., et al. (2021). Advances in the epidemiology of pancreatic cancer: trends, risk factors, screening, and prognosis. *Cancer Lett.* 520, 1–11. doi:10.1016/j.canlet.2021.06.027

Cardano, M., Magni, M., Alfieri, R., Chan, S. Y., Sabbioneda, S., Buscemi, G., et al. (2023). Sex specific regulation of TSPY-Like 2 in the DNA damage response of cancer cells. *Cell Death Dis.* 14 (3), 197. doi:10.1038/s41419-023-05722-2

Cassetta, L., and Pollard, J. W. (2023). A timeline of tumour-associated macrophage biology. *Nat. Rev. Cancer* 23 (4), 238–257. doi:10.1038/s41568-022-00547-1

Chandrasekaran, A. P., Tyagi, A., Poondla, N., Sarodaya, N., Karapurkar, J. K., Kaushal, K., et al. (2022). Dual role of deubiquitinating enzyme USP19 regulates mitotic progression and tumorigenesis by stabilizing survivin. *Mol. Ther.* 30 (11), 3414–3429. doi:10.1016/j.ymthe.2022.07.019

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, X., and Cubillos-Ruiz, J. R. (2021). Endoplasmic reticulum stress signals in the tumour and its microenvironment. *Nat. Rev. Cancer* 21 (2), 71–88. doi:10.1038/s41568-020-00312-2

Chen, G., Zhang, H., Sun, H., Ding, X., Liu, G., Yang, F., et al. (2023). Bufalin targeting BFAR inhibits the occurrence and metastasis of gastric cancer through PI3K/AKT/mTOR signal pathway. *Apoptosis*. doi:10.1007/s10495-023-01855-z

Cheng, X., Holenya, P., Can, S., Alborzinia, H., Rubbiani, R., Ott, I., et al. (2014). A TrxR inhibiting gold(I) NHC complex induces apoptosis through ASK1-p38-MAPK signaling in pancreatic cancer cells. *Mol. Cancer* 13, 221. doi:10.1186/1476-4598-13-221

Cheng, Z., Shu, H., Cui, Y., Zhang, Q., Zhao, B., Pan, D., et al. (2020). MiR-424-5p inhibits proliferation, invasion and promotes apoptosis and predicts good prognosis in glioma by directly targeting BFAR. *Pathol. Oncols. Res.* 26 (4), 2327–2335. doi:10.1007/s12253-020-00831-1

Conroy, T., Hammel, P., Hebbar, M., Ben Abdelghani, M., Wei, A. C., Raoul, J. L., et al. (2018). FOLFIRINOX or gemcitabine as adjuvant therapy for pancreatic cancer. *N. Engl. J. Med.* 379 (25), 2395–2406. doi:10.1056/NEJMoa1809775

Cubillos-Ruiz, J. R., Bettigole, S. E., and Glimcher, L. H. (2017). Tumorigenic and immunosuppressive effects of endoplasmic reticulum stress in cancer. *Cell* 168 (4), 692–706. doi:10.1016/j.cell.2016.12.004

Cubillos-Ruiz, J. R., Silberman, P. C., Rutkowski, M. R., Chopra, S., Perales-Puchalt, A., Song, M., et al. (2015). ER stress sensor XBP1 controls anti-tumor immunity by disrupting dendritic cell homeostasis. *Cell* 161 (7), 1527–1538. doi:10.1016/j.cell.2015.05.025

Dauer, P., Sharma, N. S., Gupta, V. K., Durden, B., Hadad, R., Banerjee, S., et al. (2019). ER stress sensor, glucose regulatory protein 78 (GRP78) regulates redox status in pancreatic cancer thereby maintaining "stemness". *Cell Death Dis.* 10 (2), 132. doi:10.1038/s41419-019-1408-5

Di Conza, G., Ho, P. C., Cubillos-Ruiz, J. R., and Huang, S. C. (2023). Control of immune cell function by the unfolded protein response. *Nat. Rev. Immunol.* 23, 546–562. doi:10.1038/s41577-023-00838-0

Dou, Z., Zhao, D., Chen, X., Xu, C., Jin, X., Zhang, X., et al. (2021). Aberrant Bcl-x splicing in cancer: from molecular mechanism to therapeutic modulation. *J. Exp. Clin. Cancer Res.* 40 (1), 194. doi:10.1186/s13046-021-02001-w

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

Friess, H., Lu, Z., Andrén-Sandberg, A., Berberat, P., Zimmermann, A., Adler, G., et al. (1998). Moderate activation of the apoptosis inhibitor bcl-xL worsens the prognosis in pancreatic cancer. *Ann. Surg.* 228 (6), 780–787. doi:10.1097/00000658-199812000-00009

Globocan (2020). *Cancer today*. Available at: https://gco.iarc.fr/today/home.

Grisaru-Tal, S., Rothenberg, M. E., and Munitz, A. (2022). Eosinophil-lymphocyte interactions in the tumor microenvironment and cancer immunotherapy. *Nat. Immunol.* 23 (9), 1309–1316. doi:10.1038/s41590-022-01291-2

Harnoss, J. M., Le Thomas, A., Reichelt, M., Guttman, O., Wu, T. D., Marsters, S. A., et al. (2020). IRE1α disruption in triple-negative breast cancer cooperates with antiangiogenic therapy by reversing ER stress adaptation and remodeling the tumor microenvironment. *Cancer Res.* 80 (11), 2368–2379. doi:10.1158/0008-5472.Can-19-3108

Hetz, C., Zhang, K., and Kaufman, R. J. (2020). Mechanisms, regulation and functions of the unfolded protein response. *Nat. Rev. Mol. Cell Biol.* 21 (8), 421–438. doi:10.1038/s41580-020-0250-z

Ho, W. J., Jaffee, E. M., and Zheng, L. (2020). The tumour microenvironment in pancreatic cancer - clinical challenges and opportunities. *Nat. Rev. Clin. Oncol.* 17 (9), 527–540. doi:10.1038/s41571-020-0363-5

Hoshikawa, M., Aoki, T., Matsushita, H., Karasaki, T., Hosoi, A., Odaira, K., et al. (2018). NK cell and IFN signatures are positive prognostic biomarkers for resectable pancreatic cancer. *Biochem. Biophys. Res. Commun.* 495 (2), 2058–2065. doi:10.1016/j.bbrc.2017.12.083

Hu, W., Su, Y., Fei, X., Wang, X., Zhang, G., Su, C., et al. (2020). Ubiquitin specific peptidase 19 is a prognostic biomarker and affect the proliferation and migration of clear cell renal cell carcinoma. *Oncol. Rep.* 43 (6), 1964–1974. doi:10.3892/or.2020.7565

Huang, R., Li, G., Wang, K., Wang, Z., Zeng, F., Hu, H., et al. (2021). Comprehensive analysis of the clinical and biological significances of endoplasmic reticulum stress in diffuse gliomas. *Front. Cell Dev. Biol.* 9, 619396. doi:10.3389/fcell.2021.619396

Kang, H., Choi, M. C., Kim, S., Jeong, J. Y., Kwon, A. Y., Kim, T. H., et al. (2021). USP19 and RPL23 as candidate prognostic markers for advanced-stage high-grade serous ovarian carcinoma. *Cancers (Basel)* 13 (16), 3976. doi:10.3390/cancers13163976

Kong, B., Cheng, T., Wu, W., Regel, I., Raulefs, S., Friess, H., et al. (2015). Hypoxia-induced endoplasmic reticulum stress characterizes a necrotic phenotype of pancreatic cancer. *Oncotarget* 6 (31), 32154–32160. doi:10.18632/oncotarget.5168

Kosuke, Y., Hoon, K., and Roel, G. V. (2016). *Stimate: estimate of stromal and immune cells in malignant tumor tissues from expression data. R package version 1.0.13/r21*. Available at: https://rdrr.io/rforge/estimate/#:~:text=from%20Expression%20Data-,estimate%3A%20Estimate%20of%20Stromal%20and%20Immune%20Cells%20in%20Malignant%20Tumor,tissues%20using%20gene%20expression%20data.

Lee, J. U., Hosotani, R., Wada, M., Doi, R., Kosiba, T., Fujimoto, K., et al. (1999). Role of Bcl-2 family proteins (Bax, Bcl-2 and Bcl-X) on cellular susceptibility to radiation in pancreatic cancer cells. *Eur. J. Cancer* 35 (9), 1374–1380. doi:10.1016/s0959-8049(99)00134-3

Liu, C. Y., Hsu, C. C., Huang, T. T., Lee, C. H., Chen, J. L., Yang, S. H., et al. (2018). ER stress-related ATF6 upregulates CIP2A and contributes to poor prognosis of colon cancer. *Mol. Oncol.* 12 (10), 1706–1717. doi:10.1002/1878-0261.12365

Liu, J., Fan, L., Yu, H., Zhang, J., He, Y., Feng, D., et al. (2019). Endoplasmic reticulum stress causes liver cancer cells to release exosomal miR-23a-3p and up-regulate programmed death ligand 1 expression in macrophages. *Hepatology* 70 (1), 241–258. doi:10.1002/hep.30607

Liu, W., Wang, D., Wang, X., Liu, P., and Yan, M. (2020). hsa_circ_0085539 promotes osteosarcoma progression by regulating miR-526b-5p and SERP1. *Mol. Ther. Oncolytics* 19, 163–177. doi:10.1016/j.omto.2020.09.009

Liu, X., Song, J., Zhang, H., Liu, X., Zuo, F., Zhao, Y., et al. (2023). Immune checkpoint HLA-E:CD94-NKG2A mediates evasion of circulating tumor cells from NK cell surveillance. *Cancer Cell* 41 (2), 272–287.e9. doi:10.1016/j.ccell.2023.01.001

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 (12), 550. doi:10.1186/s13059-014-0550-8

Lucianò, A. M., Di Martile, M., Pérez-Oliva, A. B., Di Caprio, M., Foddai, M. L., Buglioni, S., et al. (2023). Exploring association of melanoma-specific Bcl-xL with tumor immune microenvironment. *J. Exp. Clin. Cancer Res.* 42 (1), 178. doi:10.1186/s13046-023-02735-9

Ma, Q., Wu, X., Wu, J., Liang, Z., and Liu, T. (2017). SERP1 is a novel marker of poor prognosis in pancreatic ductal adenocarcinoma patients via anti-apoptosis and regulating SRPRB/NF-κB axis. *Int. J. Oncol.* 51 (4), 1104–1114. doi:10.3892/ijo.2017.4111

Miwa, T., Kanda, M., Koike, M., Iwata, N., Tanaka, H., Umeda, S., et al. (2017a). Identification of NCCRP1 as an epigenetically regulated tumor suppressor and biomarker for malignant phenotypes of squamous cell carcinoma of the esophagus. *Oncol. Lett.* 14 (4), 4822–4828. doi:10.3892/ol.2017.6753

Miwa, T., Kanda, M., Tanaka, H., Tanaka, C., Kobayashi, D., Umeda, S., et al. (2017b). FBXO50 enhances the malignant behavior of gastric cancer cells. *Ann. Surg. Oncol.* 24 (12), 3771–3779. doi:10.1245/s10434-017-5882-7

Mizrahi, J. D., Surana, R., Valle, J. W., and Shroff, R. T. (2020). Pancreatic cancer. *Lancet* 395 (10242), 2008–2020. doi:10.1016/s0140-6736(20)30974-0

Neoptolemos, J. P., Kleeff, J., Michl, P., Costello, E., Greenhalf, W., and Palmer, D. H. (2018). Therapeutic developments in pancreatic cancer: current and future perspectives. *Nat. Rev. Gastroenterol. Hepatol.* 15 (6), 333–348. doi:10.1038/s41575-018-0005-x

Obiedat, A., Seidel, E., Mahameed, M., Berhani, O., Tsukerman, P., Voutetakis, K., et al. (2019). Transcription of the NKG2D ligand MICA is suppressed by the IRE1/XBP1 pathway of the unfolded protein response through the regulation of E2F1. *Faseb J.* 33 (3), 3481–3495. doi:10.1096/fj.201801350RR

O'Flaherty, S. M., Sutummaporn, K., Häggtoft, W. L., Worrall, A. P., Rizzo, M., Braniste, V., et al. (2017). TLR-stimulated eosinophils mediate recruitment and activation of NK cells *in vivo*. *Scand. J. Immunol.* 85 (6), 417–424. doi:10.1111/sji.12554

Pan, H., Yang, L., Bai, H., Luo, J., and Deng, Y. (2022). Ginsenoside Rg3 increases gemcitabine sensitivity of pancreatic adenocarcinoma via reducing ZFP91 mediated TSPYL2 destabilization. *J. Ginseng Res.* 46 (5), 636–645. doi:10.1016/j.jgr.2021.08.004

Park, D. J., Sung, P. S., Kim, J. H., Lee, G. W., Jang, J. W., Jung, E. S., et al. (2020). EpCAM-high liver cancer stem cells resist natural killer cell-mediated cytotoxicity by upregulating CEACAM1. *J. Immunother. Cancer* 8 (1), e000301. doi:10.1136/jitc-2019-000301

Pommier, A., Anaparthy, N., Memos, N., Kelley, Z. L., Gouronnec, A., Yan, R., et al. (2018). Unresolved endoplasmic reticulum stress engenders immune-resistant, latent pancreatic cancer metastases. *Science* 360 (6394), eaao4908. doi:10.1126/science.aao4908

Rossi, F. A., Enriqué Steinberg, J. H., Calvo Roitberg, E. H., Joshi, M. U., Pandey, A., Abba, M. C., et al. (2021). USP19 modulates cancer cell migration and invasion and acts as a novel prognostic marker in patients with early breast cancer. *Oncogenesis* 10 (3), 28. doi:10.1038/s41389-021-00318-x

Rossi, F. A., and Rossi, M. (2022). Emerging role of ubiquitin-specific protease 19 in oncogenesis and cancer development. *Front. Cell Dev. Biol.* 10, 889166. doi:10.3389/fcell.2022.889166

Rufo, N., Garg, A. D., and Agostinis, P. (2017). The unfolded protein response in immunogenic cell death and cancer immunotherapy. *Trends Cancer* 3 (9), 643–658. doi:10.1016/j.trecan.2017.07.002

Schwarz, D. S., and Blower, M. D. (2016). The endoplasmic reticulum: structure, function and response to cellular signaling. *Cell Mol. Life Sci.* 73 (1), 79–94. doi:10.1007/s00018-015-2052-6

Sen Santara, S., Lee, D. J., Crespo, Â., Hu, J. J., Walker, C., Ma, X., et al. (2023). The NK cell receptor NKp46 recognizes ecto-calreticulin on ER-stressed cells. *Nature* 616 (7956), 348–356. doi:10.1038/s41586-023-05912-0

Sherman, M. H., and Beatty, G. L. (2023). Tumor microenvironment in pancreatic cancer pathogenesis and therapeutic resistance. *Annu. Rev. Pathol.* 18, 123–148. doi:10.1146/annurev-pathmechdis-031621-024600

Shi, C., Zhao, F., Zhang, T., Xu, D., Hao, Z., Cui, F., et al. (2022). A novel prognostic signature in osteosarcoma characterised from the perspective of unfolded protein response. *Clin. Transl. Med.* 12 (3), e750. doi:10.1002/ctm2.750

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *CA Cancer J. Clin.* 70 (1), 7–30. doi:10.3322/caac.21590

Sohal, D. P. S., Kennedy, E. B., Cinar, P., Conroy, T., Copur, M. S., Crane, C. H., et al. (2020). Metastatic pancreatic cancer: ASCO guideline update. *J. Clin. Oncol.* 38, 3217–3230. doi:10.1200/jco.20.01364

Song, M., Sandoval, T. A., Chae, C. S., Chopra, S., Tan, C., Rutkowski, M. R., et al. (2018). IRE1α-XBP1 controls T cell function in ovarian cancer by regulating mitochondrial activity. *Nature* 562 (7727), 423–428. doi:10.1038/s41586-018-0597-x

Song, P., Li, W., Guo, L., Ying, J., Gao, S., and He, J. (2022). Identification and validation of a novel signature based on NK cell marker genes to predict prognosis and immunotherapy response in lung adenocarcinoma by integrated analysis of single-cell and bulk RNA-sequencing. *Front. Immunol.* 13, 850745. doi:10.3389/fimmu.2022.850745

Sui, H., Dongye, S., Liu, X., Xu, X., Wang, L., Jin, C. Q., et al. (2022). Immunotherapy of targeting MDSCs in tumor microenvironment. *Front. Immunol.* 13, 990463. doi:10.3389/fimmu.2022.990463

Tadros, S., Shukla, S. K., King, R. J., Gunda, V., Vernucci, E., Abrego, J., et al. (2017). *De novo* lipid synthesis facilitates gemcitabine resistance through endoplasmic reticulum stress in pancreatic cancer. *Cancer Res.* 77 (20), 5503–5517. doi:10.1158/0008-5472.Can-16-3062

Terry, T., and Patricia, G. (2001). Modeling Survival Data: extending the Cox Model. Terry M. Therneau and Patricia M. Grambsch, Springer-Verlag, New York, 2000. No. of pages: xiii + 350. Price: $69.95. ISBN 0-387-98784-3. *Statistics Med.* 20 (13), 2053–2054. doi:10.1002/sim.956

Therneau, T. (2021). *A package for survival analysis in R. R package version 3.2-13*. Available at: https://cran.r-project.org/web/packages/survival/vignettes/survival.pdf.

Thummuri, D., Khan, S., Underwood, P. W., Zhang, P., Wiegand, J., Zhang, X., et al. (2022). Overcoming Gemcitabine Resistance in Pancreatic Cancer Using the BCL-X(L)-Specific Degrader DT2216. *Mol. Cancer Ther.* 21 (1), 184–192. doi:10.1158/1535-7163.Mct-21-0474

Wang, L., Yin, W., and Shi, C. (2017). E3 ubiquitin ligase, RNF139, inhibits the progression of tongue cancer. *BMC Cancer* 17 (1), 452. doi:10.1186/s12885-017-3438-7

Wei, T., and Simko, V. (2012). "R package 'corrplot': Visualization of a Correlation Matrix". *R package version 0.90 ed.*

Wiseman, R. L., and Balch, W. E. (2005). A new pharmacology--drugging stressed folding pathways. *Trends Mol. Med.* 11 (8), 347–350. doi:10.1016/j.molmed.2005.06.011

Xiang, Y., and Zeng, J. (2019). *AnnoProbe: annotate the gene symbols for probes in expression array. R package version 0.1.0*. Available at: https://rdrr.io/cran/AnnoProbe/.

Xie, X., Ma, L., Zhou, Y., Shen, W., Xu, D., Dou, J., et al. (2019). Polysaccharide enhanced NK cell cytotoxicity against pancreatic cancer via TLR4/MAPKs/NF-κB pathway *in vitro*/vivo. *Carbohydr. Polym.* 225, 115223. doi:10.1016/j.carbpol.2019.115223

Yamaguchi, A., Hori, O., Stern, D. M., Hartmann, E., Ogawa, S., and Tohyama, M. (1999). Stress-associated endoplasmic reticulum protein 1 (SERP1)/Ribosome-associated membrane protein 4 (RAMP4) stabilizes membrane proteins during stress and facilitates subsequent glycosylation. *J. Cell Biol.* 147 (6), 1195–1204. doi:10.1083/jcb.147.6.1195

Yan, W., Wang, X., Liu, T., Chen, L., Han, L., Xu, J., et al. (2019). Expression of endoplasmic reticulum oxidoreductase 1-α in cholangiocarcinoma tissues and its effects on the proliferation and migration of cholangiocarcinoma cells. *Cancer Manag. Res.* 11, 6727–6739. doi:10.2147/cmar.S188746

Yuan, Y., Jiao, P., Wang, Z., Chen, M., Du, H., Xu, L., et al. (2022). Endoplasmic reticulum stress promotes the release of exosomal PD-L1 from head and neck cancer cells and facilitates M2 macrophage polarization. *Cell Commun. Signal* 20 (1), 12. doi:10.1186/s12964-021-00810-2

Zanetti, G., Pahuja, K. B., Studer, S., Shim, S., and Schekman, R. (2011). COPII and the regulation of protein sorting in mammals. *Nat. Cell Biol.* 14 (1), 20–28. doi:10.1038/ncb2390

Zeng, H., Chen, W., Zheng, R., Zhang, S., Ji, J. S., Zou, X., et al. (2018). Changing cancer survival in China during 2003-15: a pooled analysis of 17 population-based cancer registries. *Lancet Glob. Health* 6 (5), e555–e567. doi:10.1016/s2214-109x(18)30127-x

Zhang, J., Yang, J., Lin, C., Liu, W., Huo, Y., Yang, M., et al. (2020). Endoplasmic Reticulum stress-dependent expression of ERO1L promotes aerobic glycolysis in Pancreatic Cancer. *Theranostics* 10 (18), 8400–8414. doi:10.7150/thno.45124

Zhang, Q., Guan, G., Cheng, P., Cheng, W., Yang, L., and Wu, A. (2021). Characterization of an endoplasmic reticulum stress-related signature to evaluate immune features and predict prognosis in glioma. *J. Cell Mol. Med.* 25 (8), 3870–3884. doi:10.1111/jcmm.16321

Zhang, Y., Li, T., Zhang, L., Shangguan, F., Shi, G., Wu, X., et al. (2019). Targeting the functional interplay between endoplasmic reticulum oxidoreductin-1α and protein disulfide isomerase suppresses the progression of cervical cancer. *EBioMedicine* 41, 408–419. doi:10.1016/j.ebiom.2019.02.041

Zhang, Y., Song, Y., Dai, J., Wang, Z., Zeng, Y., Chen, F., et al. (2022). Endoplasmic Reticulum Stress-Related Signature Predicts Prognosis and Drug Response in Clear Cell Renal Cell Carcinoma. *Front. Pharmacol.* 13, 909123. doi:10.3389/fphar.2022.909123

Zhou, J., Qian, W., Huang, C., Mai, C., Lai, Y., Lin, Z., et al. (2022). Combined targeting of KRT23 and NCCRP1 as a potential novel therapeutic approach for the treatment of triple-negative breast cancer. *Gland. Surg.* 11 (10), 1673–1682. doi:10.21037/gs-22-486

Zilli, F., Marques Ramos, P., Auf der Maur, P., Jehanno, C., Sethi, A., Coissieux, M. M., et al. (2021). The NFIB-ERO1A axis promotes breast cancer metastatic colonization of disseminated tumour cells. *EMBO Mol. Med.* 13 (4), e13162. doi:10.15252/emmm.202013162

frontiers | Frontiers in Oncology

TYPE Systematic Review
PUBLISHED 16 January 2024
DOI 10.3389/fonc.2024.1276549

Check for updates

OPEN ACCESS

EDITED BY
Li Chen,
Huazhong University of Science and Technology, China

REVIEWED BY
Shu Peng,
Huazhong University of Science and Technology, China
Ze-Rui Zhao,
Sun Yat-sen University Cancer Center (SYSUCC), China

*CORRESPONDENCE
Yun Han
hany@sj-hospital.org

RECEIVED 12 August 2023
ACCEPTED 02 January 2024
PUBLISHED 16 January 2024

CITATION
Wang H, Liang S, Yu Y and Han Y (2024) Efficacy and safety of neoadjuvant immunotherapy protocols and cycles for non-small cell lung cancer: a systematic review and meta-analysis.
Front. Oncol. 14:1276549.
doi: 10.3389/fonc.2024.1276549

COPYRIGHT
© 2024 Wang, Liang, Yu and Han. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Efficacy and safety of neoadjuvant immunotherapy protocols and cycles for non-small cell lung cancer: a systematic review and meta-analysis

Huaiyong Wang, Song Liang, Yue Yu and Yun Han*

Department of Thoracic Surgery, Shengjing Hospital of China Medical University, Shenyang, Liaoning, China

Objectives: This study evaluated the use of different neoadjuvant immunotherapy cycles and regimens for non-small cell lung cancer.

Materials and methods: Databases were searched for articles published up until December 2023. Data on the major pathologic response (MPR), complete pathologic response (pCR), radiological response, treatment-related adverse events (TRAEs), serious adverse events (SAEs), surgical resection, surgical complications, R0 resection, and conversion to thoracotomy were collected. A subgroup analysis was performed according to the treatment regimens and cycles. Stata/MP software was used for statistical analyses.

Results: In total, 2430 individuals were assessed from 44 studies. Compared with those following neoadjuvant immunotherapy alone (MPR/pCR/TRAEs/SAEs: ES=0.26/0.07/0.43/0.08, 95% CI: 0.18-0.34/0.04-0.10/0.28-0.58/0.04-0.14), the MPR and pCR rates, incidence of TRAEs and SAEs following neoadjuvant chemoimmunotherapy increased significantly (MPR/pCR/TRAEs/SAEs: ES=0.55/0.34/0.81/0.22, 95% CI: 0.48-0.63/0.28-0.41/0.69-0.90/0.13-0.33, P=0.001/0.002/0.009/0.034). No significant differences were found in the surgical resection, surgical complications, R0 resection, or conversion to thoracotomy. In the chemoimmunotherapy group, no statistically significant differences were found in the MPR and pCR rates, incidence of TRAEs and SAEs in the two-cycle, three-cycle and four-cycle groups (MPR/pCR/TRAEs/SAEs: ES=0.50;0.70;0.36/0.32;0.49;0.18/0.95;0.85;0.95/0.34;0.27;0.37, P=0.255/0.215/0.253/0.848). In the ICIs group, there was little change in the MPR and pCR rates, incidence of TRAEs and SAEs in the two-cycle group compared to the three-cycle group. (MPR/pCR/TRAEs/SAEs: ES=0.28;0.20/0.06;0.08/0.45;0.35/0.10;0.02, P=0.696/0.993/0.436/0.638). The neoadjuvant treatment cycle had no significant effect on surgical resection, surgical complications, R0 resection, or conversion to thoracotomy in both regimens.

Conclusion: Neoadjuvant chemoimmunotherapy significantly increased the rate of tumor pathological remission compared to neoadjuvant immunotherapy alone but also increased the incidence of TRAEs and SAEs. The efficacy and

safety of neoadjuvant chemoimmunotherapy are found to be favorable when administered for a duration of three cycles, in comparison to both two and four cycles.

Systematic review registration: https://www.crd.york.ac.uk/PROSPERO/#recordDetails, identifier CRD42023407415.

KEYWORDS

lung cancer, preoperative immunotherapy, treatment regimens, neoadjuvant cycle, major pathologic response, meta-analysis

1 Introduction

Lung cancer is second only to breast cancer in terms of incidence worldwide and has the highest mortality rate among malignant tumors (1). From 2010 to 2019, the number of new tracheal, bronchial, and lung cancer cases increased by 23.3% (2). Therefore, effective interventions for lung cancer that prolong patient survival are needed. Radical surgery combined with neoadjuvant and adjuvant therapies, when necessary, has become the mainstay of treatment for non-metastatic lung cancer.

In recent years, programmed cell death protein 1 and programmed death-ligand 1 inhibitors have demonstrated unique therapeutic benefits in the neoadjuvant treatment of melanoma, hepatocellular carcinoma, and other tumors (3, 4). In 2018, CheckMate159 (5, 6) reported a 45% major pathologic response (MPR) rate and 24% incidence of treatment-related adverse events (TRAEs) following neoadjuvant immunotherapy in non-small cell lung cancer (NSCLC), which confirmed the feasibility and safety of the treatment. This led to a series of clinical studies on preoperative immunotherapy and immunotherapy combined with chemotherapy or radiotherapy. However, the results of these studies have been inconsistent. It cannot be excluded that the differences are related to indicators such as the neoadjuvant treatment regimen, cycle, or type of immune checkpoint inhibitors (ICIs). Previous meta-analyses have confirmed that different ICIs have no significant impact on the safety and feasibility of treatment (7). Therefore, we conducted a meta-analysis of the different neoadjuvant immunotherapy regimens and cycles.

Abbreviations: CI, confidence interval; CR, complete response; CTCAE, Common Terminology Criteria for Adverse Events; ctDNA, circulating tumor DNA; EFS, event-free survival; ES, effect size; ICIs, immune checkpoint inhibitors; MPR, major pathologic response; NCT, National Clinical Trial; NSCLC, non-small cell lung cancer; OR, odds ratio; ORR, objective response rate; OS, overall survival; pCR, complete pathologic response; PR, partial response; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; RCT, randomized controlled trial; RECIST, Response Evaluation Criteria in Solid Tumors; SAEs, serious adverse events; SBRT, stereotactic body radiation therapy; TRAEs, treatment-related adverse events.

2 Materials and methods

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA statement) and was registered in the International Prospective Register of Systematic Reviews (PROSPERO CRD42023407415).

2.1 Search strategy

We retrieved relevant studies on neoadjuvant immunotherapy for lung cancer by searching seven databases, including PubMed, Embase, Cochrane Library, Ovid, Scopus, ProQuest, and Web of Science, published through December 2023. The search terms were as follows: ("Carcinoma, Non-Small-Cell Lung" OR "Lung Carcinoma, Non-Small-Cell" OR "NSCLC") AND ("Neoadjuvant Therapy" OR "Neoadjuvant Treatment" OR "Neoadjuvant Radiotherapy" OR "Neoadjuvant Chemotherapy" OR "Neoadjuvant Systemic Therapy") AND ("Immunotherapy" OR "Immunotherapies").

The inclusion criteria were as follows: 1. Patients with pathologically confirmed stage I–IV lung cancer and the possibility of surgical resection; 2. preoperative application of neoadjuvant immunotherapy or immunotherapy combined with other treatments, such as chemotherapy and radiotherapy; and 3. complete patient characteristics and inclusion of important outcome indicators, such as pathological response, radiological response, TRAEs, and surgery-related data. The exclusion criteria were as follows: 1. The primary endpoint of the study was not related to the efficacy or safety of the neoadjuvant therapy; 2. studies that have not been completed; 3. duplicate publications and data; 4. sample size <10; 5. reviews, conference abstracts, case reports, animal studies, and cytological studies; 6. non-English literature. Two investigators independently searched and screened the articles separately, resolved differences through discussions, and determined the final search results.

2.2 Data extraction

Two researchers read the original and Supplementary Materials of the included publications and extracted the following relevant data: 1. Article author(s), year of publication, National Clinical Trial (NCT) number, sample size, and primary endpoint; 2. patient age, sex ratio, smoking ratio, pathological type, tumor stage, neoadjuvant treatment regimen, and cycles; and 3. pathological response (complete pathologic response [pCR], MPR), radiological response, the incidence of TRAE, and grading; and 4. surgical resection rate, surgical delay rate, the incidence of surgical complications, surgical style, and R0 resection rate (Table 1).

2.3 Data analysis

This study used Stata/MP 17.0 software for the data analysis. The extent of data heterogeneity was determined using I^2 and Q tests. A random effects model was used if the homogeneity test results were significant; otherwise, a fixed effects model was used. The pooled effect sizes (ES) were expressed as odds ratios (ORs) and 95% confidence intervals (CIs). Meta-regression was used to determine the differences among different neoadjuvant therapies. Publication bias was evaluated using funnel plots and Egger's test, and differences were considered significant when $P < 0.05$. The stability of the results was evaluated using a sensitivity analysis.

2.4 Quality evaluation

The Cochrane Collaboration's Risk of Bias tool was used to evaluate the quality of randomized controlled trials (RCTs) (Figure 1). For single-arm and cohort studies, the MINORS scale was used for evaluation (Supplementary Table 1).

3 Results

3.1 Search results

A total of 2777 publications were retrieved. After removing duplications, 1261 publications remained. After reading records in the literature, abstracts, titles, and full texts, 44 papers were screened (Figure 2), of which 7 were RCT studies (Forde et al., 2022 (8), Heymach et al., 2023 (9), Wakelee et al., 2023 (10), Sepesi et al., 2022 (11), Cascone et al., 2021 (12), Lee et al., 2022 (25), Altorki et al., 2021 (26)). Only data from groups treated with immunotherapy alone or neoadjuvant chemoimmunotherapy were analyzed.

3.2 Primary outcomes

3.2.1 Effectiveness of treatment

MPR was defined as there being less than 10% of residual viable tumor cells in the primary tumor, and pCR was defined as there being no residual viable tumor cells in the primary tumor. Thirty-six studies have evaluated MPR (5, 6, 8–35, 37–43, 45–49), with MPR rates ranging from 8–83%. pCR rates of 0–73% have also been reported in 36 studies (5, 6, 8–31, 33–48). Because of the high heterogeneity, the random effects model analysis suggested that the preoperative application of ICIs significantly improved the proportions of MPR and pCR. The pooled ES was 0.48 for MPR (95% CI 0.41–0.55; I^2=87.96%) and 0.26 for pCR (95% CI 0.21–0.32; I^2=86.49%) (Figures 3A, B). According to the Response Evaluation Criteria in Solid Tumors (RECIST 1.1) (50), the objective response was defined as the partial response (PR) and complete response (CR). After pooling the 35 studies (5, 6, 8–26, 28, 29, 31–49), the objective response rate (ORR) was found to be 0.57 (95% CI 0.47–0.68; I^2=94.86%) (Figure 3C).

3.2.2 Safety of treatment

TRAEs were graded using the Common Terminology Criteria for Adverse Events (CTCAE), with TRAEs graded ≥3 being considered serious adverse events (SAEs). Eighteen studies have assessed TRAEs (5, 6, 8–10, 13–16, 18–21, 23, 28, 30, 30, 34, 35, 40, 43, 47, 48), and their incidence ranges from 17–100%. The pooled ES was 0.70 (95% CI 0.56–0.81; I^2=96.02%) (Figure 3D). Twenty-five studies have evaluated SAEs (5, 6, 8–16, 18–21, 25–28, 30, 31, 35–37, 39, 40, 42, 45–48), such as myelosuppression, gastrointestinal reactions, and immune-related target organ or tissue injury. The pooled ES was 0.17 (95% CI 0.10–0.25; I^2=93.35%) (Figure 3E).

3.2.3 Surgery

Thirty-seven studies have evaluated surgical resection rates (5, 6, 8–38, 40–49), with the majority of patients undergoing surgery after neoadjuvant therapy (0.94, 95% CI 0.91–0.97; I^2=88.59%) (Figure 4A). The pooled margin profile showed that most patients who underwent surgery could undergo R0 resection (0.97, 95% CI 0.96–0.99; I^2=63.30%) (8–10, 13–18, 22–24, 29–38, 40, 41, 45–49) (Figure 4B). The main surgical complications included decreased hemoglobin levels, incisional infections, and prolonged air leaks. The pooled ES was 0.27 (95% CI 0.19–0.35; I^2=80.69%) (5, 6, 8, 11, 12, 15, 16, 19–21, 24, 27, 28, 30–32, 34, 35, 42, 43, 46, 48) (Figure 4C). A few patients underwent conversions to thoracotomy due to the dense adhesion of blood vessels and lymph nodes after treatment (0.06, 95% CI 0.02–0.10; I^2=80.95%) (5, 6, 8, 11–16, 19, 20, 22, 24–27, 29–32, 36, 38, 40, 42, 45–48) (Figure 4D).

3.3 Subgroup analysis

3.3.1 Neoadjuvant immunotherapy alone and neoadjuvant chemoimmunotherapy

A subgroup analysis was performed according to the treatment regimen. In the ICIs-alone group, the heterogeneity of pCR, surgical resection, and R0 resection was significantly reduced, so the fixed effects model was used for the analysis, and the remaining indicators were analyzed using the random effects model. The pooled ES of MPR was 0.26 (95% CI 0.18-0.34; I^2=54.73%); the

TABLE 1 Summary of studies on neoadjuvant immunotherapy for NSCLC.

Study	NCT Number	Clinical trial	Study type	Neoadjuvant treatment regimen	Cycles	Male n[%]	Age	Smoking history n(%)	Clinical stage	MPR	pCR	CR +PR	TRAEs	SAEs	Surgical resection	Surgical complications	R0 resection	Conversion to thoracotomy
Bott, 2019 (5) Forde, 2018 (6)	NCT02259621	CheckMate159	Cohort study	Nivolumab	2	10 (48.0%)	66.9	18 (86.0%)	IA-IIIA	45.0%	15.0%	9.5%	23.8%	4.8%	95.2%	50.0%	NR	35.0%
Forde, 2022 (8)	NCT02998528	CheckMate816	RCT	Nivolumab+CT	3	128 (71.5%)	64.0	160 (89.4%)	IB-IV	36.9%	24.0%	63.7%	81.0%	33.0%	83.2%	41.6%	83.2%	11.4%
Heymach, 2023 (9)	NCT03800134	AEGEAN	RCT	Durvalumab+CT	4	252 (68.9%)	65.0	315 (86.1%)	IIA-IIIB	33.3%	17.2%	56.3%	86.8%	32.4%	77.6%	NR	94.7%	NR
Wakelee, 2023 (10)	NCT03425643	KEYNOTE671	RCT	Pembrolizumab+CT	4	279 (70.3%)	63.0	343 (86.4%)	IIA-IIIB	30.2%	18.1%	NR	95.7%	40.7%	82.1%	NR	92.0%	NR
Sepesi, 2022 (11) Cascone, 2021 (12)	NCT03158129	NEOSTAR	RCT	Nivolumab (+ipilimumab)	3	15 (65.0%)	66.1	18 (78.0%)	IA-IIIA	23.8%	9.5%	21.7%	NR	13.0%	91.3%	38.1%	NR	4.8%
Chaft, 2022 (13) Rusch, 2023 (14)	NCT02927301	LCMC3	Cohort study	Atezolizumab	2	88 (49.0%)	65.0	163 (90.0%)	IB-IIIB	20.3%	5.6%	6.1%	60.8%	11.0%	87.8%	NR	91.2%	9.4%
Provencio, 2020 (15) Provencio, 2022 (16)	NCT03081689	NADIM	Cohort study	Nivolumab+CT	3	34 (74.0%)	63.0	46 (100%)	IIB-IIIA	82.9%	63.4%	76.1%	93.5%	34.8%	89.1%	29.3%	100.0%	9.8%
Shu, 2020 (17)	NCT02716038	Columbia/MGH	Cohort study	Atezolizumab+CT	4	15 (50.0%)	67.0	30 (100.0%)	IIB-IIIB	65.4%	38.5%	63.3%	NR	NR	96.7%	NR	100.0%	NR
Rothschild, 2021 (18)	NCT02572843	SAKK 16/14	Cohort study	Durvalumab+CT	2	35 (52.0%)	61.0	64 (95.0%)	IIIA(N2)	61.8%	18.2%	53.7%	100.0%	88.1%	82.1%	NR	92.7%	NR
F Zhang, 2022 (19) S Gao, 2020 (20)	ChiCTR-OIC -17013726	NR	Cohort study	Sintilimab	2	33 (82.5%)	59.8	32 (80.0%)	IA-IIIB	40.5%	16.2%	20.0%	52.5%	10.0%	97.5%	10.3%	NR	0
Wislez, 2022 (21)	NCT03030131	IFCT-1601 IONESCO	Cohort study	Durvalumab	3	31 (67.4%)	60.9	45 (97.8%)	IB-IIIA	18.6%	7.0%	8.7%	34.8%	0	93.5%	32.6%	NR	NR
ZR Zhao,2021 (22)	NCT04304248	NR	Cohort study	Toripalimab+CT	3	27 (81.8%)	61.0	NR	IIIA-IIIB	66.7%	50.0%	87.9%	NR	NR	90.9%	NR	96.7%	3.3%
J Shen, 2023 (23)	NR	NR	Cohort study	Pembrolizumab/ Nivolumab/ Stintilimab (+CT)	1-3	26 (83.9%)	61.0	19 (61.3%)	IB-IIIB	46.4%	21.5%	71.0%	64.5%	NR	90.3%	NR	96.4%	NR
Jiang, 2021 (24)	NR	NR	Cohort study	Pembrolizumab/ Nivolumab+CT	2–4	29 (93.5%)	61.0	7 (22.6%)	IIA-IIIB	38.7%	9.7%	77.4%	NR	NR	100.0%	58.1%	77.4%	3.2%
Lee, 2022 (25)	NR	NR	RCT	Durvalumab	2	16 (53.0%)	71.0	24 (80.0%)	IA-IIIA	7.7%	0.0%	3.3%	NR	16.7%	86.7%	NR	NR	0

(Continued)

10.3389/fonc.2024.1276549

TABLE 1 Continued

Study	NCT Number	Clinical trial	Study type	Neoadjuvant treatment regimen	Cycles	Male n[%]	Age	Smoking history n(%)	Clinical stage	MPR	pCR	CR +PR	TRAEs	SAEs	Surgical resection	Surgical complications	R0 resection	Conversion to thoracotomy
Altorki, 2021 (26)																		
Tong, 2021 (27)	NR	TOP1501	Cohort study	Pembrolizumab	2	16 (53.0%)	72.0	26 (86.7%)	IB-IIIIA	28.0%	12.0%	NR	NR	3.3%	83.3%	48.0%	NR	20.0%
Eichhorn, 2021 (28)	NCT03197467	NEOMUN	Cohort study	Pembrolizumab	2	7 (46.7%)	59.8	NR	IIB-IIIB	26.7%	13.3%	26.7%	33.3%	20.0%	100.0%	6.7%	NR	NR
Wu, 2022 (29)	NR	NR	Cohort study	Pembrolizumab/ Nivolumab+CT	2–4	72 (95.0%)	62.0	51 (67.0%)	IB-IIIB	64.5%	36.8%	75.0%	NR	NR	100.0%	NR	100.0%	2.6%
Y Gao, 2022 (30)	ChiCTR2200057840	NR	Cohort study	Nivolumab/ Camrelizumab/ Toripalimab/ Tislelizumab/ Sintilimab/ Pembrolizumab+CT	3	33 (75.0%)	61.5	33 (75.0%)	IIIA-IIIB	81.8%	59.1%	NR	81.8%	18.2%	100.0%	11.4%	100.0%	4.5%
Hu, 2021 (31)	NR	NR	Cohort study	Pembrolizumab/ Tislelizumab/ Sintilimab/ Toripalimab+CT	2–4	18 (90.0%)	56.0	17 (85.0%)	IB-IIIB	40.0%	25.0%	75.0%	NR	0	100.0%	35.0%	100.0%	0
B Zhang, 2022 (32)	NR	NR	Cohort study	Pembrolizumab/ Sintilizumab/ Toripalimab/ Camrelizumab/ Nivolizumab/ Tislelizumab	1-5	127 (96.9%)	59.3	120 (91.6%)	IB-IIIB	53.4%	NR	78.6%	NR	NR	100.0%	21.4%	95.4%	32.1%
Faehling, 2022 (33)	NR	KOMPASSneoOP	Cohort study	Pembrolizumab/ Nivolumab+CT	NR	30 (51.0%)	63.6	53 (95.0%)	IIB-IVB	67.8%	52.5%	84.7%	NR	NR	100.0%	NR	94.9%	NR
T Chen, 2021 (34)	NR	NR	Cohort study	Nivolumab/ Pembrolizumab+CT	2/4	9 (75.0%)	61.0	7(58.3%)	IIIA-IIIB	33.3%	41.7%	50.0%	33.3%	NR	100.0%	25.0%	100.0%	NR
P Zhang, 2022 (35)	ChiCTR1900023758	NR	Cohort study	Sintilimab+CT	2–4	44 (88.0%)	64.8	38 (76.0%)	IIIA	43.3%	20.0%	46.0%	90.0%	8.0%	60.0%	3.3%	100.0%	NR
Yang, 2017 (36)	NCT01820754	TOP1201	Cohort study	Ipilimumab+CT	2	12 (50.0%)	65.0	23 (96.0%)	IIA-IIIA	NR	15.4%	58.3%	NR	45.8%	54.2%	NR	100.0%	23.1%
Y Chen, 2021 (37)	NR	NR	Cohort study	Pembrolizumab+CT	2	29 (83.0%)	62.1	27 (77.1%)	IIIA-IIIB	22.9%	51.4%	48.6%	NR	2.9%	100.0%	NR	100.0%	NR
Duan, 2021 (38)	NR	NR	Cohort study	Sintilimab/ Nivolumab/ Pembrolizumab+CT	3–4	22 (95.7%)	61.8	22 (95.7%)	IIA-IIIB	50.0%	30.0%	73.9%	NR	NR	87.0%	NR	95.0%	10.0%
Tfayli, 2020 (39)	NCT03480230	NR	Cohort study	Avelumab+CT	4	7 (46.7%)	65.0	11 (73.3%)	IB-IIIA	27.3%	9.1%	26.7%	NR	26.7%	NR	NR	NR	NR
Sun, 2023 (40)	NCT04326153	NR	Cohort study	Sintilimab+CT	2–3	18 (90.0%)	56.9	18 (90.0%)	IIIA-IIIB	62.5%	31.3%	75.0%	70.0%	35.0%	80.0%	NR	100.0%	0

(Continued)

10.3389/fonc.2024.1276549

TABLE 1 Continued

Study	NCT Number	Clinical trial	Study type	Neoadjuvant treatment regimen	Cycles	Male n[%]	Age	Smoking history n(%)	Clinical stage	MPR	pCR	CR +PR	TRAEs	SAEs	Surgical resection	Surgical complications	R0 resection	Conversion to thoracotomy
G Zhao, 2023 (41)	NR	NR	Cohort study	Pembrolizumab+CT	2-3	22 (88.0%)	65.0	9(36.0%)	IIB-IIIB	61.9%	28.6%	80.0%	NR	NR	84.0%	NR	100.0%	NR
Huang, 2021 (42)	NR	NR	Cohort study	Nivolumab	2	16 (64.0%)	62.9	15 (60.0%)	IIIA	37.5%	4.2%	32.0%	NR	12.0%	96.0%	12.5%	NR	0
Ma, 2022 (43)	NR	NR	Cohort study	Pembrolizumab+CT	3–5	27 (77.1%)	63.0	29 (82.9%)	IIB-IIIC	77.1%	48.6%	80.0%	17.1%	NR	100.0%	8.6%	NR	NR
Wang, 2021 (44)	NR	NR	Cohort study	Nivolumab/ Pembrolizumab+CT	2	66 (91.7%)	62.2	60 (83.3%)	IIIA	NR	29.2%	94.4%	NR	NR	100.0%	NR	NR	NR
Zhai, 2022 (45)	NR	NR	Cohort study	Nivolumab+CT	3	26 (56.5%)	63.0	43 (93.5%)	IIIA-IIIB	71.1%	53.3%	60.9%	NR	19.6%	97.8%	NR	95.6%	0
Hong, 2021 (46)	NR	NR	Cohort study	Pembrolizumab/ Sintilimab/ Camrelizumab+CT	2–4	23 (92.0%)	62.0	179 (68.0%)	IIA-IIIC	52.0%	32.0%	88.0%	NR	0	100.0%	52.0%	100.0%	8.0%
Shen, 2021 (47)	NR	NR	Cohort study	Pembrolizumab+CT	2	35 (94.6%)	62.8	31 (83.8%)	IIB-IIIB	64.9%	45.9%	86.5%	70.3%	10.8%	100.0%	NR	100.0%	0
Yao, 2022 (48)	NR	NR	Cohort study	Camrelizumab/ Durvalumab+CT	2–3	10 (90.9%)	67.7	10 (90.9%)	IIIA-IIIB	81.8%	72.7%	72.7%	72.7%	0	100.0%	27.3%	100.0%	9.1%
Liu, 2022 (49)	NR	NR	Cohort study	Pembrolizumab/ Sintilimab/ Camrelizumab+CT	2–5	66 (83.5%)	NR	66 (83.5%)	IB-IIIB	53.2%	NR	70.9%	NR	NR	100.0%	NR	100.0%	NR

CT, Chemotherapy; NR, Not reported.

10.3389/fonc.2024.1276549

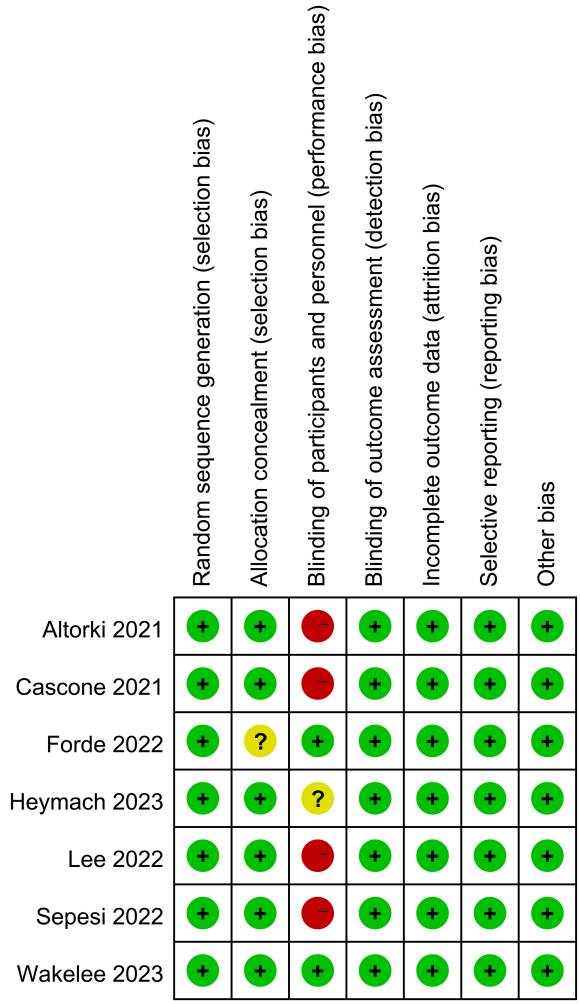


FIGURE 1
RCT literature evaluation.

pooled ES of pCR was 0.07 (95% CI 0.04-0.10; I^2=28.90%) (Supplementary Figure 1A); the pooled ES of ORR was 0.13 (95% CI 0.07-0.22; I^2=69.30%); the pooled ES of TRAEs was 0.43 (95% CI 0.28-0.58; I^2=79.91%); the pooled ES of SAEs was 0.08 (95% CI 0.04-0.14; I^2=53.85%); the pooled ES of surgical resection was 0.92 (95% CI 0.88-0.94; I^2=17.75%) (Supplementary Figure 1B); the pooled ES of R0 resection 0.95 (95% CI 0.92-0.97; I^2=16.12%) (Supplementary Figure 1C); the pooled ES of surgical complication was 0.27 (95% CI 0.14-0.41; I^2=75.49%), the pooled ES of conversion to thoracotomy was 0.06 (95% CI 0.01-0.15; I^2=79.00%).

In the chemoimmunotherapy group, each index was analyzed using the random effect model. The pooled ES of MPR was 0.55 (95% CI 0.48-0.63; I^2=88.09%); the pooled ES of pCR was 0.34 (95% CI 0.28-0.41; I^2=85.09%); the pooled ES of ORR was 0.71 (95% CI 0.65-0.77; I^2=82.06%); the pooled ES of TRAEs was 0.81 (95% CI 0.69-0.90; I^2=94.38%); the pooled ES of SAEs was 0.22 (95% CI 0.13-0.33; I^2=93.45%); the pooled ES of surgical resection was 0.95 (95% CI 0.90-0.98; I^2=91.49%); the pooled ES of R0 resection 0.98 (95% CI 0.96-0.99; I^2=70.46%); the pooled ES of surgical complication was 0.27 (95% CI 0.17-0.38; I^2=84.23%); the pooled ES of conversion to thoracotomy was 0.06 (95% CI 0.02-0.12; I^2=82.64%) (Figures 5, 6).

3.3.2 Neoadjuvant treatment cycle

The studies were divided into two-cycle, three-cycle, and four-cycle groups according to the number of neoadjuvant therapy cycles. Thirty-seven studies reported on treatment cycles, including 12 studies in the two-cycle group, 7 studies in the three-cycle group, 4 studies in the four-cycle group, and the rest of the studies could not be subgrouped. All studies in the four-cycle group were chemoimmunotherapy. Analyses will be stratified according to the neoadjuvant regimen.

For neoadjuvant immunotherapy alone, pCR, SAEs, R0 resection, and surgery resection were analyzed by the fixed effect model, and other indicators were analyzed by the random effect model. The ES for MPR in the two-cycle and three-cycle groups were 0.28 (95% CI 0.18-0.38) and 0.20 (95% CI 0.11-0.31); for pCR, they were 0.06 (95% CI 0.04-0.10) and 0.08 (95% CI 0.02-0.16); for ORR, they were 0.14 (95% CI 0.05-0.24) and 0.12 (95% CI 0.05-0.22); for TRAEs, they were 0.45 (95% CI 0.28-0.62) and 0.35 (95% CI 0.21-0.50); for SAEs, they were 0.10 (95% CI 0.07-0.14) and 0.02 (95% CI 0.00-0.07); for surgery resection, they were 0.91 (95% CI 0.88-0.94) and 0.93 (95% CI 0.85-0.98); for R0 resection, they were 0.94 (95% CI 0.91-0.97) and 0.98 (95% CI 0.92-1.00); for surgical complication, they were 0.23 (95% CI 0.08-0.44) and 0.34 (95% CI 0.23-0.47); and for conversion to thoracotomy, they were 0.06 (95% CI 0.00-0.17) and 0.05 (95% CI 0.00-0.24), respectively (Figure 7).

For neoadjuvant chemoimmunotherapy, the R0 resection rate in the two-cycle group was analyzed using the fixed effects model. The random-effects model was adopted for the rest of the indicators. No reports of conversion to thoracotomy were found in the four-cycle group, and no reports of surgical complication rates were found in the two-cycle group or the four-cycle group. Consequently, the ES for conversion to thoracotomy in the two-cycle and three-cycle group were 0.02 (95% CI 0.00-0.09) and 0.05 (95% CI 0.01-0.11), respectively; and the ES for surgical complication in the three-cycle group was 0.27 (95% CI 0.11-0.47). The ES for MPR in the two-cycle, three-cycle and four-cycle groups were 0.50 (95% CI 0.25-0.75), 0.70 (95% CI 0.53-0.84) and 0.36 (95% CI 0.27-0.46); for pCR, they were 0.32 (95% CI 0.19-0.46), 0.49 (95% CI 0.30-0.68) and 0.18 (95% CI 0.13-0.24); for ORR, they were 0.71 (95% CI 0.47-0.90), 0.72 (95% CI 0.60-0.82) and 0.52 (95% CI 0.37-0.67);for TRAEs, they were 0.95 (95% CI 0.89-0.99), 0.85 (95% CI 0.77-0.92) and 0.95 (95% CI 0.94-0.97); for SAEs, they were 0.34 (95% CI 0.01-0.83), 0.27 (95% CI 0.19-0.36) and 0.37 (95% CI 0.33-0.42); for surgery resection, they were 0.94 (95% CI 0.76-1.00), 0.93 (95% CI 0.85-0.99) and 0.83 (95% CI 0.76-0.90); for R0 resection, they were 0.99 (95% CI 0.95-1.00) (Supplementary Figure 1D), 0.97 (95% CI 0.88-1.00) and 0.95 (95% CI 0.91-0.98), respectively (Figures 8, 9).

3.4 Sensitivity analysis and publication bias

Upon examining the effects of different studies on heterogeneity within the subgroups, the heterogeneity of the

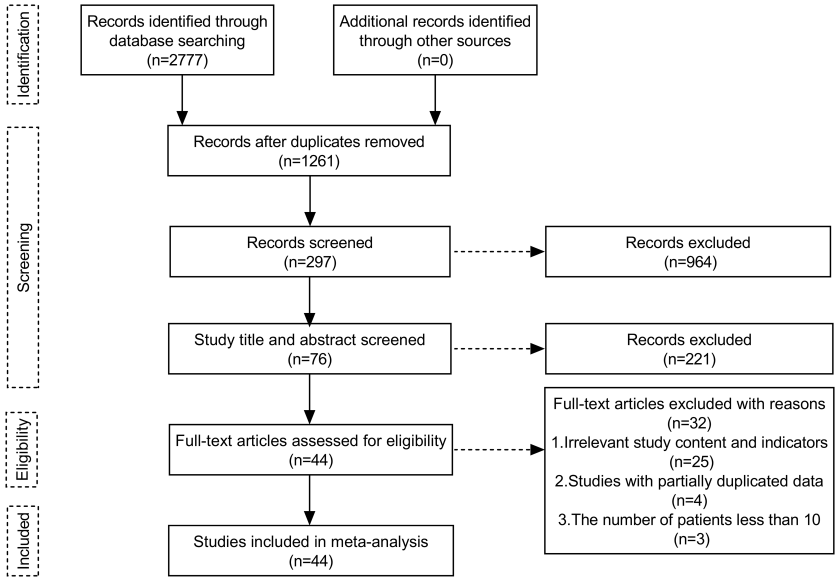


FIGURE 2
Flow chart of the meta-analysis search strategy.

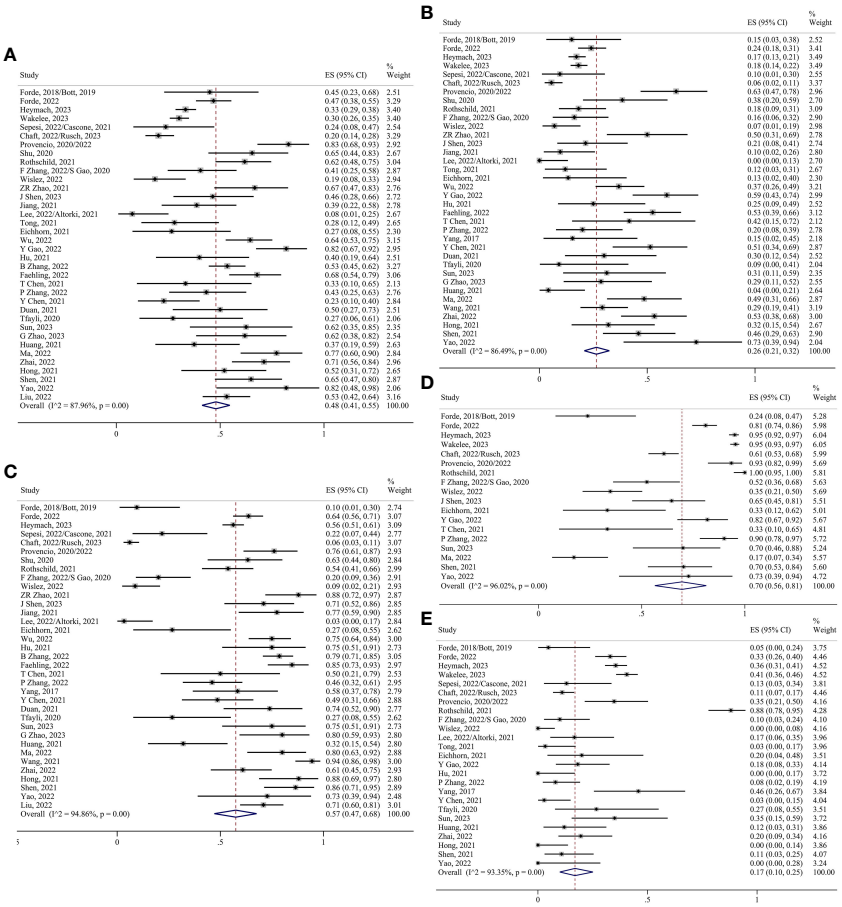


FIGURE 3
Forest plots. **(A)** MPR. **(B)** pCR. **(C)** Radiological response. **(D)** TRAEs. **(E)** SAEs.

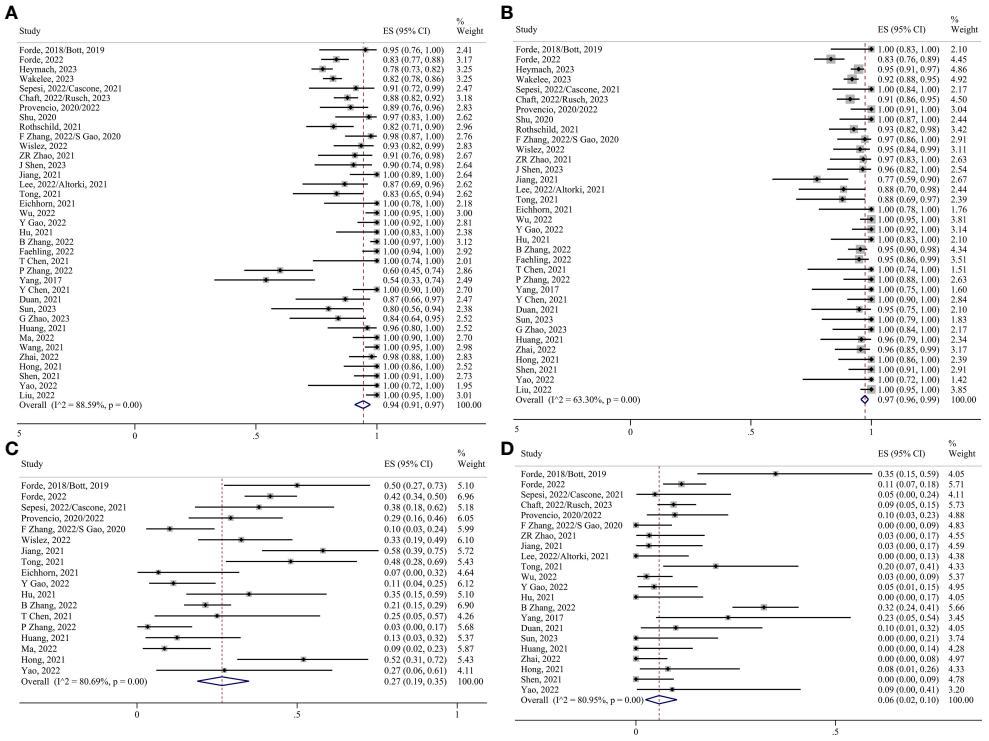


FIGURE 4
Forest plots. **(A)** Surgical resection. **(B)** R0 resection. **(C)** Surgical complications. **(D)** Conversion to thoracotomy.

chemoimmunotherapy group converted to thoracotomy after deletion of Zhang et al., 2022 (28) was significantly reduced (I^2=52.03%, P=0.01), with an effect size of 0.04 (95% CI 0.01-0.07). For the neoadjuvant cycle subgroup analysis, the study by Rothschild et al., 2021 (15) was excluded from the chemoimmunotherapy group. The ES of the TRAEs in the two-cycle group changed to 0.70 (95% CI 0.53–0.84), and that of the SAEs changed to 0.16 (95% CI 0.01–0.44). Egger's tests (Supplementary Table 2) and funnel plots (Supplementary Figure 2, 3) were performed separately within the different subgroups, and there was no marked publication bias.

4 Discussion

This study demonstrated the feasibility and safety of the preoperative application of ICIs. Subgroup and meta-regression analyses (Supplementary Table 3) showed that chemoimmunotherapy increased tumor MPR and pCR rates by 29% and 27% compared with ICIs alone (P= 0.001; P=0.002), respectively; while the ORR increased significantly (P < 0.001). However, the incidence of TRAEs and SAEs increased significantly (P = 0.009; P=0.034). All phase III large-sample clinical trials on chemoimmunotherapy reported on SAEs, which were dominated by Neutrophil count decreased, neutropenia, anemia, leukopenia, and Platelet count decreased (8–10). None of them found significant differences in TRAEs and SAEs between chemoimmunotherapy and chemotherapy. This suggests that the increased incidence of TRAEs and SAEs with chemoimmunotherapy compared to immunotherapy alone may be related to chemotherapy. AEGEAN (9) and KEYNOTE-671 (10) reported 7 and 4 deaths during the neoadjuvant therapy phase, respectively, with the main causes of death being immune-mediated lung disease, interstitial lung disease and pneumonia. Despite the low mortality associated with chemoimmunotherapy, physicians still need to be vigilant for the occurrence of immune-related diseases, especially immune-mediated lung disease. In terms of surgery, combination chemotherapy did not significantly affect surgical resection, R0 resection, conversion to thoracotomy, or surgical complications.

Among the included studies, CheckMate816 and NADIM II reported data related to circulating tumor DNA (ctDNA). There was a ctDNA clearance rate of 56% in the neoadjuvant chemotherapy group (8). Both reported that higher clearance was associated with a higher rate of pCR and longer event-free survival (EFS) (8, 16). Although follow-up data, such as five-year overall survival (OS), have not been reported, higher ctDNA clearance rates are beneficial for predicting the long-term risks of neoadjuvant immunotherapy (51, 52).

In terms of treatment cycles, for chemoimmunotherapy, MPR and pCR were improved by 20% and 17% in the three-cycle group compared with the two-cycle group, respectively, but there was no increase in the MPR or pCR for the four-cycle group. Similarly, the neoSCORE study reported a 14.5% increase in the MPR and a 4.9% increase in the pCR in a three-cycle group compared with the two-cycle group with preoperative sintilimab combined with platinum-

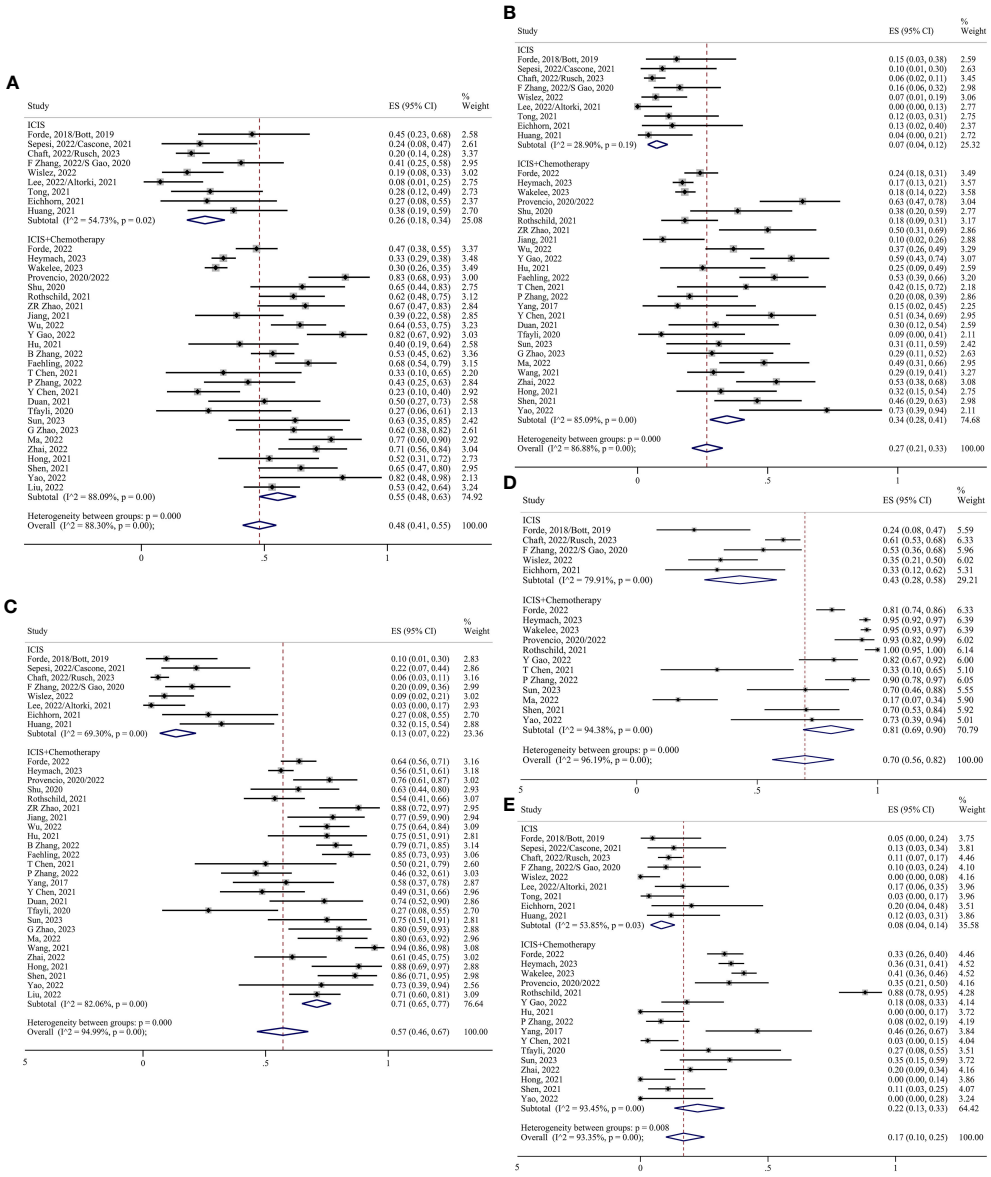


FIGURE 5
Forest plots of subgroups based on the treatment regimens. **(A)** MPR. **(B)** pCR. **(C)** Radiological response. **(D)** TRAEs. **(E)** SAEs.

based dual chemotherapy regimens (53). After sensitivity analysis, the incidence of TRAEs and SAEs in the chemoimmunotherapy group increased progressively with the number of treatment cycles, but none of them was statistically significant. Therefore, three cycles of neoadjuvant chemoimmunotherapy have an optimal efficacy and safety profile compared to two and four cycles. In the ICIs-alone group, the increase of treatment cycles had little effect on the rate of MPR and pCR, the incidence of TRAEs and SAEs. There was no significant negative effect of the increase of neoadjuvant cycles on the rate of surgical resection, the incidence of surgical complications, rate of R0 resection, or rate of conversion to thoracotomy in both treatment regimens.

The preoperative application of ICIs is not limited to combination chemotherapy. Compared with neoadjuvant chemotherapy, neoadjuvant radiation therapy combined with chemotherapy for lung cancer does not produce long-term benefits in terms of EFS and OS and has more significant side effects (54, 55). However, Lee et al., 2022 (25) and Altorki et al., 2021 (26) noted that the MPR rate in groups treated with durvalumab combined with SBRT was 53.3%, which was significantly higher than that in groups treated with durvalumab alone. The efficacy of nivolumab in combination with ipilimumab has been confirmed in NEOSTAR; however, there are few relevant studies on this treatment, and further analyses in large-sample studies are required.

This meta-analysis clarified the safety and feasibility of different neoadjuvant regimens and cycles at the present stage and provides a reference for the selection of regimens and cycles. However, there were several limitations. First, only three phase III large-sample clinical trials and a large number of phase II single-arm studies were included. The conclusions of the study are therefore

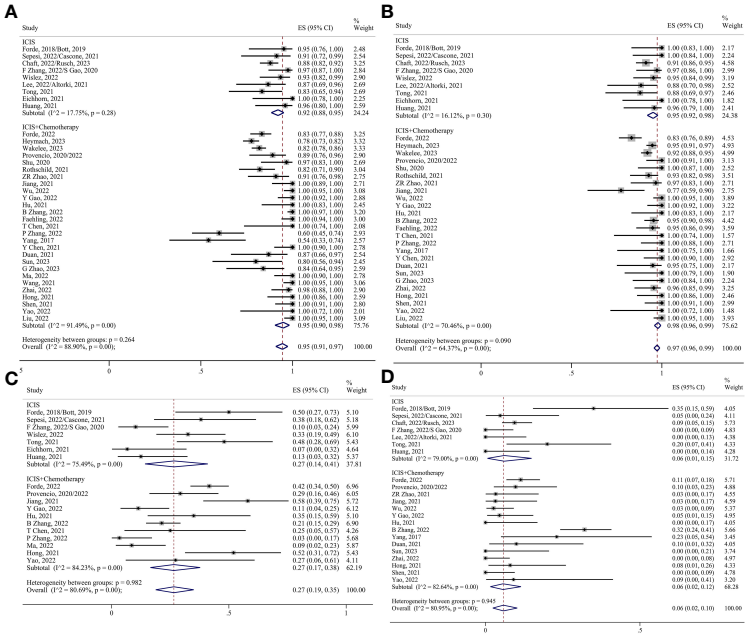


FIGURE 6
Forest plots of subgroups based on the treatment regimens. **(A)** Surgical resection. **(B)** R0 resection. **(C)** Surgical complications. **(D)** Conversion to thoracotomy.

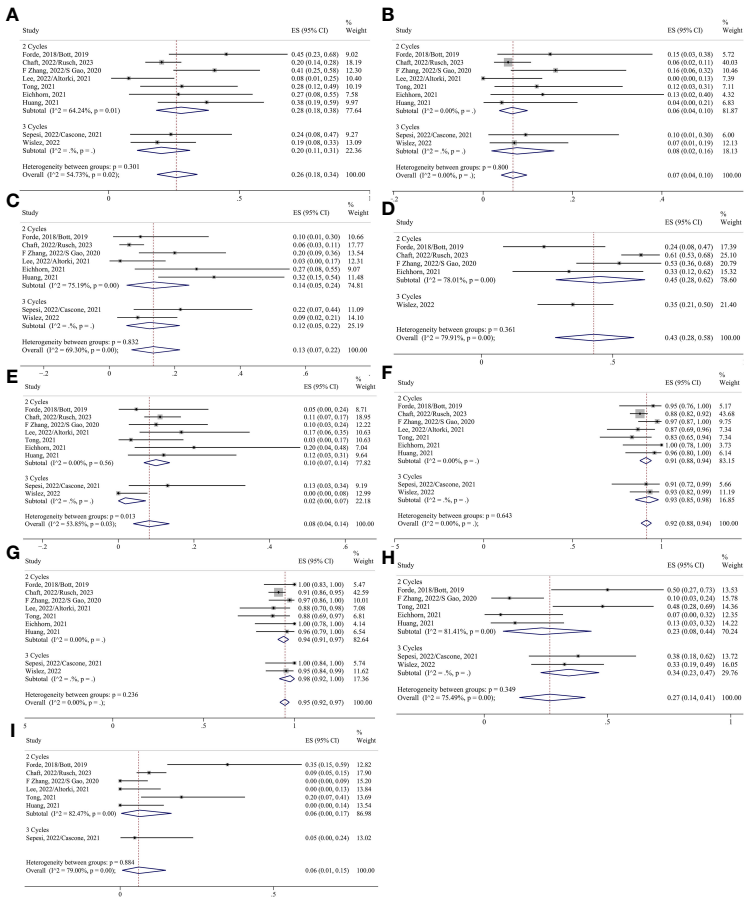


FIGURE 7
Forest plots of subgroups based on immunotherapy-alone treatment cycles. **(A)** MPR. **(B)** pCR. **(C)** Radiological response. **(D)** TRAEs. **(E)** SAEs. **(F)** Surgical resection. **(G)** R0 resection. **(H)** Surgical complications. **(I)** Conversion to thoracotomy.

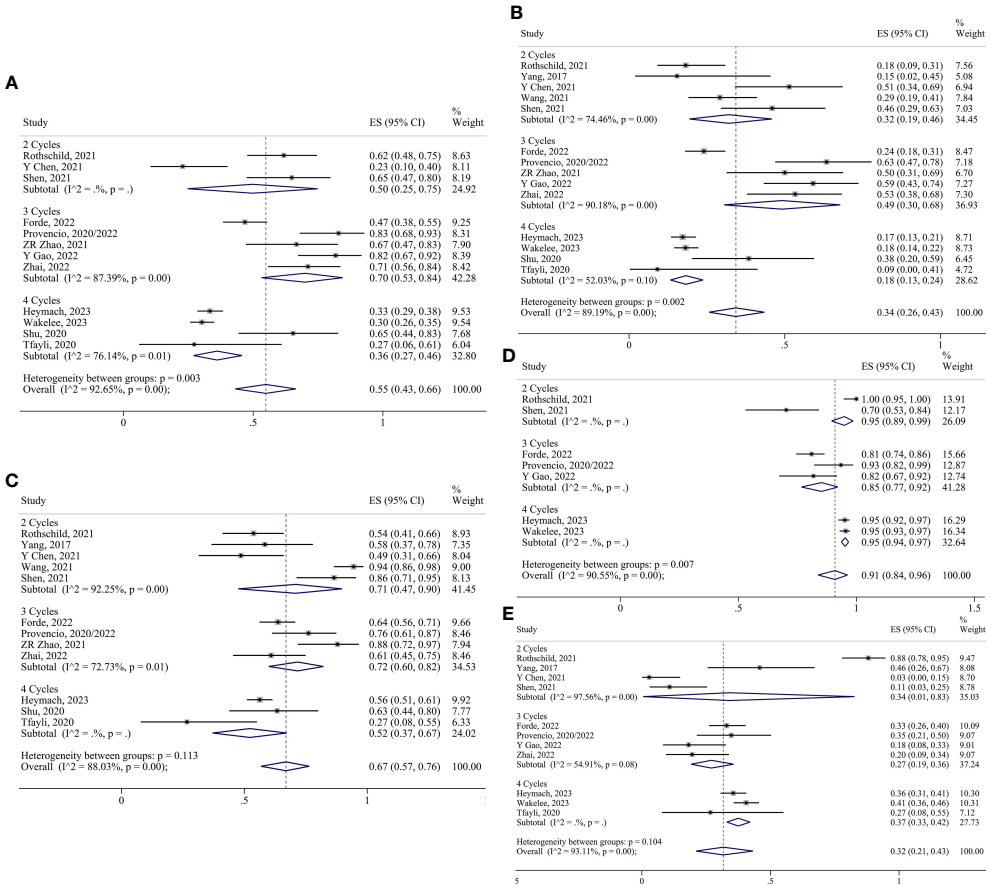


FIGURE 8
Forest plots of subgroups based on chemoimmunotherapy treatment cycles. **(A)** MPR. **(B)** pCR. **(C)** Radiological response. **(D)** TRAEs. **(E)** SAEs.

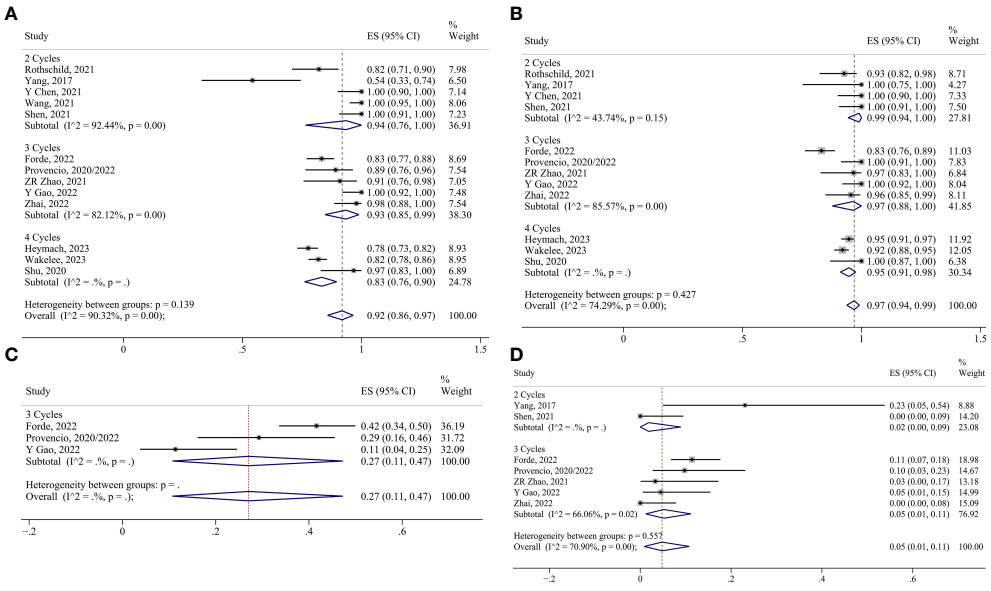


FIGURE 9
Forest plots of subgroups based on chemoimmunotherapy treatment cycles. **(A)** Surgical resection. **(B)** R0 resection. **(C)** Surgical complications. **(D)** Conversion to thoracotomy.

unrepresentative and inaccurate. Second, the heterogeneity of the outcomes was strong after pooling. The heterogeneity of some of the results decreased insignificantly after the subgroup analysis, and there was a lack of long-term follow-up data. Third, studies at this stage have mainly focused on ICIs alone and chemoimmunotherapy. We look forward to clinical studies on ICIs combined with radiotherapy, targeted therapy, or dual immunotherapy to determine the optimal neoadjuvant treatment strategy for lung cancer.

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

HW: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. SL: Data curation, Formal analysis, Investigation, Writing – review & editing. YY: Data curation, Validation, Writing – review & editing. YH: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Liaoning Province People's Livelihood Science and Technology Program joint program project [grant numbers: 2021JH2/10300088].

Acknowledgments

The authors thank the Liaoning Province People's Livelihood Science and Technology Program Joint Program for funding.

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/fonc.2024.1276549/full#supplementary-material

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660

2. GBD 2019 Respiratory. Tract Cancers Collaborators Global, regional, and national burden of respiratory tract cancers and associated risk factors from 1990 to 2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Respir Med* (2021) 9(9):1030–49. doi: 10.1016/s2213-2600(21)00164-8

3. Amaria RN, Postow M, Burton EM, Tezlaff MT, Ross MI, Torres-Cabala C, et al. Neoadjuvant relatlimab and nivolumab in resectable melanoma. *Nature* (2022) 611 (7934):155–60. doi: 10.1038/s41586-022-05368-8

4. Marron TU, Fiel MI, Hamon P, Fiaschi N, Kim E, Ward SC, et al. Neoadjuvant cemiplimab for resectable hepatocellular carcinoma: a single-arm, open-label, phase 2 trial. *Lancet Gastroenterol Hepatol* (2022) 7(3):219–29. doi: 10.1016/s2468-1253(21)00385-x

5. Bott MJ, Yang SC, Park BJ, Adusumilli PS, Rusch VW, Isbell JM, et al. Initial results of pulmonary resection after neoadjuvant nivolumab in patients with resectable non-small cell lung cancer. *J Thorac Cardiovasc Surg* (2019) 158(1):269–76. doi: 10.1016/j.jtcvs.2018.11.124

6. Forde PM, Chaft JE, Smith KN, Anagnostou V, Cottrell TR, Hellmann MD, et al. Neoadjuvant PD-1 blockade in resectable lung cancer. *N Engl J Med* (2018) 378 (21):1976–86. doi: 10.1056/NEJMoa1716078

7. Dong P, Yan Y, Yang L, Wu D, Wang H, Lv Y, et al. Neoadjuvant immunotherapy improves treatment for early resectable non-small-cell lung cancer: A systematic review and meta-analysis. *J Oncol* (2022) 2022:2085267. doi: 10.1155/2022/2085267

8. Forde PM, Spicer J, Lu S, Provencio M, Mitsudomi T, Awad MM, et al. Neoadjuvant nivolumab plus chemotherapy in resectable lung cancer. *N Engl J Med* (2022) 386(21):1973–85. doi: 10.1056/NEJMoa2202170

9. Heymach JV, Harpole D, Mitsudomi T, Taube JM, Galffy G, Hochmair M, et al. Perioperative durvalumab for resectable non-small-cell lung cancer. *N Engl J Med* (2023) 389(18):1672–84. doi: 10.1056/NEJMoa2304875

10. Wakelee H, Liberman M, Kato T, Tsuboi M, Lee SH, Gao S, et al. Perioperative pembrolizumab for early-stage non-small-cell lung cancer. *N Engl J Med* (2023) 389 (6):491–503. doi: 10.1056/NEJMoa2302983

11. Sepesi B, Zhou N, William WN, Lin HY, Leung CH, Weissferdt A, et al. Surgical outcomes after nivolumab or nivolumab with ipilimumab treatment in patients with non–small cell lung cancer. *J Thorac Cardiovasc Surg* (2022) 164(5):1327–37. doi: 10.1016/j.jtcvs.2022.01.019

12. Cascone T, William WN Jr, Weissferdt A, Leung CH, Lin HY, Pataer A, et al. Neoadjuvant nivolumab or nivolumab plus ipilimumab in operable non-small cell lung cancer: the phase 2 randomized NEOSTAR trial. *Nat Med* (2021) 27(3):504–14. doi: 10.1038/s41591-020-01224-2

13. Chaft JE, Oezkan F, Kris MG, Bunn PA, Wistuba II, Kwiatkowski DJ, et al. Neoadjuvant atezolizumab for resectable non-small cell lung cancer: an open-label, single-arm phase II trial. *Nat Med* (2022) 28(10):2155–61. doi: 10.1038/s41591-022-01962-5

14. Rusch VW, Nicholas A, Patterson GA, Waqar SN, Toloza EM, Haura EB, et al. Surgical results of the Lung Cancer Mutation Consortium 3 trial: A phase II multicenter single-arm study to investigate the efficacy and safety of atezolizumab as neoadjuvant therapy in patients with stages IB-select IIIB resectable non–small cell lung

cancer. *J Thorac Cardiovasc Surg* (2023) 165(3):828–839.e5. doi: 10.1016/j.jtcvs.2022.10.007

15. Provencio M, Nadal E, Insa A, García-Campelo MR, Casal-Rubio J, Dómine M, et al. Neoadjuvant chemotherapy and nivolumab in resectable non-small-cell lung cancer (NADIM): an open-label, multicentre, single-arm, phase 2 trial. *Lancet Oncol* (2020) 21(11):1413–22. doi: 10.1016/S1470-2045(20)30453-8

16. Provencio M, Serna-Blasco R, Nadal E, Insa A, García-Campelo MR, Casal Rubio J, et al. Overall Survival and Biomarker Analysis of Neoadjuvant Nivolumab Plus Chemotherapy in Operable Stage IIIA Non-Small-Cell Lung Cancer (NADIM phase II trial). *J Clin Oncol* (2022) 40(25):2924–33. doi: 10.1200/JCO.21.02660

17. Shu CA, Gainor JF, Awad MM, Chiuzan C, Grigg CM, Pabani A, et al. Neoadjuvant atezolizumab and chemotherapy in patients with resectable non-small-cell lung cancer: an open-label, multicentre, single-arm, phase 2 trial. *Lancet Oncol* (2020) 21(6):786–95. doi: 10.1016/s1470-2045(20)30140-6

18. Rothschild SI, Zippelius A, Eboulet EI, Prince SS, Betticher D, Bettini A, et al. SAKK 16/14: durvalumab in addition to neoadjuvant chemotherapy in patients with stage IIIA(N2) non–small-cell lung cancer—A multicenter single-arm phase II trial. *J Clin Oncol* (2021) 39(26):2872–80. doi: 10.1200/JCO.21.00276

19. Zhang F, Guo W, Zhou B, Wang S, Li N, Bin Q, et al. Three-year follow-up of neoadjuvant programmed cell death protein-1 inhibitor (Sintilimab) in NSCLC. *J Thorac Oncol* (2022) 17(7):909–20. doi: 10.1016/j.jtho.2022.04.012

20. Gao S, Li N, Gao S, Xue Q, Ying J, Wang S, et al. Neoadjuvant PD-1 inhibitor (Sintilimab) in NSCLC. *J Thorac Oncol* (2020) 15(5):816–26. doi: 10.1016/j.jtho.2020.01.017

21. Wislez M, Mazieres J, Lavole A, Zalcman G, Carre O, Egenod T, et al. Neoadjuvant durvalumab for resectable non-small-cell lung cancer (NSCLC): results from a multicenter study (IFCT-1601 IONESCO). *J Immunother Cancer* (2022) 10(10): e005636. doi: 10.1136/jitc-2022-005636

22. Zhao ZR, Yang CP, Chen S, Yu H, Lin YB, Lin YB, et al. Phase 2 trial of neoadjuvant toripalimab with chemotherapy for resectable stage III non-small-cell lung cancer. *Oncoimmunology* (2021) 10(1):1996000. doi: 10.1080/2162402X.2021.1996000

23. Shen J, Gu L, Qi Y, Yao Y, Lu S, Chen Z. Real-world outcomes of immunotherapy-based neoadjuvant therapy in resectable non-small cell lung cancer. *Front Immunol* (2023) 14:1268251. doi: 10.3389/fimmu.2023.1268251

24. Jiang L, Huang J, Jiang S, Rong W, Shen Y, Li C, et al. The surgical perspective in neoadjuvant immunotherapy for resectable non-small cell lung cancer. *Cancer Immunol Immunother* (2021) 70(8):2313–21. doi: 10.1007/s00262-021-02847-1

25. Lee B, Mynard N, Nasar A, Villena-Vargas J, Chow O, Harrison S, et al. Surgical resection after neoadjuvant durvalumab and radiation is feasible and safe in non-small cell lung cancer: Results from a randomized trial. *J Thorac Cardiovasc Surg* (2022) 165 (1):327–334.e2. doi: 10.1016/j.jtcvs.2022.07.017

26. Altorki NK, McGraw TE, Borczuk AC, Saxena A, Port JL, Stiles BM, et al. Neoadjuvant durvalumab with or without stereotactic body radiotherapy in patients with early-stage non-small-cell lung cancer: a single-centre, randomised phase 2 trial. *Lancet Oncol* (2021) 22(6):824–35. doi: 10.1016/s1470-2045(21)00149-2

27. Tong BC, Gu L, Wang X, Wigle DA, Phillips JD, Harpole DH Jr, et al. Perioperative outcomes of pulmonary resection after neoadjuvant pembrolizumab in patients with non-small cell lung cancer. *J Thorac Cardiovasc Surg* (2022) 163(2):427–36. doi: 10.1016/j.jtcvs.2021.02.099

28. Eichhorn F, Klotz LV, Kriegsmann M, Bischoff H, Schneider MA, Muley T, et al. Neoadjuvant anti-programmed death-1 immunotherapy by pembrolizumab in resectable non-small cell lung cancer: First clinical experience. *Lung Cancer* (2021) 153:150–7. doi: 10.1016/j.lungcan.2021.01.018

29. Wu J, Hou L, E H, Zhao Y, Yu X, Xu L, et al. Real-world clinical outcomes of neoadjuvant immunotherapy combined with chemotherapy in resectable non-small cell lung cancer. *Lung Cancer* (2022) 165:115–23. doi: 10.1016/j.lungcan.2022.01.019

30. Gao Y, Jiang J, Xiao D, Zhou Y, Chen Y, Yang H, et al. Robotic-assisted thoracic surgery following neoadjuvant chemoimmunotherapy in patients with stage III non-small cell lung cancer: A real-world prospective cohort study. *Front Oncol* (2022) 12:96954. doi: 10.3389/fonc.2022.96954

31. Hu Y, Ren SY, Wang RY, Zeng C, Li JN, Xiao P, et al. Surgical outcomes after neoadjuvant chemoimmunotherapy for resectable non-small cell lung cancer. *Front Oncol* (2021) 11:684070. doi: 10.3389/fonc.2021.684070

32. Zhang B, Xiao Q, Xiao H, Wu J, Yang D, Tang J, et al. Perioperative outcomes of video-assisted thoracoscopic surgery versus open thoracotomy after neoadjuvant chemoimmunotherapy in resectable NSCLC. *Front Oncol* (2022) 12:858189. doi: 10.3389/fonc.2022.858189

33. Faehling M, Witte H, Sebastian M, Ulmer M, Sätzler R, Steinestel K, et al. Real-world multicentre analysis of neoadjuvant immunotherapy and chemotherapy in localized or oligometastatic non-small cell lung cancer (KOMPASSneoOP). *Ther Adv Med Oncol* (2022) 14:17588359221085333. doi: 10.1177/17588359221085333

34. Chen T, Ning J, Campisi A, Dell'Amore A, Ciarrocchi AP, Li Z, et al. Neoadjuvant PD-1 inhibitors and chemotherapy for locally advanced NSCLC: A retrospective study. *Ann Thorac Surg* (2022) 113(3):993–9. doi: 10.1016/j.athoracsur.2021.03.041

35. Zhang P, Dai J, Sun F, Xia H, He W, Duan L, et al. Neoadjuvant sintilimab and chemotherapy for resectable stage IIIA non-small cell lung cancer. *Ann Thorac Surg* (2022) 114(3):949–58. doi: 10.1016/j.athoracsur.2022.01.039

36. Yang CJ, McSherry F, Mayne NR, Wang X, Berry MF, Tong B, et al. Surgical outcomes after neoadjuvant chemotherapy and ipilimumab for non-small cell lung cancer. *Ann Thorac Surg* (2018) 105(3):924–9. doi: 10.1016/j.athoracsur.2017.09.030

37. Chen Y, Yan B, Xu F, Hui Z, Zhao G, Liu J, et al. Neoadjuvant chemoimmunotherapy in resectable stage IIIA/IIIB non-small cell lung cancer. *Transl Lung Cancer Res* (2021) 10(5):2193–204. doi: 10.21037/tlcr-21-329

38. Duan H, Wang T, Luo Z, Tong L, Dong X, Zhang Y, et al. Neoadjuvant programmed cell death protein 1 inhibitors combined with chemotherapy in resectable non-small cell lung cancer: An open-label, multicenter, single-arm study. *Transl Lung Cancer Res* (2021) 10(2):1020–8. doi: 10.21037/tlcr-21-130

39. Tfayli A, Al Assaad M, Fakhri G, Akel R, Atwi H, Ghanem H, et al. Neoadjuvant chemotherapy and Avelumab in early stage resectable nonsmall cell lung cancer. *Cancer Med* (2020) 9(22):8406–11. doi: 10.1002/cam4.3456

40. Sun C, Liu Y, Zhang P, Wang X, Xu Y, Lin X, et al. Interim analysis of the efficiency and safety of neoadjuvant PD-1 inhibitor (sintilimab) combined with chemotherapy (nab-paclitaxel and carboplatin) in potentially resectable stage IIIA/IIIB non-small cell lung cancer: a single-arm, phase 2 trial. *J Cancer Res Clin Oncol* (2023) 149(2):819–31. doi: 10.1007/s00432-021-03896-w

41. Zhao G, Zhang H, Xu F, Lu C, Zhu Q, Grossi F, et al. Neoadjuvant pembrolizumab and chemotherapy in resectable clinical stage III non-small-cell lung cancer: a retrospective cohort study. *Transl Lung Cancer Res* (2023) 12(1):141–9. doi: 10.21037/tlcr-22-871

42. Huang Z, Wu Z, Qin Y, Zhao Y, Xuan Y, Qiu T, et al. Perioperative safety and feasibility outcomes of stage IIIA-N2 non-small cell lung cancer following neoadjuvant immunotherapy or neoadjuvant chemotherapy: A retrospective study. *Ann Transl Med* (2021) 9(8):685. doi: 10.21037/atm-21-1141

43. Ma D, Xu Y, Qin Y, Li S, Li J, Jiang Y, et al. Neoadjuvant immunotherapy followed by surgery with curative intent in 35 patients with advanced NSCLC: the retrospective experiences of a multidisciplinary team. *Ann Transl Med* (2022) 10 (10):609. doi: 10.21037/atm-22-2271

44. Wang J, Li J, Cai L, Chen S, Jiang Y. The safety and efficacy of neoadjuvant programmed death 1 inhibitor therapy with surgical resection in stage IIIA non-small cell lung cancer. *Ann Transl Med* (2021) 9(6):486. doi: 10.21037/atm-21-670

45. Zhai H, Li W, Jiang K, Zhi Y, Yang Z. Neoadjuvant nivolumab and chemotherapy in patients with locally advanced non-small cell lung cancer: A retrospective study. *Cancer Manag Res* (2022) 14:515–24. doi: 10.2147/CMAR.S344343

46. Hong T, Sun T, Zhang M, Liu X, Yuan Y, Dolo PR, et al. Surgical perspective in neoadjuvant chemoimmunotherapy for stage II-III non-small cell lung cancer. *Thorac Cancer* (2021) 12(20):2796–802. doi: 10.1111/1759-7714.14127

47. Shen D, Wang J, Wu J, Chen S, Li J, Liu J, et al. Neoadjuvant pembrolizumab with chemotherapy for the treatment of stage IIB-IIIB resectable lung squamous cell carcinoma. *J Thorac Dis* (2021) 13(3):1760–8. doi: 10.21037/jtd-21-103

48. Yao Y, Tang D, Gao W, Zhang H. Neoadjuvant immuno-chemotherapy: A new perspective for stage III NSCLC? *Front Surg* (2022) 9:843987. doi: 10.3389/fsurg.2022.843987

49. Liu Z, Gao Z, Zhang M, Wang X, Gong J, Jiang S, et al. Real-world effectiveness and prognostic factors analysis of stages I–III non-small cell lung cancer following neoadjuvant chemo-immunotherapy or neoadjuvant chemotherapy. *Ann Thorac Cardiovasc Surg* (2022) 28(2):111–20. doi: 10.5761/atcs.oa.21-00143

50. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* (2009) 45(2):228–47. doi: 10.1016/j.ejca.2008.10.026

51. Gale D, Heider K, Ruiz-Valdepenas A, Hackinger S, Perry M, Marsico G, et al. Residual ctDNA after treatment predicts early relapse in patients with early-stage non-small cell lung cancer. *Ann Oncol* (2022) 33(5):500–10. doi: 10.1016/j.annonc.2022.02.007

52. Xia L, Mei J, Kang R, Deng S, Chen Y, Yang Y, et al. Perioperative ctDNA-based molecular residual disease detection for non-small cell lung cancer: A prospective multicenter cohort study (LUNGCA-1). *Clin Cancer Res* (2022) 28(15):3308–17. doi: 10.1158/1078-0432.Ccr-21-3044

53. Qiu FM, Fan JQ, Shao M, Yao J, Zhao LF, Zhu L, et al. Two cycles versus three cycles of neoadjuvant sintilimab plus platinum-doublet chemotherapy in patients with resectable non-small-cell lung cancer (neoSCORE): A randomized, single center, two-arm phase II trial. *J Clin Oncol* (2022) 40(16_suppl):8500–0. doi: 10.1200/JCO.2022.40.16_suppl.8500

54. Pless M, Stupp R, Ris HB, Stahel RA, Weder W, Thierstein S, et al. Induction chemoradiation in stage IIIA/N2 non-small-cell lung cancer: a phase 3 randomised trial. *Lancet* (2015) 386(9998):1049–56. doi: 10.1016/s0140-6736(15)60294-x

55. Stupp R, Mayer M, Kann R, Weder W, Zouhair A, Betticher DC, et al. Neoadjuvant chemotherapy and radiotherapy followed by surgery in selected patients with stage IIIB non-small-cell lung cancer: a multicentre phase II trial. *Lancet Oncol* (2009) 10(8):785–93. doi: 10.1016/s1470-2045(09)70172-x

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 18 January 2024
DOI 10.3389/fmolb.2023.1340917

Check for updates

OPEN ACCESS

EDITED BY
Li Chen,
Huazhong University of Science and Technology, China

REVIEWED BY
Min Li,
Western Theater General Hospital, China
Yawei Hua,
Henan Provincial Cancer Hospital, China
Zhiheng Lin,
Shandong University of Traditional Chinese Medicine, China
Dongdong Zhang,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Bailin Zhang,
zhangbl@cicams.ac.cn
Gao Jidong,
ab168@cicams.ac.cn

RECEIVED 19 November 2023
ACCEPTED 29 December 2023
PUBLISHED 18 January 2024

CITATION
Wang Y, Yi K, Chen B, Zhang B and Jidong G (2024), Elucidating the susceptibility to breast cancer: an in-depth proteomic and transcriptomic investigation into novel potential plasma protein biomarkers.
Front. Mol. Biosci. 10:1340917.
doi: 10.3389/fmolb.2023.1340917

COPYRIGHT
© 2024 Wang, Yi, Chen, Zhang and Jidong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Elucidating the susceptibility to breast cancer: an in-depth proteomic and transcriptomic investigation into novel potential plasma protein biomarkers

Yang Wang[1], Kexin Yi[1], Baoyue Chen[2], Bailin Zhang[1]* and Gao Jidong[1,3]*

[1]Department of Breast Surgical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2]Department of General Surgery, Beijing Puren Hospital, Beijing, China, [3]Department of Breast Surgical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital and Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen, China

Objectives: This study aimed to identify plasma proteins that are associated with and causative of breast cancer through Proteome and Transcriptome-wide association studies combining Mendelian Randomization.

Methods: Utilizing high-throughput datasets, we designed a two-phase analytical framework aimed at identifying novel plasma proteins that are both associated with and causative of breast cancer. Initially, we conducted Proteome/Transcriptome-wide association studies (P/TWAS) to identify plasma proteins with significant associations. Subsequently, Mendelian Randomization was employed to ascertain the causation. The validity and robustness of our findings were further reinforced through external validation and various sensitivity analyses, including Bayesian colocalization, Steiger filtering, heterogeneity and pleiotropy. Additionally, we performed functional enrichment analysis of the identified proteins to better understand their roles in breast cancer and to assess their potential as druggable targets.

Results: We identified 5 plasma proteins demonstrating strong associations and causative links with breast cancer. Specifically, PEX14 (OR = 1.201, $p = 0.016$) and CTSF (OR = 1.114, $p < 0.001$) both displayed positive and causal association with breast cancer. In contrast, SNUPN (OR = 0.905, $p < 0.001$), CSK (OR = 0.962, $p = 0.038$), and PARK7 (OR = 0.954, $p < 0.001$) were negatively associated with the disease. For the ER-positive subtype, 3 plasma proteins were identified, with CSK and CTSF exhibiting consistent trends, while GDI2 (OR = 0.920, $p < 0.001$) was distinct to this subtype. In ER-negative subtype, PEX14 (OR = 1.645, $p < 0.001$) stood out as the sole protein, even showing a stronger causal effect compared to breast cancer. These associations were robustly supported by colocalization and sensitivity analyses.

Conclusion: Integrating multiple data dimensions, our study successfully pinpointed plasma proteins significantly associated with and causative of breast cancer, offering valuable insights for future research and potential new biomarkers and therapeutic targets.

KEYWORDS

proteome-wide association study, transcriptome-wide association study, plasma proteins, mendelian randomization, breast cancer

1 Introduction

In 2020, a concerning 2.3 million women were diagnosed with breast cancer, establishing it as the most common cancer among women worldwide (Sung et al., 2021). This high prevalence underscores the urgency for ongoing research; however, despite significant efforts, the precise causes of breast cancer remain elusive. The disease is marked by a wide range of biological characteristics, including diverse histological and molecular features (Prat and Perou, 2011). Among these, the estrogen receptor (ER) status stands out as a crucial biomarker, significantly influencing treatment strategies such as endocrine therapy for ER-positive breast cancers (Trayes and Cokenakes, 2021). In addition to tissue-specific protein markers, the study of proteins in circulating plasma, often found due to cellular leakage or active secretion (Anderson and Anderson, 2002), is increasingly important. Due to the ease of detection and reproducibility of plasma proteins, these proteins are suitable for biomarkers and potential therapeutic targets (Suhre et al., 2021). Recent studies have highlighted the significant relationship between a variety of circulating proteins and breast cancer, thereby providing crucial insights into the disease's prognosis (Key et al., 2010; Christopoulos et al., 2015; Rosendahl et al., 2021; Veyssière et al., 2022; Mälarstig et al., 2023). The identification of these proteins as potential biomarkers has opened new avenues for early detection and personalized medicine in breast cancer, emphasizing the importance of understanding the complex biological interactions and pathways involved in cancer progression.

Genome-wide association studies (GWAS) have been instrumental in identifying nearly 200 genetic loci associated with breast cancer, revealing insights into genetic predispositions (Michailidou et al., 2017; Shu et al., 2020; Zhang et al., 2020; Gudjonsson et al., 2022). These discoveries underscore the importance of genetic factors in breast cancer susceptibility. Particularly, SNPs located within a 500 Kb range of the transcription start sites of protein-coding genes, known as cis-acting quantitative trait loci (cis-QTLs). Among these, protein Quantity Trait Loci (pQTLs) are crucial for regulating protein levels and are valuable tools for research (Sun et al., 2018). Utilizing pQTL as genetic proxies allow us to make a deeper exploration of the role of plasma proteins in breast cancer susceptibility. Recently, Proteome-Wide Association Studies (PWAS) (Wingo et al., 2021) and Transcriptome-Wide Association Studies (TWAS) (Gusev et al., 2016) have been pivotal in understanding the functions of proteins and gene expression in disease onset and progression. Initial PWAS focused primarily on neurological contexts due to data limitations (Zhang et al., 2022a), However, recent advancements (Zhang et al., 2022a) have broadened the scope of these studies to include diverse health conditions, thereby enriching our understanding of the associations between plasma proteins and various diseases (Li et al., 2023).

Our first phase focused on identifying proteins that are inherently associated with breast cancer at both proteomic and transcriptomic levels. For PWAS analysis, we integrated plasma protein pQTL data from ARIC cohort (Zhang et al., 2022a) with breast cancer GWAS summary data, including its different ER subtypes. Additionally, we carried out a supplementary TWAS in whole blood and breast mammary tissues. This combined P/TWAS methodology revealed significant associations between plasma proteins and breast cancer. However, it is crucial to note that such associations do not automatically imply causations. To address this, in our second phase, we employed two-sample Mendelian Randomization (MR) analysis (Emdin et al., 2017), adding a causal dimension to the protein-breast cancer relationship. We further assessed shared causal variants between them by genetic Bayesian colocalization. To ensure the robustness and broader applicability of our findings, we further conducted external validations of the established causal link. These validations were achieved using 4 extensive large plasma protein pQTL datasets (Folkersen et al., 2017; Sun et al., 2018; Ferkingstad et al., 2021; Gudjonsson et al., 2022) and the eQTLGen dataset (Võsa et al., 2021).

In our study, we implemented a two-phase design that integrates P/TWAS with MR analyses. This comprehensive methodology, blending associative and causative analyses, provides valuable insights into breast cancer. Furthermore, the relative simplicity in detecting plasma proteins not only strengthens their role in development of diagnostic biomarkers but also suggests their potential value in the development of therapeutic targets for breast cancer.

2 Materials and methods

2.1 Research framework

The analysis flowchart for the study is presented in Figure 1. A two-phase analytical approach was employed in this study, merging P/TWAS for association and MR for causation. Additionally, to guarantee the validity and reliability of the findings, a discovery-confirmatory framework was implemented in both phases.

2.2 Breast cancer GWAS summary data source

The GWAS summary data from the Breast Cancer Association Consortium (BCAC), which specifically focused on individuals of European descent (https://bcac.ccge.medschl.cam.ac.uk/), was

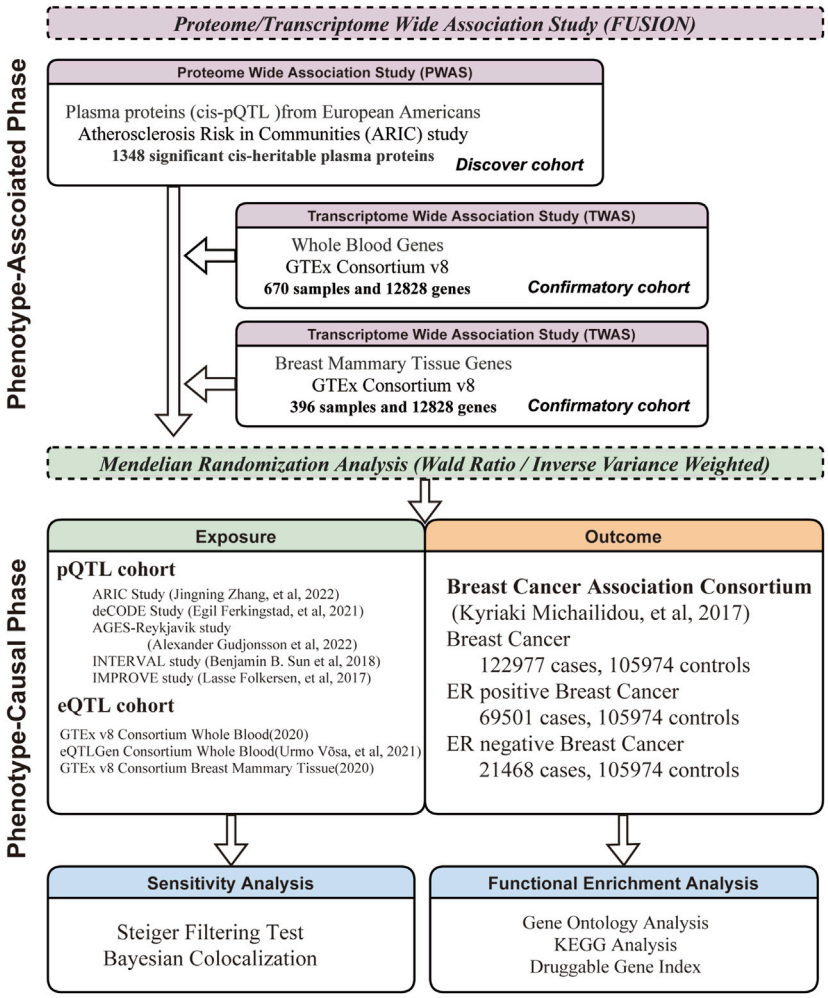


FIGURE 1
Framework of Comprehensive Research Methodology. This research methodology is divided into two phases: phenotype-association and phenotype-causation. Each phase follows a discovery-confirmatory approach.

utilized in our study. This dataset was comprised of 122,977 breast cancer cases and 105,974 controls. The same analytical approach was also applied to ER positive and negative breast cancer. The ER-positive subtype was found to consist of 69,501 cases and 105,974 controls, while the ER-negative subtype included 21,468 cases and 105,974 controls.

2.3 Quantity trait loci (QTL) dataset sources

Cis-pQTL data for European Americans' (EA) plasma proteins were obtained from the ARIC cohort (nilanjanchatterjeelab.org/pwas/), generated using PLINK2 software (Purcell et al., 2007). The SeqID file names correspond to the SOMAmers (Slow Off-rate Modified Aptamers), which are utilized for measuring protein levels in biological samples by leveraging their enhanced affinity and specificity for target proteins (Rohloff et al., 2014). For external validation, cis-pQTL data from 4 extensive plasma protein cohorts of European descent were used. Additionally, our study also explored expression quantitative trait loci (eQTLs), which influence gene expression at the transcriptome level (Zhu et al., 2016). We extracted eQTL data using the SMR toolkit (Wu et al., 2021), a tool specifically designed for genetic epidemiological research, from two major sources: the Genotype-Tissue Expression Project (GTEx Consortium, 2020) and the eQTLGen consortium (Võsa et al., 2021). Detailed descriptions of each dataset are provided in Table 1.

2.4 Proteome/transcriptome-wide association studies with fusion

FUSION (Boston, MA, United States) (Gusev et al., 2016), which is a software to establish associations between functional phenotype and GWAS phenotype, was used to conduct P/TWAS analysis. In our study, FUSION was implemented to identify associations between protein/gene expression levels and Breast Cancer susceptibility. Methodologically, FUSION takes two inputs: 1) Precomputed functional weights, and 2) GWAS summary statistics unified to a reference SNP panel. In PWAS,

TABLE 1 Detailed information about each GWAS summary data.

GWAS	Cohort	Paper title	Year	Author	PMID	Sample size	Protein/ Gene measured
pQTL							
Discovery	ARIC study	Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies	2022	Jingning Zhang, et al	35,501,419	7,213	4435 (1,318 in PWAS)
Confirmatory	Icelandic Cancer Project (52% of participants) and deCODE genetics (48% of participants)	Large-scale integration of the plasma proteome with genetics and disease	2021	Egil Ferkingstad, et al	34,857,953	35,559	4719
Confirmatory	AGES-Reykjavik study	A genome-wide association study of serum proteins reveals shared loci with common diseases	2022	Alexander Gudjonsson, et al	35,078,996	5,368	2091
Confirmatory	INTERVAL study	Genomic atlas of the human plasma proteome	2018	Benjamin B. Sun, et al	29,875,488	3,301	2,994
Confirmatory	IMPROVE study	Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease	2017	Lasse Folkersen, et al	28,369,058	3,394	83
eQTL							
Confirmatory	GTEx v8 Consortium Whole Blood	The GTEx Consortium atlas of genetic regulatory effects across human tissues	2020	GTEx Consortium	32,913,098	670	12,828
Confirmatory	eQTLGen Consortium Whole Blood	Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression	2021	Urmo Võsa, et al	34,475,573	31,684	16,987
Confirmatory	GTEx v8 Consortium Breast Mammary Tissue	The GTEx Consortium atlas of genetic regulatory effects across human tissues	2020	GTEx Consortium	32,913,098	396	12,828
Breast Cancer							
Overall Breast Cancer	Association analysis identifies 65 new breast cancer risk loci		2017	Kyriaki Michailidou, et al	29,059,683	228,951	
ER positive	Association analysis identifies 65 new breast cancer risk loci		2017	Kyriaki Michailidou, et al	29,059,683	175,475	
ER negative	Association analysis identifies 65 new breast cancer risk loci		2017	Kyriaki Michailidou, et al	29,059,683	127,442	

precomputed functional weights of plasma proteins were obtained from the ARIC study (Zhang et al., 2022b), and the reference SNP panel was derived from the European descent of the 1000G project (http://www.internationalgenome.org/faq/how-do-i-cite-1000-genomes-project). The primary outputs of FUSION are the Z-score and *p*-value, wherein Z-score quantifies the strength and direction of the associations between plasma proteins and breast cancer, while the *p*-value elucidates the statistical significance of this association. To enhance our findings at transcriptomic level, we incorporated TWAS for both whole blood and breast mammary tissues. The precomputed functional weights for TWAS, provided by Junghyun Jung from the Mancuso lab (http://gusevlab.org/projects/fusion/). A false discovery rate (FDR, Benjamini–Hochberg) threshold of 0.05 was applied to determine the statistical significance of the results.

2.5 Bayesian colocalization analysis and protein association classification

Bayesian colocalization analysis (Giambartolomei et al., 2014) was utilized to evaluate the probability that the same genetic variant affects both plasma protein and breast cancer. The default parameters set by the analysis were followed, including $p1$ = 10e−4 (the probability of a variant being a significant pQTL), $p2$ = 10e−4 (the probability of a variant associated with breast cancer), and $p12$ = 10e−5 (the probability of a variant being significant in both protein/gene and GWAS). This analysis involved five predefined hypotheses: H0, indicating no association with either trait; H1, signifying association with trait1 only; H2, implying association with trait2 only; H3, representing associations with both traits due to different SNPs;

and H4, indicating association with both traits due to a common SNP. A posterior probability of H4 (PPH4) exceeding 0.8, or in some cases 0.7, is generally interpreted as strong evidence of the same genetic variant being implicated in both traits (Giambartolomei et al., 2014).

Recent studies have investigated the causal associations between plasma proteins and diseases like colorectal cancer (Sun et al., 2023) and inflammatory bowel disease (Chen et al., 2023), utilizing a scoring system that integrates *p*-value and PPH4. Building on this approach, our research employs P/TWAS and Bayesian Colocalization analysis to systematically categorize the degrees of association between proteins. The scoring system was as follows: a significant adjusted *p*-value was awarded 1 point, and a PPH4 > 0.75 also earned 1 point. Based on the cumulative scores, associations were categorized as follows: a score between 1 and 2 indicated a "Weak" association, 3 to 4 suggested a "Moderate" association, and 5 to 6 signified a "Strong" association.

2.6 Mendelian Randomization and sensitivity analysis

In the causal analysis, we primarily conducted further analysis on proteins with strong and moderate associations. MR analysis were based on 3 essential assumptions for genetic instrumental variables: relevance, independence, and exclusion-restriction (Davies et al., 2018). We implemented a stringent selection process for SNPs to be used as instrumental variables, requiring a $p < 5e\text{-}8$, or $p < 5e\text{-}6$ in cases when SNP was absent. Clump was applied in accordance with the default parameters. The Wald Ratio (WR) method was employed when a single SNP was used as the instrumental variable, whereas the inverse-variance weighted (IVW) method was predominant when the instrumental variables involved multiple SNPs (Burgess et al., 2019). To reinforce the robustness of our findings, we conducted several sensitivity analyses. The Steiger filtering test (Deng et al., 2022) was utilized to eliminate the possibility of reverse causal associations. Additionally, heterogeneity and pleiotropy sensitivity analyses were conducted for proteins that met the criteria (Bowden et al., 2015; Greco et al., 2015). Furthermore, to improve the reliability and applicability of our results, external validation was carried out on pQTL data derived from 4 extensive plasma protein cohorts in European populations.

2.7 Enrichment analysis and potential druggable targets

To delve deeper into the intricate relationships and biological functions of significant proteins identified in our PWAS, gene ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses was performed. Given the emerging role of plasma proteins as potential therapeutic targets (Sun et al., 2018), we matched P/TWAS-MR significant proteins with the druggable genome database (Finan et al., 2017), which categorizes 4,479 genes into three druggability tiers: Tier 1 includes approved drugs and candidates in clinical trials, Tier 2 encompasses targets of biologically active molecules and those similar to approved drug targets, and Tier 3 comprises genes for secreted or extracellular proteins and other key druggable gene family members. Additionally, the significant proteins were annotated using the Therapeutic Target Database (http://db.idrblab.net/ttd/) (Zhou et al., 2022).

2.8 Statistical methods

In this study, data analysis was executed using R software (version 4.3.1). The P/TWAS analysis followed the analytical process previously described. The Benjamini–Hochberg method was employed for multiple testing correction, with adjusted *p*-values <0.05 considered statistically significant. Causations were investigated using the "TwoSampleMR" package, while Bayesian colocalization analysis was carried out using the "COLOC" package. The "ClusterProfiler" package (Wu et al., 2021) was utilized for functional enrichment analysis. Data visualization was achieved through the "Forestploter" and "ggplot2" packages, and data cleaning was performed using the "tidyverse" package.

3 Results

3.1 Identification of associations at the proteomic level

In our study, a total of 25 plasma proteins were significantly associated with breast cancer (Table 2; Figure 2A, and Supplementary Table S1). Of these proteins, 14 showed a Z-score greater than 0, denoting a positive association with breast cancer. Conversely, the remaining 11 proteins suggested an inverse association with the disease. When duplicate SOMAmers are present, we select the protein corresponding to the smallest *p*-value for subsequent analysis, such as RSPO3 (Supplementary Table S1). In ER subtypes analysis, 16 proteins were found to be significantly associated with ER-positive breast cancer and 6 with ER-negative breast cancer (Supplementary Table S2, S3). The PWAS Manhattan plot illustrates the distribution of significant genes across different chromosomes and their respective *p*-value (Figure 3A, Supplementary Figure S2A, B).

3.2 Identification of associations at the transcriptomic level

For the 25 proteins identified by PWAS, 12 showed significant associations in the whole blood TWAS analysis ($P.adj < 0.05$) (Supplementary Table S4). While, in the breast mammary tissue TWAS, 10 of these proteins were further validated ($P.adj < 0.05$) (Table 2, Supplementary Table S5). Among the 16 significant proteins in ER-positive breast cancer, with 7 were confirmed in both whole blood and breast mammary tissue analyses. Meanwhile, in ER-negative breast cancer, 2 out of the 6 significant proteins were validated (Figures 2B,C, Supplementary Table S6). It should be noted that MST1 exhibited contradictory associations in PWAS ($Z = 4.194$, $P.adj = 0.004$) and TWAS ($Z = -2.547$, $P.adj = 0.014$). This pattern was also observed in ER-positive and ER-negative subtypes. Due to the complex nature and potential biological

10.3389/fmolb.2023.1340917

TABLE 2 Integrative analysis and stratification of proteome and transcriptome associations in breast cancer.

Gene	CHR	Plasma protein PWAS (discovery cohort)			Whole blood TWAS (confirmatory cohort)			Breast tissue TWAS (confirmatory cohort)			Score	Association power
		Zscore	*P*_FDR	PPH4	Zscore	*P*_FDR	PPH4	Zscore	*P*_FDR	PPH4		
PGD	1	−9.152	7.39E-17	0.994	-	-	-	-	-	-	2	Weak
TLR1	4	6.225	3.18E-07	1	-	-	-	-	-	-	2	Weak
FBLN5	14	5.226	7.60E-05	0	0.460	0.6460	0	-	-	-	1	Weak
PEX14	1	4.839	0.0004	0.989	6.843	1.09E-10	0.195	5.341	8.14E-07	0.79	5	Strong
LAYN	11	4.499	0.0018	0.936	-	-	-	-	-	-	2	Weak
SNUPN	15	−4.413	0.0022	0.952	−5.285	8.82E-07	0.939	−5.256	8.14E-07	0.372	5	Strong
GSTM4	1	−4.273	0.0036	0.618	−3.451	0.0011	0.001	−3.736	0.0004	0.175	3	Moderate
MST1	3	4.194	0.0045	0.904	−2.547	0.0139	0.148	−3.266	0.0015	0.584	4	Moderate (inconsistent)
CSK	15	−4.147	0.0049	0.779	−3.979	0.0002	0.843	−4.613	1.46E-05	0.863	6	Strong
NTN4	12	3.938	0.0108	0	-	-	-	-	-	-	1	Weak
PAPPA	9	−3.692	0.0255	0.137	-	-	-	-	-	-	1	Weak
CTSF	11	3.681	0.0255	0.777	2.895	0.0059	0.944	4.394	0.0000	0.939	6	Strong
PARK7	1	−3.648	0.0269	0.978	−4.644	1.20E-05	0.965	−2.391	0.0185	0.014	5	Strong
NCF1	7	3.601	0.0296	0.002	4.419	2.78E-05	0.958	-	-	-	3	Moderate
COL6A3	2	−3.585	0.0296	0.026	-	-	-	-	-	-	1	Weak
RSPO3	6	−3.541	0.0319	0.272	-	-	-	-	-	-	1	Weak
HEBP1	12	3.533	0.0319	0.058	2.358	0.0215	0.015	-	-	-	2	Weak
NRP1	10	−3.508	0.0328	0.038	-	-	-	-	-	-	1	Weak
ABO	9	−3.442	0.0369	0.298	−4.953	3.41E-06	0.169	2.127	0.0334	0.994	4	Moderate
PRDX1	1	3.436	0.0369	0.004	1.993	0.0498	0.005	3.392	0.0011	0.03	3	Moderate
EMILIN3	20	3.410	0.0369	0.072	-	-	-	-	-	-	1	Weak
ANXA4	2	3.404	0.0369	0.681	2.820	0.0067	0.008	-	-	-	2	Weak
POSTN	13	3.401	0.0369	0.01	-	-	-	-	-	-	1	Weak
LDHA	11	−3.352	0.0424	0.404	−3.307	0.0016	0.416	−3.625	0.0005	0.291	3	Moderate
UROD	1	3.302	0.0487	0.254	-	-	-	2.834	0.0056	0.113	2	Weak

implications of MST1's contrasting results, we did not conduct further analysis on this protein.

3.3 Bayesian Colocalization analysis

Among 25 significant proteins, 9 exhibited strong genetic colocalization evidence. Additionally, 4 proteins–SNUPN (PPH4 = 93.9%), CSK (PPH4 = 84.3%), CTSF (PPH4 = 94.4%), and PARK7 (PPH4 = 96.5%)–also demonstrated the same strong genetic evidence at the whole blood transcriptomic level. Remarkably, CSK (PPH4 = 86.3%) and CTSF (PPH4 = 93.9%) were further validated in the breast mammary tissue transcriptomic level (Table 2). In ER-positive breast cancer, 5 proteins showed strong evidence of genetic colocalization. Notably, 2 of these proteins, CSK (PPH4 = 85.8%, 86.5%) and GDI2 (PPH4 = 97%, 97.2%), demonstrated the same strong genetic colocalization evidence in both whole blood and breast mammary tissues. In the ER-negative breast cancer, PEX14 showed strong genetic colocalization evidence in protein (PPH4 = 99.9%) and breast mammary tissue (PPH4 = 88.8%), but this pattern was not replicated at the whole blood transcriptomic level (PP4 = 8%, Supplementary Table S6).

3.4 Stratification of plasma protein association strengths

In breast cancer, 25 proteins were classified: 5 as "Strong" association (red), 6 as "Moderate" association (blue), and the

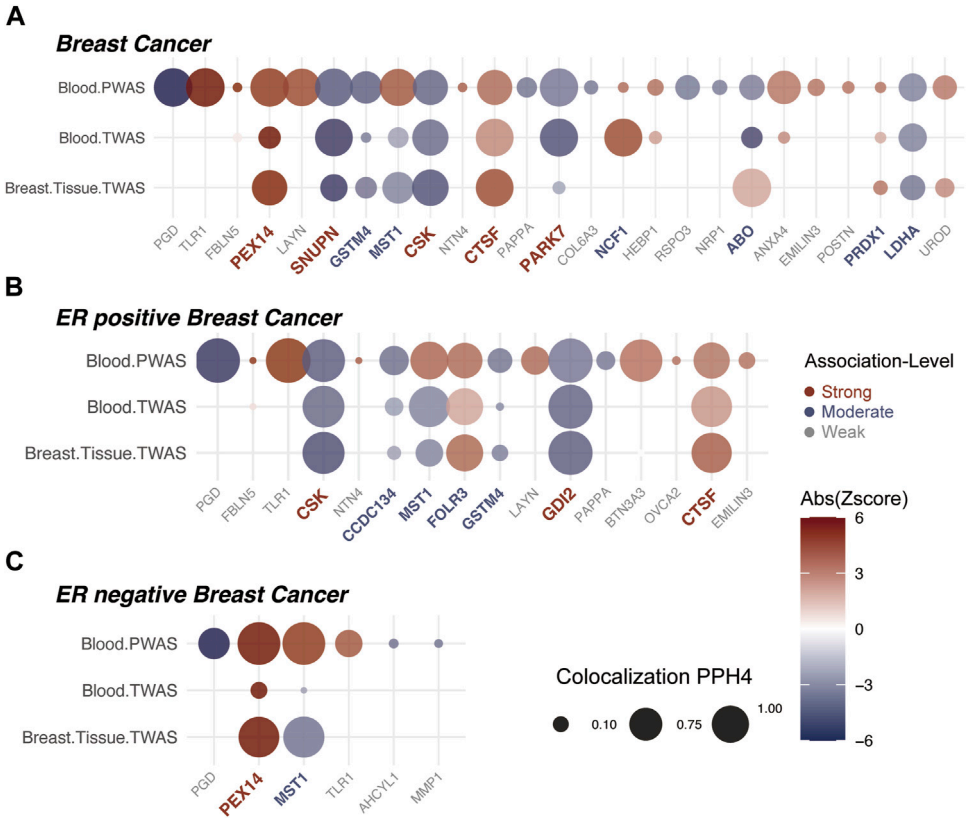


FIGURE 2
Insights from Proteome/Transcriptome-Wide Association Analyses and Bayesian Colocalization. The association strength of proteins with breast cancer is denoted by colors: red for strong, blue for moderate, and grey for weak associations. **(A)** Comprehensive P/TWAS for plasma proteins in breast cancer susceptibility. Dot size signifies results from Bayesian Colocalization analysis, with color gradient reflecting the Z-value. Proteins are sequentially arranged based on ascending *p*-value significance from left to right. **(B)** Comprehensive P/TWAS for plasma proteins in ER positive breast cancer susceptibility. **(C)** Comprehensive P/TWAS for plasma proteins in ER negative breast cancer susceptibility.

remaining as "Weak" association (grey) (Figure 2A). Among the "Strong" Tiers, PEX14 (Z = 4.839) and CTSF (Z = 3.681) had a positive association with breast cancer. Whereas, SNUPN (Z = −4.413), CSK (Z=−4.417), and PARK7 (Z = −3.648) showed negative associations (Table 2).

In ER-positive breast cancer, 16 proteins were classified: 3 proteins showed "Strong" association (red), with GDI2 (Z = −3.652) newly identified and negatively associated with ER-positive breast cancer. Additionally, CSK and CTSF followed the same trends with the findings from breast cancer. Besides, 4 proteins were "Moderate" (blue), and 9 proteins were "Weak" associations (grey) (Figure 2B). In ER-negative breast cancer, 6 proteins were classified: PEX14 and MST1 showed "Strong" associations. Notably, PEX14 not only showed the same trend as observed in breast cancer (Z = 4.839, p = 0.0004) but also exhibited a notably stronger effect (Z = 5.929, p = 2.02E-6). MST1 was not further analyzed due to inconsistent trends in P/TWAS. The other 4 proteins were categorized as "Weak" association (grey) (Figure 2C, Supplementary Table S6).

It is crucial to highlight that, although PGD and TLR1 were significant across all three outcomes in PWAS analyses (Figure 3A and Supplementary Figure S2A, B), their absence from the corresponding TWAS analysis relegated them to the "Weak" association. Moreover, the results of these two proteins were not sufficiently reliable in MR Analysis (Supplementary Figure S2C, Supplementary Table S7).

3.5 Mendelian Randomization analyses

Upon determining the strength of associations, we supplemented the causations with MR analysis (Supplementary Table S8). We primarily focused on the causal effects of "Strong" associated proteins. Among the 5 "Strong" associated proteins, PEX14 was found to have a positive causation at the proteomic (OR = 1.201, p = 0.017) and transcriptomic level (OR = 1.17, p < 0.001). CTSF demonstrated a positive causation in the ARIC cohort (OR = 1.114, p < 0.001), and the consistent trends were also external validated in INTERVAL cohort (OR = 1.144, p < 0.001) (Sun et al., 2018) and AGES-Reykjavik cohort (OR = 1.159, p < 0.001) (Gudjonsson et al., 2022) (Figure 3B). The remaining 3 proteins, SNUPN (OR = 0.905, p < 0.001), CSK (OR = 0.962, p = 0.038), and PARK7 (OR = 0.954, p < 0.001), all exhibited negative causations with breast cancer. External validations from the deCODE cohort further confirmed the causations for SNUPN (OR = 0.797, p < 0.001) and PARK7 (OR = 0.844, p = 0.017). However, CSK's causation at the whole blood transcriptomic level was somewhat unsignificant (OR = 0.84, p = 0.129) (Figure 3B, Supplementary Table S9).

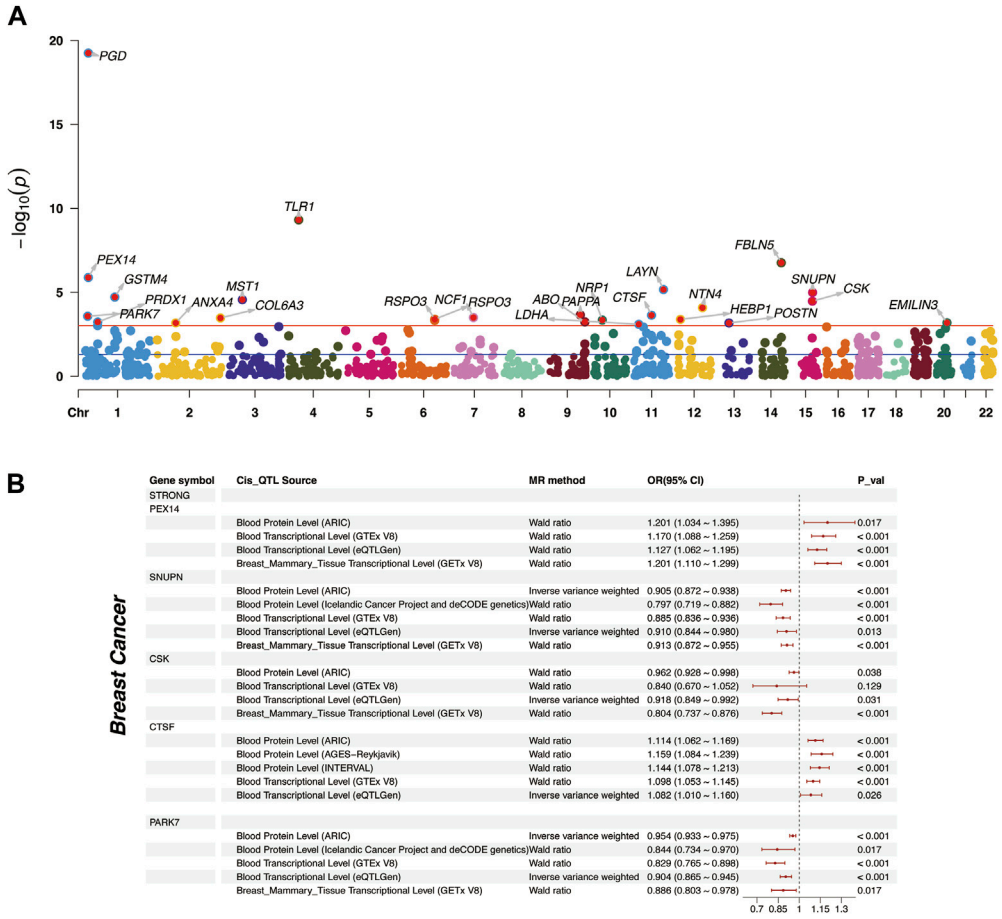


FIGURE 3
Distribution of plasma proteins and Mendelian Randomization analysis of "Strong" plasma proteins. **(A)** The Manhattan plot represented plasma proteins with significant affiliations to breast cancer. The red horizontal line indicates the FDR corrected *p*-value threshold for significance. Chromosomal designations populate the horizontal axis, contrasted with respective -log10 *p*-values on the vertical spectrum. **(B)** Two-sample Mendelian Randomization analysis for "Strong" plasma proteins to breast cancer, including external validation at proteomic and transcriptomic levels.

In ER-positive breast cancer, CSK (OR = 0.955, $p = 0.038$) and CTSF (OR = 1.125, $p < 0.001$) maintained the same causal trends as observed in breast cancer (Supplementary Table S10). Additionally, GDI2 was identified as a newly negatively significant protein (OR = 0.92, $p < 0.001$). However, its causal effect was not significant at the transcriptomic level (OR = 1.001, $p < 0.981$, Figure 4A). In ER-negative breast cancer, PEX14 stood out as the sole "Strong" protein. Notably, its causal effect in this subtype (OR = 1.645, $p < 0.001$, Figure 4B) was further pronounced compared to breast cancer (OR = 1.201, $p = 0.017$). Meanwhile, we expanded our MR analyses to include "Moderate" proteins. The results revealed that their causal effects were generally less consistent and of reduced significance compared to those of the "Strong" proteins (Supplementary Figure S1, Supplementary Table S11).

3.6 Sensitivity analysis and functional enrichment analysis

Considering that the pQTLs of most plasma proteins was a single SNP, conducting sensitivity analyses for heterogeneity and pleiotropy is typically not required. As result, in ER-positive breast cancer, BTN3A3, EMILIN3, FOLR3, and NTN4 showed heterogeneity, while in ER-negative cases, this was not observed (Supplementary Table S10). BTN3A3 in ER-positive breast cancer also displayed pleiotropy. The Steiger filtering test confirmed that MR effects were due to plasma proteins affecting breast cancer outcomes (Supplementary Table S8, S10). Importantly, our "Strong" proteins exhibited neither heterogeneity nor pleiotropy.

Furthermore, the plasma proteins identified by PWAS were subjected to Gene Ontology (GO) cluster analysis. This analysis revealed a predominant association with biological processes related to oxidative stress, such as "reactive oxygen species metabolic" and "response to reactive oxygen species" terms. Additionally, for cellular components, we observed a significant enrichment in the "collagen-containing extracellular matrix" term (Supplementary Figure S3). Besides, KEGG pathway enrichment did not reveal any significantly enriched pathways (Supplementary Table S12).

3.7 Druggable target propensity for significant proteins

Plasma proteins are not only crucial as diagnostic biomarkers but also serve as potential drug targets. In our study, we evaluated

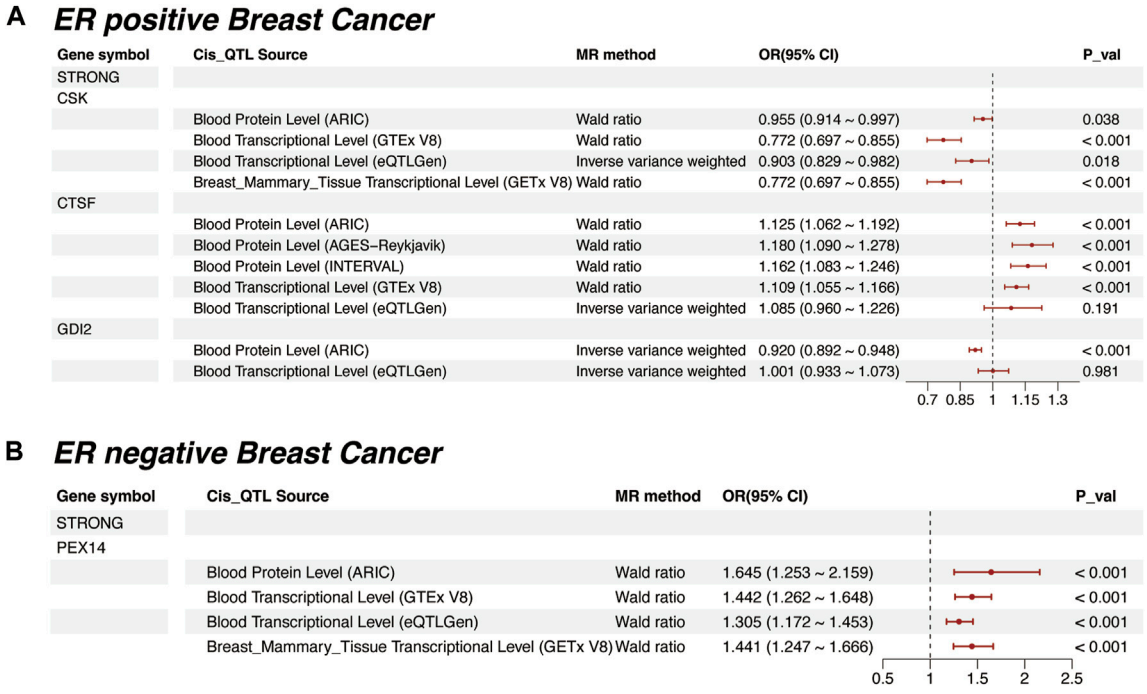


FIGURE 4
Mendelian Randomization for "Strong" plasma proteins in different ER Breast Cancer Subtypes. **(A)** Mendelian randomization results for ER-positive breast cancer, including external validation at proteomic and transcriptomic levels. **(B)** Mendelian randomization results for ER-negative breast cancer, including external validation at proteomic and transcriptomic levels.

the significant proteins for their potential as drug targets. By aligning our findings with the druggable genome database (Finan et al., 2017), we determined that 16 of the 25 proteins have druggable targets. These include 3 proteins in Tier 1; 3 in Tier 2, and 10 in Tier 3 (Supplementary Table S13, Left column). Furthermore, we compared our results with the Therapeutic Target Database (Zhou et al., 2022), 11 of these 16 proteins were identified as targets of existing or potential drugs. This group comprised 3 Successful targets, 3 Patented-recorded Targets, 1 in clinical trials, and 4 documented in literature (Supplementary Table S13 Right column). Among the "Strong" proteins, CSK and CTSF were found to be drug targets with patent records, categorized under Tiers 1 and 2 respectively. CTSF has been documented to be used in the treatment of bone cancer and chronic obstructive pulmonary disease (Li et al., 2017) (Table 3). However, the remaining "Strong" proteins have not yet been reported.

4 Discussion

Plasma proteins, due to their ease of detection and reproducibility, are increasingly utilized to distinguish between cancer patients and healthy individuals, enhancing the effectiveness of screening programs (Huijbers et al., 2010). Recent advancements in molecular technologies and techniques have shown significant potential in utilizing plasma protein biomarkers such as Adipsin and CA15-3 for early detection and quantification for diagnostic and therapeutic applications in breast cancer (Afzal et al., 2022; Rajkumar et al., 2022; Veyssière et al., 2022). A recent high-throughput study identified 61 proteins associated with various cancers (Gregga et al., 2023). While this study provided valuable insights into pan-cancer associations, it did not explore causation. Furthermore, research specifically targeting plasma protein biomarkers for breast cancer is still limited. Currently, Mendelian Randomization has emerged as an effective method to establish causation in various diseases (Emdin et al., 2017), including cholesterol-related cardiovascular disease (Kathiresan et al., 2008), inflammatory diseases (Swerdlow et al., 2012), metabolic disorders (Fall et al., 2015), and specific cancers such as small cell lung cancer and colorectal cancer (Sun et al., 2023; Wu et al., 2023).

Despite, the application of MR in identifying plasma proteins as drug targets in breast cancer is still sporadic. For instance, one study performed MR analysis on a single cohort of 732 plasma proteins, where GDI2 and CTSF were identified as potential targets for breast cancer (Ren et al., 2023), aligning with our research. However, it is important to note that this study also focused on pan-cancer research and lacked association analysis. Additionally, another study focused on the causation found a causal link between TLR1 and breast cancer (Mälarstig et al., 2023). This protein was ranked significantly in our analysis, but it is noteworthy that TLR1 lacks external cohort validation, and the study also did not perform association analyses. Therefore, current research on plasma proteins typically focuses on either association or causation, rarely addressing both. Our study bridges this gap by integrating these two approaches. We employed P/TWAS to identify associations and used MR to establish causation. This approach successfully pinpointed significant proteins related to breast cancer risk from

TABLE 3 Comprehensive evaluation of strong associated proteins as potential druggable targets or existing therapeutics.

Gene	UniProt	Description	Finan et al	Therapeutic target database		
			Tier	Target type	Drug name	Disease
Breast Cancer						
PEX14	O75381	Peroxisomal Biogenesis Factor 14	-	-	-	-
SNUPN	O95149	Snurportin 1	-	-	-	-
CSK	P41240	Tyrosine-protein kinase CSK	Tier 1	Patented-recorded Target	936,563-93-8	Not Available
CTSF	Q9UBX1	cathepsin F	Tier 2	Patented-recorded Target	PMID27998201-Compound-5	Bone cancer; Chronic obstructive pulmonary disease
PARK7	Q99497	Parkinsonism Associated Deglycase	-	-	-	-
ER pos Breast Cancer						
CSK	P41240	Tyrosine-protein kinase CSK	Tier 1	Patented-recorded Target	936,563-93-8	Not Available
CTSF	Q9UBX1	cathepsin F	Tier 2	Patented-recorded Target	PMID27998201-Compound-5	Bone cancer; Chronic obstructive pulmonary disease
GDI2	P50395	Rab GDP dissociation inhibitor beta	-	-	-	-
ER neg Breast Cancer						
PEX14	O75381	Peroxisomal Biogenesis Factor 14	-	-	-	-

thousands of candidates in 5 large proteomics cohorts. To ensure the robustness and generalizability of our findings, we adopted a "discovery-confirmatory" analytical framework at both the association and causation phases. Overall, we found 5 proteins (PEX14, CTSF, SNUPN, CSK, PARK7) with strong causal links to breast cancer. While, in ER-positive breast cancer, 3 proteins (CSK, CTSF, GDI2) were identified. In contrast, only PEX14 was linked to ER-negative breast cancer.

Among the 5 plasma proteins, SNUPN, CSK, and PARK7 emerged as "Strong" negatively causative associated proteins, indicating a protective effect against breast cancer development. A study has highlighted the potential clinical applications of SNUPN in acute lymphoblastic leukemia (Mata-Rocha et al., 2019); however, research exploring its role in solid tumors, including breast cancer, is currently limited. Despite the current research limitations, SNUPN's potential as a biomarker or tumor suppressor is promising and warrants further exploration. PARK7 is recognized for its neuroprotective role in Parkinson's disease (Kochmanski et al., 2022) and has been reported to significantly regulate cell survival and cancer progression in various cancers (Jin, 2020). It negatively regulates PTEN and PKB/Akt phosphorylation, thus influencing cell survival and death (Kim et al., 2005). In breast cancer, low PARK7 expression was correlated with pathological complete response in 79.6% of cases following neoadjuvant therapy (Kawate et al., 2013), and loss of PARK7 function is associated with increased sensitivity to doxorubicin in breast cancer cells (Zhang et al., 2015). The effect of PARK7 in balancing tumor cell survival and normal cell physiology merits further research. Lastly, as a key member of the Src family kinases (SFKs), CSK plays a vital role in combating cancer progression in various cancers (Sabe et al., 1994). Recent study indicates that CSK maintains negative regulation of Src through Tyr527 phosphorylation, inhibiting breast cancer cells growth and spread (Dias et al., 2022). Additionally, another study on ER-positive breast cancer found that in cases of endocrine therapy resistance, reduced CSK leads to enhanced PAK2 activity and subsequent non-estrogen-dependent cancer growth (Xiao et al., 2018). The dual effect of CSK in both tumor suppression and inducing endocrine treatment resistance positions it as a notable target for research.

The other two "Strong" proteins are positively associated and represent a risk factor in breast cancer onset. CTSF (cathepsin F) plays a key role in the lysosomal protein degradation pathway (Wex et al., 1999). Currently, it is reported as an effective diagnostic biomarker in cervical cancer (Vazquez-Ortiz et al., 2005), gastric cancer (Ji et al., 2018), and non-small cell lung cancer (Wei et al., 2022). A recent study reported that CTSF may act as an independent poor prognostic factor for basal-like breast cancer (Huang et al., 2021). PEX14 (Peroxisomal Biogenesis Factor 14) is essential for peroxisomal biogenesis (Neufeld et al., 2009). Our research reveals a significant causal risk association of PEX14 with breast cancer (OR = 1.201), particularly in ER-negative subtype (OR = 1.645). Notably, PEX14 has been identified as a key risk factor in triple-negative breast cancer (TNBC) (Purrington et al., 2014) and is one of the top five genes influencing adaptive anti-tumor immunity, as shown in a TNBC model study using a whole-genome RNAi screening platform

(Shuptrine et al., 2017). These insights emphasize PEX14's importance in TNBC immunotherapy and drug target research. Furthermore, PEX14 plays a crucial role in maintaining peroxisomal functions, and its deficiency leads to ROS accumulation, lipid peroxidation, and consequent cell death (Guo et al., 2023). Our functional enrichment analysis corroborates this, highlighting numerous pathways related to reactive oxygen species (ROS), which are instrumental in promoting cell growth, cancer progression, immune responses, and poorer survival outcomes in breast cancer (Oshi et al., 2022). Additionally, studies have shown that PEX14 knockdown increases intracellular H_2O_2 levels, triggering ferroptosis and cell death (Guan et al., 2022). This further underscores PEX14's pivotal role in managing oxidative stress and cell viability, marking its significance in breast cancer research. Additionally, GDI2 was identified as a protein with a "Strong" negative causal association in the ER-positive breast cancer. A study suggested that GDI2 is associated with aggressive features and poor patient survival in hepatocellular carcinoma (Zhang et al., 2021). However, the inability to confirm its role through at additional transcriptomic levels and the absence of external validation has diminished our confidence in the significance of this protein.

Given the proven effectiveness of MR in identifying drug targets (Folkersen et al., 2020), we performed a drug-target evaluation on these plasma proteins (Supplementary Table S13). Notably, CSK and CTSF emerged as Tier1 and Tier2 proteins, respectively. CSK is crucial in regulating cellular processes such as apoptosis, survival, and proliferation. Its pivotal role in cancer cell signaling earmarks CSK as a promising target for cancer therapy (Fortner et al., 2022). Similarly, CTSF, known for its significant involvement in the progression of various cancers (Wei et al., 2022), neurodegenerative diseases (van der Zee et al., 2016), and skin aging (Takaya et al., 2023), garners attention. Research on inhibitors and modulators targeting CTSF is underway. Although other strongly associated proteins currently lack clear therapeutic applications, given their strong causal relationship with breast cancer, it is worthwhile to further explore them for drug target development.

This study is currently subject to several limitations yet. First, the study only involves individuals of European descent, which necessitates caution when applying these findings to more diverse populations. Second, the precomputed functional weights for plasma proteins are currently only available from the ARIC cohort, future datasets expansion are expected to enhance the precision and breadth of such analyses. In addition, as the current BCAC molecular subtype data lacks rsID, matching chromosomes and base pair positions results in significant information loss. However, with the continuous expansion and updating of the molecular subtype database, we anticipate a deeper understanding of this content. Lastly, our analysis is primarily data-based, hence we will design related basic scientific research in the future to further investigate the etiological association between plasma proteins and breast cancer.

In summary, our study successfully identified several plasma proteins with strong association and causation to breast cancer and its distinct ER subtypes. As non-invasive and dynamic monitoring tools, plasma proteins hold significant potential as diagnostic biomarkers and therapeutic targets. They offer a comprehensive perspective on systemic health, which is crucial for early tumor detection, assessing treatment responses, and continuous disease monitoring. While these advancements are still in the early stages, they hold valuable promise for future research and practical applications in real-world scenarios.

Data availability statement

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

Author contributions

YW: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing–original draft, Writing–review and editing. KY: Data curation, Formal Analysis, Investigation, Methodology, Writing–original draft. BC: Data curation, Formal Analysis, Investigation, Methodology, Validation, Writing–review and editing. BZ: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing–review and editing.GJ: Funding acquisition, Project administration, Resources, Supervision, Validation, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by Beijing Municipal Natural Science Foundation (No.7222145) and Shenzhen Key Medical Discipline Construction Fund (SZXK095).

Acknowledgments

We would like to extend our heartfelt thanks to all the data contributors of the plasma protein cohorts referenced in our study. The pertinent literature has been cited within the article. Our sincere appreciation goes out to the participants for their commitment and to the researchers for their altruistic contributions. The breast cancer genome-wide association analyses were supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the 'Ministère de l'Économie, de la Science et de l'Innovation du Québec' through Genome Québec and grant PSR-SIIRI-701, The National Institutes of Health (U19 CA148065, X01HG007492), Cancer Research United Kingdom (C1287/A10118, C1287/A16563, C1287/A10710) and The European Union (HEALTH-F2-2009-223175 and H2020 633784 and 634935). All studies and funders are listed in Michailidou et al. (2017). We extend our profound gratitude to Junghyun Jung and the Mancuso Lab for their invaluable contribution to our study through the provision of precomputed TWAS models. Their contributions have been pivotal in the success of our research endeavors.

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/fmolb.2023.1340917/full#supplementary-material

References

Afzal, S., Hassan, M., Ullah, S., Abbas, H., Tawakkal, F., and Khan, M. A. (2022). Breast cancer; discovery of novel diagnostic biomarkers, drug resistance, and therapeutic implications. *Front. Mol. Biosci.* 9, 783450. doi:10.3389/fmolb.2022.783450

Anderson, N. L., and Anderson, N. G. (2002). The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell Proteomics* 1 (11), 845–867. doi:10.1074/mcp.r200007-mcp200

Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* 44 (2), 512–525. doi:10.1093/ije/dyv080

Burgess, S., Davey Smith, G., Davies, N. M., Dudbridge, F., Gill, D., Glymour, M. M., et al. (2019). Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open Res.* 4, 186. doi:10.12688/wellcomeopenres.15555.2

Chen, J., Xu, F., Ruan, X., Sun, J., Zhang, Y., Zhang, H., et al. (2023). Therapeutic targets for inflammatory bowel disease: proteome-wide Mendelian randomization and colocalization analyses. *EBioMedicine* 89, 104494. doi:10.1016/j.ebiom.2023.104494

Christopoulos, P. F., Msaouel, P., and Koutsilieris, M. (2015). The role of the insulin-like growth factor-1 system in breast cancer. *Mol. Cancer* 14, 43. doi:10.1186/s12943-015-0291-7

Davies, N. M., Holmes, M. V., and Davey Smith, G. (2018). Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *Bmj* 362, k601. doi:10.1136/bmj.k601

Deng, Y. T., Ou, Y. N., Wu, B. S., Yang, Y. X., Jiang, Y., Huang, Y. Y., et al. (2022). Identifying causal genes for depression via integration of the proteome and transcriptome from brain and blood. *Mol. Psychiatry* 27 (6), 2849–2857. doi:10.1038/s41380-022-01507-9

Dias, R. V. R., Ferreira, C. T. A., Jennings, P. A., Whitford, P. C., and Oliveira, L. C. (2022). Csk αC helix: a computational analysis of an essential region for conformational transitions. *J. Phys. Chem. B* 126 (50), 10587–10596. doi:10.1021/acs.jpcb.2c05408

Emdin, C. A., Khera, A. V., and Kathiresan, S. (2017). Mendelian randomization. *Jama* 318 (19), 1925–1926. doi:10.1001/jama.2017.17219

Fall, T., Hägg, S., Ploner, A., Mägi, R., Fischer, K., Draisma, H. H., et al. (2015). Age- and sex-specific causal effects of adiposity on cardiovascular risk factors. *Diabetes* 64 (5), 1841–1852. doi:10.2337/db14-0988

Ferkingstad, E., Sulem, P., Atlason, B. A., Sveinbjornsson, G., Magnusson, M. I., Styrmisdottir, E. L., et al. (2021). Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* 53 (12), 1712–1721. doi:10.1038/s41588-021-00978-w

Finan, C., Gaulton, A., Kruger, F. A., Lumbers, R. T., Shah, T., Engmann, J., et al. (2017). The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* 9 (383), eaag1166. doi:10.1126/scitranslmed.aag1166

Folkersen, L., Fauman, E., Sabater-Lleal, M., Strawbridge, R. J., Frånberg, M., Sennblad, B., et al. (2017). Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genet.* 13 (4), e1006706. doi:10.1371/journal.pgen.1006706

Folkersen, L., Gustafsson, S., Wang, Q., Hansen, D. H., Hedman Å, K., Schork, A., et al. (2020). Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat. Metab.* 2 (10), 1135–1148. doi:10.1038/s42255-020-00287-2

Fortner, A., Chera, A., Tanca, A., and Bucur, O. (2022). Apoptosis regulation by the tyrosine-protein kinase CSK. *Front. Cell Dev. Biol.* 10, 1078180. doi:10.3389/fcell.2022.1078180

Giambartolomei, C., Vukcevic, D., Schadt, E. E., Franke, L., Hingorani, A. D., Wallace, C., et al. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 10 (5), e1004383. doi:10.1371/journal.pgen.1004383

Greco, M. F., Minelli, C., Sheehan, N. A., and Thompson, J. R. (2015). Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Stat. Med.* 34 (21), 2926–2940. doi:10.1002/sim.6522

Gregga, I., Pharoah, P. D. P., Gayther, S. A., Manichaikul, A., Im, H. K., Kar, S. P., et al. (2023). Predicted proteome association studies of breast, prostate, ovarian, and endometrial cancers implicate plasma protein regulation in cancer susceptibility. *Cancer Epidemiol. Biomarkers Prev.* 32 (9), 1198–1207. doi:10.1158/1055-9965.Epi-23-0309

GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369 (6509), 1318–1330. doi:10.1126/science.aaz1776

Guan, H., Zhou, P., Qi, Y., Huang, H., Wang, J., and Liu, X. (2022). Cigarette smoke-induced trophoblast cell ferroptosis in rat placenta and the effects of L-arginine intervention. *Ecotoxicol. Environ. Saf.* 243, 114015. doi:10.1016/j.ecoenv.2022.114015

Gudjonsson, A., Gudmundsdottir, V., Axelsson, G. T., Gudmundsson, E. F., Jonsson, B. G., Launer, L. J., et al. (2022). A genome-wide association study of serum proteins reveals shared loci with common diseases. *Nat. Commun.* 13 (1), 480. doi:10.1038/s41467-021-27850-z

Guo, Y., Zhou, P., Qiao, L., Guan, H., Gou, J., and Liu, X. (2023). Maternal protein deficiency impairs peroxisome biogenesis and leads to oxidative stress and ferroptosis in liver of fetal growth restriction offspring. *J. Nutr. Biochem.* 121, 109432. doi:10.1016/j.jnutbio.2023.109432

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* 48 (3), 245–252. doi:10.1038/ng.3506

Huang, K., Wu, Y., Xie, Y., Huang, L., and Liu, H. (2021). Analyzing mRNAsi-related genes identifies novel prognostic markers and potential drug combination for patients with basal breast cancer. *Dis. Markers* 2021, 4731349. doi:10.1155/2021/4731349

Huijbers, A., Velstra, B., Dekker, T. J., Mesker, W. E., van der Burgt, Y. E., Mertens, B. J., et al. (2010). Proteomic serum biomarkers and their potential application in cancer screening programs. *Int. J. Mol. Sci.* 11 (11), 4175–4193. doi:10.3390/ijms11114175

Ji, C., Zhao, Y., Kou, Y. W., Shao, H., Guo, L., Bao, C. H., et al. (2018). Cathepsin F knockdown induces proliferation and inhibits apoptosis in gastric cancer cells. *Oncol. Res.* 26 (1), 83–93. doi:10.3727/096504017x14928634401204

Jin, W. (2020). Novel insights into PARK7 (DJ-1), a potential anti-cancer therapeutic target, and implications for cancer progression. *J. Clin. Med.* 9 (5), 1256. doi:10.3390/jcm9051256

Kathiresan, S., Melander, O., Anevski, D., Guiducci, C., Burtt, N. P., Roos, C., et al. (2008). Polymorphisms associated with cholesterol and risk of cardiovascular events. *N. Engl. J. Med.* 358 (12), 1240–1249. doi:10.1056/NEJMoa0706728

Kawate, T., Iwaya, K., Kikuchi, R., Kaise, H., Oda, M., Sato, E., et al. (2013). DJ-1 protein expression as a predictor of pathological complete remission after neoadjuvant chemotherapy in breast cancer patients. *Breast Cancer Res. Treat.* 139 (1), 51–59. doi:10.1007/s10549-013-2523-0

Key, T. J., Appleby, P. N., Reeves, G. K., and Roddam, A. W. (2010). Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies. *Lancet Oncol.* 11 (6), 530–542. doi:10.1016/s1470-2045(10)70095-4

Kim, R. H., Peters, M., Jang, Y., Shi, W., Pintilie, M., Fletcher, G. C., et al. (2005). DJ-1, a novel regulator of the tumor suppressor PTEN. *Cancer Cell* 7 (3), 263–273. doi:10.1016/j.ccr.2005.02.010

Kochmanski, J., Kuhn, N. C., and Bernstein, A. I. (2022). Parkinson's disease-associated, sex-specific changes in DNA methylation at PARK7 (DJ-1), SLC17A6 (VGLUT2), PTPRN2 (IA-2β), and NR4A2 (NURR1) in cortical neurons. *NPJ Park. Dis.* 8 (1), 120. doi:10.1038/s41531-022-00355-2

Li, H., Zhang, Z., Qiu, Y., Weng, H., Yuan, S., Zhang, Y., et al. (2023). Proteome-wide mendelian randomization identifies causal plasma proteins in venous thromboembolism development. *J. Hum. Genet.* 68, 805–812. doi:10.1038/s10038-023-01186-6

Li, Y. Y., Fang, J., and Ao, G. Z. (2017). Cathepsin B and L inhibitors: a patent review (2010 - present). *Expert Opin. Ther. Pat.* 27 (6), 643–656. doi:10.1080/13543776.2017.1272572

Mälarstig, A., Grassmann, F., Dahl, L., Dimitriou, M., McLeod, D., Gabrielson, M., et al. (2023). Evaluation of circulating plasma proteins in breast cancer using Mendelian randomisation. *Nat. Commun.* 14 (1), 7680. doi:10.1038/s41467-023-43485-8

Mata-Rocha, M., Rangel-López, A., Jiménez-Hernández, E., Morales-Castillo, B. A., González-Torres, C., Gaytan-Cervantes, J., et al. (2019). Identification and characterization of novel fusion genes with potential clinical applications in Mexican children with acute lymphoblastic leukemia. *Int. J. Mol. Sci.* 20 (10), 2394. doi:10.3390/ijms20102394

Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., et al. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature* 551 (7678), 92–94. doi:10.1038/nature24284

Neufeld, C., Filipp, F. V., Simon, B., Neuhaus, A., Schüller, N., David, C., et al. (2009). Structural basis for competitive interactions of Pex14 with the import receptors Pex5 and Pex19. *Embo J.* 28 (6), 745–754. doi:10.1038/emboj.2009.7

Oshi, M., Gandhi, S., Yan, L., Tokumaru, Y., Wu, R., Yamada, A., et al. (2022). Abundance of reactive oxygen species (ROS) is associated with tumor aggressiveness, immune response, and worse survival in breast cancer. *Breast Cancer Res. Treat.* 194 (2), 231–241. doi:10.1007/s10549-022-06633-0

Prat, A., and Perou, C. M. (2011). Deconstructing the molecular portraits of breast cancer. *Mol. Oncol.* 5 (1), 5–23. doi:10.1016/j.molonc.2010.11.003

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81 (3), 559–575. doi:10.1086/519795

Purrington, K. S., Slager, S., Eccles, D., Yannoukakos, D., Fasching, P. A., Miron, P., et al. (2014). Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. *Carcinogenesis* 35 (5), 1012–1019. doi:10.1093/carcin/bgt404

Rajkumar, T., Amritha, S., Sridevi, V., Gopal, G., Sabitha, K., Shirley, S., et al. (2022). Identification and validation of plasma biomarkers for diagnosis of breast cancer in South Asian women. *Sci. Rep.* 12 (1), 100. doi:10.1038/s41598-021-04176-w

Ren, F., Jin, Q., Liu, T., Ren, X., and Zhan, Y. (2023). Proteome-wide mendelian randomization study implicates therapeutic targets in common cancers. *J. Transl. Med.* 21 (1), 646. doi:10.1186/s12967-023-04525-5

Rohloff, J. C., Gelinas, A. D., Jarvis, T. C., Ochsner, U. A., Schneider, D. J., Gold, L., et al. (2014). Nucleic acid ligands with protein-like side chains: modified Aptamers and their use as diagnostic and therapeutic agents. *Mol. Ther. Nucleic Acids* 3 (10), e201. doi:10.1038/mtna.2014.49

Rosendahl, A. H., Björner, S., Ygland Rödström, M., Jirström, K., Borgquist, S., Ingvar, C., et al. (2021). Pre- and postoperative circulating igf-I, IGFBP-3, and IGFBP-7 levels in relation to endocrine treatment and breast cancer recurrence: a nested case-control study. *Front. Oncol.* 11, 626058. doi:10.3389/fonc.2021.626058

Sabe, H., Hata, A., Okada, M., Nakagawa, H., and Hanafusa, H. (1994). Analysis of the binding of the Src homology 2 domain of Csk to tyrosine-phosphorylated proteins in the suppression and mitotic activation of c-Src. *Proc. Natl. Acad. Sci. U. S. A.* 91 (9), 3984–3988. doi:10.1073/pnas.91.9.3984

Shu, X., Long, J., Cai, Q., Kweon, S. S., Choi, J. Y., Kubo, M., et al. (2020). Identification of novel breast cancer susceptibility loci in meta-analyses conducted among Asian and European descendants. *Nat. Commun.* 11 (1), 1217. doi:10.1038/s41467-020-15046-w

Shuptrine, C. W., Ajina, R., Fertig, E. J., Jablonski, S. A., Kim Lyerly, H., Hartman, Z. C., et al. (2017). An unbiased *in vivo* functional genomics screening approach in mice identifies novel tumor cell-based regulators of immune rejection. *Cancer Immunol. Immunother.* 66 (12), 1529–1544. doi:10.1007/s00262-017-2047-2

Suhre, K., McCarthy, M. I., and Schwenk, J. M. (2021). Genetics meets proteomics: perspectives for large population-based studies. *Nat. Rev. Genet.* 22 (1), 19–37. doi:10.1038/s41576-020-0268-2

Sun, B. B., Maranville, J. C., Peters, J. E., Stacey, D., Staley, J. R., Blackshaw, J., et al. (2018). Genomic atlas of the human plasma proteome. *Nature* 558 (7708), 73–79. doi:10.1038/s41586-018-0175-2

Sun, J., Zhao, J., Jiang, F., Wang, L., Xiao, Q., Han, F., et al. (2023). Identification of novel protein biomarkers and drug targets for colorectal cancer by integrating human plasma proteome with genome. *Genome Med.* 15 (1), 75. doi:10.1186/s13073-023-01229-9

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

Swerdlow, D. I., Holmes, M. V., Kuchenbaecker, K. B., Engmann, J. E., Shah, T., Sofat, R., et al. (2012). The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet* 379 (9822), 1214–1224. doi:10.1016/s0140-6736(12)60110-x

Takaya, K., Asou, T., and Kishi, K. (2023). Cathepsin F is a potential marker for senescent human skin fibroblasts and keratinocytes associated with skin aging. *Geroscience* 45 (1), 427–437. doi:10.1007/s11357-022-00648-7

Trayes, K. P., and Cokenakes, S. E. H. (2021). Breast cancer treatment. *Am. Fam. Physician* 104 (2), 171–178.

van der Zee, J., Mariën, P., Crols, R., Van Mossevelde, S., Dillen, L., Perrone, F., et al. (2016). Mutated CTSF in adult-onset neuronal ceroid lipofuscinosis and FTD. *Neurol. Genet.* 2 (5), e102. doi:10.1212/nxg.0000000000000102

Vazquez-Ortiz, G., Pina-Sanchez, P., Vazquez, K., Duenas, A., Taja, L., Mendoza, P., et al. (2005). Overexpression of cathepsin F, matrix metalloproteinases 11 and 12 in cervical cancer. *BMC Cancer* 5, 68. doi:10.1186/1471-2407-5-68

Veyssière, H., Bidet, Y., Penault-Llorca, F., Radosevic-Robin, N., and Durando, X. (2022). Circulating proteins as predictive and prognostic biomarkers in breast cancer. *Clin. Proteomics* 19 (1), 25. doi:10.1186/s12014-022-09362-0

Võsa, U., Claringbould, A., Westra, H. J., Bonder, M. J., Deelen, P., Zeng, B., et al. (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* 53 (9), 1300–1310. doi:10.1038/s41588-021-00913-z

Wei, S., Liu, W., Xu, M., Qin, H., Liu, C., Zhang, R., et al. (2022). Cathepsin F and Fibulin-1 as novel diagnostic biomarkers for brain metastasis of non-small cell lung cancer. *Br. J. Cancer* 126 (12), 1795–1805. doi:10.1038/s41416-022-01744-3

Wex, T., Levy, B., Wex, H., and Brömme, D. (1999). Human cathepsins F and W: a new subgroup of cathepsins. *Biochem. Biophys. Res. Commun.* 259 (2), 401–407. doi:10.1006/bbrc.1999.0700

Wingo, T. S., Liu, Y., Gerasimov, E. S., Gockley, J., Logsdon, B. A., Duong, D. M., et al. (2021). Brain proteome-wide association study implicates novel proteins in depression pathogenesis. *Nat. Neurosci.* 24 (6), 810–817. doi:10.1038/s41593-021-00832-6

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innov. (Camb)* 2 (3), 100141. doi:10.1016/j.xinn.2021.100141

Wu, Y., Wang, Z., Yang, Y., Han, C., Wang, L., Kang, K., et al. (2023). Exploration of potential novel drug targets and biomarkers for small cell lung cancer by plasma proteome screening. *Front. Pharmacol.* 14, 1266782. doi:10.3389/fphar.2023.1266782

Xiao, T., Li, W., Wang, X., Xu, H., Yang, J., Wu, Q., et al. (2018). Estrogen-regulated feedback loop limits the efficacy of estrogen receptor-targeted breast cancer therapy. *Proc. Natl. Acad. Sci. U. S. A.* 115 (31), 7869–7878. doi:10.1073/pnas.1722617115

Zhang, C., Qin, F., Li, X., Du, X., and Li, T. (2022a). Identification of novel proteins for lacunar stroke by integrating genome-wide association data and human brain proteomes. *BMC Med.* 20 (1), 211. doi:10.1186/s12916-022-02408-y

Zhang, G. Q., He, C., Tao, L., and Liu, F. (2015). Role of DJ-1 siRNA in reverse sensitivity of breast cancer cells to chemotherapy and its possible mechanism. *Int. J. Clin. Exp. Pathol.* 8 (6), 6944–6951.

Zhang, H., Ahearn, T. U., Lecarpentier, J., Barnes, D., Beesley, J., Qi, G., et al. (2020). Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat. Genet.* 52 (6), 572–581. doi:10.1038/s41588-020-0609-2

Zhang, J., Dutta, D., Köttgen, A., Tin, A., Schlosser, P., Grams, M. E., et al. (2022b). Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat. Genet.* 54 (5), 593–602. doi:10.1038/s41588-022-01051-w

Zhang, W., Liu, Z., Xia, S., Yao, L., Li, L., Gan, Z., et al. (2021). GDI2 is a novel diagnostic and prognostic biomarker in hepatocellular carcinoma. *Aging (Albany NY)* 13 (23), 25304–25324. doi:10.18632/aging.203748

Zhou, Y., Zhang, Y., Lian, X., Li, F., Wang, C., Zhu, F., et al. (2022). Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Res.* 50 (D1), D1398–d1407. doi:10.1093/nar/gkab953

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., et al. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 48 (5), 481–487. doi:10.1038/ng.3538

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 21 February 2024
DOI 10.3389/fmolb.2023.1343979

Check for updates

OPEN ACCESS

EDITED BY
Qingyu Luo,
Dana–Farber Cancer Institute, United States

REVIEWED BY
Yabing Nan,
Chinese Academy of Medical Sciences, China
Kezhi Yan,
Cystic Fibrosis Foundation, United States
Hong Yu,
The University of Texas Health Science Center at San Antonio, United States

*CORRESPONDENCE
Julia Tutzauer,
julia.tutzauer@med.lu.se

RECEIVED 24 November 2023
ACCEPTED 21 December 2023
PUBLISHED 21 February 2024

CITATION
Tutzauer J, Larsson A-M, Aaltonen K, Bergenfelz C, Bendahl P-O and Rydén L (2024), Gene expression in metastatic breast cancer—patterns in primary tumors and metastatic tissue with prognostic potential. *Front. Mol. Biosci.* 10:1343979. doi: 10.3389/fmolb.2023.1343979

COPYRIGHT
© 2024 Tutzauer, Larsson, Aaltonen, Bergenfelz, Bendahl and Rydén. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Gene expression in metastatic breast cancer—patterns in primary tumors and metastatic tissue with prognostic potential

Julia Tutzauer[1]*, Anna-Maria Larsson[1], Kristina Aaltonen[2], Caroline Bergenfelz[3], Pär-Ola Bendahl[1] and Lisa Rydén[4,5]

[1]Department of Clinical Sciences Lund, Division of Oncology, Lund University, Lund, Sweden, [2]Division of Translational Cancer Research, Department of Laboratory Medicine, Lund University, Lund, Sweden, [3]Division of Experimental Infection Medicine, Department of Translational Medicine, Lund University, Malmö, Sweden, [4]Department of Clinical Sciences Lund, Division of Surgery, Lund University, Lund, Sweden, [5]Department of Surgery, Skåne University Hospital, Malmö, Sweden

Background: Metastatic breast cancer (MBC) is the main cause of breast cancer-related death. The outcome of MBC varies, and there is a lack of biomarkers to aid in prognostication. The primary aim of this study was to evaluate the prognostic value of gene expression (GEX) signatures in the primary tumor (PT) and distant metastasis (DM) for progression-free survival (PFS) and overall survival (OS). The secondary aim was to describe GEX changes through MBC evolution and to identify MBC subtypes.

Methods: RNA was extracted from the PT, lymph node metastasis (LNM), and DM from MBC patients in a prospective observational study (*n* = 142; CTC-MBC NCT01322893) and was subjected to GEX analysis retrospectively using the NanoString Breast Cancer 360™ panel. 31 continuous GEX variables in DMs and PTs were analyzed for PFS and OS by Cox regression analysis and Kaplan-Meier estimates. Multivariable Cox regressions were adjusted for number of DM sites and CTCs, visceral metastasis, ECOG status, age at MBC diagnosis and, in additional analyses, PAM50 subtype. Differential GEX analyses and Euclidean distances were used to describe subgroup differences and visualize within-patient heterogeneity.

Results: Compared to DM GEX, GEX of the PT was at least equally useful for predicting MBC outcome. The strongest marker for a favorable PFS, both when expressed in the PT and the DM was *AR*, even after adjustment for prognostic markers including PAM50. GEX signatures related to hormone responsiveness, including *ESR1*, *FOXA1*, *PGR*, and *AR* were favorable prognostic markers, and the p53 signature was unfavorable for PFS when expressed in PT or DM. The previously published PAM50MET signature was prognostic for both PFS and OS. We established five distinct DM GEX profiles where two associated with liver and bone metastases, respectively. Finally, we identified four DM GEX profiles able to identify MBCs with poor OS in this cohort.

Conclusion: GEX of both DM and PT are useful in MBC prognostication. GEX of *AR* adds prognostic information for MBC. Our descriptive analyses illuminate the biological differences between MBCs in relation to outcome and metastatic site.

KEYWORDS

metastatic breast cancer, gene expression, prognosis, primary tumors, lymph node metastasis, distant metastasis

1 Introduction

Even though advancements in diagnostics and treatments have improved survival in primary breast cancer, 20%–30% of all breast cancer patients will eventually develop a metastatic disease (Valachis et al., 2022). Metastatic breast cancer (MBC) is for most patients a disease without curative treatment options, and remains the foremost cause of breast cancer-related death. While the median overall survival of MBC is estimated to approximately 3 years (Grinda et al., 2021; Valachis et al., 2022), a considerable variability in prognosis and disease progression underscores a significant biological heterogeneity between patients. Despite this, there is a lack of prognostic biomarkers specifically tailored for MBC. The molecular biomarkers estrogen receptor alpha (ER) and human epidermal growth factor receptor 2 (HER2), initially linked to prognosis in primary breast cancer, are routinely used for prognostication and treatment prediction in MBC. The addition of gene expression (GEX) profiling, including the PAM50 subtypes, Prosigna®, OncoTypeDX®, and Mammaprint®, are useful in prognostication of primary breast cancer, but these profiles have not been validated for MBC. Therefore, there is a need for new biomarkers to improve prognostication and enable more personalized treatment approaches.

The marked diversity characterizing the clinical course of MBC is mirrored in its biological landscape. Another relevant contributing factor to this heterogeneity is the metastatic site (Brasó-Maristany et al., 2022), which furthermore is known to carry prognostic information. Notably, patients with bone metastasis as the exclusive metastatic site generally exhibit more favorable outcomes across all established molecular breast cancer subtypes (Bertho et al., 2021). Various models addressing the relationship between GEX and metastatic site of breast cancer have been formulated, serving both predictive purposes (Albaradei et al., 2021) and for exploratory elucidation of the biological underpinnings of the metastatic niches (Brasó-Maristany et al., 2022). However, study cohorts have overall been limited in size, methodologies have varied between studies, and findings have rarely been validated in independent cohorts. As a result, the molecular identities of MBC remain poorly elucidated and are of limited clinical use in prognostication of the individual patient.

In this study, we used a unique MBC cohort encompassing biopsies from both the primary tumor (PT), lymph node metastasis (LNM; defined as regional lymph node metastasis diagnosed and retrieved at the time of primary surgery for the PT), and distant metastasis (DM), to address the primary aim of evaluating the prognostic significance of GEX in MBC as candidates for clinical implementation.

This unique material also provides an opportunity to investigate MBC in terms of DM subtypes and tumor evolution over time. Thus, the secondary aim was exploratory: to decipher GEX in the MBC setting; to describe its successive changes through MBC evolution, from PT and LNM to the DM. Finally, using data from the secondary aim, we sought to identify biologically relevant molecular subtypes identified in MBC and investigate their prognostic potential.

The identification of novel prognostic markers and an enhanced comprehension of the molecular variants of MBC have great potential to enhance prognostic accuracy for the individual patient, to enable personalized treatment strategies, and could lead to the identification of novel targets for drug development.

2 Materials and methods

2.1 Study population

The current study is based on a cohort of MBC patients enrolled in the prospective observational CTC-MBC trial focusing on circulating tumor cells (CTCs), available at Clinical-Trials.gov, NCT01322893 (Larsson et al., 2018). The inclusion criteria were a diagnosis of MBC with a life expectancy of >2 months, ECOG performance status 0–2, and an age of 18 years or older. Patients were excluded if unable to understand the study information, if they had been diagnosed with other malignant disease in the past 5 years, or if they had undergone prior systemic treatment for metastatic disease. A detailed flowchart for the patients and material used in the current study is shown in Figure 1. In total, 156 patients fulfilled the inclusion criteria and were enrolled in the original study.

LNMs were defined as regional lymph node metastasis diagnosed at the time of diagnosis of the PT. Information regarding the biopsy sites of the DMs were acquired from pathology reports. Metastatic biopsies labeled “Lung” included DMs from the lung or pleura, whereas the label “Bone” included metastases from the bone or the bone marrow. Only non-regional lymph node metastasis that were obtained at the time of the diagnosis of DM were included to the group of metastatic biopsies classified to originate from “Lymph node.”

Immunohistochemical staining of ER, PR, and HER2 on tumor samples were performed according to clinical standard practice and assessed by board-certified pathologists (Larsson et al., 2018). HER2 status was considered positive if amplified by *in situ* hybridization or defined as 3+ by immunohistochemistry.

2.2 GEX analysis

Macrodissection of tumor tissue and RNA isolation has been described in detail previously (Jørgensen et al., 2021). Briefly, areas with representative invasive breast carcinoma tissue were selected from formalin-fixed, paraffin-embedded tumor sections and extracted by microdissection. After RNA extraction, GEX was quantified on a NanoString nCounter® SPRINT Profiler (NanoString Technologies Inc., Seattle, WA, United States) using the NanoString Breast Cancer 360™ assay (BC360), which has been described in detail earlier (Jørgensen et al., 2021).

In total, GEX data were successfully acquired for $n = 269$ tumor samples: $n = 124$ PT, $n = 71$ LNM, and $n = 74$ DM samples from $n = 142$ patients (Figure 1). In cases where GEX data acquired from duplicate biopsies from the same tumor was available, the mean GEX data of the duplicates was used ($n = 21$ PT and $n = 1$ LNM). When patients had GEX data from biopsies of multiple DMs ($n = 7$), only the biopsy from the chronologically first diagnosed DM was included. Matched PT-LNM pairs were available for 62 patients, PT-DM pairs for 62 patients, LNM-DM pairs for 39 patients, and PT-LNM-DM triplets for 36 patients.

The expression of GEX signatures and single genes were normalized to z-scores using the sample mean and standard deviation (SD). At the time of data acquisition, the BC360 assay covered 757 individual genes (where six could not be evaluated; *CETN2, CDK4, NEIL2, PMS2, ERCC1,* and *PDCD1*) and 22 multigene signatures, including the four PAM50 signatures.

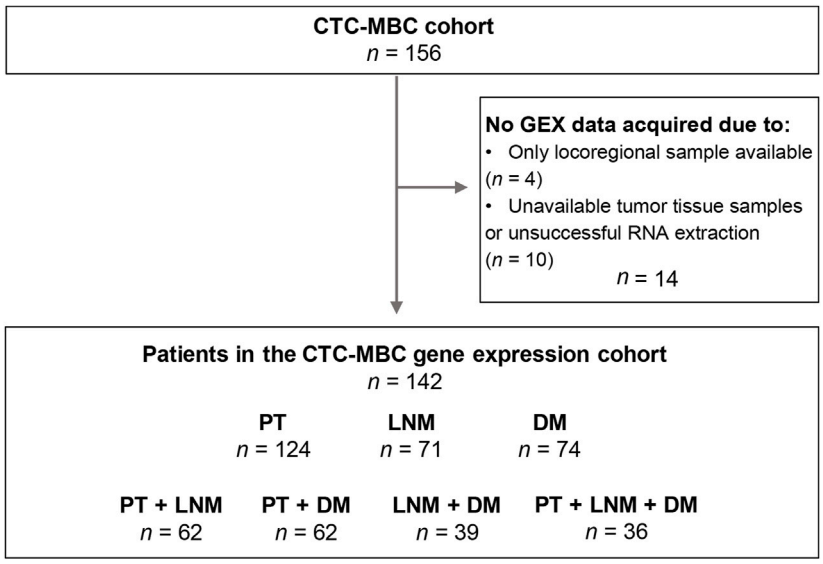


FIGURE 1
Inclusion flowchart of the study. CTC-MBC, circulating tumor cells in metastatic breast cancer; GEX, gene expression; PT, primary tumor; LNM, lymph node metastasis; DM, distant metastasis.

The PAM50 signatures were excluded from exploratory and prognostic analyses as their prognostic value in this cohort has been described previously (Jørgensen et al., 2021). In total, 31 GEX variables were included in the prognostic analyses; 19 multigene signatures and 12 single genes, as defined by the BC360 biological signature set. Supplementary Table S1 contains additional information on the BC360 panel, including a description of the multigene signatures and their acronyms (Supplementary Table S1, Sheet 1), all the individual genes included in the panel (Supplementary Table S1, Sheet 2), and all genes constituting the GEX signatures of the biological signature set (Supplementary Table S1, Sheet 3).

2.3 Correlation plots

Correlation plots were created using the *corrplot* R package.

2.4 Gene ontology enrichment analyses

Gene Ontology (GO) enrichment analyses were performed using the R packages *clusterProfiler* and *org. Hs.eg.d*, from the biological processes subontology using a cutoff at $p = 0.01$. The functional analyses were visualized by dot plots and gene-concept networks using the R package *enrichplot*.

2.5 GEX clustering and heatmaps

To identify GEX patterns, we employed visualization by heatmap and k-means clustering. For more stable clustering of single genes, we focused on the 100 genes with the highest standard deviation (SD). The heatmap was generated using the R package *ComplexHeatmap* based on DM GEX data transformed to *z*-scores. The clusters in rows and columns of heatmaps as well as the cluster profiles used for prognostic evaluation were rendered using k-means clustering, based on 1000 repeated runs for optimal reproducibility.

2.6 Phylogenetic trees and enrichment analyses

GEX data from triplets of PT, LNM, and DM ($n = 36$) were extracted and transformed to *z*-scores. Based on all 751 genes, distance metrics between PT, LNM, and DM were calculated for each patient using Euclidean distance. Phylogenetic trees were generated by neighbor-joining (NJ) estimation using the R package *ape* and visualized using *ggtree*. To illustrate the PAM50 status of the tumors in the phylogenetic trees, we used the scores from the PAM50 analysis provided by NanoString. These scores range between 0–1, with a high score indicating that the tumor is similar to the subtype, and a low score indicate that the tumor is less similar to the subtype.

2.7 Differential gene expression analysis

Differential GEX analysis by linear models were performed using the R package *limma*. Genes with an FDR-adjusted *p*-value (the *q*-value) below 0.05 were considered differentially expressed. Additional methodological details are found in the Supplementary Tables S2, S3.

2.8 Stepwise logistic regression model predicting bone metastases

Following differential GEX analysis, differentially expressed genes were subjected to a stepwise logistic regression algorithm minimizing the models' Akaike information criterion (AIC) by backward selection, using the R package *MASS*, allowing 10 000 steps. The binary outcome variable was whether the tumor metastasized to bone or not. Luminal status of the PT was included in the stepwise regression, but not forced into the final model. The discriminatory performance of the final model was illustrated using a ROC curve and summarized as the area under curve (AUC). To address the likelihood that this model was overfitted, we repeated the prediction model development pipeline 100 000 times with random permutations of the outcome variable. A higher AUC for the true model compared to the mean of the 100 000 mock models was interpreted as indicative of a real signal in the true model. As references, several additional models including a reduced number of descriptors were also fitted.

2.9 Calculation of the PAM50MET score

The PAM50MET score was calculated as described in the original article (Prat et al., 2020). The coefficients used in the calculations are found enclosed in Supplementary Table S4. Primarily, the PAM50MET score was calculated based on DM GEX. For patients where DM GEX data was not available, GEX data from PT was used ($n = 61$), as described by the authors (Prat et al., 2020). Complete data for calculation based on DM GEX was available for almost all DM samples ($n = 72$). Additional prognostic analyses were performed in the ER+, HER2- subgroup, where the calculations were made based on $n = 56$ DM, and $n = 25$ PT. Subtype was primarily determined by the ER and HER2 status of the DM, but in lack of DM data, the subtype of PT was used.

2.10 Statistical analysis

For testing successive changes in GEX through tumor progression, GEX of PTs, LNMs, and DMs were compared using ANOVA. The relationship between GEX signatures and the number of CTCs measured prior to treatment for MBC was evaluated using linear regression models. GEX was assessed in relation to clinical outcome using the Cox proportional hazards model where *z*-transformed GEX variables were entered as a continuous score. Cox proportional hazards models were fitted both for the full follow-up (FU) and using a time variable truncated at 2 years after diagnosis to better meet the proportional hazards assumption. The cutoff of 2 years was based on the median overall and progression-free survival of this cohort at the most recent follow-up. The endpoints were progression-free survival (PFS) and overall survival (OS). Progression was defined as progressive disease and non-progression was defined as stable disease, partial response, or complete response according to modified Response Evaluation Criteria In Solid Tumors (RECIST) 1.1 (Eisenhauer et al., 2009; Larsson et al., 2018). The multivariable Cox regressions were adjusted for the number of metastatic sites, the presence of visceral metastasis, CTC count of ≥5 cells per 7.5 mL blood at the time of MBC diagnosis, ECOG performance status, and age at MBC diagnosis. Output from multivariable analyses adjusting for PAM50 subtype of the DM is available as Supplementary Tables S2, S3. A $p < 0.05$ were considered statistically interesting, but given the exploratory nature of the study, *p*-values should be carefully interpreted.

3 Results

3.1 Patient and tumor characteristics

Patient and tumor characteristics of the original cohort (Larsson et al., 2018) and the subcohort subjected to GEX analyses were compared in Table 1. The characteristics of the original cohort and the GEX cohort were well balanced, indicating that the GEX cohort is representative of the full cohort (Table 1).

No GEX signature correlated to number of CTCs measured prior to treatment for MBC (data not shown).

3.2 Relationships between GEX signatures in the PT, LNM, and DM

To improve our understanding of the interplay between the biological signatures in the BC360 panel in the context of MBC, and consequently establish a foundation that could aid in interpreting the findings of this study, we wanted to assess the relationship between the GEX signatures of the BC360 panel. To this end, we constructed correlation plots showing the GEX of genes and multigene signatures included in the biological signature set of BC360 for all three tumor sites (Figure 2). In general, the patterns were similar at all three sites. The GEX signature representing differentiation correlated strongly with a cluster of hormone-related GEX variables, including the genes *FOXA1*, *ESR1*, *AR*, and *PGR*, and the multigene signature ER signaling. In the PT, these variables all correlated strongly with the mast cell GEX signature, a trend that was not as pronounced in the LNM or DM. A second cluster represented immune-related genes, including the genes *CD274* (from here on referred to as PDL1), *PDCD1* (from here on referred to as PD1), *PDCD1LG2* (from here on referred to as PDL2), *TIGIT*, *IDO1*, and the signatures representing APM, T-reg, cytotoxic cells, macrophages, and CD8 T-cell signatures. Interestingly, the hormone-dependent GEX cluster and the immune-related gene cluster correlated negatively at all three tumor sites. The hormone-dependent genes also correlated negatively to GEX signatures related to the TNBC subtype and genetic instability, including BRCAness, claudin low, p53, and BRCAness/DNA scar, which is the BRCAness signature combined with a homologous recombination deficiency (HRD) signature. In general, the relationships between the GEX signatures followed a similar pattern at the different tumor sites. Due to the high level of correlation between the GEX signatures explored in this study, results must be carefully interpreted. More information on the multigene signatures and abbreviations used in multigene signature names is found in Supplementary Table S1.

TABLE 1 Patient and tumor characteristics in the full CTC-MBC cohort and the GEX cohort.

	Original cohort *n* (%)	GEX cohort *n* (%)
Patients in the cohorts	156	142
Age at PT diagnosis (years)		
<50	42 (26.9)	39 (27.5)
50–65	41 (26.3)	36 (25.4)
≥65	73 (46.8)	67 (47.2)
Tumor size PT		
1	57 (38.8)	49 (36.6)
2	51 (34.7)	48 (35.8)
3	20 (13.6)	20 (14.9)
4	19 (12.9)	17 (12.7)
Missing	9	8
Subtype PT (IHC-based)		
Hormone receptor-positive (ER+, HER2-)	92 (73.0)	81 (72.3)
HER2 positive	17 (13.5)	16 (14.3)
Triple-negative (ER-, HER2-)	17 (13.5)	15 (13.4)
Missing	30	30
ER status PT		
Positive	123 (83.7)	111 (83.5)
Negative	24 (16.3)	22 (16.5)
Missing	9	9
HER2 status PT		
Positive	18 (14.2)	17 (15.0)
Negative	109 (85.8)	96 (85.0)
Missing	29	29
Endocrine treatment PT		
Yes	96 (61.5)	87 (61.3)
No	60 (38.5)	55 (38.7)
Chemotherapy PT		
Yes	71 (45.5)	66 (46.5)
No	85 (54.5)	76 (53.5)
Antibody treatment PT		
Yes	5 (3.2)	4 (2.8)
No	151 (96.8)	138 (97.2)
Age at DM diagnosis (years)		
<60	48 (30.8)	45 (31.7)
60–70	45 (28.8)	39 (27.5)
≥70	63 (40.4)	58 (40.8)
ECOG status at DM diagnosis		
0	91 (60.7)	84 (60.4)
1	37 (24.7)	35 (25.2)
2	22 (14.7)	20 (14.4)
Missing	6	3
Subtype DM (IHC-based)		
Hormone receptor positive, HER2- (ER+, HER2-)	80 (70.2)	75 (70.1)
HER2+	16 (14.0)	16 (15.0)
Triple negative (ER-, HER2-)	18 (15.8)	16 (15.0)
Missing	42	35
Metastasis-free interval (MFI) (years)		
0 years (*De novo*)	31 (19.9)	29 (20.4)
>0, ≤3 years	28 (17.9)	25 (17.6)
>3 years	97 (62.2)	88 (62.0)

(Continued on following page)

TABLE 1 (*Continued*) Patient and tumor characteristics in the full CTC-MBC cohort and the GEX cohort.

	Original cohort *n* (%)	GEX cohort *n* (%)
Number of metastatic sites		
<3	109 (69.9)	97 (68.3)
≥3	47 (30.1)	45 (31.7)
Visceral metastases		
Yes	91 (58.3)	84 (59.2)
No	65 (41.7)	58 (40.8)
PFS follow-up for event-free patients (median, range) (years)	6.3 (0.64–7.4)	5.5 (0.64–7.0)
OS follow-up for event-free patients (median, range) (years)	7.4 (6.2–11)	8.1 (6.2–11)

Abbreviations: CI, confidence interval; DM, distant metastasis; ER, estrogen receptor; GEX, gene expression; HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; OS, overall survival; PFS, progression-free survival; PT, primary tumor.
*Defined as a DM, diagnosed together with the PT, i.e., with MFI = 0.

3.3 Diverse GEX patterns through tumor progression

To explore the biological profile at each step of tumor progression in MBC, we compared the GEX at each tumor site including all available data from the PTs, LNMs, and DMs. First, to address broader patterns at the different sites, we plotted the biological signatures of the BC360 panel (Supplementary Figures S1A, B). The patterns for GEX of the genes and multigene signatures were in general either that GEX; 1) successively decreased through tumor progression, which was observed mainly in signatures associated to a less aggressive tumor type, including mast cells, ER signaling, and *PGR*, or 2) increase in GEX in LNM compared to PT, often to decrease below the PT level in DM, the pattern of most immune-related signatures, including APM, cytotoxic cells, *IDO1*, PD1, PDL1/2, *TIGIT*, TIS, and T-reg, as well as the apoptosis signature. A similar pattern, but where the LNM GEX instead were lower than both PT and DM, was observed for the B7H3 gene. The pattern of downregulation of *ESR1* GEX was unclear in the full cohort (Supplementary Figure S1B). In line with previous reports (Encarnación et al., 1993; Johnston et al., 1995), the *ESR1* GEX decreased through tumor progression in patients with luminal PT that had undergone adjuvant endocrine treatment for their primary breast cancer (Supplementary Figures S2, n = 65). However, this decrease was noticed already in the LNM, where the tumor cells are still treatment naïve.

To address differences in GEX at the different sites at the gene level, we performed differential GEX analysis including all 751 individual genes of the BC360 panel. Genes with a $q < 0.05$ were considered differentially expressed. A total of 252 genes were differentially expressed in the DM compared to the PT, 341 genes in the DM compared to the LNM, and 333 in the LNM compared to the PT. The up- and downregulated genes in the LNM and DM compared to the PT are shown as gene-concept networks in Figure 3. Between the PT and the LNM, the functions of upregulated genes were mainly lymphocyte and mononuclear cell differentiation, cytokine signaling, adaptive immune response, and cell-cell adhesion (Figure 3A), possibly reflecting an increased immune infiltration at this site. In line with this, the most highly differentially expressed gene between PT and LNM was the frequently used B cell marker *CD19* (Wang et al., 2012), for which the expression was 8 times higher in the LNM than in the PT. The functions of downregulated genes in LNM compared to PT included, e.g. epithelial cell proliferation and gland development (Figure 3B). With regards to differences between the PT and the DM, genes upregulated in the DM compared to the PT were associated with functions such as epithelial cell proliferation, cell cycle progression, cell fate commitment, and morphogenesis of a branching epithelium (Figure 3C). Interestingly, some genes that were downregulated in the DM compared to the PT were also active in epithelial cell proliferation. Another biological function suggested to be downregulated in the DM compared to the PT was angiogenesis (Figure 3D). The most prominent GEX fold change in DM compared to PT was seen for *SFRP2*, which is involved in angiogenesis (van Loon et al., 2021), where DM expression was found to be around 6% of that observed in PTs. All genes that are differentially expressed between the sites with a $q < 0.05$ are found in Supplementary Table S2.

To summarize the GEX changes through tumor evolution of MBC, the patterns of the BC360 biological signatures illustrate how traits related to differentiation and less aggressive tumor behavior are successively downregulated through MBC progression. Both in terms of BC360 biological signatures and on the single gene level, it is clear that the tumor microenvironment (TME) of the LNM harbors an increased level of immune-related events compared to the PT. Furthermore, on the single gene level, a pattern appears that several angiogenesis-associated genes are downregulated in the DM compared to the PT.

3.4 Phylogenetic relationships between triplets of PT, LNM, and DM

To further explore the dynamic changes in GEX throughout tumor progression in different MBCs where GEX data was available for PT, LNM, and DM (n = 36), we calculated the GEX relatedness as Euclidean distance measures, and visualized the patterns as phylogenetic trees (Figure 4; Supplementary Figures S3, S4). Figure 4 presents representative MBC cases selected to cover a range of clinical and biological conditions. The PAM50 subtypes are presented both with the label and the subtype score calculated in the PAM50 analysis, the latter to illustrate how similar the tumor´s GEX is to that of the subtype. As previously reported for this cohort (Jørgensen et al., 2021), changes in PAM50 subtype were common through tumor progression, with a higher tendency of shifts toward

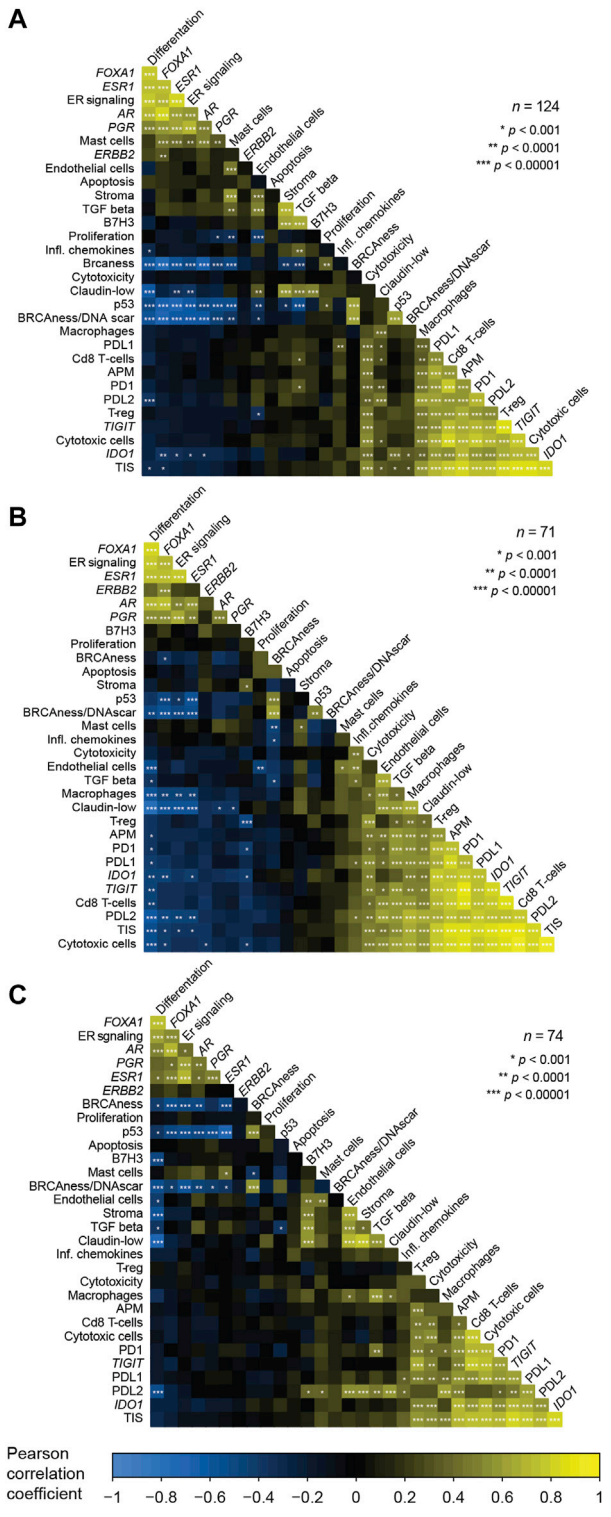


FIGURE 2
Correlations between the expression of genes and multigene signatures included in the BC360 signature set. Analyzed for each tumor site separately; the PT **(A)**, the LNM **(B)**, and the DM **(C)**. The correlation plots are ordered based on the first principal component and represent Pearson correlations between GEX of a total of 19 multigene signatures and 12 single genes standardized to z-values. Information of the abbreviations used for names of multigene signatures and which genes are included in each signature are found in Supplementary Table S1. Abbreviations: DM, *(Continued)*

FIGURE 2 (Continued)
distant metastasis; GEX, gene expression; LNM, lymph node metastasis; PT, primary tumor.

a more aggressive subtype. These shifts were observed both in cases where the PT displayed a low subtype score, indicating a relatively lower similarity to the subtype, (e.g., Figures 4E,L), and cases where the PT had a high subtype GEX score, indicating a higher similarity to the subtype, (e.g., Figures 4A,C). In most patients, the PT displayed a closer relationship to the LNM than to the DM. An interesting pattern was seen in one of the *de novo* MBCs (Figure 4I), where the PT and DM were basal-like, but the LNM had a luminal B subtype and showed considerably lower GEX resemblance to the PT than the DM. Furthermore, both *de novo* MBCs with GEX data from both PT, LNM, and DM exhibited subtype shifts (Figures 4F,I). It is interesting to note that the Euclidean distance between PT and LNM of one of the *de novo* MBC was longer than that of a PT and DM diagnosed 19 years apart (Figures 4B,I). Figures 4G,H illustrates that similar metastatic site, MFI, and PAM50 subtype of all three tumors, two MBCs can exhibit different patterns in terms of GEX evolution from the time of the PT and LNM to the DM. No statistically convincing relationship was found between Euclidean GEX distance of the MBC and clinical factors such as MFI, PAM50 subtype, or age at diagnosis (data not shown). In summary, these data illustrate the versatility of MBC in terms of evolutionary relationships between tumors, that MBC often shifts to more aggressive subtypes, and suggest that the PT and LNM most often share a higher level of similarity than the PT and DM.

3.5 Site-predictive potential of PT gene expression for bone metastasis

Considering that the metastatic site carries prognostic information in MBC, it is of clinical interest to predict which PTs are at risk of recurring at specific sites. Therefore, to identify potential PT GEX profiles related to the site of metastatic spread, we performed differential GEX analyses for each of the more prevalent metastatic sites, stratifying on a binary variable of whether the PT spread to a particular site or not. The sites included in the analyses were skin, liver, lung, lymph nodes, and bone. After adjusting for FDR, the only metastatic site where any differential GEX had a $q < 0.05$ was bone, where 85 genes were differentially expressed in PTs that later recurred in bone at initial MBC diagnosis compared to PTs that did not recur in bone (Supplementary Table S3).

To further evaluate the site-predictive potential of the 85 genes associated to bone metastasis, we included these genes in a stepwise logistic regression in relation to the binary outcome of whether the PT metastasized to bone. As the majority of the PTs in this cohort are luminal breast cancers, which are prone to spread to bone (Kennecke et al., 2010), we adjusted the model for luminal status of the PT to prevent the risk of the final model only reflecting luminal vs. non-luminal genes. The final model is available in Supplementary Table S5. Interestingly, luminal subtype was not included in the final model, suggesting that the genes had a stronger association to bone metastasis than a luminal subtype does. GO

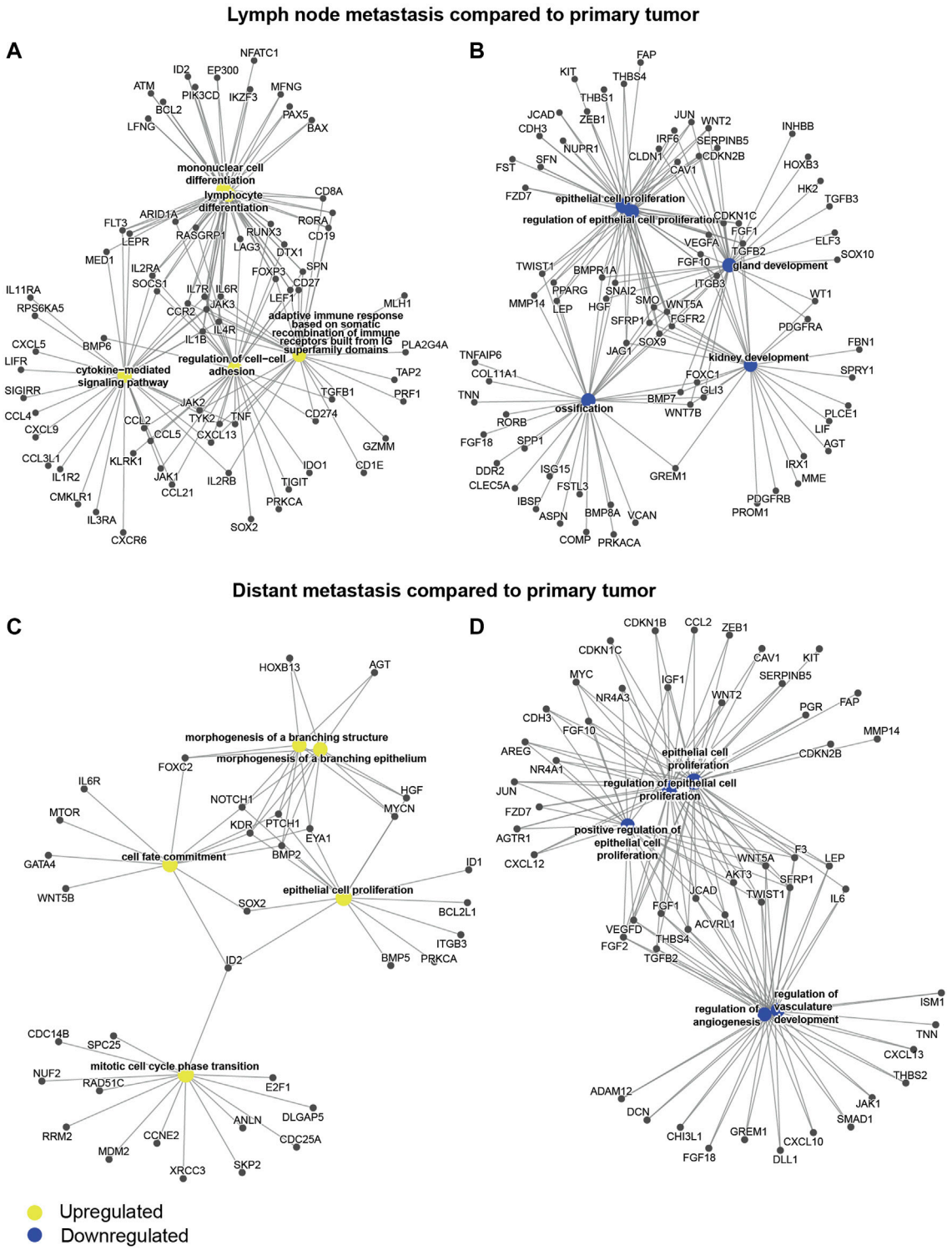


FIGURE 3
Differential gene expression through tumor evolution. Gene-concept networks presenting the results from differential gene expression analysis between tumor sites: **(A)** genes upregulated in the LNM compared to the PT, **(B)** genes downregulated in the LNM compared to the PT, **(C)** genes upregulated in the DM compared to the PT, **(D)** genes downregulated in the DM compared to the PT. The networks were based on all genes with a $q < 0.05$ from *limma* models, where the log fold change was >0 for upregulated genes **(A, C)**, or <0 for downregulated genes **(B, D)**. Abbreviations: DM, distant metastasis; LNM, lymph node metastasis; PT, primary tumor.

enrichment analysis of biological function showed that the genes included in the model were mainly involved in cell cycle regulation and mitosis (Supplementary Table S5). After comparing the resulting model with several other models restricted to between four and thirty predictors, as well as 100 000 mock models developed after random permutations of the outcome variables, i.e. under the

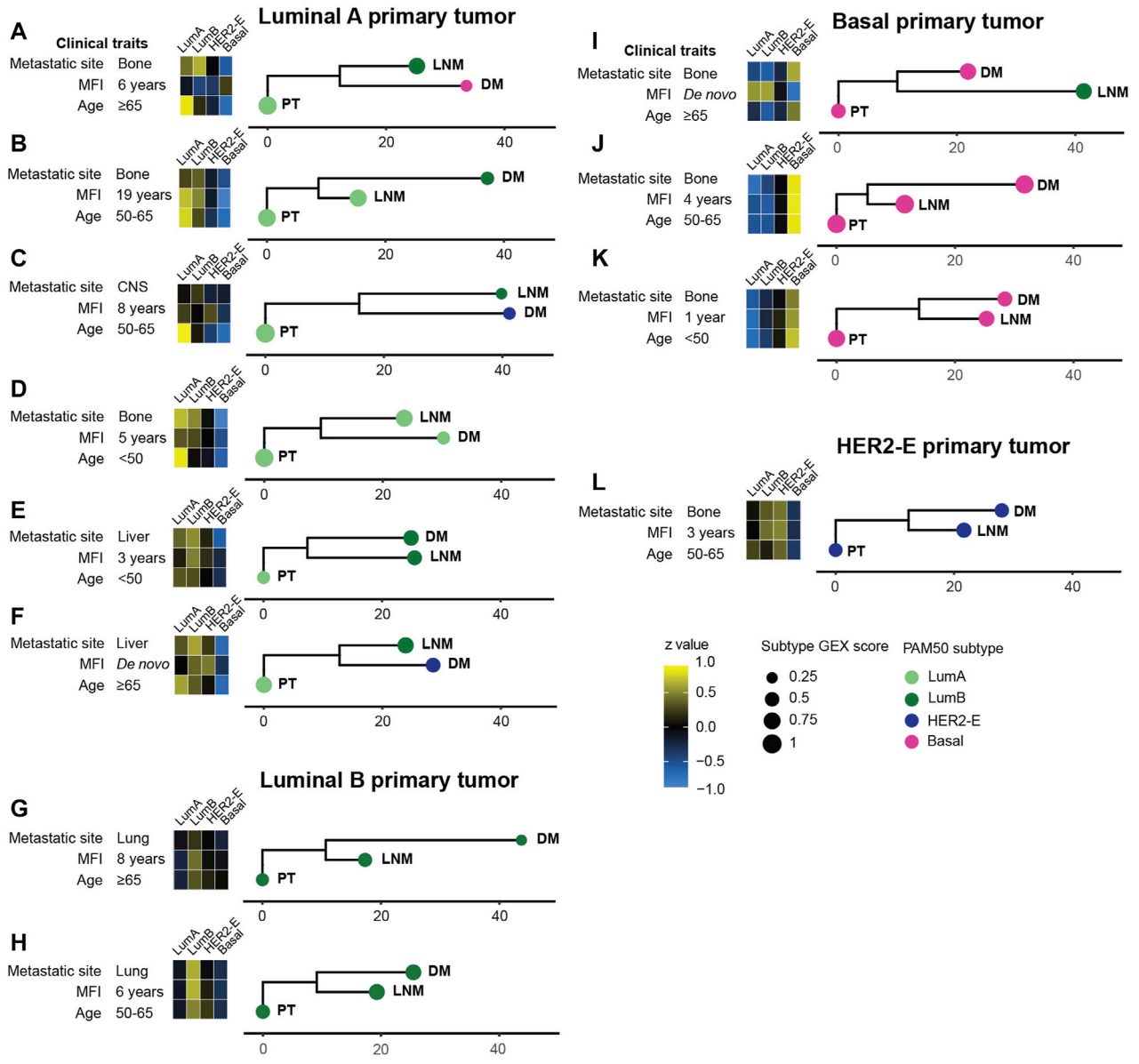


FIGURE 4
Phylogenetic trees for a selected subset of MBC triplets **(A–L)**. The relationship between tumors within patients was determined using the neighbor joining (NJ) algorithm from Euclidian distances of all 751 genes in the BC360 gene set, and manually rooted in the primary tumor. The horizontal scale bars represent the calculated Euclidian distance. Heatmaps show the gene expression of the PAM50 subtypes. The score of the PAM50 subtype, i.e., how well the GEX profile of the tumor agrees with the determined PAM50 subtype, is shown as the size of the bullet. Phylogenetic trees for the remaining MBC triplets are found in Supplementary Figures S3, S4. Abbreviations: DM, distant metastasis; LNM, lymph node metastasis; Lum, Luminal; MFI, metastasis-free interval; NJ, neighbor joining; PT, primary tumor.

null hypothesis of no association between expression of the selected genes and outcome (Supplementary Table S5), we concluded that the final large model is most likely overfitted, but that a high model complexity is required to accurately predict bone metastases in this material. In the light of this observation and the fact that the PT population in our dataset is relatively small and does not represent a clinical setting for patients with newly diagnosed primary breast cancer, it is appropriate to interpret the data from this model as hypothesis-generating, rather than truly predictive of bone metastases.

3.6 GEX of DMs at different metastatic sites

To address if the GEX of DMs is related to the metastatic site, we constructed a heatmap representing the GEX of the 100 genes of highest expression variability in the DMs (Figure 5). Four stable gene clusters were identified (Figure 5, column clusters). Gene set enrichment analysis showed that the core enrichment genes, i.e., the genes that account for the enrichment signal (Subra et al., 2005) of gene cluster 1 were related to epidermal development, embryonic processes, epithelial to mesenchymal

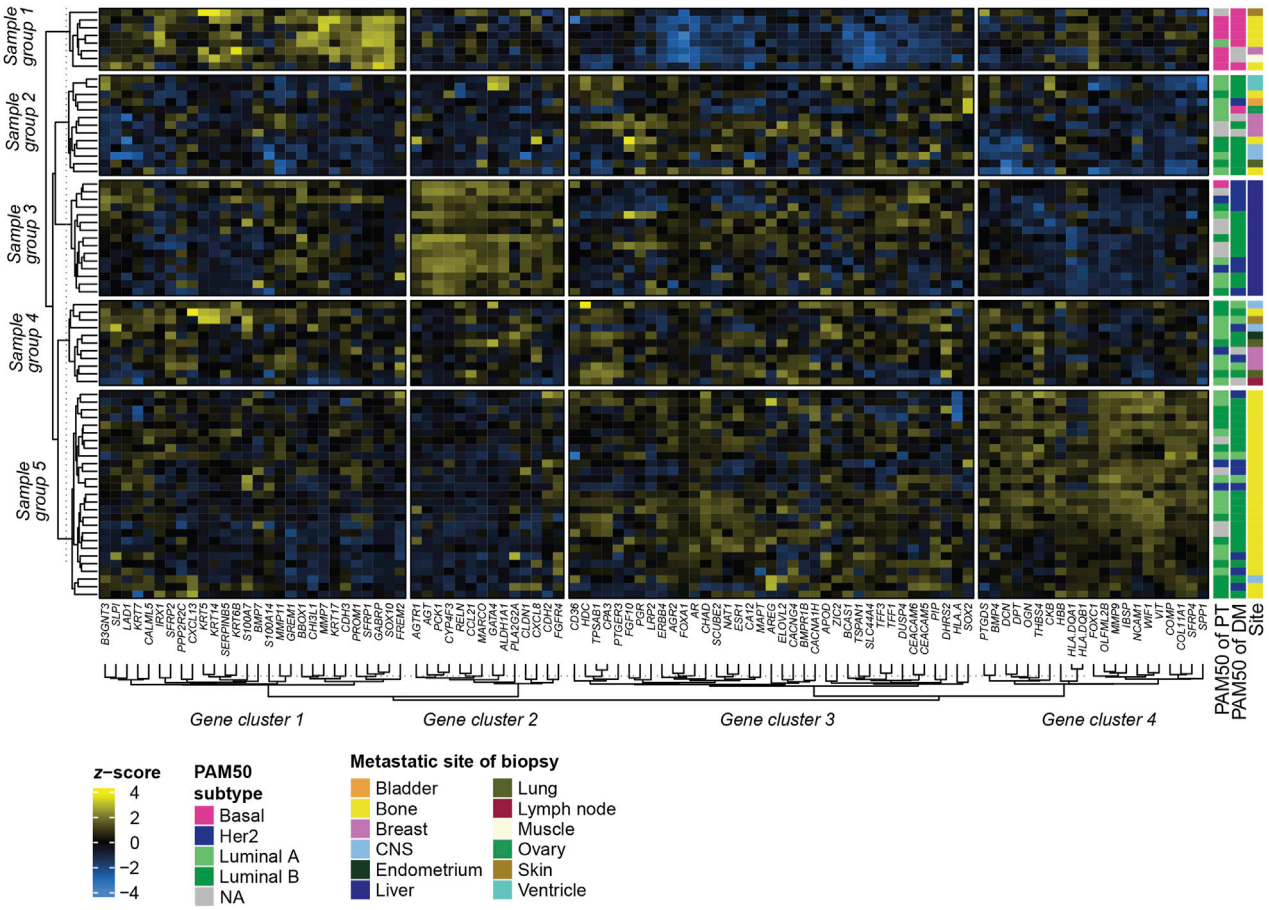


FIGURE 5
Heatmap of GEX in metastatic biopsies. Heatmap based on the 100 genes with the highest variability (SD) among the DMs. Rows represent different samples, and columns represent genes. Metastatic site refers to the site of the metastatic biopsy, not excluding that the patient may have had multiple metastatic sites. Clusters of rows (samples) and columns (genes) were rendered using k-means clustering and represent the consensus of 1000 repeated runs for optimal reproducibility. Abbreviations: DM, distant metastasis; GEX, gene expression; Lum, luminal; SD, standard deviation.

transition (EMT), and extracellular matrix (Figure 6A). Gene cluster 2 had several genes involved in MAPK signaling, macrophage-derived foam cell differentiation, and metabolic processes involved in, e.g., lipid metabolism (Figure 6B). Cluster 3 was the largest group and most notably involved in female sex differentiation and the development of the reproductive system and mammary glands. This cluster included several genes from the hormone sensitive cluster found in the correlation plots (Figure 2), *ESR1*, *PGR*, and *FOXA1* (Figure 6C). Finally, gene cluster 4 was involved in processes concerning skeletal bone, including ossification, skeletal system morphogenesis, cartilage development, and chondrocyte differentiation, but also processes in the extracellular matrix (Figure 6D).

In terms of sample (row) groups, it is highly interesting to note that regardless of the PAM50 profile of the PT, all DMs residing in the liver ended up in the same group (Figure 5, Sample group 3). This group had high expression of gene cluster 2, which consists of several metabolism-regulating genes such as *PCK1*, which is considered the master regulator of gluconeogenesis, and *AGTR1* and *AGT*, which are involved in NAD(P)H oxidase activity. Similarly, sample group 5 identified a large portion of the DMs located in bone, while only including one DM from a different site. In contrast, sample group 2 and 4 included metastases from a variety of sites. Two sample groups (Albaradei et al., 2021; Valachis et al., 2022) consisting of mainly bone metastases formed in separate parts of the heatmap. Sample group 1 comprises mostly basal DMs and is highly active in gene cluster 1 and inactive in gene cluster 3, whereas sample group 5 has a high expression in gene cluster 4. It is also interesting to note that the only two ventricle DMs, both of luminal B subtype originating from a luminal A PT, clustered as nearest neighbors in sample group 2, indicating related GEX profiles. In summary, these results show that among the most variably expressed genes in our gene set, some are highly associated with the metastatic site, and to some degree with the PAM50 subtype.

3.7 Prognostic role of GEX in metastatic breast cancer DM

The results of univariable and multivariable Cox regression models of the prognostic value of continuous DM GEX of the biological signatures in the BC360 panel are found in Figure 7. In

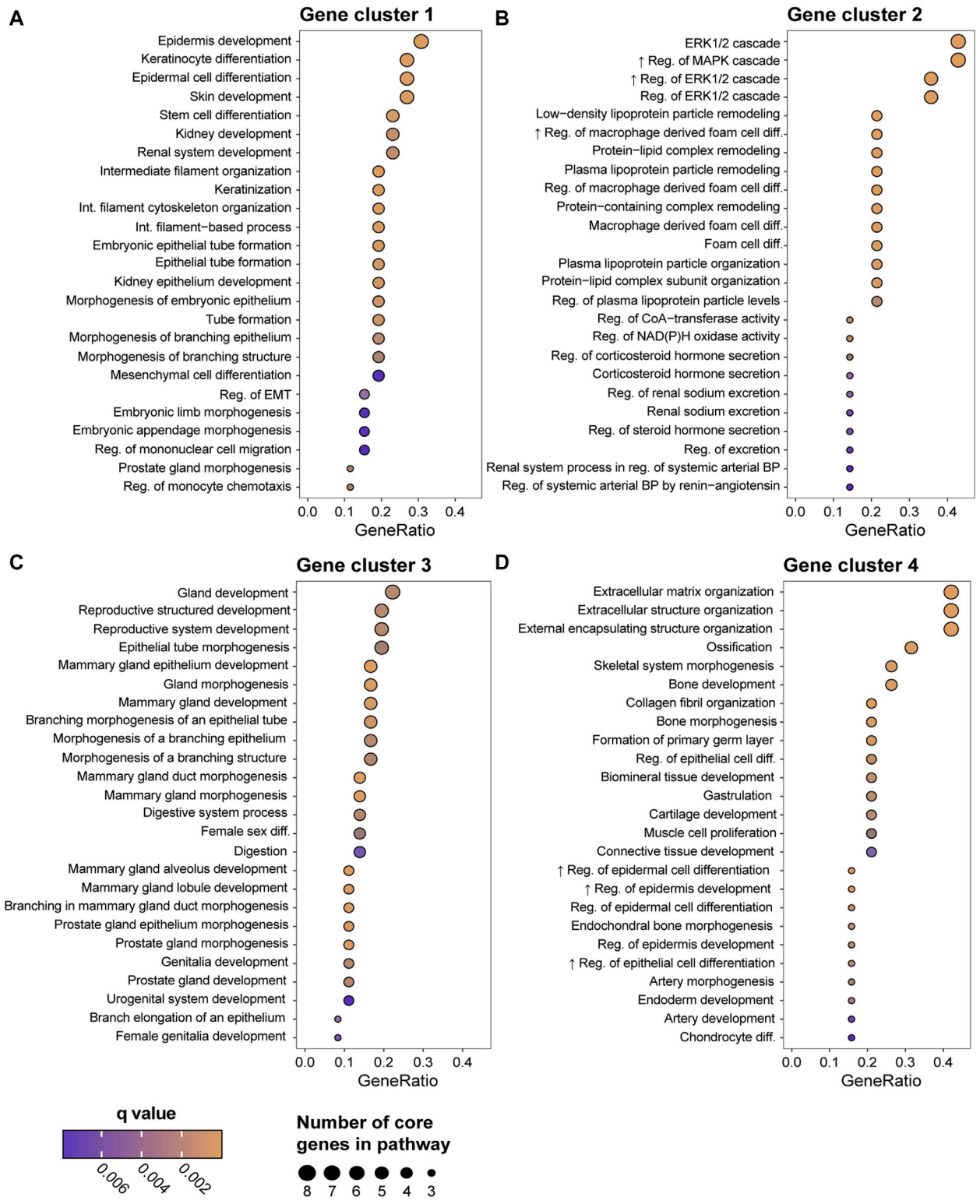


FIGURE 6
Gene Ontology (GO) enrichment analysis of the biological function of the four gene clusters generated in the Figure 5 heatmap; group 1 **(A)**, group 2 **(B)**, group 3 **(C)**, and group 4 **(D)**. Showing the top 25 biological functions based on GeneRatio, which is the ratio between the core enrichment genes and the total number of genes in the pathway and illustrates to which level the gene cluster is altered in the pathway. The number of core enrichment genes is indicated by the dot size. The *q* values represent *p* values adjusted for FDR. Abbreviations: FDR, false discovery rate; GO, Gene Ontology.

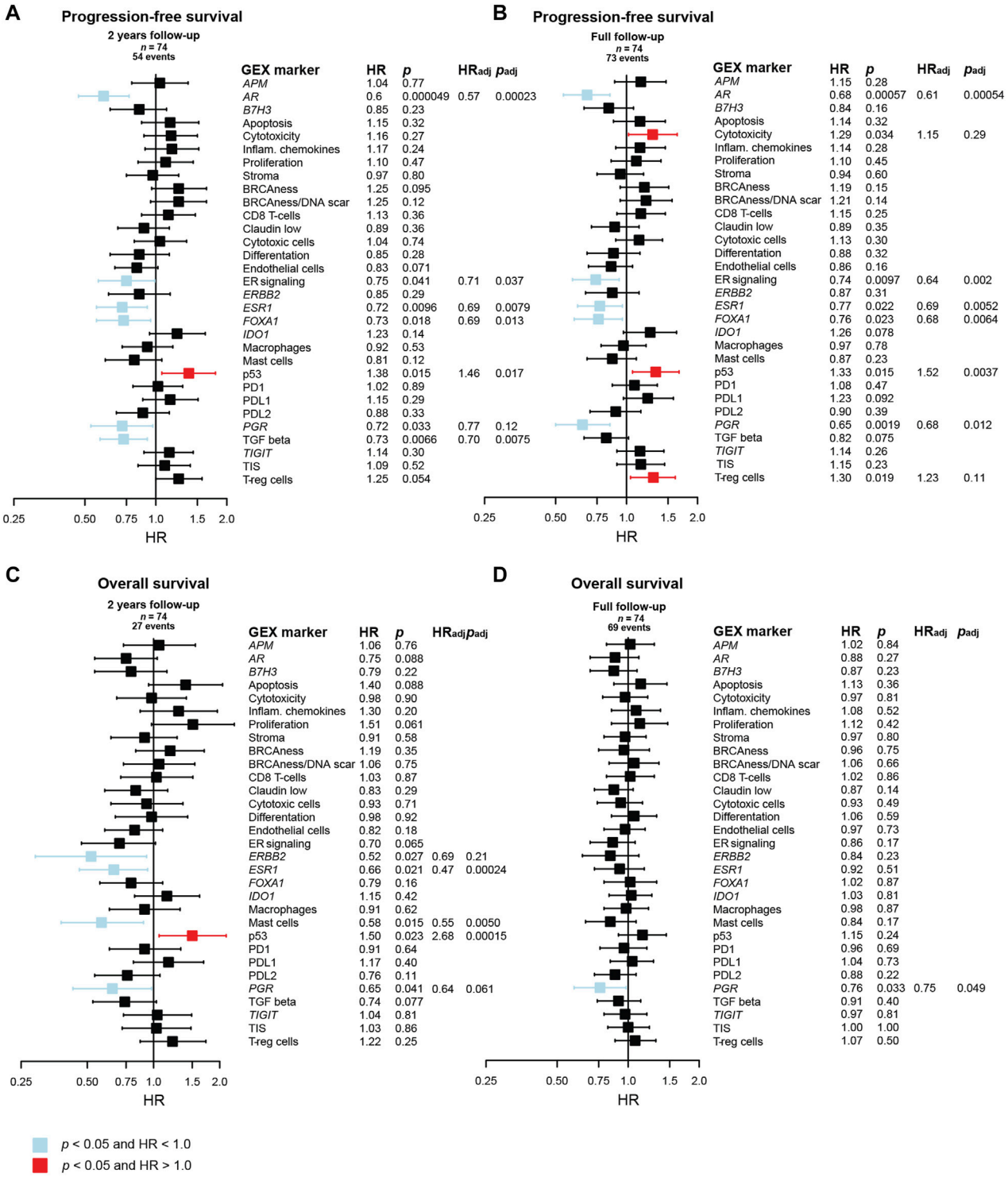


FIGURE 7
Prognostic value of the BC360 biological GEX signatures when expressed in the DM. Forest plots presenting the relationship between DM expression of the signatures and **(A)** PFS after 2 years of follow-up, **(B)** PFS after full follow-up, **(C)** OS after 2 years of follow-up, **(D)** OS after full follow-up. The plots, as well as the number of patients (*n*) and events are based on univariable Cox regression models with z-transformed GEX entered as continuous variables. HR is plotted with 95% CI. Colored based on the *p*-value. Presented *p*-values are crude. Multivariable Cox regression models were calculated for signatures with a $p < 0.05$ in the univariable Cox regression model. Adjustments were made for number of metastatic sites, visceral metastasis, ECOG performance status, number of CTC (≥5 per 7.5 mL blood at baseline) at MBC diagnosis, and age at MBC diagnosis, and the results are presented as HR_{adj} and p_{adj}. Abbreviations: CI, confidence interval; CTC, circulating tumor cells; DM, distant metastasis; GEX, gene expression; PFS, progression-free survival; MBC, metastatic breast cancer; OS, overall survival.

line with previous reports for primary breast cancer (Lundgren et al., 2023), a cluster of genes related to hormone-responsiveness were associated with a decreased risk in terms of PFS, both within the first 2 years after MBC diagnosis, and when considering the full follow-up (Figures 7A,B). These genes included *ESR1* ($HR_{Full\ FU}$ = 0.77, 95% CI = 0.62–0.96, p = 0.022), *PGR* ($HR_{Full\ FU}$ = 0.65, 95% CI = 0.50–0.85, p = 0.0019), as well as *FOXA1* ($HR_{Full\ FU}$ = 0.76, 95% CI = 0.60–0.96, p = 0.023) and *AR* ($HR_{Full\ FU}$ = 0.68, 95% CI = 0.54–0.85, p = 0.00057), as well as the signature representing ER signaling ($HR_{Full\ FU}$ = 0.74, 95% CI = 0.59–0.93, p = 0.0097). The p53 signature was found to associate with an increased risk of early progression within both 2 years and during the full follow-up ($HR_{Full\ FU}$ = 1.33, 95% CI = 1.06–1.67, p = 0.015). Interestingly, two other GEX signatures that associated to a worse PFS when considering the full follow-up were cytotoxicity ($HR_{Full\ FU}$ = 1.29, 95% CI = 1.02–1.64, p = 0.034) and T-reg cells ($HR_{Full\ FU}$ = 1.30, 95% CI = 1.04–1.61, p = 0.019).

In a multivariable Cox model adjusting for number of metastatic sites, visceral metastasis, ECOG performance status, number of CTCs (≥5 per 7.5 mL blood at baseline), and age at MBC diagnosis, *ESR1*, *AR*, *FOXA1*, ER signaling, and p53 remained prognostic for PFS with $p < 0.05$ both for 2 years and full follow-up. TGF beta was prognostic for 2 years PFS after adjustment, and *PGR* was prognostic for PFS at full follow-up after adjustment (Figures 7A,B). Many genes in the BC360 GEX panel are included in the PAM50 subclassification, why no multivariable adjustment was made for PAM50 subtype in the initial Cox regression model. However, to identify factors that may add prognostic information beyond PAM50, we also fitted the multivariable Cox model including PAM50 subtype of the DM in addition to adjusting for the clinical variables mentioned above. In this analysis, *AR* and TGF beta were identified as independent prognostic factors for favorable PFS at 2 years of follow-up, and only *AR* for the full follow-up (Supplementary Figure S5).

In relation to OS after 2 years of follow-up (Figure 7C), a better outcome was observed for patients with high DM expression of *ERBB2* (HR_{2yrs} = 0.52, 95% CI = 0.29–0.93, p = 0.027), *ESR1* (HR_{2yrs} = 0.66, 95% CI = 0.46–0.94, p = 0.021), mast cells (HR_{2yrs} = 0.58, 95% CI = 0.38–0.90, p = 0.015), p53 (HR_{2yrs} = 1.50, 95% CI = 1.06–2.14, p = 0.023), and *PGR* (HR_{2yrs} = 0.65, 95% CI = 0.43–0.98, p = 0.041). In multivariable analysis, *ESR1*, mast cells, and p53 remained prognostic for 2 years OS. The only prognostic factor found for OS with full follow-up was *PGR* (HR_{Full} = 0.76, 95% CI = 0.59–0.98, p = 0.033), which was also prognostic in multivariable analysis (Figure 7D). When including PAM50 in the multivariable Cox model, only the p53 signature remained prognostic of OS, and only after 2 years of follow-up (Supplementary Figure S5).

To further explore the prognostic value of *AR* GEX of the DM, *AR* quartiles were plotted in relation to PFS and OS in Kaplan-Meier curves (Supplementary Figure S6). In line with the results from the Cox regressions of the linear *AR* variable (Figure 7), the *AR* quartiles were ordered with the highest quartile (Q4) being visually associated with the best PFS. The pattern was particularly clear during the first 2 years after diagnosis for PFS ($HR_{2\ yrs}$ = 0.44, 95% CI = 0.22–0.91, p = 0.027), but not for OS ($HR_{2\ yrs}$ = 0.77, 95% CI = 0.27–1.94, p = 0.52).

3.8 Prognostic role of GEX in the PT of metastatic breast cancer

As tumor tissue from the DM is not always available, we also wanted to determine the prognostic value of GEX of the BC360 panel in PT in MBC (Figure 8). To address this, univariable and multivariable Cox regression models were fitted with the GEX of the biological signatures in the BC360 panel as continuous variables in relation to PFS and OS. Similar patterns to what was observed for DM GEX emerged, but interestingly, the prognostic value was even more pronounced in relation to the PT GEX.

The favorable prognostic value of the cluster of genes and GEX signatures related to hormone-responsiveness, including *AR*, ER signaling, *ESR1*, *FOXA1*, and *PGR*, was strong in relation to both PFS and OS after both 2 years and full follow-up. Similarly, the BRCAness signature emerged as a prognostically unfavorable marker for poor PFS and OS after both follow-up intervals. Similar results were observed for p53, but the effect was not as pronounced for OS after full follow-up. *ERBB2* and differentiation also associated to better PFS and OS after both follow-up intervals. Similar results were seen for the mast cell signature.

In a multivariable Cox model adjusting for number of metastatic sites, visceral metastasis, ECOG performance status, number of CTCs (≥5 per 7.5 mL blood) at baseline, and age at MBC diagnosis, all the genes and GEX signatures mentioned above remained associated to outcome with $p < 0.05$. Importantly, the strongest statistical evidence of prognostic value of PT GEX was observed for *AR*, which had p-values between 3.9×10^{-8} and 2.3×10^{-5} in both univariable and multivariable for both endpoints and both follow-up intervals.

As many genes in the BC360 GEX panel are included in the PAM50 subclassification, no multivariable adjustment was made for PAM50 subtype in the initial Cox regression model. However, to address if the prognostic value of the genes found to be associated to outcome was independent on PAM50 status, additional multivariable Cox regression analyses were performed. In these models, *AR* was the only gene or GEX signature that remained prognostic for 2-year PFS. Both *AR* and *ESR1* were prognostic for favorable PFS at full follow-up. Genes and GEX signatures associated to 2-year OS after adding PT PAM50 status to the Cox model were *ESR1* and p53, and after full follow-up, apoptosis, proliferation, *ESR1*, mast cells, and PDL1 remained associated to outcome (Supplementary Figure S7).

3.9 Prognostic performance of PAM50MET

The PAM50MET scores in the current study were primarily calculated based on DM GEX (n = 72) and based on PT GEX when DM data was not available (n = 61), as described by Prat et al. (2020) To evaluate the prognostic value of PAM50MET, we first assessed the scores in terms of quartiles, where a linear relationship was found between outcome and PAM50MET quartile, both for PFS (Figure 9A; logrank test for trend p = 0.0055) and OS (Figure 9B; logrank test for trend p = 0.0049). The relationship was more pronounced during the first 2 years of follow-up for both PFS (logrank test for trend p = 0.0008) and OS (logrank test for trend

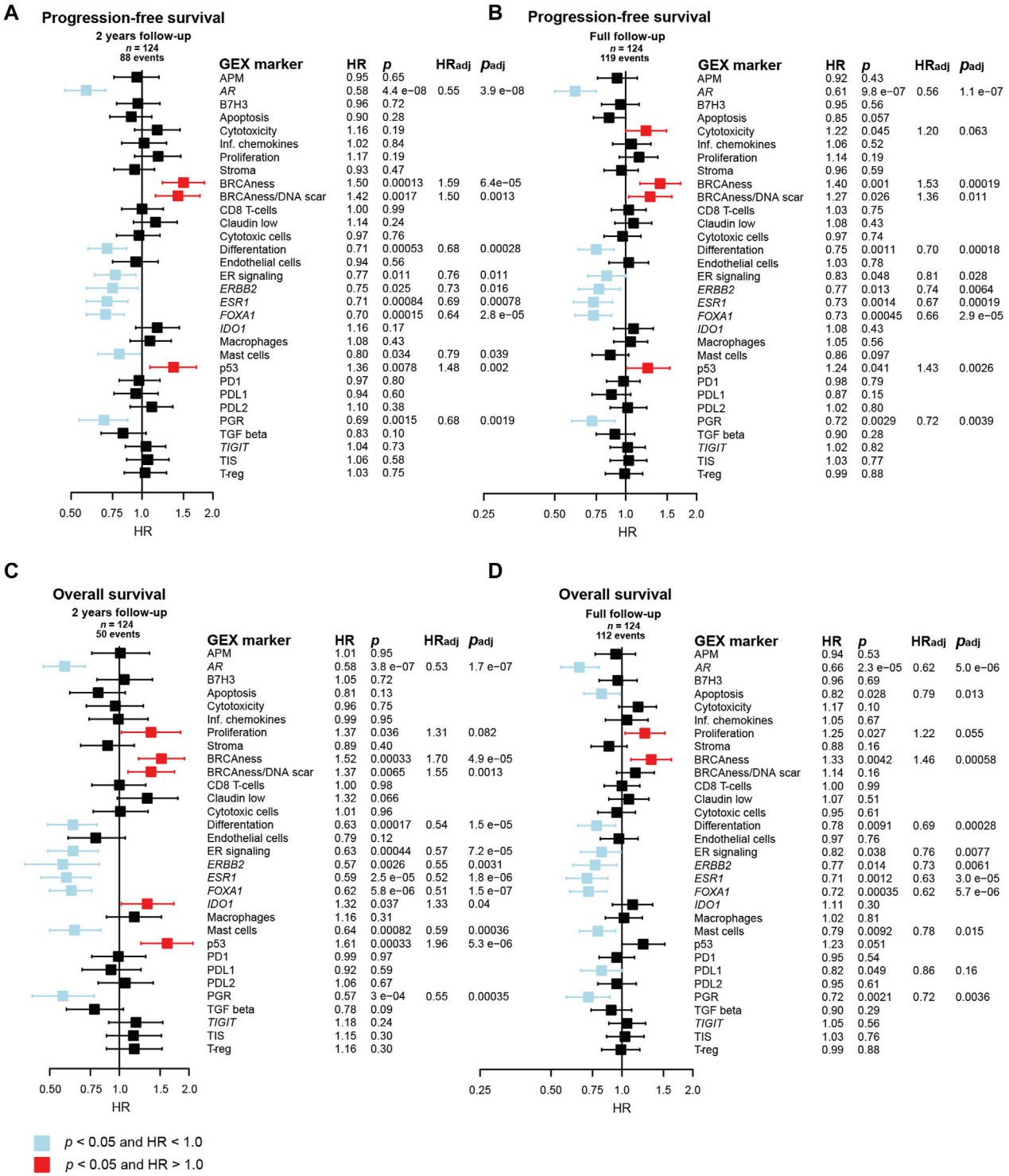


FIGURE 8
Prognostic value of the BC360 biological GEX signatures when expressed in the PT. Forest plots presenting the relationship between PT expression of the signatures and **(A)** PFS after 2 years of follow-up, **(B)** PFS after full follow-up, **(C)** OS after 2 years of follow-up, **(D)** OS after full follow-up. The plots, as well as the number of patients (*n*) and events are based on univariable Cox regression models with z-transformed GEX entered as continuous variables. HR is plotted with 95% CI. Colored based on the *p*-value. Presented *p*-values are crude. Multivariable Cox regression models were calculated for signatures with a $p < 0.05$ in the univariable Cox regression model. Adjustments were made for number of metastatic sites, visceral metastasis, ECOG performance status, number of CTC (≥5 per 7.5 mL blood at baseline) at MBC diagnosis, and age at MBC diagnosis, and the results are presented as HR_{adj} and p_{adj}. Abbreviations: CI, confidence interval; CTC, circulating tumor cells; GEX, gene expression; PFS, progression-free survival; PT, primary tumor; MBC, metastatic breast cancer; OS, overall survival.

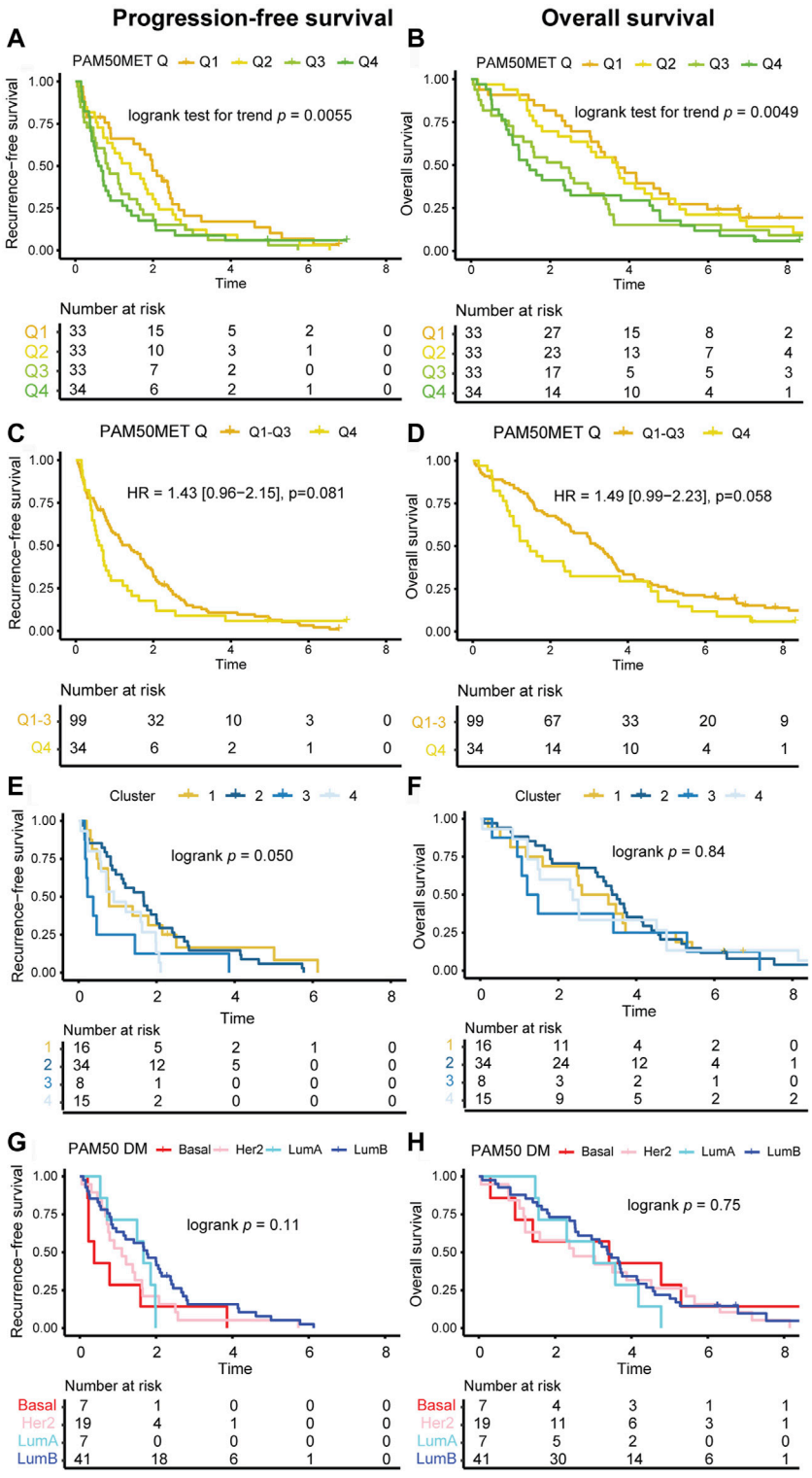


FIGURE 9
Kaplan-Meier estimates of GEX signatures. Kaplan-Meier estimates of the prognostic value in terms of PFS (left panels) and OS (right panels). Outcome in relation to **(A, B)** PAM50MET score at four levels defined by quartiles, calculated based on *n* = 72 DMs and *n* = 61 PTs, **(C, D)** PAM50MET stratified as quartile 1–3 (Q1-3) and quartile 4 (Q4), calculated based on *n* = 72 DMs and *n* = 61 PTs, **(E, F)** clusters aqcuired by applying k-means clustering on the Euclidean distance calculated based on the BC360 biological GEX signatures, **(G, H)** PAM50 subtype of the DM. Differences between groups were tested using either logrank tests (categorical nominal variables), logrank tests for trend (categorical ordinal variables), or Cox proportional hazards models (binary or continuous variables). HRs are presented with 95% CI and crude *p* values and correspond to analyses of the full follow-up. Abbreviations: CI, confidence interval; DM, distant metastasis; GEX, gene expression; HR, hazard ratio; PFS, progression-free survival; PT, primary tumor; MBC, metastatic breast cancer; OS, overall survival; Q, quartile.

p = 0.0001). We also compared PAM50MET as quartile four compared to quartiles one to three, as reported in the original article (Figures 9C,D) (Prat et al., 2020). The prognostic value during the full follow-up period was limited, but during the first 2 years of follow-up, quartile four associated to an inferior prognosis both in terms of PFS ($HR_{2\ yrs\ FU}$ = 1.70, 95% CI = 1.09–2.66, p = 0.019) and OS ($HR_{2\ yrs\ FU}$ = 2.28, 95% CI = 1.30–4.00, p = 0.0040). As the PAM50MET model was trained in ER+/HER2- MBC, we also performed the prognostic analyses including only this subgroup. In this subgroup, the PAM50MET calculations were based on n = 56 DMs and n = 25 PTs. In general, the patterns were similar to what was observed in the full cohort (Supplementary Figure S8). In summary, PAM50MET showed a promising potential as a prognostic tool for MBC in this material.

3.10 Prognostic performance of the clusters

To address if the biological information provided by the 31 biological signatures in the BC360 panel can be used for prognostic purposes, we performed k-means clustering of the Euclidean distance based on the DM GEX of the BC360 panel. We detected four sample clusters, which were analyzed in relation to outcome. When analyzing these groups in prognosticating PFS and OS with Kaplan-Meier estimates, the clusters had a clear separation (Figures 9E,F). Visually, the ability of the clusters to identify DMs with a poor OS was higher than the DM PAM50 subtype (Figures 9E–H). In summary, the application of unsupervised machine learning revealed distinct metastatic breast cancer subtypes with prognostic implications, underscoring the crucial link between gene expression patterns and clinical outcomes.

4 Discussion

Using a unique material representing a timeline from primary tumor (PT) to lymph node metastasis (LNM), and distant metastasis (DM), we delineate the dynamics and prognostic relevance of gene expression (GEX) in MBC. We find that *ESR1*, *AR*, and *FOXA1* GEX expression in DM, as well as in PT, are of favorable prognostic value, and that the multigene signature representing mutant p53 is unfavorable, independently of other established prognostic factors. Notably, we find that the prognostic performance of PT GEX was at least equal to that of the DM in this cohort, indicating that the PT can provide prognostic information for MBC. As DM GEX cannot always be acquired due to practical and financial limitations, these findings are of high clinical relevance. We confirm the prognostic value of the PAM50MET model (Prat et al., 2020), which incorporates both clinical and GEX-based variables and can be performed based on GEX from the DM as well as the PT.

We validate the prognostic utility of the PAM50MET model in MBC, in terms of both PFS and OS as endpoints. Importantly, we demonstrate its applicability beyond the ER+, HER2-patient subgroup for which it was initially developed (Prat et al., 2020). In contrast to the multigene signatures evaluated in this manuscript, the PAM50MET model also integrates clinical data. While our study highlights the promising prognostic value of GEX data in MBC, it appears unlikely that GEX will entirely replace or be surrogate of clinical prognostic indicators such as the patients' ECOG score and number of metastatic sites. This adds to the advantage of PAM50MET, as the panel illustrates how GEX can serve as a complement to clinical parameters.

In agreement with a pattern recently reported in primary breast cancer (Lundgren et al., 2023), several genes from the highly intercorrelating gene cluster associated to hormone responsiveness (*AR*, *PGR*, ER signaling, *ESR1*, and *FOXA1*) emerged as interesting prognostic markers for MBC, both when expressed in PT and in DM. These genes exhibited a higher expression in tumors with a favorable prognosis, independently on established prognostic markers. The strongest prognostic GEX marker in this study was *AR*, most notably in relation to PFS, and independently of other prognostic markers including DM PAM50 subtype. This suggests that *AR* GEX provides supplementary prognostic information and may have the potential to enhance the precision of MBC prognostication when integrated with the clinically established set of prognostic biomarkers. In line with prior findings (Encarnación et al., 1993; Johnston et al., 1995), we observed a downregulation of *ESR1* throughout tumor progression in endocrine-treated patients with luminal PT. However, our data reveal a decreased *ESR1* GEX in the LNM, where tumor cells remain naïve to endocrine treatment. This challenges the idea that ESR1 downregulation in breast cancer is propelled by selection pressure from endocrine treatment, suggesting that the shift from a hormone-stimulated phenotype serves additional biological purposes. One interesting idea is that growth stimulation by ER is limited in efficiency, and that prioritizing other proliferative signaling is evolutionary favorable as the tumor becomes more aggressive. Such a pattern may be more pronounced in PT and LNM of MBCs, as these could be expected to be more aggressive than the average PT and LNM.

The GEX data of this study is derived from bulk RNA. Although the tumor fractions used for GEX analyses in this study underwent macrodissection to isolate tumor tissue exclusively, this method does not exclude non-tumor cells present in the TME. This is evident in Supplementary Figure S1A, where a noticeable elevation in immune cell-related signatures in the LNM samples suggests a more prominent immune cell infiltration in this tissue. However, the relevance of non-tumor cells within the TME has gained increasing recognition, with tumor stroma playing a pivotal role in both tumor formation and progression (Xu et al., 2022). Notably, a high stromal component associates with poor outcome in several cancer types (Mesker et al., 2007; Almangush et al., 2021) including breast cancer (Roeke et al., 2017; Vangangelt et al., 2018; Vangangelt et al., 2020), underscoring the relevance of considering GEX of non-malignant, tumor-associated cells in tumor GEX analyses. Indeed, in line with previous data, the multigene signature representing T-reg cells associated to worse PFS (Walens et al., 2021). In contrast, the mast cell signature correlated positively to hormone-related genes and a favorable outcome, especially when expressed in the PT, aligning with observations in primary breast cancer (Lundgren et al., 2023). Further GEX studies employing single cell-resolution techniques retaining positional information, such as spatial transcriptomics, should be conducted to delve deeper into these concepts at a single-cell level.

The biological profile of DMs at different sites is poorly understood. We identified DM genes strongly associated to the liver and bone as metastatic sites. Interestingly, these genes align with the results from a study by Brasó-Maristany et al., which based on the PAM50 BC360 panel, identified an 18-gene signature specific for breast cancer liver metastases (Brasó-Maristany et al., 2022) that shares a total of nine genes with the 14-gene signature from the present study. Brasó-Maristany *et al.* also reported a 36-gene signature for bone metastases (Brasó-Maristany et al., 2022) sharing nine genes with the 21-gene cluster related to bone metastasis identified in this material. The overlap in results between studies using different methodologies strengthens the findings. While DMs at the same site expressing similar GEX profiles may not be surprising, the observation is in line with the "seed and soil" hypothesis: that distinct tumor cell clones possess varying selective affinities for metastatic sites (Paget, 1989; Langley and Fidler, 2011). However, our data cannot address the factor of causality–whether the similarities in GEX are due to natural selection, where the disseminated tumor cells must match the "soil" in order to adhere and thrive at the new site, or if the tumor cells are influenced by local factors inherent to the microenvironment once adhered to the new site. Regardless, these consistent patterns of expression at distinct metastatic sites potentially testify to a site-specific biological reliance on their functions, which could unlock new avenues for targeted treatments. In contrast, some sample clusters included DMs of various sites. This indicates that while some DM GEX profiles appear highly adapted or specialized to the metastatic site, other are tissue-agnostic and lack preferential affinity in this regard.

Our data on the relatedness of tumor triplets in terms of GEX suggest that despite similar clinical manifestations such as metastatic site, MFI, and PAM50 subtype of all three tumors, two MBCs can exhibit different patterns in terms of GEX evolution. This underlines the vast heterogeneity of MBC, and the gap of knowledge needed to be filled to be able to consider each MBC individually in outcome prediction and tailoring of systemic therapy. The international AURORA study, an initiative collecting MBC samples from paired PT and DM, circulating tumor DNA, and clinical data from 11 European countries and the United States, will provide a larger knowledgebase of temporal and spatial heterogeneity in MBC. Hopefully, this study will be a valuable contribution to the understanding MBC tumor evolution and heterogeneity, shedding light on how this can be applied to individualize treatment (Caballero et al., 2023).

After discovering that PT GEX carries relevant prognostic information regarding the outcome of MBC, we wanted to address if the PT GEX could predict future metastatic sites. Intriguingly, we identified 85 genes that were differentially expressed in PTs among patients with bone metastases at MBC diagnosis, suggesting predictive value. Therefore, we fitted a model utilizing PT GEX as predictors. The finalized model included a number of descriptors approximately tenfold higher than recommended by the Steyerberg rule-of-thumb to prevent overfitting clinical prediction models (Steyerberg, 2008). However, although the AUC estimation penalizes based on the number of descriptors, the AUC of our model drastically declined when we reduced the number of descriptors (Supplementary Table S5), indicating that a high number of genes contributed to the effective discrimination of the model in this cohort. To conclude, it is imperative to note that this is an exploratory prediction model that must be further validated in an independent dataset–preferably in a primary breast cancer cohort, as this would better simulate a clinical scenario where risk prediction of bone metastases holds practical significance. A prediction model of future metastatic site from PT GEX would be an intriguing addition to today's prognostic factors, paving the way for individualized follow-up strategies and adjuvant therapies.

Strengths of this study include the unique cohort of 142 MBC patients previously untreated for systemic disease. The cohort has 36 PT-LNM-DM matched tumor triplets with complete GEX data, providing a unique opportunity to describe GEX dynamics through tumor progression. One weakness of the study is the cohort size. Due to limited statistical power, adjustment for false discovery rate was not performed for all analyses. Furthermore, comparison between the prognostic performances of GEX from the PT and DM is complicated by the unequal statistical power of PT samples and DM. Thus, our data should be validated in a larger MBC cohort. Another weakness is the targeted GEX analyses, as the data reported in this study only reflects a selected fraction of the full tumor transcriptomics. The BC360 panel is curated to mirror biological events of relevance for primary breast cancer, but it is not optimized for MBC. As illustrated in this study, there are profound shifts in terms of GEX from the PT to its metastases. Furthermore, with each metastatic site exhibiting such distinct GEX profiles, it is likely that the transcriptomic landscapes of each site are profoundly different, and that the GEX coverage of LNM and DM may be suboptimal.

To conclude, the study fulfilled its primary aim: to assess the prognostic value of GEX in MBC–and the secondary aim: to describe the dynamics of GEX through MBC progression. The study results suggest GEX from both PT and DM as markers of potential prognostic utility. Notably, PT GEX emerges as a potent prognostic tool, offering a valuable alternative in patients where DM samples cannot be collected or analyzed. The most profound prognostic GEX markers were *AR*, *ESR1*, and *FOXA1* for a favorable prognosis, and the multigene signature p53 for an unfavorable prognosis. We also validate the performance of the PAM50MET model in predicting PFS and OS, again supporting the use of PT GEX for predicting MBC outcome in patients without available DM biopsies. The identification of specific genes associated with liver and bone metastases, along with the tissue-agnostic GEX patterns, underscores the complexity and heterogeneity of MBC. While our findings clearly support the use of GEX expression for prognostication of MBC they also contribute to our understanding of MBC. Future research exploring MBC GEX with spatial and single-cell resolution would be highly interesting to paint a more comprehensive picture.

Data availability statement

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

Ethics statement

The studies involving humans were approved by the Lund University Ethics Committee (2010/135). 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

JT: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. A-ML: Conceptualization, Funding acquisition, Methodology, Writing–review and editing. KA: Investigation, Methodology, Writing–review and editing. CB: Funding acquisition, Writing–review and editing. P-OB: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing–review and editing. LR: Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Funding acquisition, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Lund University provided Open access funding. This study was also supported by grants from the Swedish Cancer Society (Nos 19 0388 Pj and 22 2137 Pj), the Governmental Funding of Clinical Research within the Swedish National Health Service (LR: No. 2022-40304, A-ML: No. 2020-40634), Anna and Edwin Berger Foundation (No. 2019-2022), the Berta Kamprad Foundation (AML: no FBKS-2021-25-325 LR: nos FBKS-2020-21 142593; FBKS-2021-22 11336). Malmö Allmänna Sjukhuset (MAS) stiftelse för bekämpande av cancer, the Medical Faculty at Lund University.

Acknowledgments

Charlotte Levin Tykjær Jørgensen, Carina Forsare and the pathologist Anna Ehinger for expert technical contribution. All the patients in the CTC-MBC cohort and study nurses at Skåne University Hospital and Halmstad hospital. A preliminary version of the manuscript was uploaded as a preprint at Research Square (https://doi.org/10.21203/rs.3.rs-3691482/v1) (Tutzauer et al., 2022).

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/fmolb.2023.1343979/full#supplementary-material

References

Albaradei, S., Uludag, M., Thafar, M. A., Gojobori, T., Essack, M., and Gao, X. (2021). Predicting bone metastasis using gene expression-based machine learning models. *Front. Genet.* 12, 771092. doi:10.3389/fgene.2021.771092

Almangush, A., Alabi, R. O., Troiano, G., Coletta, R. D., Salo, T., Pirinen, M., et al. (2021). Clinical significance of tumor-stroma ratio in head and neck cancer: a systematic review and meta-analysis. *BMC Cancer* 21 (1), 480. doi:10.1186/s12885-021-08222-8

Bertho, M., Fraisse, J., Patsouris, A., Cottu, P., Arnedos, M., Pérol, D., et al. (2021). Real-life prognosis of 5041 bone-only metastatic breast cancer patients in the multicenter national observational ESME program. *Ther. Adv. Med. Oncol.* 13, 1758835920987657. doi:10.1177/1758835920987657

Brasó-Maristany, F., Paré, L., Chic, N., Martínez-Sáez, O., Pascual, T., Mallafré-Larrosa, M., et al. (2022). Gene expression profiles of breast cancer metastasis according to organ site. *Mol. Oncol.* 16 (1), 69–87. doi:10.1002/1878-0261.13021

Caballero, C., Irrthum, A., Goulioti, T., Cameron, D., Norton, L., and Piccart, M. (2023). International research to address the challenges of metastatic breast cancer: the AURORA Program (BIG 14-01). *npj Breast Cancer* 9 (1), 42. doi:10.1038/s41523-023-00548-9

Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., et al. (2009). New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* 45 (2), 228–247. doi:10.1016/j.ejca.2008.10.026

Encarnación, C. A., Ciocca, D. R., McGuire, W. L., Clark, G. M., Fuqua, S. A. W., and Osborne, C. K. (1993). Measurement of steroid hormone receptors in breast cancer patients on tamoxifen. *Breast Cancer Res. Treat.* 26 (3), 237–246. doi:10.1007/BF00665801

Grinda, T., Antoine, A., Jacot, W., Blaye, C., Cottu, P. H., Diéras, V., et al. (2021). Evolution of overall survival and receipt of new therapies by subtype among 20 446 metastatic breast cancer patients in the 2008-2017 ESME cohort. *ESMO Open* 6 (3), 100114. doi:10.1016/j.esmoop.2021.100114

Johnston, S. R., Saccani-Jotti, G., Smith, I. E., Salter, J., Newby, J., Coppen, M., et al. (1995). Changes in estrogen receptor, progesterone receptor, and pS2 expression in tamoxifen-resistant human breast cancer. *Cancer Res.* 55 (15), 3331–3338.

Jørgensen, C. L. T., Larsson, A. M., Forsare, C., Aaltonen, K., Jansson, S., Bradshaw, R., et al. (2021). PAM50 intrinsic subtype profiles in primary and metastatic breast cancer show a significant shift toward more aggressive subtypes with prognostic implications. *Cancers (Basel)* 13 (7), 1592. doi:10.3390/cancers13071592

Kennecke, H., Yerushalmi, R., Woods, R., Cheang, M. C. U., Voduc, D., Speers, C. H., et al. (2010). Metastatic behavior of breast cancer subtypes. *J. Clin. Oncol.* 28 (20), 3271–3277. doi:10.1200/JCO.2009.25.9820

Langley, R. R., and Fidler, I. J. (2011). The seed and soil hypothesis revisited--the role of tumor-stroma interactions in metastasis to different organs. *Int. J. Cancer* 128 (11), 2527–2535. doi:10.1002/ijc.26031

Larsson, A.-M., Jansson, S., Bendahl, P.-O., Levin Tykjaer Jörgensen, C., Loman, N., Graffman, C., et al. (2018). Longitudinal enumeration and cluster evaluation of circulating tumor cells improve prognostication for patients with newly diagnosed metastatic breast cancer in a prospective observational trial. *Breast Cancer Res.* 20 (1), 48. doi:10.1186/s13058-018-0976-0

Lundgren, C., Tutzauer, J., Church, S. E., Stål, O., Ekholm, M., Forsare, C., et al. (2023). Tamoxifen-predictive value of gene expression signatures in premenopausal breast cancer: data from the randomized SBII:2 trial. *Breast Cancer Res.* 25 (1), 110. doi:10.1186/s13058-023-01719-z

Mesker, W. E., Junggeburt, J. M., Szuhai, K., de Heer, P., Morreau, H., Tanke, H. J., et al. (2007). The carcinoma-stromal ratio of colon carcinoma is an independent factor for survival compared to lymph node status and tumor stage. *Cell Oncol.* 29 (5), 387–398. doi:10.1155/2007/175276

Paget, S. (1989). The distribution of secondary growths in cancer of the breast. 1889. *Cancer Metastasis Rev.* 8 (2), 98–101.

Prat, A., Tsai, Y. H., Pascual, T., Pare, L., Adamo, B., Vidal, M., et al. (2020). A prognostic model based on PAM50 and clinical variables (PAM50MET) for metastatic hormone receptor-positive HER2-negative breast cancer. *Clin. Cancer Res.* 26 (23), 6141–6148. doi:10.1158/1078-0432.CCR-20-2793

Roeke, T., Sobral-Leite, M., Dekker, T. J. A., Wesseling, J., Smit, V., Tollenaar, R., et al. (2017). The prognostic value of the tumour-stroma ratio in primary operable invasive cancer of the breast: a validation study. *Breast Cancer Res. Treat.* 166 (2), 435–445. doi:10.1007/s10549-017-4445-8

Steyerberg, E. W. (2008). *Clinical prediction models: a practical approach to development, validation, and updating*. Cham: Springer International Publishing.

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 (43), 15545–15550. doi:10.1073/pnas.0506580102

Tutzauer, J., and GitHub (2023). Data related to GEX analyses of the CTC-MBC study (NCT01322893). Available at: https://github.com/jtutz/CTC-MBC-GEX.

Tutzauer, J., Larsson, A. M., Aaltonen, K., Bergenfelz, C., Bendahl, P. O., and Rydén, L., (2022). *Gene expression in metastatic breast cancer – patterns with prognostic potential in primary tumors and metastatic tissue*. Durham, NC: © Research Square. doi:10.21203/rs.3.rs-3691482/v1

Valachis, A., Carlqvist, P., Ma, Y., Szilcz, M., Freilich, J., Vertuani, S., et al. (2022). Overall survival of patients with metastatic breast cancer in Sweden: a nationwide study. *Br. J. Cancer* 127 (4), 720–725. doi:10.1038/s41416-022-01845-z

Vangangelt, K. M. H., Green, A. R., Heemskerk, I. M. F., Cohen, D., van Pelt, G. W., Sobral-Leite, M., et al. (2020). The prognostic value of the tumor-stroma ratio is most discriminative in patients with grade III or triple-negative breast cancer. *Int. J. Cancer* 146 (8), 2296–2304. doi:10.1002/ijc.32857

Vangangelt, K. M. H., Tollenaar, L. S. A., van Pelt, G. W., de Kruijf, E. M., Dekker, T. J. A., Kuppen, P. J. K., et al. (2018). The prognostic value of tumor-stroma ratio in tumor-positive axillary lymph nodes of breast cancer patients. *Int. J. Cancer* 143 (12), 3194–3200. doi:10.1002/ijc.31658

van Loon, K., Huijbers, E. J. M., and Griffioen, A. W. (2021). Secreted frizzled-related protein 2: a key player in noncanonical Wnt signaling and tumor angiogenesis. *Cancer Metastasis Rev.* 40 (1), 191–203. doi:10.1007/s10555-020-09941-3

Walens, A., Olsson, L. T., Gao, X., Hamilton, A. M., Kirk, E. L., Cohen, S. M., et al. (2021). Protein-based immune profiles of basal-like vs. luminal breast cancers. *Lab. Invest.* 101 (6), 785–793. doi:10.1038/s41374-020-00506-0

Wang, K., Wei, G., and Liu, D. (2012). CD19: a biomarker for B cell development, lymphoma diagnosis and therapy. *Exp. Hematol. Oncol.* 1 (1), 36. doi:10.1186/2162-3619-1-36

Xu, M., Zhang, T., Xia, R., Wei, Y., and Wei, X. (2022). Targeting the tumor stroma for cancer therapy. *Mol. Cancer* 21 (1), 208. doi:10.1186/s12943-022-01670-1

frontiers | Frontiers in Oncology

TYPE Systematic Review
PUBLISHED 19 March 2024
DOI 10.3389/fonc.2024.1267596

Check for updates

OPEN ACCESS

EDITED BY
Li Chen,
Huazhong University of Science and Technology, China

REVIEWED BY
Chen Li,
King's College London, United Kingdom
Nguyen Quoc Khanh Le,
Taipei Medical University, Taiwan

*CORRESPONDENCE
Danke Su
sudanke33@sina.com

RECEIVED 03 November 2023
ACCEPTED 07 March 2024
PUBLISHED 19 March 2024

CITATION
Liu L, Liao H, Zhao Y, Yin J, Wang C, Duan L, Xie P, Wei W, Xu M and Su D (2024) CT-based radiomics for predicting lymph node metastasis in esophageal cancer: a systematic review and meta-analysis. *Front. Oncol.* 14:1267596. doi: 10.3389/fonc.2024.1267596

COPYRIGHT
© 2024 Liu, Liao, Zhao, Yin, Wang, Duan, Xie, Wei, Xu and Su. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

CT-based radiomics for predicting lymph node metastasis in esophageal cancer: a systematic review and meta-analysis

Liangsen Liu[1,2], Hai Liao[1], Yang Zhao[1], Jiayu Yin[1,3], Chen Wang[1], Lixia Duan[1], Peihan Xie[1], Wupeng Wei[4], Meihai Xu[3] and Danke Su[1]*

[1]Department of Medical Imaging Center, Guangxi Medical University Cancer Hospital, Nanning, China, [2]Department of Nuclear Medicine, The Fifth Affiliated Hospital of Guangxi Medical University, Nanning, China, [3]Department of Radiology, The Fifth Affiliated Hospital of Guangxi Medical University, Nanning, China, [4]Department of Radiology, The Second Affiliated Hospital of Guangxi Medical University, Nanning, China

Objective: We aimed to evaluate the diagnostic effectiveness of computed tomography (CT)-based radiomics for predicting lymph node metastasis (LNM) in patients diagnosed with esophageal cancer (EC).

Methods: The present study conducted a comprehensive search by accessing the following databases: PubMed, Embase, Cochrane Library, and Web of Science, with the aim of identifying relevant studies published until July 10th, 2023. The diagnostic accuracy was summarized using the pooled sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), and area under the curve (AUC). The researchers utilized Spearman's correlation coefficient for assessing the threshold effect, besides performing meta-regression and subgroup analysis for the exploration of possible heterogeneity sources. The quality assessment was conducted using the Quality Assessment of Diagnostic Accuracy Studies-2 and the Radiomics Quality Score (RQS).

Results: The meta-analysis included six studies conducted from 2018 to 2022, with 483 patients enrolled and LNM rates ranging from 27.2% to 59.4%. The pooled sensitivity, specificity, PLR, NLR, DOR, and AUC, along with their corresponding 95% CI, were 0.73 (0.67, 0.79), 0.76 (0.69, 0.83), 3.1 (2.3, 4.2), 0.35 (0.28, 0.44), 9 (6, 14), and 0.78 (0.74, 0.81), respectively. The results demonstrated the absence of significant heterogeneity in sensitivity, while significant heterogeneity was observed in specificity; no threshold effect was detected. The observed heterogeneity in the specificity was attributed to the sample size and CT-scan phases ($P < 0.05$). The included studies exhibited suboptimal quality, with RQS ranging from 14 to 16 out of 36. However, most of the enrolled studies exhibited a low-risk bias and minimal concerns relating to applicability.

Conclusion: The present meta-analysis indicated that CT-based radiomics demonstrated a favorable diagnostic performance in predicting LNM in EC. Nevertheless, additional high-quality, large-scale, and multicenter trials are warranted to corroborate these findings.

Systematic Review Registration: Open Science Framework platform at https://osf.io/5zcnd.

KEYWORDS

esophageal cancer, lymph node metastasis, computerized tomography, radiomics, diagnosis, meta-analysis

Introduction

Esophageal cancer (EC), a prevalent and deadly neoplasm, has been identified as the seventh most commonly diagnosed cancer globally in 2020, with 604,000 new cases reported and the sixth leading cause of death, with 544,000 fatalities (1). Patients with EC are often diagnosed at advanced stages, making them unsuitable for surgery and leading to a poor prognosis with low 5-year survival rates of only 20%-30% (2–4). Lymph node metastasis (LNM) has been identified as a vital prognostic determinant for patient survival (5–7). The eighth edition of the AJCC's International Staging Standard for Esophageal Cancer introduced a clinical staging system based on preoperative imaging; it incorporated the count of lymph node metastases in postoperative staging (8, 9). While pathological findings continue to be the gold standard for diagnosing LNM, lymph node biopsy represents an invasive procedure with a non-trivial incidence of complications (10). Therefore, accurate evaluation of lymph node status using non-invasive imaging methods is imperative for making informed treatment decisions and essential for precise prognostication.

The computed tomography (CT) scans is a widely used non-invasive imaging technique for acquiring preoperative and postoperative tumor-related data to assess lymph node status in EC patients (11). However, relying solely on morphological standards, including short diameter and shape measured by physicians with varying levels of clinical diagnostic expertise, conventional CT scans are inadequate for accurately identifying LNM. The accuracy of conventional CT scans for precise detection is suboptimal, with sensitivity ranging from 37.3% to 67.2% and specificity ranging from 63.9% to 96.4% (12). Furthermore, these criteria have limitations culminating in a markedly low diagnostic accuracy for normal-sized lymph nodes. Consequently, conventional evaluation of lymph node status through CT scans remains challenging.

Radiomics is an innovative technique that swiftly extracts numerous quantitative features from conventional medical images using high-throughput computation, yielding invaluable information for diagnostic and prognostic purposes (13–15). In recent years, radiomics has been extensively employed in detecting, grading, assessing the therapeutic response, and prognostic evaluation for patients with EC (16–20). CT-based radiomics have been utilized to predict LNM in patients with EC (11, 21–23). These findings suggest that this approach has great potential as an accurate and reproducible tool for non-invasive preoperative evaluation of LNM. Radiomics techniques offer a promising solution that overcomes some of the limitations of conventional CT imaging. However, owing to variations in imaging protocols, study design, sample size, modeling techniques, and software used for radiomics analysis across various studies on EC, the reported diagnostic efficacy of radiomics in preoperative identification of LNM has demonstrated significant variability. These inconsistencies have led to uncertainty regarding the effectiveness of using radiomics for this purpose in clinical practice. Therefore, we aimed to conduct a comprehensive meta-analysis that thoroughly assesses the diagnostic accuracy of CT-based radiomics in predicting LNM in individuals diagnosed with EC.

Materials and methods

This study followed the guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (24). The protocol for this review has been registered on the Open Science Framework (OSF) platform, with the registration link available at https://osf.io/5zcnd.

Literature search

In attempting to comprehensively identify all studies that may be related to our question, an independent search was conducted by two authors (L.S.L. and Y.Z.) in four databases, namely PubMed, Embase, Web of Science, and Cochrane Library, which was limited to studies published until July 10th, 2023. Various keywords, including "artificial intelligence," "machine learning," "radiomics,"

"deep learning," "esophageal neoplasms," "esophageal cancer," "lymph node metastasis," "lymph node," and "LNM" were used for the search. MeSH terms and variations of each keyword were utilized during the search to ensure inclusivity. Any disagreements during the literature selection process were resolved through discussion and consensus among the research team, with the assistance of a designated third-party reviewer (H.L.).

Study selection

The selected studies had to meet specific criteria: (1) original research studies with sample sizes greater than 40; (2) histopathological diagnosis of EC and LNM; (3) LNM detected using CT-based radiomics, and (4) data sufficient for reconstructing 2×2 contingency table, aiming at determining diagnostic sensitivity and specificity. Meanwhile, the criteria used to exclude ineligible studies were: (1) reviews, case reports, consensus statements, guidelines, animal studies, letters, and editorials, and (2) multiple studies using the same study population (in such cases, the most recent or comprehensive report was also included).

Quality assessment

The quality assessment and data extraction were independently evaluated by two reviewers, L.S.L. and Y.Z. Any disagreements were resolved by the third reviewer, H.L. Four domains of the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) were customized to evaluate the potential bias in the selected studies, including patient selection, index testing, reference standards, and flow and timing (25). The researchers utilized the Radiomics Quality Score (RQS) to evaluate the methodological quality of the included studies. The RQS comprises five components: imaging protocol, feature extraction from radiological images, data modeling, model validation, and data sharing (13). Additional details can be found in the Supplementary Table S1. The concordance between the two primary reviewers was determined through the calculation of the intra-class correlation coefficient (ICC). The ICC values were classified as excellent (≥ 0.85), good (0.70–0.84), moderate (0.55–0.69), and weak or poor (≤ 0.54) (26).

Data extraction

All pertinent data was acquired from the entirety of the incorporated full-text articles. The information that was obtained through extraction were: first author, publication year, country, study type, total number of patients and LNM cases, CT machine type, segmentation details, feature selection method, algorithms, information about radiomics and deep learning, data source (single or multiple institutions), sensitivity, and specificity. The researchers calculated the numbers of true positive/negative and false positive/negative cases using the reported sensitivity, specificity, LNM-present, and LNM-absent values in each study. If a single study presented multiple models derived from the same patient cohort, only the model demonstrating superior diagnostic accuracy in the validation cohort (or the training cohort if a validation cohort was unavailable) was considered for inclusion in this meta-analysis (27).

Statistical analysis

Stata 16.0, Meta-Disc 1.4, and Review Manager 5.3 were employed for statistical analysis of meta-analysis. Diagnostic accuracy was evaluated by pooling sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), and diagnostic odds ratio (DOR) with their respective 95% CI. The summary receiver operating characteristic curve (SROC) and its corresponding area under the curve (AUC) were used to summarize the findings. Heterogeneity assessment of the studies incorporated in the analysis was performed through Cochran's Q-test and Higgins' I2 test. A $P < 0.05$ (Cochran's Q-test) or a Higgins' I2 value >50% indicated significant heterogeneity between the studies (28). Sensitivity analysis were conducted by systematically eliminating individual studies from the meta-analysis calculations to evaluate their influence on the overall estimation. Deeks' test assessed the publication bias by analyzing the effective sample size funnel plot (29).

Clinical utility

The study employed Fagan plot analysis for assessing the clinical effectiveness of CT-based radiomics in predicting LNM. This method calculated the LNM post-test probability (P-post) based on the pre-test probabilities (P-pre), signifying a suspicion of LNM (30).

Results

Study selection

By using the aforementioned search strategy, 163 studies were initially identified; only 96 remained after duplicate removal. After reviewing the titles and abstracts, only 24 studies were deemed eligible for further analysis and potential inclusion. After carefully reading the full-text articles, six studies (21, 31–35) were considered eligible for inclusion. In contrast, seven were excluded due to insufficient data, and eleven did not meet the intended purpose of investigating radiomics for predicting LNM (Figure 1).

Features of the enrolled studies

Table 1 lists an overview of the six enrolled studies spanning from 2018 to 2022. A total of 483 patients were enrolled, with LNM

10.3389/fonc.2024.1267596

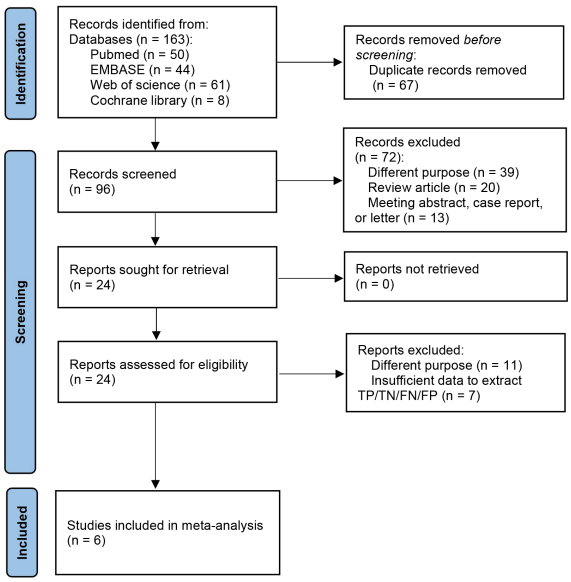


FIGURE 1
Flow diagram of study selection for meta-analysis according to PRISMA.

rates varying between 27.2% to 59.4%. The included studies were conducted retrospectively and exclusively in China. Additionally, all the studies were based on single-center data. Three studies focused on esophageal squamous cell carcinoma, while the remaining three did not specify the cancer subtype. Manual segmentation was utilized for radiomics analysis in all the studies. Furthermore, only one study combined feature extraction with deep learning methods, while the other five exclusively relied on radiomics. The feature selection methods employed in radiomics analysis included the least absolute shrinkage and selection operator (LASSO), t-test, analysis of variance, and ridge regression. In terms of the radiomics diagnostic model developed using machine learning algorithms, logistic regression (LR) was used in four studies, while random forest (RF) and support vector machine (SVM) were used in one study each.

Quality assessment and publication bias

The detailed assessments of RQS and QUADAS-2 for each study are provided in Supplementary Tables S2 and S3. The agreement between primary reviewers was excellent, with ICC values of 0.94 (95% CI 0.64-0.99) for RQS and 0.92 (95% CI 0.56-0.99) for QUADAS-2. The RQS varied between 14 and 16 across studies, with two studies (33, 34) achieving the highest RQS percentage at 44.4%. However, all the selected studies lacked the use of phantoms to evaluate robustness, prospective research design, discussion of potential biological correlates, or comprehensive cost-effectiveness analysis. The qualitative assessment using the QUADAS-2 tool indicated that most studies had a low risk of bias and minimal concerns regarding their applicability (Figure 2). Deeks' funnel plot analysis revealed no evidence of publication bias, suggesting a low risk of bias among the included studies. (Figure 3; P = 0.78).

TABLE 1 Key characteristics of included studies in the meta-analysis.

Author	Year	Country	Study design	Tumor type	Sample size	LNM, n(%)	Scanner	Phases	Segmentation method	ROI	Feature selection method	Algorithms	Combine deep learning (Yes/No)	Data source
Chen	2022	China	Re	ESCC	92	46 (50.0)	NA	CE	manual	2D	LASSO, t-test	RF	YES	Single institution
Li	2021	China	Re	EC	60	25 (41.7)	Siemens	NCE, CE	manual	3D	LASSO, Analysis of variance	LR	NO	Single institution
Ou	2021	China	Re	ESCC	101	60 (59.4)	GE	CE	manual	3D	LASSO, t-test, Mann-Whitney U test	LR	NO	Single institution
Peng	2022	China	Re	ESCC	81	41 (50.6)	GE	NCE, CE	manual	3D	LASSO	LR	NO	Single institution
Shen	2018	China	Re	EC	57	19 (33.3)	GE, Phillips	CE	manual	3D	LASSO、Ridge regression	LR	NO	Single institution
Yu	2021	China	Re	EC	92	25 (27.2)	GE	NCE	manual	3D	LASSO	SVM	NO	Single institution

NA, not available; Re, retrospective; ESCC, esophageal squamous cell carcinoma; EC, esophageal carcinoma; LNM, lymph node metastasis; TP, true positive; FP, false positive; FN, false negative; TN, true negative; CE, contrast-enhanced; NCE, non-contrast-enhanced; ROI, region of interest; 2D, two−dimensional; 3D, three−dimensional; LASSO, least absolute shrinkage and selection operator; LR, logistic regression; RF, random forest; SVM, support vector machine.

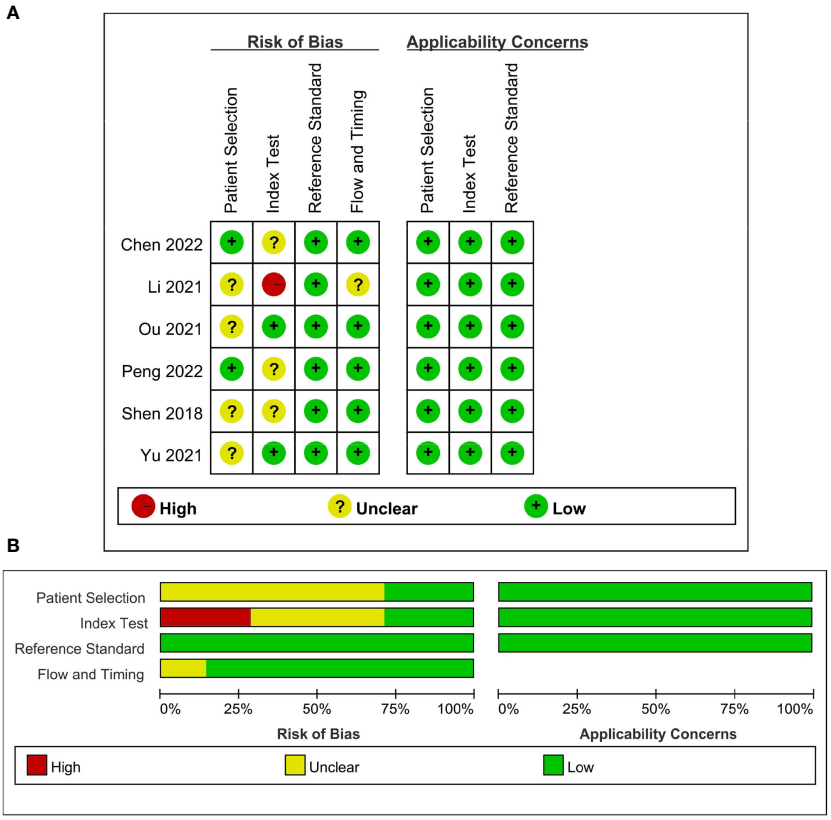


FIGURE 2
Quality assessment of included studies according to Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) criteria. **(A)** Individual studies, **(B)** summary.

Diagnostic accuracy of CT-based radiomics

Across all six selected studies, the pooled sensitivity and specificity (as displayed by the forest plots in Figure 4) for the CT-based radiomics in evaluating LNM in EC were determined to be 0.73 (95% CI, 0.67-0.79) and 0.76 (95% CI, 0.69-0.83), respectively. The PLR, NLR, and DOR were found to be 3.1 (95% CI, 2.3-4.2), 0.35 (95% CI, 0.28-0.44), and 9 (95% CI, 6-14), respectively. Furthermore, the SROC analysis yielded an AUC of 0.78 (95% CI, 0.74-0.81), indicating significant overall diagnostic efficacy (Figure 5).

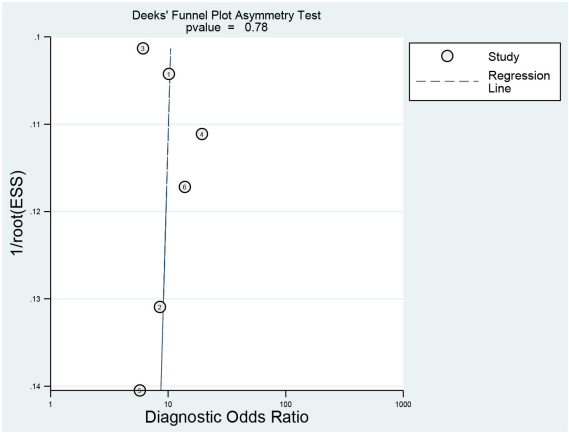


FIGURE 3
Deeks' funnel plot asymmetry test for publication bias.

Heterogeneity assessment

Figure 4 presents the forest plot, which indicates no significant heterogeneity among the studies included in the research when considering sensitivity (P=0.85; I2 = 0). However, specificity exhibited a significant heterogeneity (P=0.07) with a moderate level of heterogeneity indicated by an I2 value of 51.56% (95% CI, 6.97%-96.14%). The Spearman's correlation coefficient was 0.543, with a non-significant P-value of 0.266, suggesting the absence of a threshold effect.

Meta-regression

The sources of heterogeneity were identified through the application of univariate meta-regression analysis. Table 2 presents the results of both subgroup analysis and univariate meta-regression, revealing that several categories, including tumor type, LNM ratio, feature selection method, and algorithms, significantly contributed to the heterogeneity observed in the pooled sensitivity ($P < 0.05$). Moreover, the sample size and CT-

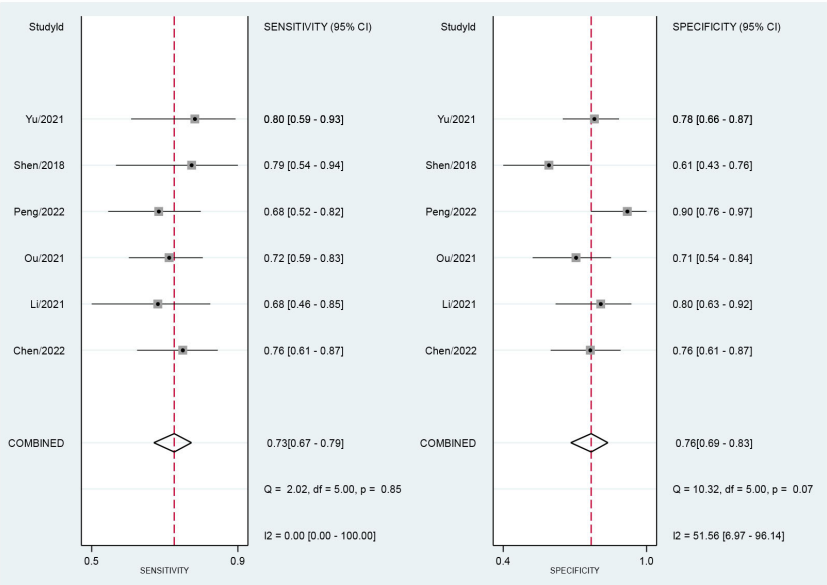


FIGURE 4
Forest plot of sensitivity and specificity based on radiomics for predicting LNM in esophageal carcinoma.

scan phases used were identified as contributors to the heterogeneity in specificity analysis (P < 0.05). The forest plot is presented in Supplementary Figure S1.

Subgroup analysis

Studies on esophageal squamous cell carcinoma (n=3) demonstrated equivalent sensitivity (72%, 95% CI, 65-79 vs. 75%; 95% CI, 65-86) and higher specificity (79%, 95% CI, 70-88 vs. 74%; 95% CI, 64-83) compared to studies on esophageal carcinoma (n=3). Studies with ≥90 patients (n=3) showed equivalent sensitivity (75%, 95% CI 67-82 vs. 71%, 95% CI 62-81) and specificity (75%, 95% CI 66-85 vs. 78%, 95% CI 68-88) compared to studies with <90 patients (n=3). Studies with an LNM ratio ≥50% (n=3) had equivalent sensitivity (72%, 95% CI, 65-79 vs. 75%; 95% CI, 65-86) and higher specificity (79%, 95% CI, 70-88 vs. 74%; 95% CI, 64-83) when compared to studies with an LNM ratio of <50% (n=3).

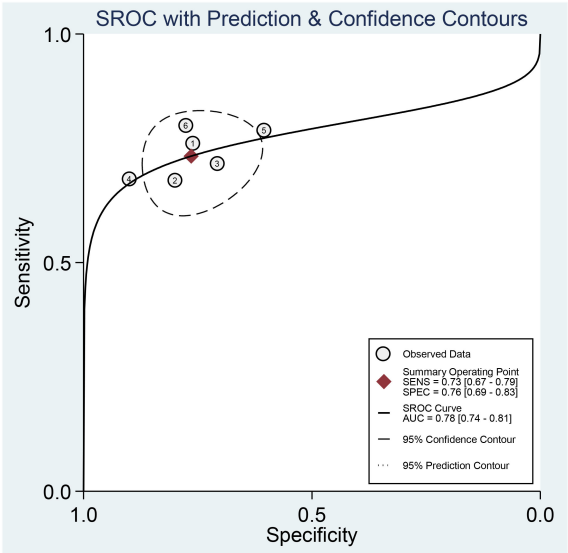


FIGURE 5
Summary receiver operating characteristic curves (SROC) based on radiomics for predicting LNM in esophageal carcinoma.

Three studies using only General Electric(GE) equipment showed similar sensitivity (72%, 95% CI, 64-80 vs. 73%, 95% CI, 60-86) and higher specificity (80%, 95% CI, 71-88 vs. 70%, 95% CI, 57-84) to two studies using other equipment. Three studies using contrast-enhanced CT only had equivalent sensitivity (74%, 95% CI, 67-82 vs. 71%; 95% CI, 62-81) and lower specificity (70%, 95% CI, 62-78 vs. 82%; 95% CI, 75-88) than three studies using other methods. In terms of ROI selection, only one study using the 2D method showed similar sensitivity (76%, 95% CI, 63-89 vs. 72%, 95% CI, 66-79) and specificity (76%, 95% CI, 60-93 vs. 76%, 95% CI, 69-84) compared to five studies that employed the 3D method. Two studies using the LASSO method had similar sensitivity (73%, 95% CI, 62-84 vs. 73%, 95% CI, 66-81) and higher specificity (83%, 95% CI, 74-91 vs. 72%, 95% CI, 64-80) compared to four other studies using different methods. Studies (n=4) utilizing LR for their model had lower sensitivity (71%, 95% CI, 64-79 vs. 77%, 95% CI, 68-87) and equivalent specificity (76%, 95% CI, 67-85 vs. 77%, 95% CI, 66-88) compared to studies (n=2) using different algorithms. A study that combined deep learning features (n=1) showed similar sensitivity (76%, 95% CI, 63-89 vs. 72%, 95% CI, 66-79) and specificity (76%, 95% CI, 60-93 vs. 76%, 95% CI, 69-84), when compared to studies solely utilizing radiomics (n=5).

TABLE 2 Results of univariate meta-regression and subgroup analyses.

Subgroup	Category	Studies (n)	Sensitivity (95%CI)	P value	Specificity (95%CI)	P value
Tumor type	ESCC	3	0.72 (0.65-0.79)	0.02	0.79 (0.70-0.88)	0.16
	EC	3	0.75 (0.65-0.86)		0.74 (0.64-0.83)	
Sample size	≥90	3	0.75 (0.67-0.82)	0.09	0.75 (0.66-0.85)	0.04
	<90	3	0.71 (0.62-0.81)		0.78 (0.68-0.88)	
LNM ratio	≥50%	3	0.72 (0.65-0.79)	0.02	0.79 (0.70-0.88)	0.16
	<50%	3	0.75 (0.65-0.86)		0.74 (0.64-0.83)	
Scanner	GE	3	0.72 (0.64-0.80)	0.13	0.80 (0.71-0.88)	0.53
	Other	2	0.73 (0.60-0.86)		0.70 (0.57-0.84)	
Phases	CE	3	0.74 (0.67-0.82)	0.06	0.70 (0.62-0.78)	0.00
	Other	3	0.71 (0.62-0.81)		0.82 (0.75-0.88)	
ROI	2D	1	0.76 (0.63-0.89)	0.20	0.76 (0.60-0.93)	0.27
	3D	5	0.72 (0.66-0.79)		0.76 (0.69-0.84)	
Feature selection method	LASSO	2	0.73 (0.62-0.84)	0.05	0.83 (0.74-0.91)	0.27
	Other	4	0.73 (0.66-0.81)		0.72 (0.64-0.80)	
Algorithms	LR	4	0.71 (0.64-0.79)	0.01	0.76 (0.67-0.85)	0.06
	Other	2	0.77 (0.68-0.87)		0.77 (0.66-0.88)	
Combine deep learning	Yes	1	0.76 (0.63-0.89)	0.20	0.76 (0.60-0.93)	0.27
	No	5	0.72 (0.66-0.79)		0.76 (0.69-0.84)	

ESCC, esophageal squamous cell carcinoma; EC, esophageal carcinoma; LNM, lymph node metastasis; CE, contrast-enhanced; NCE, non-contrast-enhanced; ROI, region of interest; 2D, two –dimensional; 3D, three–dimensional; LASSO, least absolute shrinkage and selection operator; LR, logistic regression.

Sensitivity analysis

Supplementary Table S1 lists the sensitivity analysis results for each of the six chosen studies. Our findings demonstrated the robustness of results as no significant changes were observed when excluding each study individually; this suggested that any particular study did not significantly influence the overall outcome.

Clinical utility

A CT-based radiomics model can substantially increase the P-post from 20% to 44% with a PLR of 3 for positive pre-tests. Conversely, it can decrease the P-post to 8% with an NLR of 0.35 for negative pre-tests (Figure 6). This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

Discussion

According to our current awareness, this study represented the first systematic review and meta-analysis investigating the efficacy of CT-based radiomics in assessing LNM among patients diagnosed with EC. Our findings demonstrated that the pooled sensitivity, specificity, and AUC for CT-based radiomics were 0.73 (95% CI, 0.67-0.79), 0.76 (95% CI, 0.69-0.83), and 0.78 (95% CI, 0.74-0.81), respectively, showcasing its significantly effective performance. The good sensitivity was crucial for accurately identifying most patients with LNM, thus reducing the chances of missed diagnoses. Furthermore, the high specificity was key in lowering false-positive rates, offering a reliable way to rule out LNM in patients and helping clinicians avoid unnecessary treatments and their potential complications. The inclusion of positive and negative likelihood ratios (PLR and NLR) of 3.1 and 0.35, respectively, further improved the diagnostic utility of CT-based radiomics. A PLR of 3.1 meant that patients with LNM were over three times more likely to have a positive test result than those without, greatly increasing the accuracy of identifying affected individuals. On the other hand, an NLR of 0.35 indicated that a negative result significantly reduced the chance of LNM by 65%, lowering the risk of false negatives. These ratios adjusted the post-test probability based on the initial probability, with a positive test increasing the probability from 20% to 44%, and a negative test decreasing it to about 8%. Such adjustments significantly changed how test results were interpreted clinically, boosting the predictive accuracy. Therefore, our study validated the precision and effectiveness of CT-based radiomics in predicting LNM in EC, facilitating personalized treatment plans. By precisely detecting LNM and improving cancer staging, it facilitated more personalized care, optimizing treatment for high-risk patients while avoiding unnecessary interventions for others.

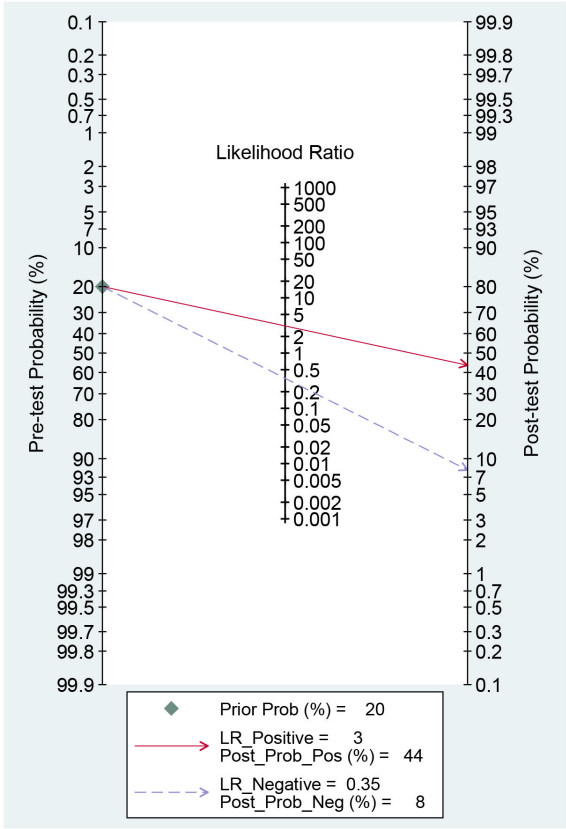


FIGURE 6
Fagan plots for assessing clinical utility.

A moderate degree of heterogeneity in terms of specificity was indicated by the meta-analysis of the studies included, which is worth noting. To determine the possible sources of heterogeneity, a univariate meta-regression analysis and subgroup analysis were conducted, as the Spearman's correlation coefficient test revealed that heterogeneity was not associated with threshold effects. The results suggested that the heterogeneity in specificity could be attributed to the utilization of different CT scan phases and variations in sample size. However, upon conducting a subgroup analysis, it was observed that the sensitivity and specificity were not significantly affected by these factors. It is important to acknowledge that due to the variations in methodologies employed among the included studies, it was challenging to identify all the factors contributing to the observed heterogeneity. Interestingly, while there was no heterogeneity in the pooled sensitivity across all studies, several subgroups showed significant results ($P < 0.05$) in the univariate meta-regression analysis. Therefore, further research with methodological standardization is necessary to improve accuracy, eliminate heterogeneity, and provide more robust evidence for using CT-based radiomics in predicting LNM in patients with EC.

LASSO regression is widely used for feature selection and dimensionality reduction, aiming to reduce the number of features and eliminate irrelevant ones (36, 37). In the subgroup analysis of this study, using LASSO alone for dimensionality reduction resulted in slightly higher specificity than combining it with other methods. This observation may be attributed to the distribution of data features and the division of subgroups. Most studies on radiomics diagnostic models used LR due to the binary nature of LNM status. Nevertheless, studies utilizing other algorithms, including SVM and RF, showed higher sensitivity rates than those using LR. Additionally, previous studies have shown that neural network models or RF based on clinical features could more effectively predict LNM than traditional LR, exhibiting higher AUC, specificity, positive predictive value, and accuracy (38, 39). Regrettably, owing to limited available literature, only one article each for SVM and RF was retrieved, making it difficult to draw reliable conclusions regarding the comparison between SVM or RF and LR.

Image segmentation is a pivotal element of radiomic analysis, incorporating manual delineation using 2D or 3D images, as well as semi-automatic and fully automatic techniques. However, a universally accepted standard for tumor segmentation remains elusive (40). Although manual segmentation offers high precision, it is labor-intensive, subjective, and lacks standardization, leading to limited reproducibility and elevated time and labor expenses. Semi-automatic segmentation necessitates manual refinement, whereas automatic segmentation employs sophisticated computer algorithms for efficient and reproducible lesion boundary identification (41, 42). However, it's crucial to mention that the studies incorporated in this research exclusively used manual delineation for image segmentation. Moreover, only one study in this meta-analysis employed the 2D method, and the subgroup analysis did not reveal a significant difference in sensitivity and specificity between the 2D and 3D methods. Nonetheless, the majority of previous studies have recognized that radiomics-based 3D imaging traits offer a wider and more diverse range of specific information, covering the entire tumor volume and providing a more comprehensive and accurate representation of its shape, size, and texture. Furthermore, 3D segmentation enhances reproducibility by reducing interobserver variability and offers a standardized approach to tumor delineation (43, 44). Hence, future research could concentrate on investigating the advantages and limitations of manual, semi-automatic, and fully automatic delineation in both 2D and 3D methods in radiomics-based imaging analysis, with the goal of determining the most suitable imaging technique for specific clinical situations and enhancing the accuracy and reproducibility of radiomics-based tumor characterization.

Previous studies have highlighted the potential impact of variations in manufacturers and devices on the reproducibility of radiomics features, which could affect the precision of image diagnosis (13, 45). Similarly, the subgroup analysis results of this study revealed that radiomics features derived from distinct CT devices had an impact on the pooled specificity. Nonetheless, it is important to interpret these results cautiously, considering the limited number of studies included in the meta-analysis and the potential for bias due to the small sample size. Multicenter studies can validate the generalization ability of radiomics models by overcoming data differences across regions and devices, thereby improving the stability and reliability of the model (46). However, it is worth noting that all the studies included in this analysis were conducted in the same geographical region, China, which introduces a potential geographical bias. To gain a deeper understanding of the value of radiomics in diagnosing LNM in EC, further analysis is required through more prospective, multi-regional, and high-quality studies.

To assess the robustness of our study, we conducted a sensitivity analysis by sequentially removing one literature source at a time. The results showed no significant changes in the combined DOR after each exclusion, indicating that individual studies did not significantly influence our meta-analysis and that the conclusions were stable and reliable. Moreover, the lack of publication bias, as evidenced by Deeks' funnel plot, further supports the credibility of our findings.

To bolster the robustness and reproducibility of radiomics methodologies, Lambin et al. introduced the RQS guidelines in 2017 (13), aiming to establish a benchmark for quality in radiomics research. However, the absence of standardized quality thresholds remained a notable gap. In response, Wesdorp et al. (47) suggested adopting a 30% cut-off score to enhance clarity and consistency across studies. Despite the RQS percentage of included studies in this meta-analysis ranging from 38.9% to 44.4%, surpassing the 30% threshold, and the pooled diagnostic efficacy demonstrating commendable performance in detecting LNM, the methodological quality of included studies remained a concern. This was because none of the studies utilized phantoms to assess robustness against inter-scansner discrepancies and vendor-specific characteristics. Additionally, comprehensive cost-effectiveness analysis, discussions on potential biological correlations, and a prospective study design were lacking in these studies. Therefore, caution is advised when interpreting the study outcomes.

Several constraints should be considered in the meta-analysis. Firstly, a constrained number of studies met our selection criteria. Secondly, the exclusively retrospective studies analyzed, all conducted in China and solely encompassing English-language publications, may have introduced selection biases and affected quality assessment, thereby potentially constraining the generalizability of our findings. Thirdly, despite conducting various analyses, heterogeneity persisted, emphasizing the need for cautious interpretation of the pooled quantitative results. During data extraction, the highest diagnostic performance model was chosen among multiple models, potentially leading to overestimating the radiomics diagnostic accuracy. Lastly, radiomics could be influenced by factors such as imaging equipment technology and protocols, contributing to heterogeneity. Therefore, establishing standardized presentation protocols in future radiomics research papers is necessary.

Conclusions

Our findings indicated that the CT-based radiomics demonstrated good diagnostic accuracy in predicting LNM in EC, with commendable sensitivity and specificity levels. However, considering the suboptimal RQS and observed heterogeneity among the included studies, it is essential to conduct additional high-quality, multicenter, and large-scale prospective trials to establish more robust and conclusive evidence for the findings presented in this research.

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/s.

Author contributions

LL: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Writing – original draft, Writing – review & editing, Project administration. HL: Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. YZ: Data curation, Methodology, Software, Writing – original draft, Writing – review & editing. JY: Formal analysis, Writing – original draft, Writing – review & editing. CW: Methodology, Visualization, Writing – original draft, Writing – review & editing. LD: Methodology, Writing – original draft, Writing – review & editing. PX: Software, Writing – original draft, Writing – review & editing. WW: Supervision, Visualization, Writing – review & editing. MX: Supervision, Visualization, Writing – review & editing. DS: Conceptualization, Funding acquisition, Project administration, Supervision, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Guangxi Clinical Research Center for Medical Imaging Construction (Grant No. Guike AD20238096), Guangxi Key Clinical Specialty (Medical Imaging Department), Dominant Cultivation Discipline of Guangxi Medical University Cancer Hospital (Medical Imaging Department), and self-financing research of the Health Department of Guangxi Autonomous Region (Grant No. Z-A20231142).

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/fonc.2024.1267596/full#supplementary-material

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2021) 71:209–49. doi: 10.3322/caac.21660

2. Jin X, Zheng X, Chen D, Jin J, Zhu G, Deng X, et al. Prediction of response after chemoradiation for esophageal cancer using a combination of dosimetry and CT radiomics. *Eur Radiol*. (2019) 29:6080–8. doi: 10.1007/s00330-019-06193-w

3. Li X, Zhao J, Liu M, Zhai F, Zhu Z, Yu F, et al. Determination of radiotherapeutic target zones for thoracic esophageal squamous cell cancer with lower cervical lymph node metastasis according to CT-images. *Oncotarget*. (2016) 7:35865–73. doi: 10.18632/oncotarget.9094

4. Smyth EC, Lagergren J, Fitzgerald RC, Lordick F, Shah MA, Lagergren P, et al. Oesophageal cancer. *Nat Rev Dis Primers*. (2017) 3:17048. doi: 10.1038/nrdp.2017.48

5. Ding M, Cui H, Li B, Zou B, Fan B, Ma L, et al. Integrating preoperative computed tomography and clinical factors for lymph node metastasis prediction in esophageal squamous cell carcinoma by feature-wise attentional graph neural network. *Int J Radiat Oncol Biol Phys*. (2023) 116:676–89. doi: 10.1016/j.ijrobp.2022.12.050

6. Qu J, Shen C, Qin J, Wang Z, Liu Z, Guo J, et al. The mr radiomic signature can predict preoperative lymph node metastasis in patients with esophageal cancer. *Eur Radiol*. (2019) 29:906–14. doi: 10.1007/s00330-018-5583-z

7. Ji X, Cai J, Chen Y, Chen LQ. Lymphatic spreading and lymphadenectomy for esophageal carcinoma. *World J Gastrointest Surg*. (2016) 8:90–4. doi: 10.4240/wjgs.v8.i1.90

8. Chen M, Li X, Chen Y, Liu P, Chen Z, Shen M, et al. Proposed revision of the 8th edition ajcc clinical staging system for esophageal squamous cell cancer treated with definitive chemo-imrt based on CT imaging. *Radiat Oncol*. (2019) 14:54. doi: 10.1186/s13014-019-1258-4

9. D'Journo XB. Clinical implication of the innovations of the 8(Th) edition of the tnm classification for esophageal and esophago-gastric cancer. *J Thorac Dis*. (2018) 10: S2671–S81. doi: 10.21037/jtd.2018.03.182

10. Huang YL, Yan C, Lin X, Chen ZP, Lin F, Feng ZP, et al. The development of a nomogram model for predicting left recurrent laryngeal nerve lymph node metastasis in esophageal cancer based on radiomics and clinical factors. *Ann Transl Med*. (2022) 10:1282. doi: 10.21037/atm-22-5628

11. Zhao B, Zhu HT, Li XT, Shi YJ, Cao K, Sun YS. Predicting lymph node metastasis using computed tomography radiomics analysis in patients with resectable esophageal squamous cell carcinoma. *J Comput Assist Tomogr*. (2021) 45:323–9. doi: 10.1097/RCT.0000000000001125

12. Liu J, Wang Z, Shao H, Qu D, Liu J, Yao L. Improving CT detection sensitivity for nodal metastases in oesophageal cancer with combination of smaller size and lymph node axial ratio. *Eur Radiol*. (2018) 28:188–95. doi: 10.1007/s00330-017-4935-4

13. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141

14. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169

15. Le VH, Kha QH, Minh TNT, Nguyen VH, Le VL, Le NQK. Development and validation of CT-based radiomics signature for overall survival prediction in multi-organ cancer. *J Digit Imaging*. (2023) 36:911–22. doi: 10.1007/s10278-023-00778-0

16. Shi YJ, Zhu HT, Yan S, Li XT, Zhang XY, Sun YS. A CT-based radiomics nomogram model for differentiating primary Malignant melanoma of the esophagus from esophageal squamous cell carcinoma. *BioMed Res Int*. (2023) 2023:6057196. doi: 10.1155/2023/6057196

17. Takahashi N, Tanaka S, Umezawa R, Takanami K, Takeda K, Yamamoto T, et al. Development and validation of an [(18)F]Fdg-pet/CT radiomic model for predicting progression-free survival for patients with stage ii - iii thoracic esophageal squamous cell carcinoma who are treated with definitive chemoradiotherapy. *Acta Oncol*. (2023) 62:159–65. doi: 10.1080/0284186X.2023.2178859

18. Du KP, Huang WP, Liu SY, Chen YJ, Li LM, Liu XN, et al. Application of computed tomography-based radiomics in differential diagnosis of adenocarcinoma and squamous cell carcinoma at the esophagogastric junction. *World J Gastroenterol*. (2022) 28:4363–75. doi: 10.3748/wjg.v28.i31.4363

19. Wang J, Yu X, Zeng J, Li H, Qin P. Radiomics model for preoperative prediction of 3-year survival-based CT image biomarkers in esophageal cancer. *Eur Arch Otorhinolaryngol*. (2022) 279:5433–43. doi: 10.1007/s00405-022-07510-8

20. Xie C-Y, Pang C-L, Chan B, Wong EY-Y, Dou Q, Vardhanabhuti V. Machine learning and radiomics applications in esophageal cancers using non-invasive imaging methods-a critical review of literature. *Cancers*. (2021) 13(10):2469. doi: 10.3390/cancers13102469

21. Chen L, Ouyang Y, Liu S, Lin J, Chen C, Zheng C, et al. Radiomics analysis of lymph nodes with esophageal squamous cell carcinoma based on deep learning. *J Oncol*. (2022) 2022:8534262. doi: 10.1155/2022/8534262

22. Lee HN, Kim JI, Shin SY, Kim DH, Kim C, Hong IK. Combined CT texture analysis and nodal axial ratio for detection of nodal metastasis in esophageal cancer. *Br J Radiol*. (2020) 93:20190827. doi: 10.1259/bjr.20190827

23. Tan X, Ma Z, Yan L, Ye W, Liu Z, Liang C. Radiomics nomogram outperforms size criteria in discriminating lymph node metastasis in resectable esophageal squamous cell carcinoma. *Eur Radiol*. (2019) 29:392–400. doi: 10.1007/s00330-018-5581-1

24. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the prisma-dta statement. *JAMA*. (2018) 319:388–96. doi: 10.1001/jama.2017.19163

25. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. (2011) 155:529–36. doi: 10.7326/0003-4819-155-8-201110180-00009

26. Schober P, Mascha EJ, Vetter TR. Statistics from a (Agreement) to Z (Z score): A guide to interpreting common measures of association, agreement, diagnostic accuracy, effect size, heterogeneity, and reliability in medical research. *Anesth Analg*. (2021) 133:1633–41. doi: 10.1213/ANE.0000000000005773

27. Deantonio L, Garo ML, Paone G, Valli MC, Cappio S, La Regina D, et al. 18f-fdg pet radiomics as predictor of treatment response in oesophageal cancer: A systematic review and meta-analysis. *Front Oncol*. (2022) 12:861638. doi: 10.3389/fonc.2022.861638

28. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. (2003) 327:557–60. doi: 10.1136/bmj.327.7414.557

29. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*. (2005) 58:882–93. doi: 10.1016/j.jclinepi.2005.01.016

30. Hellmich M, Lehmacher W. A ruler for interpreting diagnostic test results. *Methods Inf Med*. (2005) 44:124–6. doi: 10.1055/s-0038-1633930

31. Li X, Liu Q, Hu B, Xu J, Huang C, Liu F. A computed tomography-based clinical-radiomics model for prediction of lymph node metastasis in esophageal carcinoma. *J Cancer Res Ther*. (2021) 17:1665–71. doi: 10.4103/jcrt.jcrt_1755_21

32. Ou J, Wu L, Li R, Wu CQ, Liu J, Chen TW, et al. CT radiomics features to predict lymph node metastasis in advanced esophageal squamous cell carcinoma and to discriminate between regional and non-regional lymph node metastasis: A case control study. *Quant Imaging Med Surg*. (2021) 11:628–40. doi: 10.21037/qims-20-241

33. Peng G, Zhan Y, Wu Y, Zeng C, Wang S, Guo L, et al. Radiomics models based on CT at different phases predicting lymph node metastasis of esophageal squamous cell carcinoma (Gasto-1089). *Front Oncol*. (2022) 12:988859. doi: 10.3389/fonc.2022.988859

34. Shen C, Liu Z, Wang Z, Guo J, Zhang H, Wang Y, et al. Building CT radiomics based nomogram for preoperative esophageal cancer patients lymph node metastasis prediction. *Transl Oncol*. (2018) 11:815–24. doi: 10.1016/j.tranon.2018.04.005

35. Yu L, Huang L, Yuan Z, Zhao Q, Yin H, Yu P, et al. Radiomics model based on plain CT for predicting lymph node metastasis of esophageal cancer. *Chin J Med Imaging Technol*. (2021) 37:1333–7. doi: 10.13929/j.issn.1003-3289.2021.09.014

36. Shao B, Bjaanæs MM, Helland Å, Schütte C, Conrad T. Emt network-based feature selection improves prognosis prediction in lung adenocarcinoma. *PloS One*. (2019) 14:e0204186. doi: 10.1371/journal.pone.0204186

37. Fonti V, Belitser E. *Feature Selection Using Lasso. VU Amsterdam research paper in business analytics*, Vol. 30. (2017). pp. 1–25. Available at: https://api.semanticscholar.org/CorpusID:41727607.

38. Chen H, Zhou X, Tang X, Li S, Zhang G. Prediction of lymph node metastasis in superficial esophageal cancer using a pattern recognition neural network. *Cancer Manag Res*. (2020) 12:12249–58. doi: 10.2147/CMAR.S270316

39. Gao M, Huang S, Pan X, Liao X, Yang R, Liu J. Machine learning-based radiomics predicting tumor grades and expression of multiple pathologic biomarkers in gliomas. *Front Oncol*. (2020) 10:1676. doi: 10.3389/fonc.2020.01676

40. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magnetic resonance Imaging*. (2012) 30:1234–48. doi: 10.1016/j.mri.2012.06.010

41. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics*. (2019) 9:1303–22. doi: 10.7150/thno.30309

42. Le NQK. Hematoma expansion prediction: still navigating the intersection of deep learning and radiomics. *Eur Radiol*. (2024) 22. doi: 10.1007/s00330-024-10586-x

43. Watzenboeck ML, Heidinger BH, Rainer J, Schmidbauer V, Ulm B, Rubesova E, et al. Reproducibility of 2d versus 3d radiomics for quantitative assessment of fetal lung development: A retrospective fetal mri study. *Insights Imaging*. (2023) 14:31. doi: 10.1186/s13244-023-01376-y

44. Adelsmayr G, Janisch M, Kaufmann-Bühler AK, Holter M, Talakic E, Janek E, et al. CT texture analysis reliability in pulmonary lesions: the influence of 3d vs. 2d lesion segmentation and volume definition by a hounsfield-unit threshold. *Eur Radiol*. (2023) 33:3064–71. doi: 10.1007/s00330-023-09500-8

45. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: A systematic review. *Int J Radiat Oncol Biol Phys*. (2018) 102:1143–58. doi: 10.1016/j.ijrobp.2018.05.053

46. Dou Q, So TY, Jiang M, Liu Q, Vardhanabhuti V, Kaissis G, et al. Federated deep learning for detecting covid-19 lung abnormalities in CT: A privacy-preserving multinational validation study. *NPJ Digit Med*. (2021) 4:60. doi: 10.1038/s41746-021-00431-6

47. Wesdorp NJ, Hellingman T, Jansma EP, van Waesberghe JTM, Boellaard R, Punt CJA, et al. Advanced analytics and artificial intelligence in gastrointestinal cancer: A systematic review of radiomics predicting response to treatment. *Eur J Nucl Med Mol Imaging*. (2021) 48:1785–94. doi: 10.1007/s00259-020-05142-w

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 26 April 2024
DOI 10.3389/fmolb.2024.1402498

Check for updates

OPEN ACCESS

EDITED BY
Payam Zarrintaj,
University of Montana, United States

REVIEWED BY
Jianhua Zhou,
Sun Yat-sen University Cancer Center (SYSUCC), China
Fei Xiang,
First Medical Center of PLA General Hospital Bei Jing China, China
Hong Yang,
Guangxi Medical University, China

*CORRESPONDENCE
Yong Wang,
wangyong@cicams.ac.cn

RECEIVED 17 March 2024
ACCEPTED 15 April 2024
PUBLISHED 26 April 2024

CITATION
Liu M, Cui N, Sun C, Gong X, Wang B, Yang D and Wang Y (2024), A prospective study on using shear wave elastography to predict the ypT0 stage of rectal cancer after neoadjuvant therapy: a new support for the watch-and-wait approach?
Front. Mol. Biosci. 11:1402498.
doi: 10.3389/fmolb.2024.1402498

COPYRIGHT
© 2024 Liu, Cui, Sun, Gong, Wang, Yang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A prospective study on using shear wave elastography to predict the ypT0 stage of rectal cancer after neoadjuvant therapy: a new support for the watch-and-wait approach?

Mengjia Liu, Ningyi Cui, Chao Sun, Xuantong Gong, Bo Wang, Di Yang and Yong Wang*

Department of Ultrasound, National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Introduction: The diagnostic accuracy of traditional imaging examination in predicting ypT stage of rectal cancer after neoadjuvant therapy is significantly reduced, which would affect patients' subsequent treatment choices. This study aimed to investigate the use of endorectal shear wave elastography (SWE) for diagnosing ypT0 stage of rectal cancer after neoadjuvant chemoradiotherapy (nCRT).

Methods: Sixty patients with rectal cancer were prospectively recruited in this study. Data on endorectal ultrasound (ERUS) and SWE parameters were collected before nCRT and 6–8 weeks after nCRT. Postoperative pathological results were the gold standard for evaluating the diagnostic accuracy of SWE and ERUS in predicting the ypT0 stage of rectal cancer after nCRT. Receiver operating characteristic (ROC) curve analysis was used to determine the cut-off values of the SWE parameters that best corresponded to the ypT0 stage and analyze the sensitivity, specificity, and accuracy.

Results: The diagnostic accuracies of using ERUS to predict the ypT and ypT0 stages of rectal cancer after nCRT were 58.1% (18/31) and 64.3% (9/14), respectively. The ROC curve was constructed with the lesion's Emean, Emean corrected (EC), Emean difference (ED), Emean corrected differencede (ECD), Emean descendding rate (EDR) and Emean corrected descendding rate (ECDR) values after nCRT, the cut-off values of diagnosing the ypT0 stage were 64.40 kPa, 55.45 kPa, 72.55 kPa, 73.75 kPa, 50.15%, and 55.93%, respectively; the area under the curve (AUC) for diagnosing the ypT0 stage was 0.924, 0.933, 0.748, 0.729, 0.857 and 0.861, respectively. The EC value showed the best diagnostic performance.

Conclusion: SWE could improve the accuracy of conventional ERUS in diagnosing the ypT0 stage of rectal cancer after nCRT. It is expected to become a new method to help predict pathological complete responses in clinical practice and provide new evidence for the watch-and-wait approach.

KEYWORDS

rectal cancer, endorectal ultrasound, shear wave elastography, neoadjuvant chemoradiotherapy, pathological complete response

Introduction

Rectal cancer is significant among newly occurring cancers worldwide, based on the latest report by the Global Cancer Observatory (Sung et al., 2021). The National Comprehensive Cancer Network guidelines (Benson et al., 2022) recommend neoadjuvant chemoradiotherapy (nCRT) combined with total mesorectal excision (TME) as the gold standard in patients with locally advanced rectal cancer (stage T3-4b). This is also recommended in patients with stage T2 low-position rectal cancer with a strong desire for anal preservation. However, the watch-and-wait (W&W) approach is a better choice in patients with clinical complete remission (cCR) after nCRT (Glynne-Jones et al., 2018). With the improvement in individualized treatment and recent development and application of new drugs, most patients experience tumor regression and downstaging after nCRT. Approximately 13.5%–40% (Fernández-Martos et al., 2010; Nilsson et al., 2013; Garcia-Aguilar et al., 2015; Perez et al., 2017; Zhang et al., 2020) can achieve pathological complete response (pCR). Systematic reviews (Li et al., 2016; Dossa et al., 2017) have found no significant differences regarding non-regrowth recurrence, cancer-specific mortality, disease-free survival, distant metastasis rates, and overall survival between patients managed with W&W and those with cCR undergoing surgery. Therefore, some patients may have been overtreated. Accurate diagnosis of ypT0 stage is the main aspect of confirming cCR. Conventional methods for ypT0 stage diagnosis after nCRT include magnetic resonance imaging (MRI), endoscopy, and digital rectal examination (MAAS et al., 2011). In patients with ypT0 stage, MRI shows substantial downsizing without residual tumor or residual fibrosis (with low signal on high b-value diffusion-weighted imaging [DWI]). Endoscopy shows no residual tumor or only a small residual erythematous ulcer or scar. Moreover, analysis of biopsies from scar, ulcer, or former tumor location reveals negative results. No palpable tumor is detected, when initially palpated with digital rectal examination. However, the current criteria cannot accurately guide for predicting ypT0 stage. Systematic reviews have shown that among patients diagnosed with cCR, the proportion of those achieving pCR was only 30% (Glynne-Jones et al., 2008).

The guidelines (Glynne-Jones et al., 2018; Diagnosis & Treatment Guidelines For Colorectal Cancer Working Group, 2019; Benson et al., 2022) recommend endorectal ultrasound (ERUS) as one of rectal cancer's main imaging evaluation methods. A previous study (Wang et al., 2012) has shown satisfactory accuracy in diagnosing the T-staging of preoperative rectal cancer without neoadjuvant therapy using ERUS based on a technique that involves a sterile coupling gel filling the rectum, particularly in patients with early cancer (Tis and T1). However, pathological changes such as inflammation, edema, fibrosis, and necrosis occur in tumors and surrounding tissues and significantly decrease the accuracy of T staging in cancers after nCRT using ERUS (Dickman et al., 2013; Zhao et al., 2014; Tang et al., 2024).

Shear wave elastography (SWE) is a new elastic imaging technology that reflects the biomechanical characteristics of tissues in real-time and quantitatively measures lesion stiffness. SWE can provide information on stiffness changes in lesions before and after chemoradiotherapy to evaluate its efficacy, make up for the deficiency in conventional ERUS for differentiating inflammation from tumor, and improve the accuracy of ERUS in predicting ypT0. SWE is easy to perform, by which the results are more objective and reproducible compared with traditional strain elastography, which requires manual pressure. The application of SWE to patients with rectal cancer is in the preliminary research stage, and current studies mainly focus on the stiffness difference between benign and malignant rectal tumors, the initial stiffness value of tumors in different T stages, or the stiffness difference judgment of different T stages during downstaging (Fan et al., 2019; Cong et al., 2021; Loft et al., 2022a; Dong et al., 2023; Tang et al., 2024). There are few relevant studies on diagnosing the ypT0 stage using SWE after nCRT. Based on our previous study, the current study analyzed changes in stiffness before and after neoadjuvant therapy using SWE parameters and aimed to explore the application value of diagnosing the ypT0 stage and improving the diagnostic accuracy of ERUS in diagnosing rectal cancer after nCRT and provide more evidence for diagnosing patients with cCR in clinical practice.

Materials and methods

Patients

This study prospectively recruited patients with rectal cancer treated in our hospital between January 2023 and January 2024.

The inclusion criteria were:(1) rectal cancer confirmed by colonoscopic biopsy, (2) lower tumor margin <15 cm from the anal margin, and (3) consent to receive nCRT and TME treatment. The exclusion criteria were: (1) not undergoing TME surgical treatment; (2) tumor bleeding, obstruction, or intestinal stenosis and not completing ERUS examination; and (3) not undergoing timely examinations. Notably, all patients underwent ERUS and SWE within 1 week before nCRT initiation, 6–8 weeks after nCRT or 2 weeks preoperatively (Figure 1). The hospital Ethics Committee approved the ERUS and SWE examinations. Furthermore, two experienced sonographers jointly performed the ERUS and SWE assessments.

Instruments and methods

All ERUS and SWE examinations were performed using a GE LOGIQ E11 (GE Healthcare,WI, USA) diagnostic apparatus equipped with an IC5-9-D end-fire type endorectal probe at a frequency of 5–9 MHz.

ERUS examination

Patients underwent enemas to remove all air or stool from the rectum before the ERUS examination. Thereafter, the patients

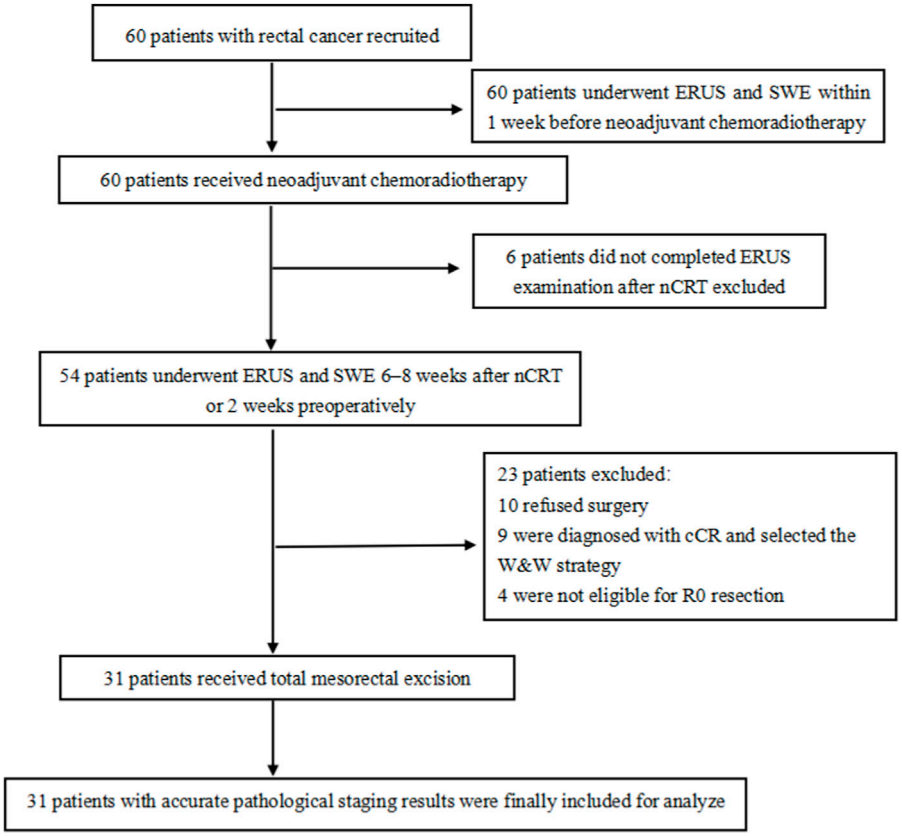


FIGURE 1
Flow chart of patients' inclusion and exclusion. SWE: shear wave elastography, ERUS: endorectal ultrasound, nCRT: neoadjuvant chemoradiotherapy, cCR: clinical complete remission, W&W: watch-and-wait.

remained in the left lateral decubitus position, and approximately 100–150 mL coupling gel was injected directly into the rectum to ensure that the five layers of the intestinal wall and lesion could be clearly seen.

The tumors were evaluated for their location, length, thickness, echo pattern, color Doppler flow imaging, and the depth of the rectal wall layer invasion disrupted by the tumor. Ultrasound staging of the depth of rectal tumor invasion (uT) was performed using the Beynon staging criteria (Beynon et al., 1986): uT1 is a tumor confined to the mucosa or submucosa; uT2 is a tumor that breaks through the submucosa and invades the hypoechoic muscularis propria, whereas the adventitia is hyperechoic intact; uT3 is a tumor that breaks through the muscularis propria and reaches the serosa; and UT4 is a tumor that breaks through the complete layer and invades adjacent tissues or organs. The dynamic and static images of the scanned lesions were stored on a workstation.

SWE examination

After the ERUS examination, we switched to the elastic mode with a dual display of grayscale and elastography for the SWE examination. As shown in Figure 2B, the figure on leftside is the quality control image of gray-scale ultrasound, while the rightside image is the SWE image. A clearly displayed tumor section was selected, the probe was fixed, and the elastic sampling frame was adjusted to an appropriate range. The images were frozen after automatic scanning, the hardest part of the lesion and the normal intestinal wall were selected as the region of interest (ROI). Additionally, the mean Young's modulus value (Emean, kPa) was measured using the Q-box, a quantitative measurement tool of the ultrasound instrument. Color coding of grayscale images reflects the quality control of elastography; yellow indicates a higher measurement quality, and red indicates a lower measurement quality. Color coding on the SWE images reflects tissue stiffness: blue represents low stiffness and low Emean values, and red represents high stiffness and high Emean values. The Q-box was round, with a diameter of 5 mm. The hardest area of the lesion was selected for measurement, which was repeated thrice; the average value was the lesion's final Emean value. The intestinal wall above 1 cm from the edge of the tumor was selected as a normal wall and measured thrice; the average value was the intestinal wall's Emean (EW, kPa).

The Emean value is the absolute value of the lesion measured using SWE. Owing to the differences in the normal intestinal wall stiffness values before and after nCRT in different patients, we calculated the difference as the Emean corrected value (EC, kPa), the Emean value subtracted from the EW value, to exclude the potential interference factors of the bowel wall.

The relative values of lesion stiffness changes before and after NCRT were also calculated. Emean difference (ED, kPa) was

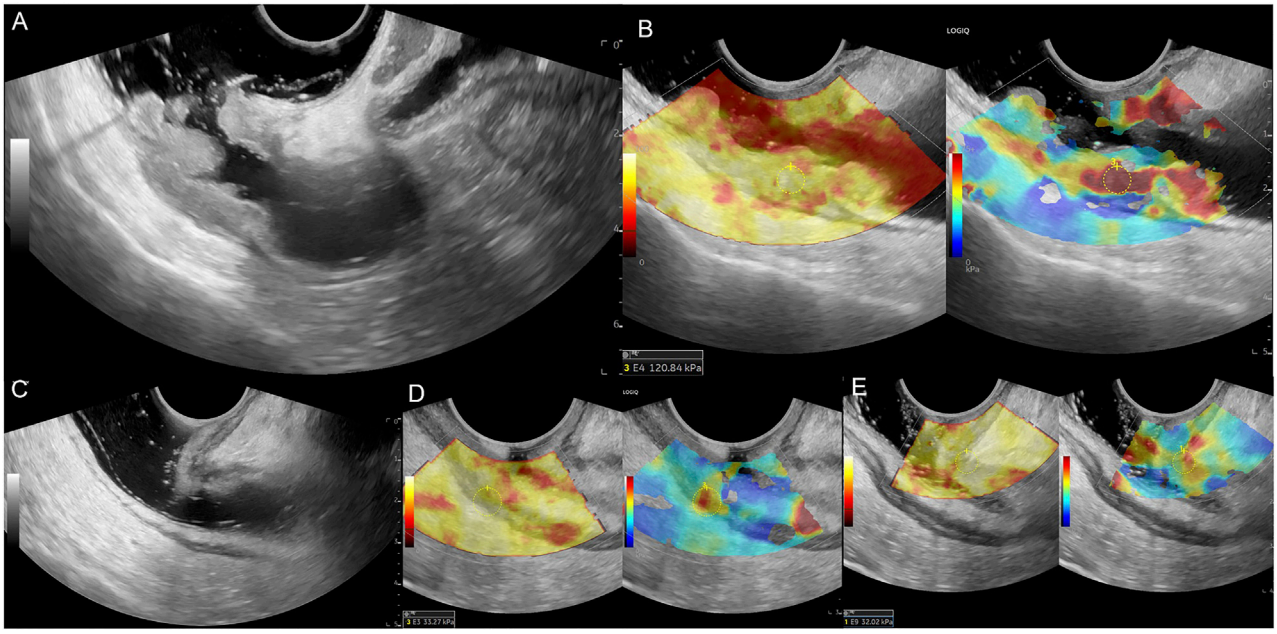


FIGURE 2
Male patient, 69 years old, **(A)** Middle and upper rectum circumferential protuberant lesion, gray-scale shown disruption of the rectal mucosa, submucosa, muscularis and serosa, ERUS staged uT3; **(B)** SWE showed the lesion is hard, pre-nCRT Emean = 120.84 kPa, pre-nCRT EC = 107.84kPa; **(C)** after nCRT, the wall of the middle and upper rectum was slightly thickened, especially the area of second rectal fold of the anterior wall, but the intestinal wall is clearly layered, ERUS staged uT0; **(D, E)** SWE showed the lesion area turn to be soft, choose the higher stiffness of the posterior rectal wall for calculating, post-nCRT Emean = 33.27kPa, post-nCRT EC = 26.27kPa, ED = 87.57kPa, ECD = 81.57kPa, EDR = 72.47%, ECDR = 75.64%, pathological result is ypT0N0.

calculated by subtracting the post-nCRT Emean from the pre-nCRT Emean. Emean corrected difference (ECD, kPa) was calculated by subtracting post-nCRT EC from pre-nCRT EC. Emean descending rate (EDR, %) was defined as ED/pre-nCRT Emean × 100%. Emean corrected descending rate (ECDR, %) was defined as ECD/pre-nCRT EC × 100%.

Statistical analysis

Statistical analysis was performed using SPSS 25.0 software, and the measurement data are expressed as mean ± standard deviation. A t-test was used to compare the quantitative data between two groups, and a one-way analysis of variance was used to compare quantitative data between multiple groups. The non-parametric rank-sum and chi-square tests were used to analyze data that did not conform to the normal distribution. Receiver operating characteristic (ROC) curves were used to evaluate the diagnostic efficacy of Emean, EC, ED, ECD, EDR, and ECDR in diagnosing the ypT stage of rectal cancer. Statistical significance was set at $p \leq 0.05$.

Results

Identification initiative

This study prospectively enrolled 60 patients with rectal cancer who received nCRT. The patients aged 21–74 years (mean, 57.5 ± 11.8 years), with 45 male and 15 female individuals. There were 55 cases of T3 stage, two of T2 stage, and three of T4 stage before nCRT staging using diagnostic imaging.

Thirty-one patients who underwent TME and had accurate pathological staging results were included. They all had rectal adenocarcinoma. Of the other 29 patients, six patients did not complete the preoperative ERUS examination, 10 patients opted for conservative treatment and refused surgery, nine patients were diagnosed with cCR and selected the W&W strategy, and four patients were not eligible for R0 resection.

Of the 31 patients with surgical pathology, 23 (74.19%, 23/31) had downstaging, and eight (25.81%, 8/31) did not. Fourteen patients (45.16%, 14/31) yielded pCR of the ypT0 stage, whereas 17 (54.84%, 17/31) had a non-yield-pathological complete response (nypCR), including one (3.22%, 1/31) of the ypT1 stage, seven (22.58%, 7/31) of the ypT2 stage, eight (48.40%, 8/31) of the ypT3 stage, and one (3.22%, 1/31) of the ypT4 stage.

ERUS evaluates T stage

Table 1 presents the results of the uT stage diagnosed using ERUS after nCRT compared with the postoperative pathological ypT stage. Eighteen patients (18/31, 58.1%) were correctly staged using ERUS, whereas six cases were overstaged and seven were understaged. The accuracy of conventional ERUS in predicting the ypT stage after nCRT was 58.1% (18/31). The accuracy in predicting the ypT0 stage was 64.3%; of the 14 ypT0 stage patients, nine were correctly diagnosed (Figure 2). However, among the five

TABLE 1 Thirty-one patients' ultrasound T (uT) stage after neoadjuvant chemoradiotherapy and pathological T (ypT) stage after surgery.

Pathological T stage	Ultrasound T stage				
	uT0	uT1	uT2	uT3	uT4
ypT0	9	2	2	1	0
ypT1	0	1	0	0	0
ypT2	1	2	3	1	0
ypT3	0	2	1	5	0
ypT4	0	1	0	0	0

uT:ultrasound T stage, ypT:yield pathological T stage.

misdiagnosed patients, four were misdiagnosed as the uT1-T2 stage (Figure 3) and one as the uT3 stage (Figure 4).

SWE evaluates T stage

The Emean and EC values decreased in 29 patients and increased slightly in two patients after nCRT. The Emean values of lesions before and after nCRT were 126.43 ± 21.78 kPa and 71.21 ± 30.13 kPa, respectively, and the difference was statistically significant (t = 10.51, $p < 0.001$). The EC values of lesions before and after nCRT were 117.57 ± 22.76 kPa and 61.64 ± 30.35 kPa, respectively (t = 10.43, $p < 0.001$). The Emean values of the normal intestinal wall were 8.85 ± 3.13 kPa and 9.58 ± 2.12 kPa, respectively; the difference was not statistically significant (t = −1.19, $p = 0.24$).

Only one patient each had ypT1 and ypT4 stage, respectively. Therefore, for analysis, ypT1 and ypT2 with tumors confined to the intestinal wall were grouped into the ypT1-2 group, and ypT3 and ypT4 with tumors invading the outside of the intestinal wall were grouped into the ypT3-4 group. Table 2 lists the Emean and EC values in each group after nCRT. After nCRT, the Emean and EC values increased with increasing tumor stage, and the Emean and EC values of lesions at different pathological stages were significantly different ($p = 0.006$). There were statistically significant differences between the ypT0 and ypT1-2 groups ($p = 0.031$) and between the ypT0 and ypT3-4 groups ($p < 0.001$), whereas the ypT1-2 and ypT3-4 groups showed no significant difference ($p = 0.686$).

The ROC curves constructed with the lesions' post-nCRT Emean and post-nCRT EC values showed that the best diagnostic cut-off values for diagnosing the ypT0 stage were 64.40 kPa and 55.45 kPa, respectively. The area under the curves (AUC) were 0.924 (95% confidence interval [CI]: 0.832–1) and 0.933 (95% CI: 0.844–1), respectively (Figure 5). Emean <64.40 kPa was used for stage uT after nCRT, and the sensitivity, specificity, and accuracy of the ypT0 stage were 94.12%, 78.60%, and 92.86% (13/14), respectively. EC < 55.45 kPa was used for stage uT after nCRT, and the sensitivity, specificity, and accuracy of the ypT0 stage were 94.12%, 85.70%, and 92.86% (13/14), respectively.

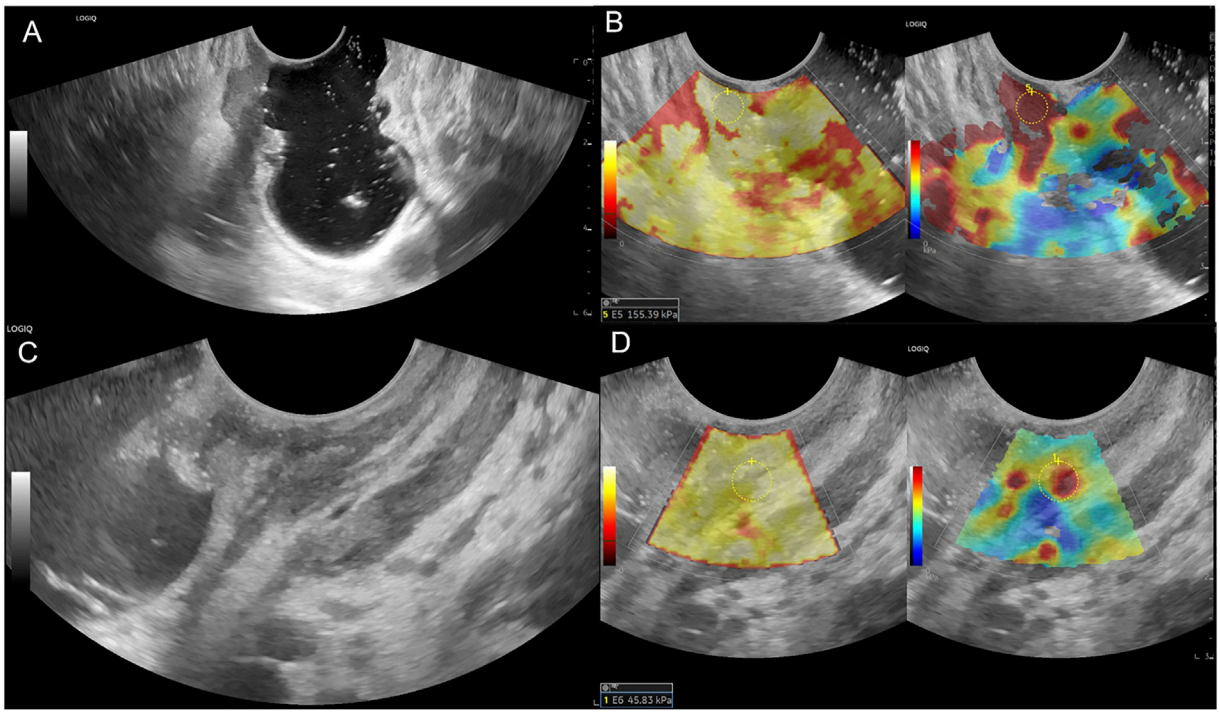


FIGURE 3
Female patient, 69 years old, **(A)** Lower rectum circumferential ulcerative lesion, gray-scale shown disruption of the rectal mucosa, submucosa, muscularis and serosa, ERUS staged uT3; **(B)** SWE showed the lesion is tough, pre-nCRT Emean = 155.39 kPa, pre-nCRT EC = 149.19 kPa; **(C)** after nCRT, rectal mucosa, submucosa and muscularis were thickened in the front wall, the intestinal wall is unclearly layered, ERUS staged uT2; **(D)** SWE showed the lesion area turn to be soft, post-nCRT Emean = 45.83kPa, post-nCRT EC = 34.33kPa, ED = 109.56 kPa, ECD = 114.86kPa, EDR = 70.51%, ECDR = 76.99%, pathological result is ypT0N0.

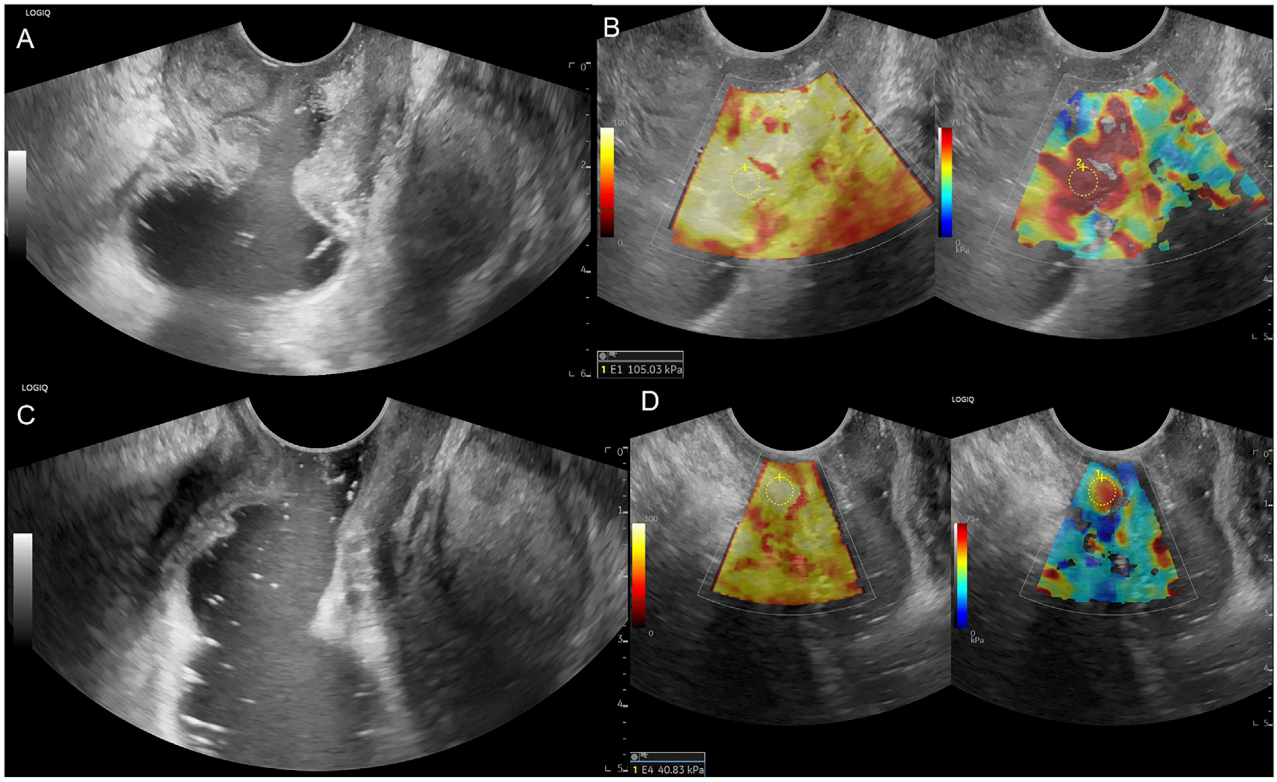


FIGURE 4
Male patient, 54 years old, **(A)** Low rectum circumferential protuberant lesion, gray-scale shown disruption of the rectal mucosa, submucosa, muscularis, serosa and periintestinal fat tissue, ERUS staged uT3; **(B)** SWE showed the lesion is almost in red colour, pre-nCRT Emean = 105.03 kPa, pre-nCRT EC = 101.03kPa; **(C)** after nCRT, the wall of low rectum was circumferential thickened, the intestinal mucosa, submucosa, muscularis, serosa seems unclearly layered, ERUS staged uT3; **(D)** SWE showed the lesion area turn soft, post-nCRT Emean = 40.83kPa, post-nCRT EC = 29.83kPa, ED = 64.20 kPa, ECD = 71.20 kPa, EDR = 61.13%, ECDR = 70.47%, pathological result is ypT0N0.

TABLE 2 post-nCRT Emean values and post-nCRT EC values of rectal tumor lesions in different T staging after nCRT.

	Pathological T staging after surgery		
	ypT0	ypT1-2	ypT3-4
post-nCRT Emean (kPa)	48.14 ± 19.11	80.13 ± 21.80	99.17 ± 22.49
post-nCRT EC (kPa)	38.40 ± 18.40	70.04 ± 21.74	90.31 ± 23.45

pre-nCRT: before neoadjuvant chemoradiotherapy, post-nCRT: after neoadjuvant chemoradiotherapy, EC (kPa): Emean corrected value.

The ROC curves constructed with the lesions' ED, ECD, EDR, and ECDR showed that the best diagnostic cut-off values for diagnosing the ypT0 stage were 72.55 kPa, 73.75 kPa, 50.15%, and 55.93%, respectively (Figure 6). Table 3 summarizes the sensitivity, specificity, and AUC for diagnosing the ypT0 stage. ED > 72.55 kPa was used for stage ypT0 after nCRT, and the sensitivity, specificity, and accuracy were 57.14%, 94.11%, and 85.71% (12/14), respectively. ECD >73.75 kPa was used for stage ypT0 after nCRT, and the sensitivity, specificity, and accuracy were 50.00%, 94.11%, and 85.71%, respectively. EDR >50.15% was used for stage ypT0 after nCRT, and the sensitivity, specificity, and accuracy were 78.57%, 94.11%, and 92.86% (13/14), respectively. ECDR >55.93% was used for stage ypT0 after nCRT, and the sensitivity, specificity, and accuracy were 78.57%, 100%, and 92.86% (13/14), respectively.

Analyzing patients with cCR

We also evaluated the patients with a clinical diagnosis of cCR and showed that the average post-nCRT Emean and EC values of the lesions were 40.57 ± 20.76 kPa and 31.89 ± 20.39 kPa, respectively. Seven patients had the coincident Emean and EC values with the cut-off values for diagnosing the ypT0 stage, and the accuracy was 77.78% (7/9), respectively. The ED, ECD, EDR, and ECDR values were 95.67 ± 21.98 kPa, 95.04 ± 22.06 kPa, 70.53% ± 13.28%, and 75.25% ± 14.42%, respectively. Eight patients had the coincident ED,

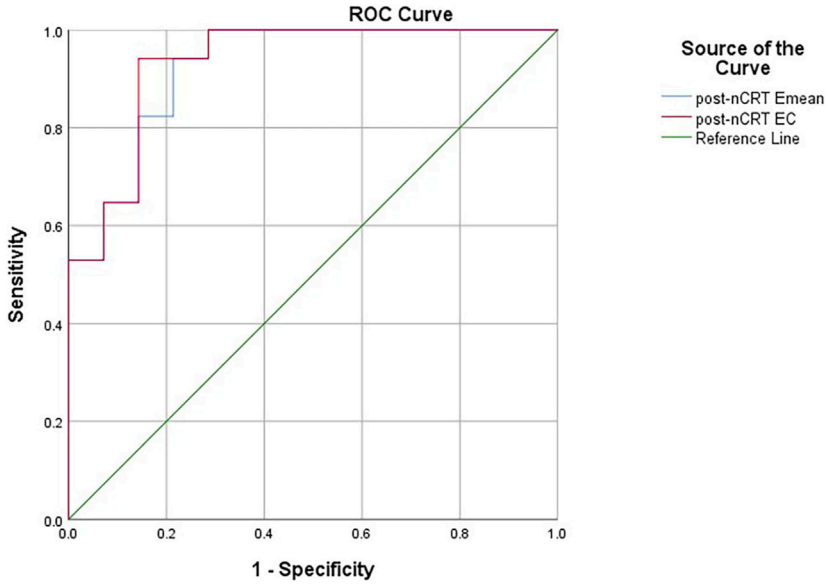


FIGURE 5
The receiver operating characteristic curves constructed with the lesions' post-nCRT Emean and post-nCRT EC values for diagnosing the ypT0 stage. pre-nCRT: before neoadjuvant chemoradiotherapy, post-nCRT: after neoadjuvant chemoradiotherapy, EC (kPa): Emean corrected value.

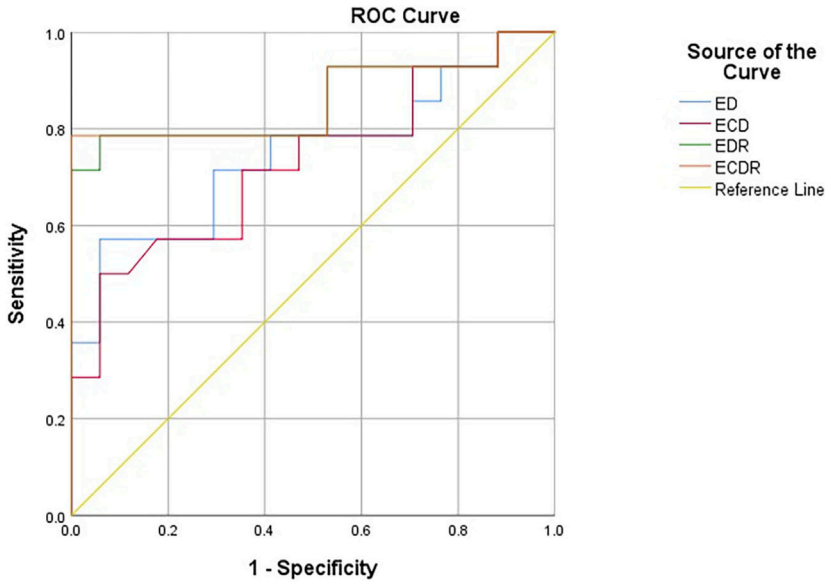


FIGURE 6
The receiver operating characteristic curves constructed with ED, ECD, EDR and ECDR for diagnosing the ypT0 stage. ED (kPa): Emean difference, ECD (kPa): Emean corrected differencede, EDR (%): Emean descendding rate, ECDR (%):Emean corrected descendding rate.

ECD, EDR, and ECDR values with the cut-off values for diagnosing the ypT0 stage, and the diagnostic accuracy was all 88.89% (8/9).

Discussion

The recent gradual individualization and precision of treatment regimens and the development and application of new drugs, such as immunotherapy (programmed death-1/programmed death-ligand 1 monoclonal antibody), have improved rectal cancer treatment. Approximately 13.5%–40% of patients with rectal cancer who initially cannot undergo surgery or achieve R0 resection (Fernández-Martos et al., 2010; Nilsson et al., 2013; Garcia-Aguilar et al., 2015; Perez et al., 2017; Zhang et al., 2020) can achieve pCR after nCRT. This study was conducted simultaneously with a clinical

TABLE 3 Results of ROC curves of ED, ECD, EDR and ECDR.

SWE	Cut-off value (kPa)	Sensitivity (%)	Specificity (%)	AUC
ED (kPa)	72.55	57.14	94.11	0.748(95% CI:0.565–0.931)
ECD (kPa)	73.75	50.00	94.11	0.729(95% CI:0.544–0.914)
EDR(%)	50.15	78.57	94.11	0.857(95% CI:0.705–1)
ECDR (%)	55.93	78.57	100	0.861(95% CI:0.709–1)

ED (kPa): Emean difference, ECD (kPa): Emean corrected differencede, EDR (%): Emean descendding rate, ECDR (%): Emean corrected descendding rate.

randomized controlled trial at our hospital's Radiation Oncology Department. Among them, 45.16% of patients achieved pCR after nCRT, a higher proportion compared with previous studies. This may be associated with the clinical application of new drugs and our hospital's precise delineation of radiation therapy target volumes.

The consensus guidelines (Chinese Watch & Wait Database Research Cooperation Group et al., 2020; Temmink et al., 2023) suggest no significant difference in survival rates between patients with cCR after nCRT who received the W&W strategy and those with pCR. Compared with radical surgery, the W&W strategy can significantly improve patients' quality of life without compromising treatment efficacy (Renehan et al., 2016). Currently, there are varying standards for determining cCR, including endoscopy, digital rectal examination, transrectal ultrasonography, rectal MRI T2 weighted image/DWI sequences, and serum carcinoembryonic antigen (CEA) (Habr-Gama et al., 2013; Glynne-Jones et al., 2018; Temmink et al., 2023). However, currently established cCR criteria cannot accurately determine pCR. There was only a partial overlap between the cCR and pCR groups. A systematic review reported that among patients diagnosed with cCR preoperatively, the proportion of those with pCR postoperatively was only 30% (Glynne-Jones et al., 2008), indicating a high rate of misdiagnosis. Among the 14 patients with the ypT0 stage in this study, colonoscopy results suggested residual tumor in seven patients, and high signals on the DWI sequence of MRI indicated residual tumor in seven patients. In misdiagnosed patients, radical surgery may cause overtreatment and affect their long-term quality of life. Therefore, improving the accuracy of the yielded pCR prediction has become crucial.

SWE has recently become a popular imaging modality. It has been widely used for the differential diagnosis of benign and malignant breast lesions, thyroid nodules, prostate lesions, musculoskeletal diseases, and liver cirrhosis (Zhou et al., 2014; Taljanovic et al., 2017; Anbarasan et al., 2021; Gao et al., 2022; Luo et al., 2022). However, there are relatively few studies on its application to T-staging diagnosis of rectal cancer. SWE does not require manual compression by an operator compared with traditional strain elastography. It measures the Young's modulus value (Emean), representing the stiffness. In this study, we performed ERUS by filling the rectum with sterile coupling gel (Wang et al., 2012). The probe did not need to press the bowel wall, and a clear view of the rectum and tumor was obtained without compression. This approach minimizes the operator's influence and makes the measurements more objective. Notably, most studies on SWE for tumor diagnosis have used Emean or Emax values as the diagnostic criteria (Zhou et al., 2014; Taljanovic et al., 2017; Anbarasan et al., 2021; Gao et al., 2022; Luo et al., 2022). There are differences in the selection of diagnostic indicators for different tumors. Emax has shown a better diagnostic performance for breast cancer diagnosis (Zhou et al., 2014). Currently, there is no consensus on the choice of stiffness measurement for patients with rectal cancer. A previous study using the AixPlorer ultrasound instrument from Supersonic Imaging demonstrated that Emax has good diagnostic value for assessing the T-staging of rectal cancer after nCRT (Cui et al., 2020). Another study suggested that Emean had a better diagnostic value than did Emax (Loft et al., 2022b). Therefore, in this study, we selected the Emean as the research indicator to further validate the diagnostic value of different E values. The GE LOGIQ E11 diagnostic apparatus used in this study is popular in hospitals in China, making it easier to perform ERUS and SWE in patients and promoting multicenter cooperation in the future.

After nCRT, tumor lesions undergo pathological changes, such as tumor cell necrosis, infiltration of lymphocytes and megakaryocytes, and proliferation of connective tissue around the lesion (Dworak et al., 1997). These changes alter the tumor's physical properties, including its stiffness. In this study, 29 patients had decreased post-nCRT Emean and EC values, whereas two showed a slight increase. Consistent with previous studies (Cui et al., 2020; Tang et al., 2024), the overall post-nCRT Emean values were significantly lower compared with those measured before treatment. Regarding the two patients with a slight increase in post-nCRT Emean values, postoperative pathology showed no decrease in the ypT stage. Therefore, the increase in Emean value may be associated with tumor progression, increased number of tumor cells, or fibrosis within the lesion. In this study, the Emean and EC values increased with the tumor T stage after nCRT. Specifically, there were significant differences in these values between the ypT0 and ypT1-2 groups and between the ypT0 and ypT3-4 groups. However, no significant differences were observed between the ypT1-2 and ypT3-4 groups. This could be because the sample size was relatively small, and the statistical power was insufficient, resulting in no statistical difference.

In this study, the accuracy of conventional ERUS for T staging after nCRT was only 58.1%, similar to previous studies (Dickman et al., 2013; Zhao et al., 2014; Cong et al., 2017; Chen et al., 2019; Cong et al., 2019; Tang et al., 2024). Among them, six patients, mainly patients with ypT0 (5/6), were overstaged. This may be because after radiotherapy and chemotherapy, fibrotic tissue replaces the tumor, and grayscale ultrasound shows poor definition and irregular borders of the bowel wall layers (Figures 3C, 4C), which cannot distinguish between fibrotic tissue and tumor lesions, leading to its misdiagnosis as a residual tumor.

Seven cases were understaged, possibly due to the presence of only a very small amount of residual tumor infiltrating the muscularis propria or mucosal layer after radiotherapy and chemotherapy, which was difficult to detect using grayscale ultrasound. This study's accuracy of ERUS in diagnosing the ypT0 stage was 64.3%. Among the 14 patients who achieved a complete response after nCRT, only nine were accurately diagnosed using ERUS, whereas five were misdiagnosed. Among them, four were misdiagnosed as the uT1-T2 stage and one as the uT3 stage. The diagnostic accuracy was low, consistent with previous studies (Fan et al., 2019; Cong et al., 2021). The Emean and EC values represent the lesion's absolute stiffness. Using a cut-off Emean value of <64.4 kPa after nCRT to diagnose the ypT0 stage, we had a sensitivity, specificity, and accuracy of 94.12%, 78.60%, and 92.86% (13/14), respectively.

Similarly, using a cut-off value of EC < 55.45 kPa after nCRT to diagnose ypT0, we had a sensitivity, specificity, and accuracy of 94.12%, 85.70%, and 92.86% (13/14), respectively. The diagnostic accuracy of EURS combined with Emean and EC values was significantly improved compared with that of conventional ERUS. We measured the Emean values of a normal bowel wall before and after nCRT in the same patient to exclude individual differences in bowel wall stiffness, and found that the post-nCRT EC value, which excluded the interference of bowel wall stiffness, had a higher sensitivity, specificity, and AUC in the ROC curve, indicating better diagnostic performance. We derived relative stiffness parameters, including the lesion ED, ECD, EDR, and ECDR, by comparing the stiffness of lesions before and after nCRT. The AUC of the ROC curves were 0.748, 0.729, 0.857, and 0.861, respectively. These relative stiffness parameters had slightly lower diagnostic performances than did the absolute stiffness values. However, using the cut-off values of these SWE parameters to re-stage uT significantly improved the accuracy of diagnosing the ypT0 stage to 85.71%–92.86%. As shown in Figures 3C, 4C, these two patients were in the ypT0N0 stage postoperatively, ERUS showed thickening and unclear layers of the lesions on grayscale after nCRT, and they were diagnosed with uT2 and uT3 stages, respectively. As shown in Figures 3D, 4D, the SWE examination showed that the lesions were soft. After Emean measurement and calculation, SWE parameters were consistent with the diagnostic cut-off value of the ypT0 stage, indicating that ERUS combined with SWE could achieve an accurate re-staging diagnosis. The SWE parameters in our study could improve the accuracy of predicting ypT0 stage of rectal cancer after neoadjuvant therapy, help clinicians judge the cCR status in patients after nCRT, and provide strong support for the W&W strategy. In addition, SWE is easy to operate; provides objective data; and easily promotes and applies to different hospitals and equipment, which can be directly applied to clinical work and benefit patients.

Nine patients were clinically diagnosed with cCR and did not undergo surgery. Statistical analysis showed that the SWE parameters after nCRT in these patients were 77.78%–88.89%, consistent with our study's diagnostic cut-off values. No local recurrences or distant metastases were observed.

This study has some limitations. First, it was a single-center study, and there may have been bias in case selection. Second, the sample sizes for this study's different T stages were unbalanced. Notably, some patients with uT1 refused surgical treatment due to the inability to preserve the anus. However, patients with uT4 were often unsuitable for TME surgery owing to the lack of radical resection indications. Therefore, only one patient each had ypT1 and ypT4 stage, respectively. Consequently, it was impossible to analyze the diagnostic efficacy of SWE in each T stage. Third, this was a prospective study. However, the sample size was relatively small, and some patients refused surgery or could not undergo radical resection. Only 31 of 60 patients received surgical treatment, resulting in a small number of patients with pathological outcomes. In the future, with enough long duration of time and large number of patients, the diagnostic effect of SWE can be evaluated more comprehensively to further validate the conclusions in this study. The application of SWE in patients with rectal cancer after nCRT is not only limited to the prediction of ypT0, but also to the diagnostic cut-off values of other T stages and monitoring for local recurrence in patients with cCR.

In conclusion, the SWE parameters in this study improved the diagnostic accuracy of ERUS for predicting ypT0 stage after nCRT; the EC value showed the best diagnostic performance. These findings provide more evidence for clinically and accurately diagnosing patients with cCR. The results can be used as a pilot study. In future studies, we will expand the sample size, promote multicenter cooperation, and further validate the accuracy of the relevant parameters.

Data availability statement

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

Ethics statement

The studies involving humans were approved by GCP center Ethics Committee, National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital, Chinese Academy of Medical Sciences. 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. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

ML: Writing–review and editing, Writing–original draft, Supervision, Data curation. YW: Writing–review and editing, Supervision, Funding acquisition, Data curation. NC: Writing–review and editing, Methodology. CS: Writing–review and editing. XG: Writing–review and editing. BW: Writing–review and editing. DY: Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the following grants: National Natural Science Foundation of China (grant numbers 81974268).

Acknowledgments

We are grateful to doctor Yuan Tang, the member of the Department of Radiotherapy of National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital, Beijing 100021, China, for the patients recruitment. We would like to thank Editage (www.editage.cn) for English language editing.

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

Anbarasan, T., Wei, C., Bamber, J. C., Barr, R. G., and Nabi, G. (2021). Characterisation of prostate lesions using transrectal shear wave elastography (SWE) ultrasound imaging: a systematic review. *Cancers (Basel)* 13, 122. doi:10.3390/cancers13010122

Benson, A. B., Venook, A. P., Al-Hawary, M. M., Azad, N., Chen, Y. J., Ciombor, K. K., et al. (2022). Rectal cancer, version 2.2022, NCCN clinical practice guidelines in Oncology. *J. Natl. Compr. Canc Netw.* 20, 1139–1167. doi:10.6004/jnccn.2022.0051

Beynon, J., Foy, D. M., Roe, A. M., Temple, L. N., and Mortensen, N. J. (1986). Endoluminal ultrasound in the assessment of local invasion in rectal cancer. *Br. J. Surg.* 73, 474–477. doi:10.1002/bjs.1800730618

Chen, L., Liu, X. Y., Zhang, W. J., Jiang, Q. L., Tan, S., Yu, J. L., et al. (2019). Clinical value of endorectal ultrasonography in predicting neoadjuvant treatment, response for locally advanced rectal cancer. *Chin. J. Ultrasonogr.* 28 (8), 691–695. doi:10.3760/cma.j.issn.1004-4477.2019.08.009

Chinese Watch & Wait Database Research Cooperation Group(CWWD), Chinese Association of Surgeons, Chinese Society of Coloproctology, Chinese Medical Doctor Association, Chinese Society of Colorectal Surgery, Chinese Medical Association, Colorectal Cancer Physician Specialty Committee, Chinese Medical Doctor Association, Radiation Therapy Specialty Committee, Chinese Anticancer Association (2020). Consensus on the Watch and Wait policy in rectal cancer patients after neoadjuvant treatment (2020 version). *Zhonghua Wei Chang. Wai Ke Za Zhi* 23, 1–9. doi:10.3760/cma.j.issn.1671-0274.2020.01.001

Cong, Y., Fan, Z., Dai, Y., Zhang, Y., and Yan, K. (2019). Use of shear wave elastography in the evaluation of local advanced rectal cancer after neoadjuvant radiochemotherapy: the initial experience. *Chin. J. Ultrasonogr.* 28, 901–906. doi:10.3760/cma.j.issn.1004-4477.2019.10.015

Cong, Y., Fan, Z., Dai, Y., Zhang, Z., and Yan, K. (2021). Application value of shear wave elastography in the evaluation of tumor downstaging for locally advanced rectal cancer after neoadjuvant chemoradiotherapy. *J. Ultrasound Med.* 40, 81–89. doi:10.1002/jum.15378

Cong, Y., Liao, R., Fan, Z. H., Zhang, Z., Li, R., and Yan, K. (2017). Application of endorectal elastography to evaluate local advanced rectal cancer after neoadjuvant radiochemotherapy: compared with pathology. *Chin. J. Ultrasonogr.* 26, 882–886. doi:10.3760/cma.j.issn.1004-4477.2017.10.013

Cui, N. Y., Wang, Y., Tang, Y., Zhang, R., and Liu, M. J. (2020). Value of endorectal shear wave elastography in T staging of rectal cancer after neoadjuvant radiochemotherapy. *Chin. J. Colorectal Dis. Electron. Ed.* 09, 586–591. doi:10.3969/j.issn.1008-6978.2022.03.020

Diagnosis & Treatment Guidelines For Colorectal Cancer Working Group (2019). Chinese Society of Clinical Oncology (CSCO) diagnosis and treatment guidelines for colorectal cancer 2018 (English version). *Chin. J. Cancer Res.* 31, 117–134. doi:10.21147/j.issn.1000-9604.2019.01.07

Dickman, R., Kundel, Y., Levy-Drummer, R., Purim, O., Wasserberg, N., Fenig, E., et al. (2013). Restaging locally advanced rectal cancer by different imaging modalities after preoperative chemoradiation: a comparative study. *Radiat. Oncol.* 8, 278. doi:10.1186/1748-717X-8-278

Dong, X. Y., Li, Q. M., Xue, W. L., Sun, J. W., Zhou, H., Han, Y., et al. (2023). Diagnostic performance of endorectal ultrasound combined with shear wave elastography for rectal tumors staging. *Clin. Hemorheol. Microcirc.* 84, 399–411. doi:10.3233/CH-231716

Dossa, F., Chesney, T. R., Acuna, S. A., and Baxter, N. N. (2017). A watch-and-wait approach for locally advanced rectal cancer after a clinical complete response following neoadjuvant chemoradiation: a systematic review and meta-analysis. *Lancet Gastroenterol. Hepatol.* 2, 501–513. doi:10.1016/S2468-1253(17)30074-2

Dworak, O., Keilholz, L., and Hoffmann, A. (1997). Pathological features of rectal cancer after preoperative radiochemotherapy. *Int. J. Colorectal Dis.* 12, 19–23. doi:10.1007/s003840050072

Fan, Z., Cong, Y., Zhang, Z., Li, R., Wang, S., and Yan, K. (2019). Shear wave elastography in rectal cancer staging, compared with endorectal ultrasonography and magnetic resonance imaging. *Ultrasound Med. Biol.* 45, 1586–1593. doi:10.1016/j.ultrasmedbio.2019.03.006

Fernández-Martos, C., Pericay, C., Aparicio, J., Salud, A., Safont, M., Massuti, B., et al. (2010). Phase II, randomized study of concomitant chemoradiotherapy followed by surgery and adjuvant capecitabine plus oxaliplatin (CAPOX) compared with induction CAPOX followed by concomitant chemoradiotherapy and surgery in magnetic resonance imaging-defined, locally advanced rectal cancer: Grupo cancer de recto 3 study. *J. Clin. Oncol.* 28, 859–865. doi:10.1200/JCO.2009.25.8541

Gao, X. Q., Ma, Y., Peng, X. S., Wang, L. L., Li, H. X., Zheng, X. L., et al. (2022). Diagnostic performance of C-TIRADS combined with SWE for the diagnosis of thyroid nodules. *Front. Endocrinol. (Lausanne)* 13, 939303. doi:10.3389/fendo.2022.939303

Garcia-Aguilar, J., Chow, O. S., Smith, D. D., Marcet, J. E., Cataldo, P. A., Varma, M. G., et al. (2015). Effect of adding mFOLFOX6 after neoadjuvant chemoradiation in locally advanced rectal cancer: a multicentre, phase 2 trial. *Lancet Oncol.* 16, 957–966. doi:10.1016/S1470-2045(15)00004-2

Glynne-Jones, R., Wallace, M., Livingstone, J. I., and Meyrick-Thomas, J. (2008). Complete clinical response after preoperative chemoradiation in rectal cancer: is a "wait and see" policy justified? *Dis. Colon Rectum* 51, 10–19. doi:10.1007/s10350-007-9080-8

Glynne-Jones, R., Wyrwicz, L., Tiret, E., Brown, G., Rödel, C., Cervantes, A., et al. (2018). Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* 29, iv263. doi:10.1093/annonc/mdy161

Habr-Gama, A., Sabbaga, J., Gama-Rodrigues, J., São Julião, G. P., Proscurshim, I., Bailão Aguilar, P., et al. (2013). Watch and wait approach following extended neoadjuvant chemoradiation for distal rectal cancer: are we getting closer to anal cancer management? *Dis. Colon Rectum* 56, 1109–1117. doi:10.1097/DCR.0b013e3182a25c4e

Li, J., Li, L., Yang, L., Yuan, J., Lv, B., Yao, Y., et al. (2016). Wait-and-see treatment strategies for rectal cancer patients with clinical complete response after neoadjuvant chemoradiotherapy: a systematic review and meta-analysis. *Oncotarget* 7, 44857–44870. doi:10.18632/oncotarget.8622

Loft, M. K., Pedersen, M. R. V., Grimm, P., Lauritzen, A. H., Dam, C., and Rafaelsen, S. R. (2022a). Intra- and interobserver variability of shear wave elastography in rectal cancer. *Cancers (Basel)* 14, 2633. doi:10.3390/cancers14112633

Loft, M. K., Pedersen, M. R. V., Lindebjerg, J., Rahr, H. B., and Rafaelsen, S. R. (2022b). Endorectal ultrasound shear-wave elastography of complex rectal adenoma and early rectal cancer. *Diagn. (Basel)* 12, 2166. doi:10.3390/diagnostics12092166

Luo, Q. T., Zhu, Q., Zong, X. D., Li, M. K., Yu, H. S., Jiang, C. Y., et al. (2022). Diagnostic performance of transient elastography versus two-dimensional shear wave elastography for liver fibrosis in chronic viral hepatitis: direct comparison and a meta-analysis. *Biomed. Res. Int.* 2022, 1960244. doi:10.1155/2022/1960244

Maas, M., Beets-Tan, R. G., Lambregts, D. M., Lammering, G., Nelemans, P. J., Engelen, S. M., et al. (2011). Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. *J. Clin. Oncol.* 29, 4633–4640. doi:10.1200/JCO.2011.37.7176

Nilsson, P. J., Van Etten, B., Hospers, G. A., Påhlman, L., Van De Velde, C. J., Beets-Tan, R. G., et al. (2013). Short-course radiotherapy followed by neo-adjuvant chemotherapy in locally advanced rectal cancer--the RAPIDO trial. *BMC Cancer* 13, 279. doi:10.1186/1471-2407-13-279

Perez, K., Safran, H., Sikov, W., Vrees, M., Klipfel, A., Shah, N., et al. (2017). Complete neoadjuvant treatment for rectal cancer: the Brown university Oncology group CONTRE study. *Am. J. Clin. Oncol.* 40, 283–287. doi:10.1097/COC.0000000000000149

Renehan, A. G., Malcomson, L., Emsley, R., Gollins, S., Maw, A., Myint, A. S., et al. (2016). Watch-and-wait approach versus surgical resection after chemoradiotherapy for patients with rectal cancer (the OnCoRe project): a propensity-score matched cohort analysis. *Lancet Oncol.* 17, 174–183. doi:10.1016/S1470-2045(15)00467-2

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

Taljanovic, M. S., Gimber, L. H., Becker, G. W., Latt, L. D., Klauser, A. S., Melville, D. M., et al. (2017). Shear-wave elastography: basic physics and musculoskeletal applications. *Radiographics* 37, 855–870. doi:10.1148/rg.2017160116

Tang, X., Yuan, H., Mu, X., Gu, P., and Kong, P. (2024). Endosonography elastography and magnetic resonance imaging in the restaging and response assessment of rectal cancer after neoadjuvant therapy. *Ultrasound* Q. doi:10.1097/RUQ.0000000000000676

Temmink, S. J. D., Peeters, K., Bahadoer, R. R., Kranenbarg, E. M., Roodvoets, A. G. H., Melenhorst, J., et al. (2023). Watch and wait after neoadjuvant treatment in rectal cancer: comparison of outcomes in patients with and without a complete response at first reassessment in the International Watch & Wait Database (IWWD). *Br. J. Surg.* 110, 676–684. doi:10.1093/bjs/znad051

Wang, Y., Zhou, C. W., Hao, Y. Z., Li, L., Liu, S. M., Feng, X. L., et al. (2012). Improvement in T-staging of rectal carcinoma: using a novel endorectal ultrasonography technique with sterile coupling gel filling the rectum. *Ultrasound Med. Biol.* 38, 574–579. doi:10.1016/j.ultrasmedbio.2011.12.020

Zhang, X. Y., Lu, Q. Y., and Sun, Y. S. (2020). Application of imaging diagnosis in watch and wait strategy for locally advanced rectal cancer patients after neoadjuvant therapy. *Zhonghua Wei Chang. Wai Ke Za Zhi* 23, 243–247. doi:10.3760/cma.j.cn.441530-20200224-00078

Zhao, R. S., Wang, H., Zhou, Z. Y., Zhou, Q., and Mulholland, M. W. (2014). Restaging of locally advanced rectal cancer with magnetic resonance imaging and endoluminal ultrasound after preoperative chemoradiotherapy: a systemic review and meta-analysis. *Dis. Colon Rectum* 57, 388–395. doi:10.1097/DCR.0000000000000022

Zhou, J., Zhan, W., Chang, C., Zhang, X., Jia, Y., Dong, Y., et al. (2014). Breast lesions: evaluation with shear wave elastography, with special emphasis on the "stiff rim" sign. *Radiology* 272, 63–72. doi:10.1148/radiol.14130818

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 30 April 2024
DOI 10.3389/fmolb.2024.1385140

Check for updates

OPEN ACCESS

EDITED BY
Qingyu Luo,
Dana–Farber Cancer Institute, United States

REVIEWED BY
Lin Han,
Dana–Farber Cancer Institute, United States
Kezhi Yan,
Cystic Fibrosis Foundation, United States

*CORRESPONDENCE
Alba Lucía Combita,
acombita@cancer.gov.co
Yulieth Torres-Llanos,
yxtorresl@unal.edu.co

RECEIVED 12 February 2024
ACCEPTED 28 February 2024
PUBLISHED 30 April 2024

CITATION
Torres-Llanos Y, Zabaleta J, Cruz-Rodriguez N, Quijano S, Guzmán PC, de los Reyes I, Poveda-Garavito N, Infante A, Lopez-Kleine L and Combita AL (2024), *MIR4435-2HG* as a possible novel predictive biomarker of chemotherapy response and death in pediatric B-cell ALL.
Front. Mol. Biosci. 11:1385140.
doi: 10.3389/fmolb.2024.1385140

COPYRIGHT
© 2024 Torres-Llanos, Zabaleta, Cruz-Rodriguez, Quijano, Guzmán, de los Reyes, Poveda-Garavito, Infante, Lopez-Kleine and Combita. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

MIR4435-2HG as a possible novel predictive biomarker of chemotherapy response and death in pediatric B-cell ALL

Yulieth Torres-Llanos[1]*, Jovanny Zabaleta[2], Nataly Cruz-Rodriguez[3], Sandra Quijano[4], Paula Carolina Guzmán[5], Iliana de los Reyes[5], Nathaly Poveda-Garavito[1], Ana Infante[6], Liliana Lopez-Kleine[7] and Alba Lucía Combita[1,8]*

[1]Cancer Biology Group, Instituto Nacional de Cancerología, Bogotá, Colombia, [2]Department of Interdisciplinary Oncology, Louisiana State University Health Sciences Center, New Orleans, LA, United States, [3]Versiti Blood Research Institute, Milwaukee, WI, United States, [4]Department of Microbiology, Pontificia Universidad Javeriana, Bogotá, Colombia, [5]Department of Pediatrics, Hospital Militar Central, Bogotá, Colombia, [6]Department of Pediatrics, Hospital Universitario San Ignacio, Bogotá, Colombia, [7]Department of Statistics, Universidad Nacional de Colombia, Bogotá, Colombia, [8]Department of Microbiology, School of Medicine, Universidad Nacional de Colombia, Bogotá, Colombia

Introduction: Although B-cell acute lymphoblastic leukemia (B-cell ALL) survival rates have improved in recent years, Hispanic children continue to have poorer survival rates. There are few tools available to identify at the time of diagnosis whether the patient will respond to induction therapy. Our goal was to identify predictive biomarkers of treatment response, which could also serve as prognostic biomarkers of death, by identifying methylated and differentially expressed genes between patients with positive minimal residual disease (MRD+) and negative minimal residual disease (MRD-).

Methods: DNA and RNA were extracted from tumor blasts separated by immunomagnetic columns. Illumina MethlationEPIC and mRNA sequencing assays were performed on 13 bone marrows from Hispanic children with B-cell ALL. Partek Flow was used for transcript mapping and quantification, followed by differential expression analysis using DEseq2. DNA methylation analyses were performed with Partek Genomic Suite and Genome Studio. Gene expression and differential methylation were compared between patients with MRD−/− and MRD$^{+/+}$ at the end of induction chemotherapy. Overexpressed and hypomethylated genes were selected and validated by RT-qPCR in samples of an independent validation cohort. The predictive ability of the genes was assessed by logistic regression. Survival and Cox regression analyses were performed to determine the association of genes with death.

Results: *DAPK1*, *BOC*, *CNKSR3*, *MIR4435-2HG*, *CTHRC1*, *NPDC1*, *SLC45A3*, *ITGA6*, and *ASCL2* were overexpressed and hypomethylated in MRD$^{+/+}$ patients. Overexpression was also validated by RT-qPCR. *DAPK1*, *BOC*, *ASCL2*, and *CNKSR3* can predict refractoriness, but *MIR4435-2HG* is the best predictor. Additionally, higher expression of *MIR4435-2HG* increases the probability of non-response, death, and the risk of death. Finally, *MIR4435-2HG* overexpression,

together with MRD+, are associated with poorer survival, and together with overexpression of *DAPK1* and *ASCL2*, it could improve the risk classification of patients with normal karyotype.

Conclusion: *MIR4435-2HG* is a potential predictive biomarker of treatment response and death in children with B-cell ALL.

KEYWORDS

B-cell acute lymphoblastic leukemia, biomarkers, MRD, gene expression, DNA methylation, prognosis, treatment response

1 Introduction

B-cell acute lymphoblastic leukemias (B-cell ALL) are the most frequent neoplasms in children (Pui et al., 2008). Cure rates for acute lymphoblastic leukemias (ALL) have improved remarkably in the last 4 decades; however, while developed countries achieve 80% cure rates, those rates are around 60% in developing countries (Vera et al., 2012). Some studies have shown that, even under the same treatment protocols, Hispanic children have worse survival and treatment response compared to White and Asian children (Matasar et al., 2006; Walsh et al., 2013; Walsh et al., 2014). The mechanisms underlying these differences in survival rates are still unknown.

Currently, clinical parameters such as leukocyte count, age, extramedullary infiltration, chromosomal translocations, and minimal/measurable residual disease (MRD) classify patients into risk groups. MRD is the most used variable to define treatment response (van Dongen Jacques et al., 1998; Van Dongen JJM et al., 2015). However, due to low survival rates in our patients, it is possible to propose that those variables do not fully define risk groups, which leads to incorrect selection of chemotherapy protocol, affecting patient survival (Sok et al., 2022).

In ALL, gene expression alterations not only result from mutations; alterations at the epigenetic level also play a relevant role in this pathology (Garcia-Manero et al., 2009; Newton et al., 2014; Hu and Shilatifard, 2016; Nordlund and Syvänen, 2018). Thus, epigenetic alterations, including aberrant DNA methylation, could act as important molecular mechanisms in developing resistance to treatment of ALL (Newton et al., 2014). In bone marrow (BM), DNA methylation patterns change during normal hematopoiesis and play an essential role in lineage differentiation (Cullen et al., 2014; Wainwright and Scaffidi, 2017). As in normal cells, tumor cells may also depend on specific DNA methylation patterns to acquire their phenotype and maturation patterns (Patel and Vanharanta, 2017; Wainwright and Scaffidi, 2017; Poli et al., 2018). Therefore, the characterization of aberrant patterns in DNA methylation in tumors can provide important clues about how gene expression is regulated in these pathologies (Nordlund and Syvänen, 2018). Hogan et al., 2015 found that patients with relapses presented promoter hypermethylation and identified a clear signature of differentially expressed genes at the time of diagnosis and relapse; moreover, this signature differs between early-relapse patients and to late-relapse patients. Similarly, aberrant promoter methylation has been associated with MRD. For example, aberrant methylation of the promoters of the *RASSF6* and *RASSF10* genes has been observed in adults with B-cell ALL, which can be detected in peripheral blood and could be useful as potential biomarkers to measure MRD (Younesian et al., 2019). Furthermore, it has been reported that promoter methylation of the *TLX3* and *FOXE3* genes in children with B-cell ALL differentiates MRD + patients from MRD-patients (Chatterton et al., 2014).

Although differential methylation and gene expression patterns have been observed between samples at diagnosis and in relapse, whether these variables could be tools to predict treatment response, including relapse or death, is yet to be determined. Also, a CpG island methylation analysis identified candidate genes as biomarkers of pediatric ALL subgroups and their correlation with disease prognosis (Stumpel et al., 2009). Identifying genomic markers, derived from methylation and gene expression analysis, could improve risk classification, and define patient prognosis.

We hypothesized that gene expression and DNA methylation of blasts obtained at diagnosis differ between MRD+ and MRD- patients and that by comparing these two conditions, candidate genes predictive of treatment response and death could be identified. We collected BM samples obtained at diagnosis, purified leukemic blasts, and compared gene expression and DNA methylation profiles between MRD+ and MRD-patients at the end of induction, looking for overexpressed and hypomethylated genes in MRD + patients. Subsequently, we evaluated if the selected genes could predict response to induction chemotherapy, or death. The search for new genomic biomarkers will improve risk classification and, in the future, patient survival.

2 Materials and methods

2.1 Patient samples

Forty-three patients with B-ALL who attended the Instituto Nacional de Cancerología, Hospital Militar Central and Hospital Universitario San Ignacio (Bogotá, Colombia) between 2017 and 2021 were included. The discovery cohort consisted of 13 BM samples taken at the time of the diagnosis in which RNA-seq/DNA methylation protocols were performed. Sequencing data from 12 patients was used to enrich the survival analyses. Eighteen BM samples taken at the time of the diagnosis were included in the validation cohort by RT-qPCR.

Newly diagnosed patients were included in the study when they entered to the institutions for symptomatology associated with ALL and after verification of the inclusion criteria (not having received chemotherapy, not having another type of cancer, not having genetic diseases and being younger than 18 years old). The diagnosis was confirmed using flow cytometry (Van Dongen JJM et al., 2012) and morphological analysis of BM. This study was conducted following

the recommendations of the Colombian Regulation for Research in Humans (Resolution 8430 of 1993, Ministry of Health of Colombia) and in accordance with the Declaration of Helsinki and approved by each participating institution's Institutional Review Boards (IRB). All methods for nucleic acid analysis were approved by the LSUHSC Translational Genomics Core's Institutional Biosafety Committee protocol number 17370. Informed consent was signed by the parents of all participants. Each patient was treated according to the assigned risk and in accordance with the Berlin-Frankfurt-Munich protocol (Stary et al., 2014). Patients with treatment abandonment or non-adherence to it were excluded.

According to the Berlin-Frankfurt-Munich protocol, response to induction therapy was evaluated by flow cytometry detecting MRD at day 15, where patients with <0.1% residual blasts in BM were MRD-, and patients with >0.1% residual blasts were considered MRD+. At day 33, patients with <0.01% residual blasts in BM were MRD-, while patients with >0.01% residual blasts were MRD+ (Stary et al., 2014). Therefore, we considered patients with MRD-day15 and MRD-day33 as MRD−/− and patients with MRD + day15 and MRD + day33 as refractory patients or $MRD^{+/+}$.

2.2 Blasts isolation and purification

BM samples were collected by a hemato-oncologist and processed within 24 h after sample collection. First, mononuclear cells were separated from BM by density-gradient centrifugation (Lymphoprep, Lonza). The blasts were separated using magnetic microbeads coated with anti-CD19 or anti-CD34 antibodies, followed by MACS column enrichment (Miltenyi, Bergisch Gladbach, Germany). The purity of sorted blasts was assessed with CD34-PERCPCy5.5, CD45-V500, CD19-PECy7, and CD10 APC antibodies. Data was acquired in a FACSCanto II flow cytometer (Becton/Dickinson Biosciences, San Jose, CA), using the FACSDiva software program. Infinicyt software (Cytognos SL, Salamanca, Spain) was used for data analysis (Cruz-Rodriguez et al., 2016).

2.3 DNA and RNA extraction

DNA and RNA were extracted from MACS-sorted blasts using the Allprep mini kit and the robotic workstation QIAcube (Qiagen, Hilden, Germany). RNA quality was evaluated using the Agilent RNA 6000 Nano and Pico kits in the Agilent 2100 Bioanalyzer. RNA concentration was calculated using the Qubit™ RNA High Sensitivity and Broad Range kits, while DNA concentration was calculated using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific).

2.4 Library preparation and RNA sequencing

Samples with RIN >6 and purity by flow cytometry with >90% blasts were selected for RNA-seq. For RNA library preparation, 300 ng of total RNA was used. TruSeq Stranded mRNA RNA libraries were prepared following Illumina's protocol. Resulting libraries were sequenced at 2 × 75 bp on a NextSeq550 sequencer system at the Stanley S. Scott Cancer Center's Translational Genomics Core at LSUHSC-New Orleans. On average, more than 50 million reads per sample were obtained. FASTQ files were uploaded to Partek Flow for analysis. First, removal of contaminant sequences (rDNA, mtrDNA, tRNA) was done with Bowtie 2.0 v2.2.5. Reads were aligned to the hg38 version of the human genome, using STAR 2.7.3a. Genes were quantified with RefSeq 96. For the analysis, genes with less than 5 reads in at least 80% of the samples were excluded. One sample with a low correlation (<0.4) with respect to the others was removed. Normalization was done with the Median Ratio and differential expression analysis was assessed with DESEQ2. Hierarchical clustering, pathways (KEEG) and GO terms were all analyzed in Partek Flow.

2.5 DNA methylation assay

Bisulfite conversion was performed in 500 ng of DNA for each sample following the recommendations of the EZ DNA Methylation-Startup Kit (Catalogue number D5001, Zymo Research, United States). Bisulfite-converted DNA was amplified and hybridized to the Infinium Methylation EPIC Kit chips and scanned on the Illumina's iScan. Analysis of the methylation assays was done in Partek Genomic Suite. Low-confidence probes (p-value >0.05) and probes mapped to X and Y chromosomes were excluded. Normalization was done using NOOB (normalization for Illumina Infinium methylation arrays).

2.6 RT-qPCR

Total RNA from sorted blasts was treated with DNase I Amplification Grade (Invitrogen, United States) prior to reverse transcription. cDNA was synthesized using the SuperScript III First-Strand Synthesis SuperMix Kit (Invitrogen, United States), following the manufacturer's procedures. TaqMan probes were used to quantify mRNA expression levels of candidate genes obtained by RNA-seq analysis (Assay IDs: *DAPK1* Hs00234489_m1; *NPDC1* Hs00209870_m1; *CNKSR3* Hs00295109_m1; *SLC18A2* Hs00996835_m1; *CTHRC1* Hs00298917_m1; *BOC* Hs00264408_m1; *SLC45A3* Hs00263832_m1; *GAPDH* Hs99999905_m1, *ASCL2* Hs00270888_s1; *MIR4435-2HG* Hs03680374_m1). The reaction was amplified in a QuantStudio 12 K plex Real-Time PCR machine (Applied Biosystems). The $2^{-\Delta\Delta CT}$ method was used to estimate the fold induction of each gene using *GAPDH* and Ct values to determine the fold change (FC) for each sample. A pool of samples was used as internal calibrator, as well as water as negative control. Assays were done in duplicate.

2.7 Statistical analysis

2.7.1 Transcriptomic and methylation data analysis

Normalization and differential expression analysis were performed using the Deseq2 library in RStudio. Differentially expressed genes (DEGs) were selected if they had a p-value <0.05 and a FC > 2. ggplot library was used to generate

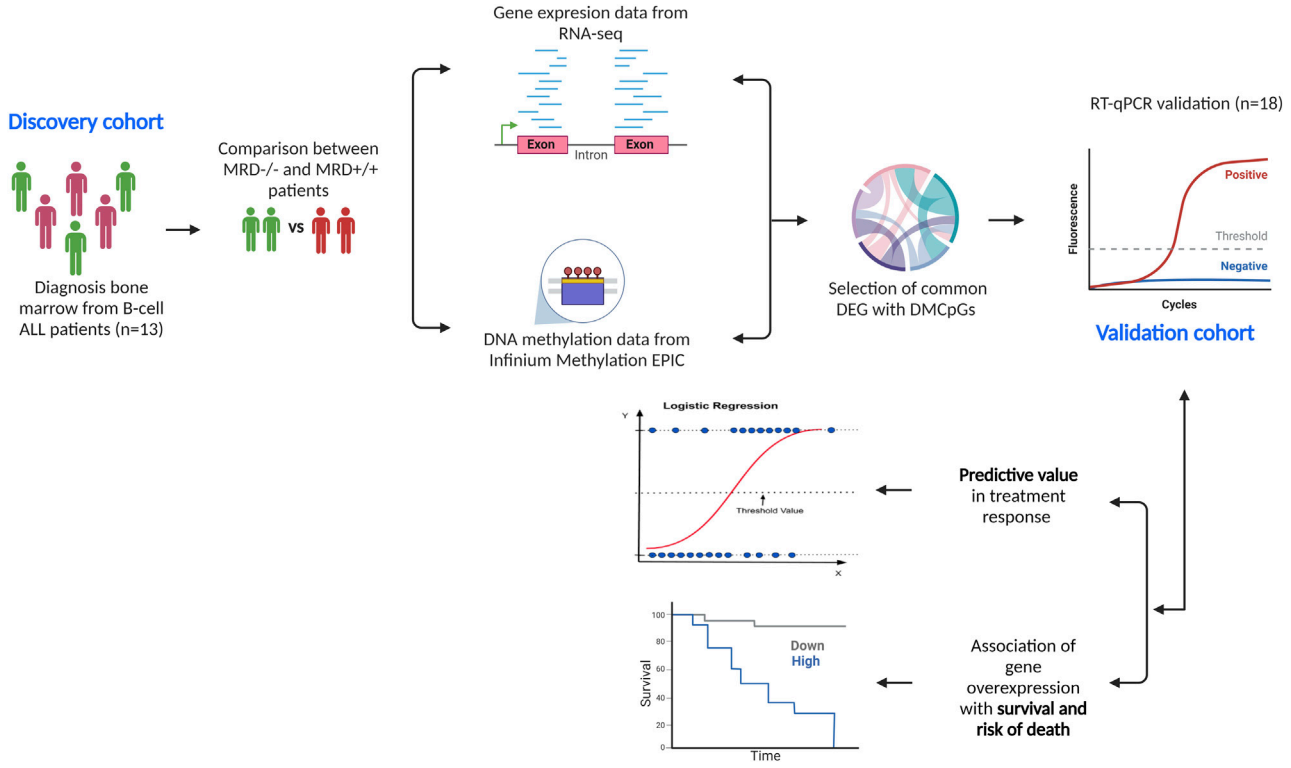


FIGURE 1
Methodological design of the study. We describe the descriptive cohort (n = 13) in which DNA methylation and gene expression analyses were performed separately, comparing the results between MRD−/− vs. MRD$^{+/+}$. From the comparisons, 10 DEGs with differentially methylated CpGs were selected and verified in the validation cohort (n = 18) via RT-qPCR. Subsequently, the predictive ability of genes on outcomes such as achievement of complete remission at end of induction and death was tested. The relationship between gene overexpression and overall survival and risk of death was also assessed.

heatmaps, and GenomeStudio to calculate beta values for each hybridized probe. The Partek Genomic Suite was used for differential methylation analysis to identify genes with differentially methylated CpGs (GDMCpGs). GDMCpGs were chosen if they had FC > 2 and FDR <0.05. Enrichment analysis and functional gene annotation were performed Clusterprofiler in RStudio. Pearson correlation was used to determine correlation between overexpressed genes and hypomethylated probes; those with an inverse correlation less than −0.50 and a p-value <0.05 were selected.

2.7.2 Experimental design

Figure 1 describes the methodological design of the study. Gene and methylation profiles of induction treatment were compared between MRD−/− vs. MRD$^{+/+}$. Treatment response was the only variable used to define profiles in each comparison.

2.7.3 RT-qPCR analysis

Spearman correlation was used to determine any correlation between normalized RNA-seq counts and FC values for RT-qPCR. Genes with p-value <0.05 and r > 0.72 were selected. Mann-Whitney test was used to compare FC between MRD−/− and MRD$^{+/+}$ patients. GraphPad software was used for statistical tests and graphic images. In both analyses, outliers were identified by the ROUT method (Q = 1%) and excluded from the analyses.

2.7.4 Clinical data analysis

To compare clinical variables between patient cohort, t-test and chi-square tests were used. The follow-up time for relapse and death was 2 years. Logistic regression analysis was performed to evaluate whether candidate genes could predict treatment response. Survival analyses were estimated according to gene expression using Kaplan-Meier curves. Cox regression was used to determine whether gene expression conferred a higher risk of death. The Youden index of normalized RNA-seq counts was used to define the cutoff threshold for overexpression for each gene.

3 Results

Table 1 describes the clinicopathological characteristics of the patients included in the discovery cohort, and Supplementary Material S1 shows the clinical variables of the validation cohort. As can be observed, no differences in clinical variables between MRD$^{+/+}$ and MRD−/− patients, except for risk, were found. However, this was to be expected because MRD$^{+/+}$ patients are considered intermediate to high risk, whereas MRD−/− patients may be low to intermediate risk. Interestingly, more than seventy percent of patients had normal karyotype. The MRD−/− group had 1 death related to relapse and progression and another one due to febrile neutropenia. Similarly, the MRD$^{+/+}$ group had one death due

TABLE 1 Clinical characteristics of the DISCOVERY cohort.

Clinical characteristics	MRD$^{+/+}$ (n = 6)		MRD−/− (n = 7)		
	n (%)	Mean (range)	n (%)	Mean (range)	p-value
AGE (years)		11 (3–17)		10.7 (3–15)	0.91
SEX					
Female	2 (33.3)		2 (28.5)		0.85
Male	4 (66.7)		5 (71.5)		
WBC (cel/mL)		90.7 (9.7–292)		35.99 (6.22–91.8)	0.22
RISK					
Low	0 (0)		1 (14.2)		0.03
Intermediate	2 (33.3)		6 (85.8)		
High	4 (66.7)		0 (0)		
EXTRAMEDULLAR INFILTRATION					
Yes	1 (16.6)		2 (28.5)		0.61
No	5 (83.4)		5 (71.5)		
CORTICOID RESPONSE					
Yes	5 (83.4)		7 (100)		0.26
No	1 (16.6)		0 (0)		
CARIOTYPE/MOLECULAR ALTERATIONS					
Normal	4 (66.8)		5 (71.4)		0.56
t (1; 19)	1 (16.6)		1 (14.3)		
t (9; 22)	1 (16.6)		0 (0)		
t (3; 14)	0 (0)		1 (14.3)		
RELAPSE					
Yes	1 (16.6)		1 (14.2)		0.90
No	5 (83.4)		6 (85.8)		
DEATH					
Yes	4 (66.7)		2 (28.5)		0.16
No	2 (33.3)		5 (71.5)		

MRD$^{+/+}$+/+ indicates minimal residual disease positive. MRD−/−-/- indicates minimal residual disease negative.

to relapse and progression and 4 deaths during the induction phase (very aggressive disease).

3.1 Identification of DEGs

To identify genes that could differentiate MRD$^{+/+}$ patients from MRD−/− patients, we performed RNA-seq and MethylationEPIC in nucleic acids extracted from immunomagnetic column-enriched leukemic blasts obtained at the time of diagnosis. MRD status was obtained from medical charts at day 15 and 33. We then compared the gene expression and DNA methylation profiles between MRD−/− vs. MRD$^{+/+}$ patients. Unsupervised hierarchical cluster analysis showed 117 upregulated and 36 downregulated DEGs MRD$^{+/+}$ vs. MRD−/− patients (Figure 2A). Among the biological processes with the highest number of genes involved are neutrophil activation, serine/threonine membrane receptors, extracellular matrix organization, among others ((Figure 2B). The cellular components with the highest number of genes involved include cell-cell junction, adhesion, vesicles, among others ((Figure 2C).

3.2 Identification of GDMCpGs

Additionally, a total of 2726 GDMCpGs were identified between MRD$^{+/+}$ and MRD−/− patients (Figure 3A). To establish a correlation between DEGs and their corresponding methylation

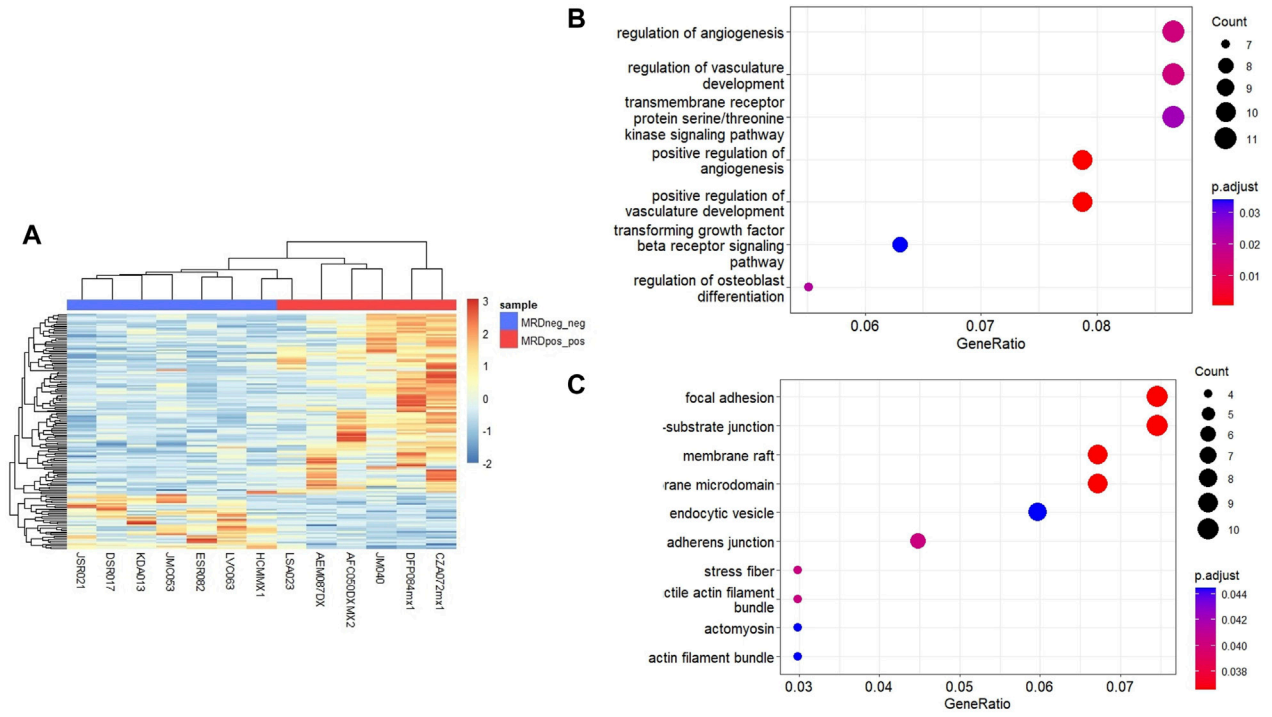


FIGURE 2
Differentially expressed genes between MRD+ and MRD-patients. **(A)** Heatmap of DEGs between MRD−/− and MRD$^{+/+}$. Each column represents an individual patient and horizontal axis indicates each differentially expressed gene. In blue scale downregulated genes, and the red scale shows upregulated genes (FC > 2 and <−2, *p*-value <0.05). **(B)** Dot plot showing the 7 most significant biological processes in which differentially expressed genes between conditions are grouped. **(C)** Dot plot showing the 10 most significant cellular components in which differentially expressed genes between conditions are grouped.

levels, we compared DEGs and GDMCpGs to determine if there were common genes between the two techniques. This comparison revealed 40 common genes between MRD +/+ and MRD −/− patients (Figure 3B). Notably, we observed a significative inverse correlation involving the overexpression of 10 genes and their associated CpGs hypomethylation (Table 2).

3.3 Gene verification by RT-qPCR

Subsequently, gene expression was verified by RT-qPCR and all genes showed correlation between the normalized read counts (RNA-seq) and the $2^{-\Delta\Delta CT}$ values obtained from RT-qPCR. Remarkably, *CTHRC1*, *CNKSR3*, *MIR4435-2HG*, *DAPK1*, and *ITGA6* demonstrated correlations exceeding 0.80 (Figure 4). Although *SLC18A2* was the gene with the best concordance, it was excluded from the analysis because only 6 patients were used for this analysis.

Then, we wanted to test if the expression of these genes remained differential between MRD- and MRD + patients in the validation cohort. Due to the low incidence of the disease (Katz et al., 2015) and the small number of MRD + patients, samples from the MDR + patients in the discovery cohort were pooled with those from the validation cohort for RT-qPCR analyses. Consistent with RNA-seq results, all genes except *CTHRC1* were overexpressed in MRD$^{+/+}$ patients by RT-PCR (Figure 5).

3.4 Predictive value of genes

Afterward, we used logistic regression to evaluate whether genes could predict response to induction chemotherapy. Remarkably, *MIR4435-2HG* was found to be the best predictor of whether a patient would be MRD−/−, MRD$^{+/+}$ (Figures 6A,B). It was observed that genes *DAPK1*, *BOC*, *ASCL2*, and *CNKSR3* could also predict whether the patient would be MRD−/− or MRD$^{+/+}$ (Supplementary Material S2). To assess whether *MIR4435-2HG* could predict the risk of death, we performed logistic regression using our normalized read counts. We observed that *MIR4435-2HG* can predict death with good sensitivity and specificity (Figures 6C,D). Interestingly, we found that patients with counts >5.1 had a 66% probability of being MRD$^{+/+}$ to treatment (refractory), and this probability increased proportionally to increases in gene expression. Similarly, the probability of death increased when counts were >7.0 (Figure 6E).

3.5 Relationship between gene expression and risk of death

With the aim of testing the potential of the identified genes as predictive biomarkers of mortality, four Cox regression models were performed to determine the clinical variables that influence patient survival (white blood cell count at diagnosis, age, extramedullary infiltration, response on day 8 of corticosteroid treatment, MRD on

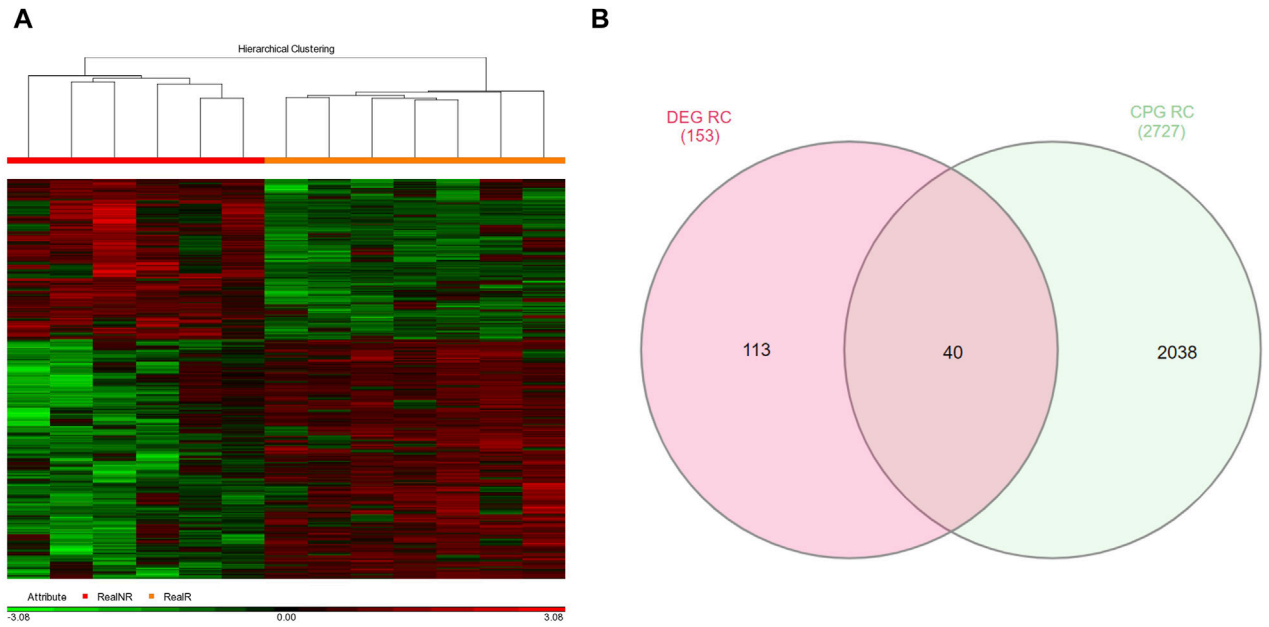


FIGURE 3
Genes with differentially methylated CpGs between MRD+ and MRD-patients. **(A)** Heatmap of differentially methylated GDMCpGs between MRD−/− and MRD$^{+/+}$. Each column represents an individual patient and horizontal axis indicates the beta values of each differentially methylated CpG. Hypomethylated probes are shown in green and hypermethylated probes in red. The MRD−/− patient group is shown in orange and the MRD$^{+/+}$ patient group in red. FC > 2 and <−2, *p*-value <0.05. **(B)** Venn diagram identifying common DEGs and GDMCpGs after MRD−/− vs. MRD$^{+/+}$ comparison.

day 15, and at the end of treatment). induction) and overexpression of the identified genes.

The initial model incorporated current clinical variables used to determine the risk of death, but none of these variables demonstrated a significant association with increased risk of mortality. In the second model, both clinical variables and gene overexpression were considered, with none of these variables increasing the risk of death. In the third model, MRD and overexpression of selected genes were included, revealing that overexpression of *MIR4435-2HG* emerged as the unique variable that elevated the risk of death 74-fold. Similarly, the fourth model, which evaluated the complete gene profile, indicated that overexpression of *MIR4435-2HG* significantly elevated the risk of death. (Table 3).

In line with the previous result, survival analysis showed that patients with *MIR4435-2HG* overexpression had worse survival; however, it is important to validate this result in a larger cohort of patients (Figure 7A). Since MRD is the current variable most commonly used to define the risk of death, however, for the survival analysis we first considered MRD at the end of induction, effectively demonstrating that MRD-patients have better survival than MRD + patients. (Figure 7B). Importantly, a more accurate separation of survival curves was achieved when we compared the survival of patients combining MRD+ with *MIR4435-2HG* overexpression *versus* MRD-patients with *MIR4435-2HG* down-expression. This revealed that patients with MRD+ and *MIR4435-2HG* overexpression experienced markedly worse survival (Figure 7C).

Finally, given that more than half of our patients had normal karyotype, we evaluated whether selected genes could improve risk classification in this subgroup of patients. Remarkably, the simultaneous overexpression of *MIR4435-2HG*, *DAPK1*, and *ASLC2* was associated with worse survival in patients with normal karyotype compared to those who did not overexpress them (Figure 7D).

4 Discussion

Given the wide genetic and epigenetic heterogeneity inherent in ALL, there is a critical need for new biomarkers to improve the prognosis of patients (Lejman et al., 2022). The present study conducted an integrative analysis of genome-wide DNA methylation and gene expression by RNA-seq in a cohort of 14 pediatric patients with B-cell ALL to explore whether differential DNA methylation genes and gene expression patterns could be proposed as potential predictive biomarkers that differentiate responder from non-responder patients and confer risk of death in pediatric patients with B-cell ALL.

Aberrant DNA methylation has been considered a hallmark in different types of cancer, including ALL (McCabe et al., 2009; Hanahan, 2022). Consistent with the findings of Borssén et al., 2018, our study demonstrated a clear separation in both DNA methylation and gene expression profiles between MRD- and MRD + patients. Notably, the overexpression of genes was associated with a more aggressive phenotype. Previously, Figueroa et al., 2013 reported that aberrant DNA methylation in childhood ALL could play a crucial role as a determinant of gene expression in disease-specific alterations. In our study, we observed a negative correlation between hypomethylation of CpGs and overexpression of genes *DAPK1*, *CNKSR3*, *MIR4435-HG2*, *CTHRC1*, *NPDC1*, *SLC45A3*, *ITGA6*, *ASCL2*, and *BOC*, supporting the idea that changes in DNA methylation have the

TABLE 2 Relationship between beta values of CPGS sites and normalized RNASEQ counts of differentially expressed genes between MRD$^{+/+}$ and MRD−/− patients.

	Relation to CpGs island	AFC050	AEM087	LSA023	JM040	CZA072	DFP084	KDA013	DSR017	JSR021	HCMMX1	JMC053	LVC063	ESR082	*p*-value	Rho
CNKSR3																
Normalized RNAseq counts		5.39	4.49	7.01	8.20	11.94	8.41	4.20	4.28	4.58	5.90	4.77	6.10	3.63	0.04	−0.57
cg00460149		0.10	0.22	0.07	0.06	0.09	0.09	0.24	0.60	0.13	0.21	0.20	0.07	0.48		
CTHRC1																
Normalized RNAseq counts		6.00	6.17	3.63	10.76	11.57	10.42	4.29	4.09	4.24	4.35	7.22	5.33	4.28	0,001	−0.78
cg01224234		0.47	0.47	0.22	0.05	0.06	0.17	0.79	0.83	0.81	0.82	0.07	0.73	0.83		
NPDC1																
Normalized RNAseq counts		11.79	11.51	5.83	7.91	13.17	12.45	4.66	5.68	5.32	4.71	6.42	8.90	9.16	0,003	−0.74
cg14190761	Island	0.07	0.15	0.20	0.09	0.05	0.08	0.29	0.24	0.18	0.18	0.20	0.26	0.18		
DAPK1																
Normalized RNAseq counts		13.20	10.94	5.75	12.47	12.19	13.07	10.23	4.65	7.59	9.04	4.98	8.69	5.37		
cg08719486	N_Shore	0.74	0.43	0.55	0.43	0.14	0.35	0.82	0.72	0.88	0.47	0.82	0.85	0.47	0.14	−0.42
cg11518830		0.07	0.13	0.35	0.07	0.17	0.10	0.13	0.52	0.44	0.16	0.55	0.18	0.75	0,0001	−0.87
SLC45A3																
Normalized RNAseq counts		8.83	8.50	6.99	9.67	12.74	12.34	7.11	5.95	6.40	5.27	9.32	7.30	5.46		
cg01455178	N_Shore	0.32	0.72	0.52	0.71	0.15	0.24	0.74	0.71	0.75	0.80	0.78	0.81	0.39	0.02	−0.60
cg04896348		0.37	0.69	0.63	0.09	0.15	0.41	0.61	0.58	0.62	0.61	0.61	0.67	0.65	0,007	−0.69
ITGA6																
Normalized RNAseq counts		14.90	10.68	8.75	10.51	15.81	15.34	10.62	11.99	8.73	10.83	10.01	10.64	8.06		

(Continued on following page)

10.3389/fmolb.2024.1385140

TABLE 2 (*Continued*) Relationship between beta values of CPGS sites and normalized RNASEQ counts of differentially expressed genes between MRD$^{+/+}$ and MRD−/− patients.

	Relation to CpGs island	AFC050	AEM087	LSA023	JM040	CZA072	DFP084	KDA013	DSR017	JSR021	HCMMX1	JMC053	LVC063	ESR082	*p*-value	Rho
cg07592198		0.12	0.46	0.46	0.82	0.29	0.36	0.73	0.80	0.70	0.55	0.75	0.75	0.74	0,008	−0.69
cg13586889		0.05	0.14	0.08	0.08	0.05	0.06	0.39	0.19	0.37	0.16	0.22	0.10	0.20	0.05	−0.53
MIR4435-2HG																
Normalized RNAseq counts		6.30	5.54	4.90	6.04	9.43	7.20	4.55	4.74	4.89	4.55	5.15	4.44	4.42	0.04	−0.56
cg24783876		0.08	0.69	0.53	0.08	0.26	0.06	0.22	0.68	0.73	0.79	0.65	0.42	0.67		
SLC18A2																
Normalized RNAseq counts		4.08	4.3	4.85	6.1	9.46	11.4	4.29	4.28	4.48	4.05	4.25	4	3.63		
cg03570973	N_Shore	0.81	0.5	0.4	0.15	0.67	0.07	0.79	0.59	0.79	0.71	0.82	0.81	0.8	0.01	−0.63
cg08521987	Island	0.51	0.11	0.2	0.22	0.1	0.07	0.67	0.4	0.5	0.26	0.44	0.48	0.78	0.01	−0.64
cg27186877		0.12	0.1	0.08	0.09	0.08	0,007	0.17	0.17	0.16	0.06	0.54	0.34	0.36	0.09	−0.47
cg00512279	Island	0.63	0.15	0.19	0.18	0.22	0,006	0.7	0.47	0.47	0.2	0.59	0.53	0.74	0.01	−0.63
ASCL2																
Normalized RNAseq counts		5.52	7.08	4.19	7.8	8.71	8.15	5.18	4.09	4.37	3.63	3.63	4.36	4.27		
cg11644479	Island	0.74	0.57	0.23	0.30	0.47	0.44	0.75	0.74	0.73	0.84	0.78	0.72	0.74	0.03	−0.59
cg13762320	Island	0.91	0.74	0.61	0.60	0.68	0.71	0.93	0.77	0.83	0.86	0.94	0.91	0.91	0.03	−0.59
cg19284039	Island	0.83	0.62	0.29	0.36	0.56	0.45	0.85	0.50	0.87	0.89	0.85	0.78	0.72	0.09	−0.48
cg26051413	Island	0.74	0.62	0.32	0.30	0.54	0.43	0.69	0.46	0.82	0.65	0.79	0.73	0.69	0.12	−0.45
cg13930892	Island	0.61	0.23	0.43	0.17	0.21	0.23	0.55	0.32	0.42	0.28	0.74	0.78	0.74	0.01	−0.63
BOC																
Normalized RNAseq counts		4.52	6.09	4.50	4.59	8.17	7.60	4.10	4.09	4.06	3.63	5.04	4.52	4.27		
cg22413784		0.55	0.75	0.76	0.12	0.42	0.41	0.75	0.74	0.78	0.81	0.74	0.76	0.80	0,008	−0.70

10.3389/fmolb.2024.1385140

10.3389/fmolb.2024.1385140

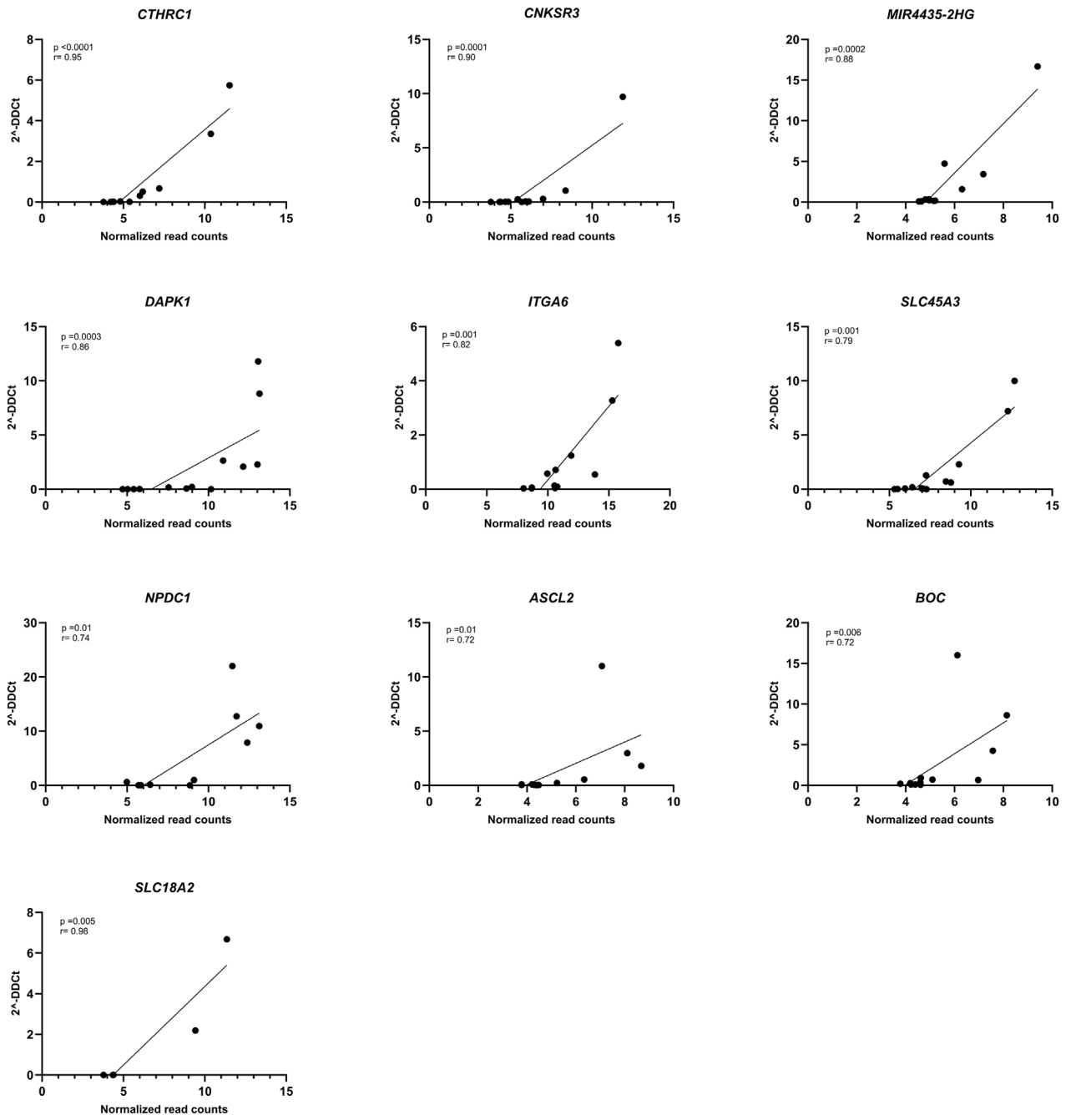


FIGURE 4
Correlation of RNA-seq data with RT-qPCR. Spearman correlation plot between normalized RNA-seq counts (*X*-axis) and $2^{-\Delta\Delta CT}$ (*Y*-axis) of *DAPK1* (*n* = 13), *CNKSR3* (*n* = 12), *MIR4435-HG2* (*n* = 13), *CTHRC1* (*n* = 12), *NPDC1* (*n* = 10), *SLC45A3* (*n* = 13), *ITGA6* (*n* = 12), *ASCL2* (*n* = 11), *BOC* (*n* = 13), and *SLC18A2* (*n* = 6) genes. *p*-value <0.05 and r < 0.70.

potential to influence gene expression. While it is widely recognized that promoter methylation can influence gene expression (Moore et al., 2013), the specific biological mechanisms driving this alteration in leukemias remain unclear. Some researchers have proposed several approaches that may be linked to mutations in the epigenetic machinery. For instance, mutations in DNMTAs have been associated with a gain of function in the protein, leading to either global or segmented hypermethylation (Schulze et al., 2016; Brunetti et al., 2017). Conversely, alterations in DNA demethylation mechanisms, such as gain-of-function mutations in TET enzymes, can result in zones of hypomethylation (Huang et al., 2013; Bowman and Levine, 2017; Wu and Zhang, 2017). Additionally, some studies have suggested a correlation between altered methylation states and the availability of the substrate SAM (S-adenosylmethionine), indicating that a low dietary intake of SAM-containing foods could impact an individual's methylation states (Mentch et al., 2015).

10.3389/fmolb.2024.1385140

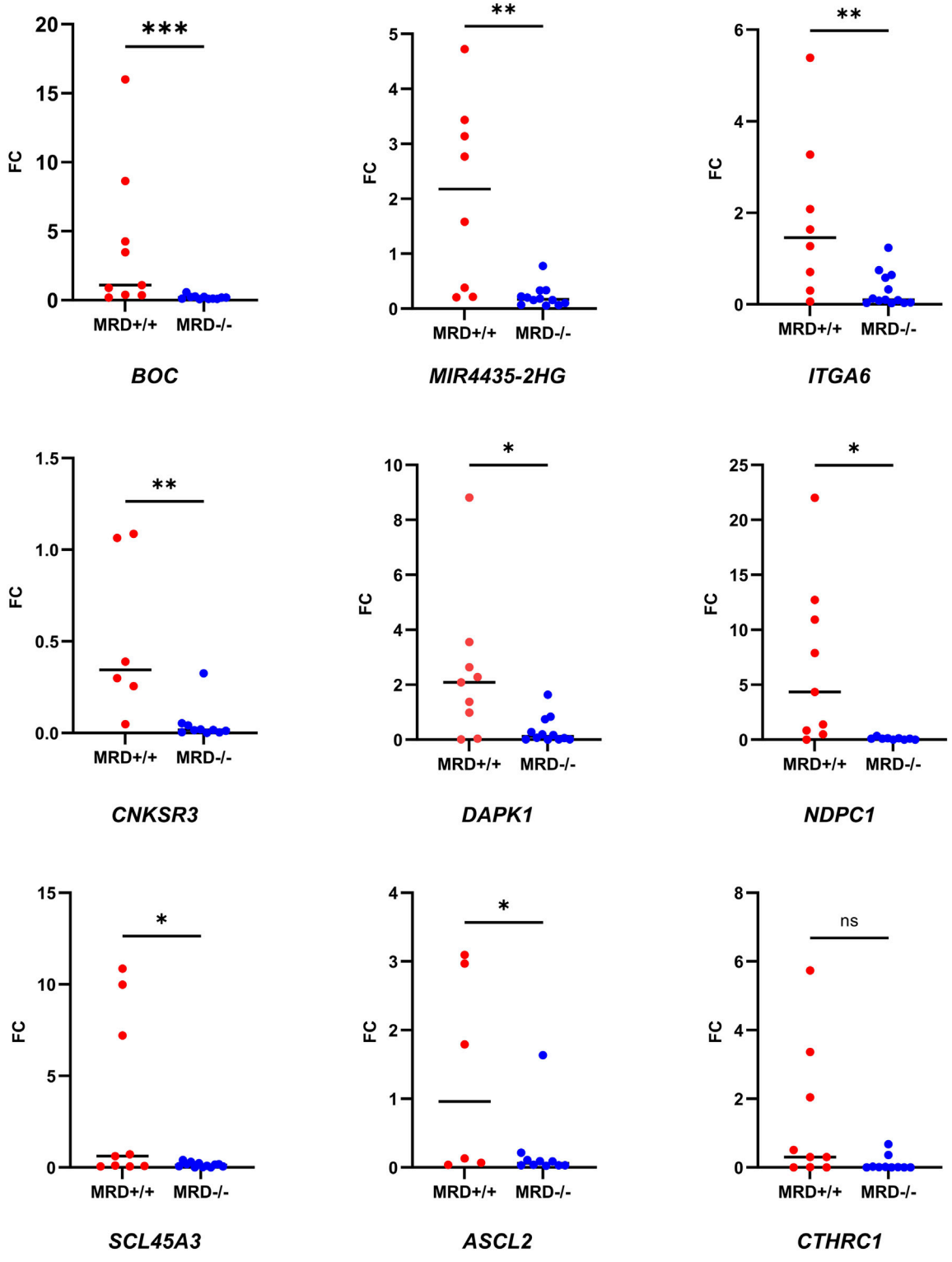


FIGURE 5
Comparison of fold change of the genes between MRD- and MRD + by RT-qPCR. Dot plot showing the comparison of FC of the nine selected genes between MRD−/− and MRD$^{+/+}$ patients. Mann-Whitney test was used to compare FC between the groups (*p*-value = 0.03 (*), 0.0021 (**); 0.0008 (***), ns = >0.05)). Number of samples corresponding to each analysis: *MIR4435-HG2* (MRD−/− = 12 vs. MRD$^{+/+}$ = 8), *DAPK1* (MRD−/− = 12 vs. MRD$^{+/+}$ = 9), *CNKSR3* (MRD−/− = 10 vs. MRD$^{+/+}$ = 6), *CTHRC1* (MRD−/− = 12 vs. MRD$^{+/+}$ = 9), *NPDC1* (MRD−/− = 9 vs. MRD$^{+/+}$ = 9), *SLC45A3* (MRD−/− = 12 vs. MRD$^{+/+}$ = 9), *ITGA6* (MRD−/− = 13 vs. MRD$^{+/+}$ = 8), *ASCL2* (MRD−/− = 10 vs. MRD$^{+/+}$ = 6) and *BOC* (MRD−/− = 11 vs. MRD$^{+/+}$ = 9).

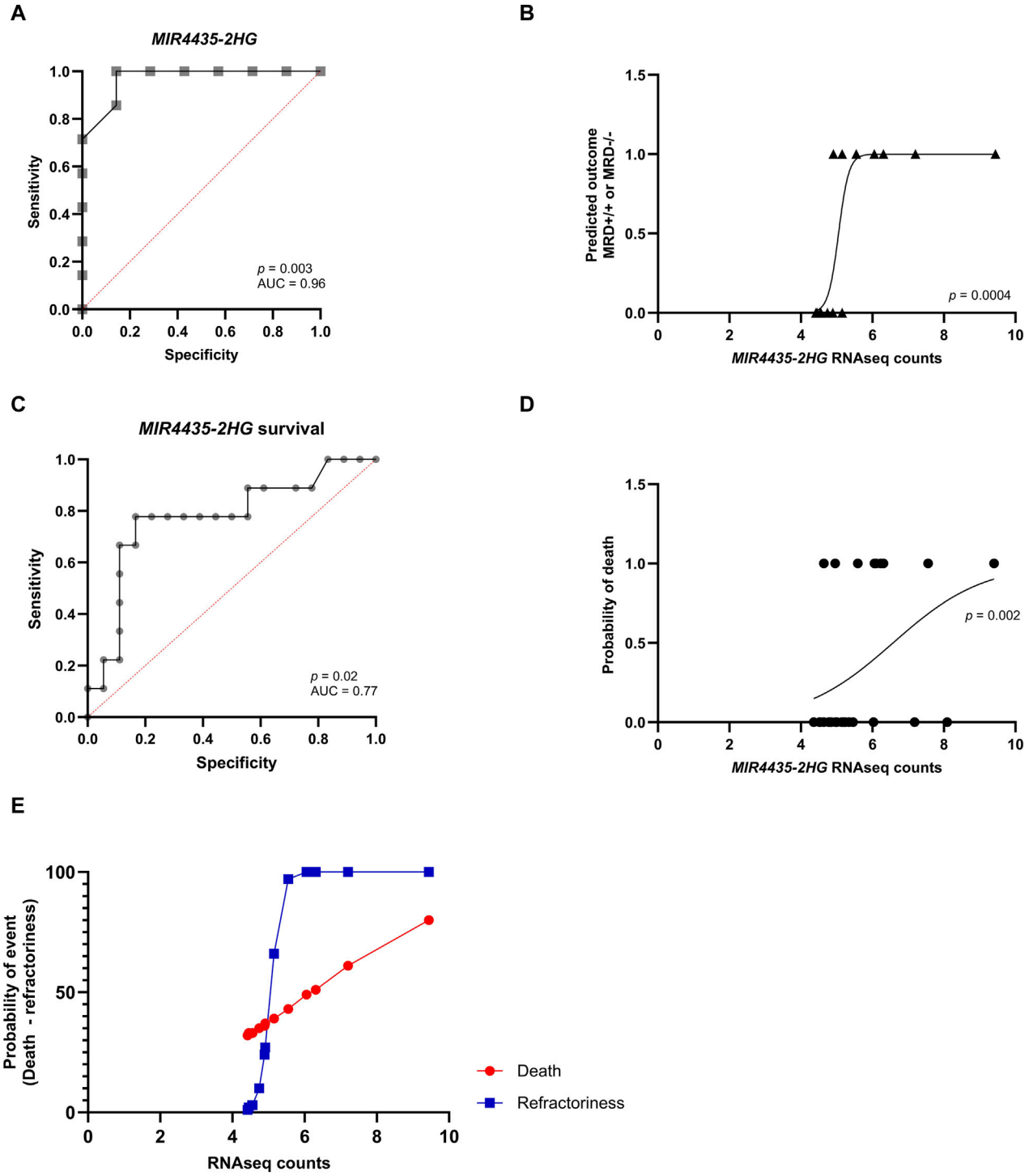


FIGURE 6
ROC curves and logistic regression graphs of *MIR4435-2HG* as a predictor of chemotherapy response and survival. Logistic regression analysis and ROC curves to determine the predictive capacity of the *MIR4435-2HG* and to determine its sensitivity and specificity. In logistic regressions, the probability of having one of the outcomes is represented on the *Y*-axis by numbers between 0 and 1, where 1 indicates that the patient does not respond to treatment and 0 indicates that the patient responds to treatment. On the *X*-axis, RNA-seq counts for the *MIR4435-2HG* gene are observed. **(A,B)** Prediction of if patient will be MRD−/− or MRD$^{+/+}$ and **(C,D)** whether they will survive. Scatter plot of the relationship of RNAseq counts to the probability of death (red) or being refractory (blue) **(E)**. Likelihood ratio test, *p*-value <0.05 and area under the ROC curve >0.77 and *p*-value <0.05.

However, this study does not fully elucidate the biological mechanism underlying hypomethylation associated with gene overexpression. Nevertheless, these findings provide valuable insights that can help generate new hypotheses to further understand the underlying biological mechanisms.

TABLE 3 Multiple Cox regression using currently clinical variables and gene profile.

Parameter	Model 1. Currently used clinical variables		Model 2. Currently used clinical variables and gene profile		Model 3. Gene profile and MRD		Model 4. Gene profile	
	p-value	OR (95%CI)	*p*-value	OR (95%CI)	*p*-value	OR (95%CI)	*p*-value	OR (95%CI)
Age	0.62	NC	0.89	0.20 (0.001–4.79)				
WBC	0.72	NC	>0.99	NC				
Extramedullar infiltration	>0.99	NC	>0.99	NC				
Corticoid response	0.18	NC	>0.99	NC				
MRD + day 15	0.90	NC	0.39	0.14 (0.0002–9.03)	0.29	0.20 (0.002–3.24)		
MRD + end of induction	0.31	NC	0.27	3.71 (0.22–258.3)	0.14	6.82 (0.50–189.3)		
MIR4435-2HG			0.17	NC	**0.01**	74.38 (4.18–6502)	**0.007**	22.7 (2.73–327.7)
DAPK1			>0.99	NC	0.37	3.12 (0.16–96.95)	0.27	3.23 (0.30–28.07)
ITGA6			0.15	0.01 (0.0006–5.48)	0.06	0.03 (0.00002–1.59)	0.11	0.06 (0.001–1.89)
NPDC1			>0.99	NC	0.06	13.22 (0.68–631)	0.07	12.62 (1.02–342.6)
CNKSR3			0.78	1.53 (0.04–36.46)	0.61	1.17 (0.11–9.55)	0.93	1.09 (0.12–8.72)
ASCL2			>0.99	NC	0.60	1.58 (0.08–29.20)	0.99	0.98 (0.05–19.78)
CTHRC1			>0.99	NC	0.59	0.43 (0.02–5.49)	0.22	0.25 (0.02–2.42)
SCL45A3			>0.99	NC	0.63	2.55 (0.08–45.41)	0.97	1.04 (0.02–72.07)
BOC			0.50	0.19 (0.0004–19.70)	0.16		0.51	0.39 (0.02–10.47)

MRD + indicates minimal residual disease positive. NC, No calculable;. WBC, white blood counts. Statistically significant results are shown in bold.

Aberrant patterns of DNA methylation have been linked to clinical outcome in patients with ALL; however, further research is required to evaluate the clinical utility of some of these findings (Tsellou et al., 2005; Roman-Gomez et al., 2007; Kuang et al., 2008; Musialik et al., 2015; Mai et al., 2016; Ogawa et al., 2016). In this study, we selected *MIR4435-2HG*, *DAPK1*, *ASCL2*, *BOC*, and *CNKSR3* as potential biomarkers of treatment response. Their overexpression reliably predicts treatment failure or refractoriness with high sensitivity and specificity. Notably, among these biomarkers, *MIR4435-2HG* stands out as the most robust predictor of therapeutic failure.

Recently, selected genes had been described as possible diagnostic and prognostic biomarkers in different types of cancer and other non-neoplastic diseases (Table 4).

In particular, *MIR4435-2HG*, which is a long non-coding RNA, is also known as *LncRNA-AWPPH*, *LINC00978*, or *MORRBID* (Ghasemian et al., 2022). Interestingly, *MIR4435-2HG* overexpression has previously been associated with hypomethylation in gliomas (Zhong et al., 2022). In patients with T-cell ALL, *MIR4435-2HG* showed an elevated expression compared to healthy individuals and has been linked to the promotion of proliferation as well as the inhibition of apoptosis of ALL cell lines (Li et al., 2020). Although the precise biological role of *MIR4435-2HG* is still under investigation, it is known to contribute by deregulating different signaling pathways associated with proliferation, invasion, migration, epithelial-mesenchymal transition, and apoptosis. Specifically, it plays a role in signaling pathways such as TGF-β, WNT-β catenin, MDM2/p53, PI3K/AKT, Hippo, and MAPK/ERK (Ouyang et al., 2019; Zhong et al., 2022).

Unfortunately, no clinical variable was identified as risk factor for death in our population; however, *MIR4435-2HG* overexpression was found to significantly increases risk of death, predicts death, and correlated with poorer survival. Similar findings have been reported in acute myeloid leukemia by Zhigang Cai et al., 2020. Moreover, in other cancer models, *MIR4435-2HG* overexpression has consistently been associated with worse progression-free survival and overall survival (Ouyang et al., 2019; Zhu et al., 2020; Zhong et al., 2022).

No studies have explored the relationship between aberrant DNA methylation, gene overexpression, and MRD during induction chemotherapy. However, certain authors have reported differences in the methylation profiles of patients who experienced relapse compared to those who did not (Borssén et al., 2018), where patients with a less methylated CpG island methylator phenotype at diagnosis exhibited inferior overall survival compared to those with more methylated CpG island phenotype. In a prior study, Sandoval et al., 2012 reported hypomethylation in various genome regions, including Polycomb target genes, and its association with poor survival and relapse. Similarly, Hogan et al. reported epigenetic dysregulation in the acquisition of chemoresistance during relapse, involving genes *CDKN2A*, *COL6A2*, *PTPRO*, and *CSMD1* (Hogan et al., 2015).

The search for biomarkers in the transcriptome or methylome of patients is very valuable, especially when 25% of patients with pediatric leukemia lack detectable genetic alterations and have a low mutation rate, which is a challenge for risk classification

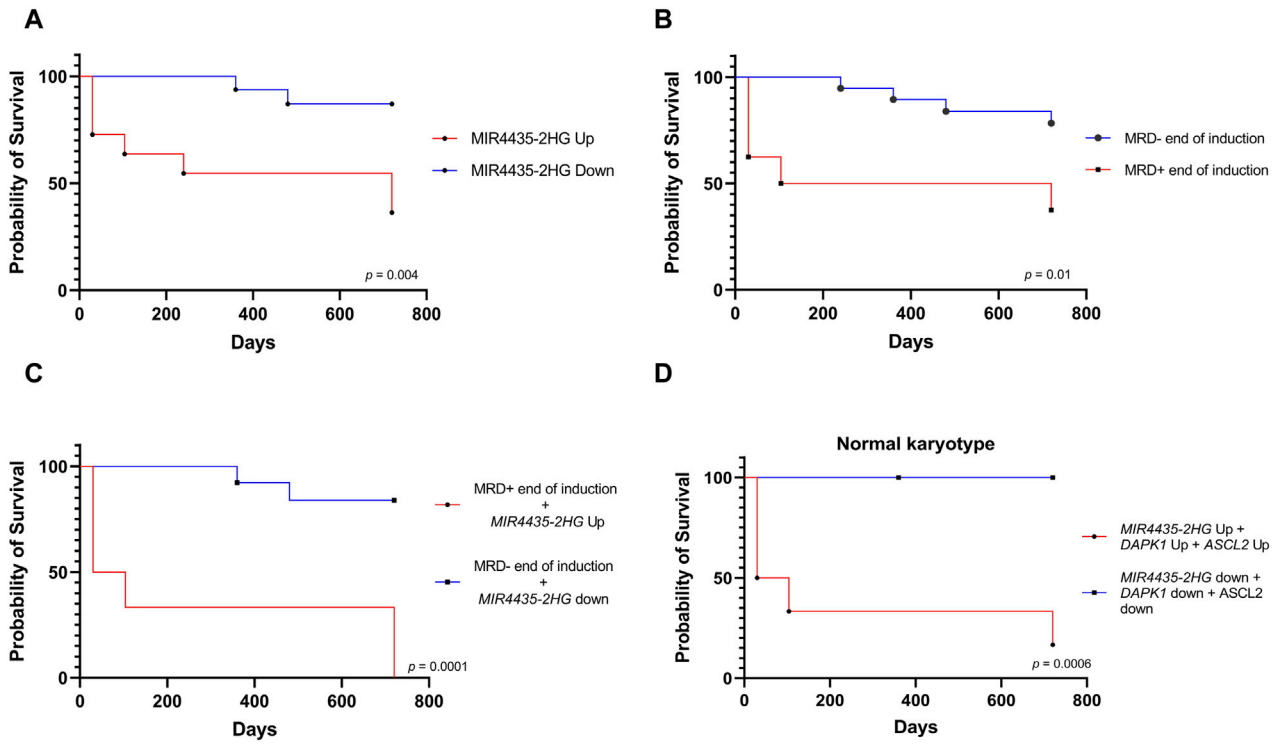


FIGURE 7
Relationship of the *MIR4435-2HG* expression with death. Kaplan-Meier analysis comparing 2-year survival with respect **(A)** *MIR4435-2HG* expression (n = 26); **(B)** MRD—or MRD + at the end of induction treatment (n = 26); **(C)** patients with overexpression of *MIR4435-2HG* and MRD + at the end of induction vs patients with low expression of *MIR4435-2HG* and MRD-at the end of induction (n = 26); and **(D)** patients with normal karyotype and overexpression or not of *MIR4435-2HG*, *ASCL2*, and *DAPK1* (n = 15). p-value = <0.05.

TABLE 4 List of genes choose as possible predictive biomarkers of induction chemotherapy response.

Gene name	FC	*p*-value	Rho	Cancer association	Ref
DAPK1	3.9	0,000009	−0.91	Gastric, pancreatic, head and neck, thyroid, brain, uterine, lung, esophageal cancers, CLL, AML and MDS	Calmon et al. (2007), Greco et al. (2010), Qin et al. (2014), Wang et al. (2014), Yuan et al. (2017), Wei et al. (2020), Gasimli et al. (2022), Guru et al. (2022), Movahhed et al. (2022)
CNKSR3	5.7	0,0003	−0.85	Melanoma	Lake et al. (2013)
SLC18A2	8.7	0,0009	−0.8	Prostate cancer and AML	Sørensen et al. (2009), Lebedev et al. (2019)
CTHRC1	5.5	0,003	−0.76	Renal, head and neck, liver, stomach, lung, endometrial and colorectal cancers	Sial et al. (2021), Meng et al. (2022)
NPDC1	4.4	0,003	−0.74	Gastric, neuroblastoma, pancreatic neuroendocrine tumors and AML	Bloomston et al. (2004), Tong et al. (2013), Nguyen et al. (2019), Dong et al. (2022)
BOC	4.6	0.01	−0.68	Gastric and pancreatic cancers	Mathew et al. (2014), Fattahi et al. (2021)
ASCL2	5.4	0,004	−0.66	Colorectal, gastric and lung cancer. Risk of ALL in pregnancy	Kwon et al. (2013), Hu et al. (2016), Potter et al. (2018), Zuo et al. (2018), Wu et al. (2022)
SLC45A3	3.9	0.03	−0.59	Prostate cancer	Esgueva et al. (2010)
		0,04	−0.57		
ITGA6	3.8	0.02	−0.63	Multiple myeloma and AML	Yamakawa et al. (2012), Song et al. (2021)
MIR4435-2HG	4.4	0.02	−0.55	Digestive, reproductive, respiratory, nervous, and urinary tumors, AML and T-cell ALL	Ouyang et al. (2019), Shen et al. (2020), Zhu et al. (2020), Ghasemian et al. (2022), Zhong et al. (2022)

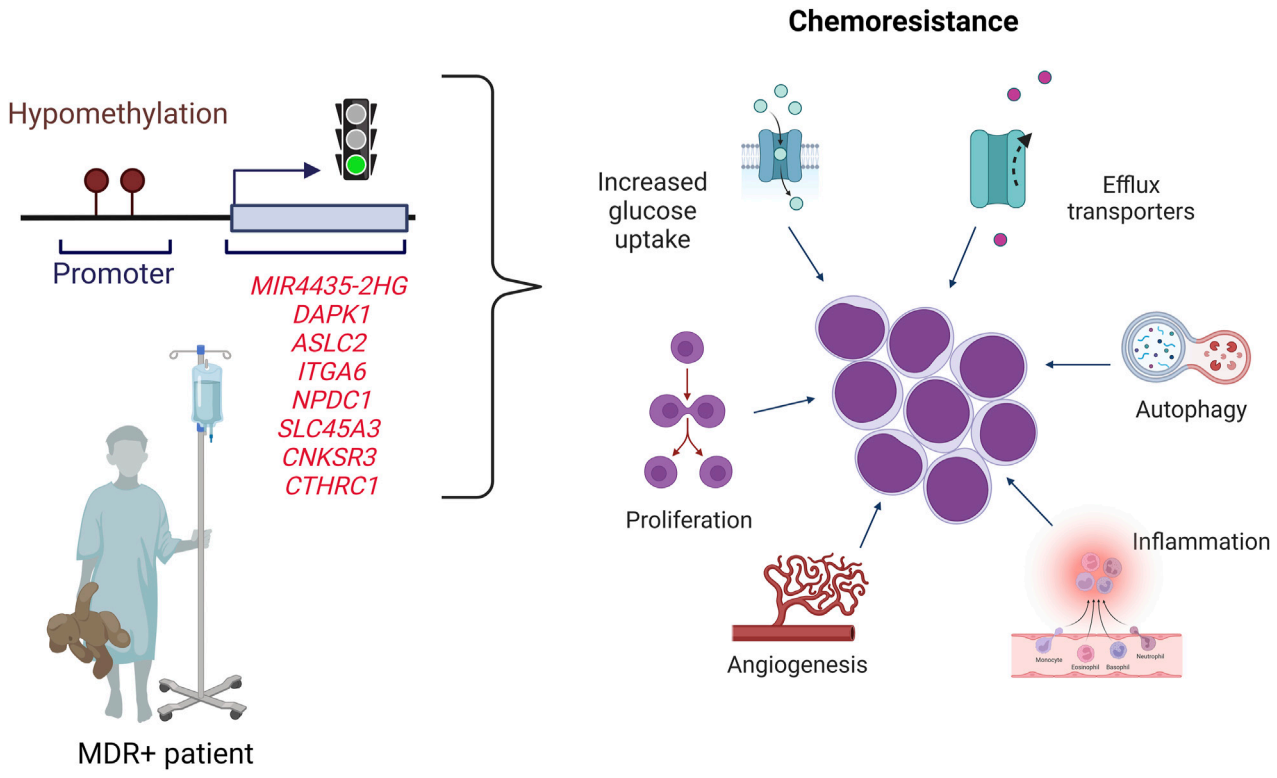


FIGURE 8
Proposed scheme for pediatric B-cell ALL patients. Patients with B-ALL who overexpress the *MIR4435-2HG*, *DAPK1*, *ASCL2*, *ITGA6*, *NPDC1*, *SLC45A3*, *CNKSR3* and *CTHRC1* genes have a high probability of being MRD+ and dying. The overexpression could be related to hypomethylation of the CpGs sites of these genes. The genes overexpressed are related to different hallmarks of cancer.

(Iacobucci and Mullighan, 2017). Interestingly, in our cohort, more than 70% of our patients showed no genetic alterations but displayed overexpressed of *MIR4435-2HG*, *DAPK1* and *ASCL2*, which correlated with worse survival. These results suggest that assessing the expression of these genes by RT-qPCR could improve risk classification, especially in patients without genetic alterations.

While *MIR4435-2HG* overexpression appears to be a poor prognostic factor, the association of *DAPK1* expression with poor prognosis is controversial (San Jose-Eneriz et al., 2013). In this study, we associated *DAPK1* overexpression and hypomethylation with therapeutic failure, and poorer survival in patients with normal karyotype. *DAPK1* has also been associated with resistance to imatinib in chronic myeloid leukemia (Guru et al., 2022), autophagia (Singh et al., 2016), alterations in the p53 signaling pathway in chronic lymphocytic leukemia (Wang et al., 2014) and methylated in myelodysplastic syndrome (Greco et al., 2010).

Here we propose a gene profile that can predict treatment response in children with B-ALL In particular, we demonstrate for the first time that *MIR4435-2HG* is overexpressed and hypomethylated in MRD + patients, and that it has the ability to predict treatment response and confer an increased risk of death in those patients who overexpress it (Figure 8). The detection of *MIR4435-2HG* could be combined with MRD analysis to improve risk classification, particularly in patients with normal karyotype. The proposed genetic profile offers the possibility of expanding research into new biomarkers predictive of response to treatment, which, in the future, would be a valuable tool to improve risk classification. A great contribution of this study is that the genes can be identified by RT-qPCR, which is efficient, fast, and cost-effective at the clinical level.

The limitations of this study are associated with the relatively low number of newly diagnosed patients eligible for this investigation, as well as the limited number of MRD + patients. Another limitation is associated with the follow-up duration for the patients, typically limited 2 years in most cases, which does not allow us to generate a solid conclusion about patient survival (death). Therefore, it is crucial to validate these results in a larger cohort of patients with a prolonged follow-up period for a more comprehensive evaluation.

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 below: Gene Expression Omnibus (GEO), accession number GSE229056.

Ethics statement

The studies involving humans were approved by the Comité de investigaciones y ética de la facultad de medicina de la Pontificia Universidad Javeriana (Hospital Universitario San Ignacio). Comité

de ética en investigación (Hospital Militar Central). Comité de ética en investigaciones del Instituto Nacional de Cancerología (Instituto Nacional de Cancerología). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

YT-L: Writing–review and editing, Writing–original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal Analysis, Data curation, Conceptualization. JZ: Writing–review and editing, Supervision, Software, Methodology, Investigation, Formal Analysis, Data curation. NC-R: Writing–review and editing, Methodology, Investigation, Formal Analysis, Conceptualization. SQ: Writing–review and editing, Methodology, Formal Analysis, Data curation. PG: Writing–review and editing, Methodology, Formal Analysis, Conceptualization. IR: Writing–review and editing, Methodology, Formal Analysis. NP-G: Writing–review and editing, Methodology. AI: Writing–review and editing, Methodology, Funding acquisition. LL: Writing–review and editing, Methodology, Investigation, Formal Analysis, Data curation. Alba Lucia AC: Writing–review and editing, Writing–original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal Analysis, Conceptualization.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Instituto Nacional de Cancerología and Hospital Universitario San Ignacio. JZ received funding from NIH grants P20 GM121288-06 and R25GM1148309-01.

Acknowledgments

We are mainly grateful to the patients who agreed to participate in the study, as well as to their tutors. We also thank the institutions that recruited the patients (Instituto Nacional de Cancerología, Hospital Militar Central and Hospital Universitario San Ignacio), especially Dr. Amaranto Suárez who provided support in the diffusion of the project within the Instituto Nacional de Cancerología, as well as Dr. David Garay, Dr. Eddie Pabón, Dr. Camila Prada, Dr. Jorge Buitrago, Andrea Naranjo, Giovanna Bedon, Bibiana Martinez, Paula Toro and Cindy Arevalo. Finally, special thanks to the Fundación Colombiana de Leucemia y Linfoma for the financial support to Yulieth Torres-Llanos in her PhD studies.

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/fmolb.2024.1385140/full#supplementary-material

References

Bloomston, M., Durkin, A., Yang, I., Rojiani, M., Rosemurgy, A. S., Enkmann, S., et al. (2004). Identification of molecular markers specific for pancreatic neuroendocrine tumors by genetic profiling of Core biopsies. *Ann. Surg. Oncol.* 11 (4), 413–419. doi:10.1245/ASO.2004.03.077

Borssén, M., Nordlund, J., Haider, Z., Landfors, M., Larsson, P., Kanerva, J., et al. (2018). DNA methylation holds prognostic information in relapsed precursor B-cell acute lymphoblastic leukemia. *Clin. Epigenetics* 10 (1), 31–37. doi:10.1186/s13148-018-0466-3

Bowman, R. L., and Levine, R. L. (2017). TET2 in normal and malignant hematopoiesis. *Cold Spring Harb. Perspect. Med.* 7 (8), a026518. doi:10.1101/cshperspect.a026518

Brunetti, L., Gundry, M. C., and Goodell, M. A. (2017). DNMT3A in leukemia. *Cold Spring Harb. Perspect. Med.* 7 (2), a030320. doi:10.1101/cshperspect.a030320

Cai, Z., Aguilera, F., Ramdas, B., Daulatabad, S. V., Srivastava, R., Kotzin, J. J., et al. (2020). Targeting bim via a lncRNA morrbid regulates the survival of preleukemic and leukemic cells. *Cell Rep.* 31 (12), 107816. doi:10.1016/j.celrep.2020.107816

Calmon, M. F., Colombo, J., Carvalho, F., Souza, F. P., Filho, J. F. G., Fukuyama, É. E., et al. (2007). Methylation profile of genes CDKN2A (p14 and p16), DAPK1, CDH1, and ADAM23 in head and neck cancer. *Cancer Genet. Cytogenet* 173 (1), 31–37. doi:10.1016/j.cancergencyto.2006.09.008

Chatterton, Z., Burke, D., Emslie, K. R., Craig, J. M., Ng, J., Ashley, D. M., et al. (2014). Validation of DNA methylation biomarkers for diagnosis of acute lymphoblastic leukemia. *Clin. Chem.* 60 (7), 995–1003. doi:10.1373/clinchem.2013.219956

Cruz-Rodriguez, N., Combita, A. L., Enciso, L. J., Quijano, S. M., Pinzon, P. L., Lozano, O. C., et al. (2016). High expression of ID family and IGJ genes signature as predictor of low induction treatment response and worst survival in adult Hispanic patients with B-Acute lymphoblastic leukemia. *J. Exp. Clin. Cancer Res.* 35 (1), 64–14. doi:10.1186/s13046-016-0333-z

Cullen, S. M., Mayle, A., Rossi, L., and Goodell, M. A. (2014). Hematopoietic stem cell development: an epigenetic journey. *Curr. Top. Dev. Biol.* 107, 39–75. doi:10.1016/B978-0-12-416022-4.00002-0

Dong, R., Chen, S., Lu, F., Zheng, N., Peng, G., Li, Y., et al. (2022). Models for predicting response to immunotherapy and prognosis in patients with gastric cancer: DNA damage response genes. *Biomed. Res. Int.* 2022, 4909544. doi:10.1155/2022/4909544

Esgueva, R., Perner, S., J Lafargue, C., Scheble, V., Stephan, C., Lein, M., et al. (2010). Prevalence of TMPRSS2-ERG and SLC45A3-ERG gene fusions in a large prostatectomy cohort. *Mod. Pathol.* 23 (4), 539–546. doi:10.1038/modpathol.2009.193

Fattahi, S., Nikbakhsh, N., Ranaei, M., Sabour, D., and Akhavan-Niaki, H. (2021). Author Correction: association of sonic hedgehog signaling pathway genes IHH, BOC, RAB23a and MIR195-5p, MIR509-3-5p, MIR6738-3p with gastric cancer stage. *Sci. Rep.* 11 (1), 16027. doi:10.1038/s41598-021-95379-8

Figueroa, M., Chen, S., Andersson, A. K., Phillips, L. A., Li, Y., Sotzen, J., et al. (2013). Integrated genetic and epigenetic analysis of childhood acute lymphoblastic leukemia. *J. Clin. Invest.* 123 (7), 3099–3111. doi:10.1172/JCI66203

Garcia-Manero, G., Yang, H., Kuang, S. Q., O'Brien, S., Thomas, D., and Kantarjian, H. (2009). Epigenetics of acute lymphocytic leukemia. *Semin. Hematol.* 46 (1), 24–32. doi:10.1053/j.seminhematol.2008.09.008

Gasimli, K., Raab, M., Becker, S., Sanhaji, M., and Strebhardt, K. (2022). The role of DAPK1 in the cell cycle regulation of cervical cancer cells and in response to topotecan. *J. Cancer* 13 (3), 728–743. doi:10.7150/jca.66492

Ghasemian, M., Rajabibazl, M., Sahebi, U., Sadeghi, S., Maleki, R., Hashemnia, V., et al. (2022). Long non-coding RNA MIR4435-2HG: a key molecule in progression of cancer and non-cancerous disorders. *Cancer Cell Int.* 22, 215. doi:10.1186/s12935-022-02633-8

Greco, M., D'Alò, F., Scardocci, A., Criscuolo, M., Fabiani, E., Guidi, F., et al. (2010). Promoter methylation of DAPK1, E-cadherin and thrombospondin-1 in *de novo* and therapy-related myeloid neoplasms. *Blood Cells Mol. Dis.* 45 (3), 181–185. doi:10.1016/j.bcmd.2010.05.008

Guru, S. A., Sumi, M. P., Mir, A. R., Beg, M. M. A., koner, B. C., and Saxena, A. (2022). Aberrant hydroxymethylation in promoter CpG regions of genes related to the cell cycle and apoptosis characterizes advanced chronic myeloid leukemia disease, poor imatinib respondents and poor survival. *BMC Cancer* 22 (1), 405. doi:10.1186/s12885-022-09481-9

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

Hogan, L. E., Meyer, J. A., Yang, J., Wang, J., Wong, N., Yang, W., et al. (2015). Integrated genomic analysis of relapsed childhood acute lymphoblastic leukemia reveals therapeutic strategies. *Blood* 118 (19), 5218–5226. doi:10.1182/blood-2011-04-345595

Hu, D., and Shilatifard, A. (2016). Epigenetics of hematopoiesis and hematological malignancies. *Genes Dev.* 30 (18), 2021–2041. doi:10.1101/gad.284109.116

Hu, X. G., Chen, L., Wang, Q. L., Zhao, X. L., Tan, J., Cui, Y. H., et al. (2016). Elevated expression of ASCL2 is an independent prognostic indicator in lung squamous cell carcinoma. *J. Clin. Pathol.* 69 (4), 313–318. doi:10.1136/jclinpath-2015-203025

Huang, H., Jiang, X., Li, Z., Li, Y., Song, C. X., He, C., et al. (2013). TET1 plays an essential oncogenic role in MLL-rearranged leukemia. *Proc. Natl. Acad. Sci. U. S. A.* 110 (29), 11994–11999. doi:10.1073/pnas.1310656110

Iacobucci, I., and Mullighan, C. G. (2017). Genetic basis of acute lymphoblastic leukemia. *J. Clin. Oncol.* 35 (9), 975–983. doi:10.1200/JCO.2016.70.7836

Katz, A. J., Chia, V. M., Schoonen, W. M., and Kelsh, M. A. (2015). Acute lymphoblastic leukemia: an assessment of international incidence, survival, and disease burden. *Cancer Causes Control* 26 (11), 1627–1642. doi:10.1007/s10552-015-0657-6

Kuang, S. Q., Tong, W. G., Yang, H., Lin, W., Lee, M. K., Fang, Z. H., et al. (2008). Genome-wide identification of aberrantly methylated promoter associated CpG islands in acute lymphocytic leukemia. *Leukemia* 22 (8), 1529–1538. doi:10.1038/leu.2008.130

Kwon, O. H., Park, J. L., Baek, S. J., Noh, S. M., Song, K. S., Kim, S. Y., et al. (2013). Aberrant upregulation of ASCL2 by promoter demethylation promotes the growth and resistance to 5-fluorouracil of gastric cancer cells. *Cancer Sci.* 104 (3), 391–397. doi:10.1111/cas.12076

Lake, S. L., Damato, B. E., Kalirai, H., Dodson, A. R., Taktak, A. F. G., Lloyd, B. H., et al. (2013). Single nucleotide polymorphism array analysis of uveal melanomas reveals that amplification of CNKSR3 is correlated with improved patient survival. *Am. J. Pathology* 182 (3), 678–687. doi:10.1016/j.ajpath.2012.11.036

Lebedev, T. D., Vagapova, E. R., Popenko, V. I., Leonova, O. G., Spirin, P. V., and Prassolov, V. S. (2019). Two receptors, two isoforms, two cancers: comprehensive analysis of kit and trka expression in neuroblastoma and acute myeloid leukemia. *Front. Oncol.* 9, 1046. doi:10.3389/fonc.2019.01046

Lejman, M., Chałupnik, A., Chilimoniuk, Z., and Dobosz, M. (2022). Genetic biomarkers and their clinical implications in B-cell acute lymphoblastic leukemia in children. *Int. J. Mol. Sci.* 23, 2755. doi:10.3390/ijms23052755

Li, X., Song, F., and Sun, H. (2020). Long non-coding RNA AWPPH interacts with ROCK2 and regulates the proliferation and apoptosis of cancer cells in pediatric T-cell acute lymphoblastic leukemia. *Oncol. Lett.* 20 (5), 239. doi:10.3892/ol.2020.12102

Mai, H., Liu, X., Chen, Y., Li, C., Cao, L., Chen, X., et al. (2016). Hypermethylation of p15 gene associated with an inferior poor long-term outcome in childhood acute lymphoblastic leukemia. *J. Cancer Res. Clin. Oncol.* 142 (2), 497–504. doi:10.1007/s00432-015-2063-6

Matasar, M. J., Ritchie, E. K., Consedine, N., Magai, C., and Neugut, A. I. (2006). Incidence rates of the major leukemia subtypes among U.S. Hispanics, Blacks, and non-Hispanic Whites. *Leuk. Lymphoma* 47 (11), 2365–2370. doi:10.1080/10428190600799888

Mathew, E., Zhang, Y., Holtz, A. M., Kane, K. T., Song, J. Y., Allen, B. L., et al. (2014). Dosage-dependent regulation of pancreatic cancer growth and angiogenesis by Hedgehog signaling. *Cell Rep.* 9 (2), 484–494. doi:10.1016/j.celrep.2014.09.010

McCabe, M. T., Brandes, J. C., and Vertino, P. M. (2009). Cancer DNA methylation: molecular mechanisms and clinical implications. *Clin. Cancer Res.* 15, 3927–3937. doi:10.1158/1078-0432.CCR-08-2784

Meng, C., Zhang, Y., Jiang, D., and Wang, J. (2022). CTHRC1 is a prognosis-related biomarker correlated with immune infiltrates in colon adenocarcinoma. *World J. Surg. Oncol.* 20 (1), 89. doi:10.1186/s12957-022-02557-7

Mentch, S. J., Mehrmohamadi, M., Huang, L., Liu, X., Gupta, D., Mattocks, D., et al. (2015). Histone methylation dynamics and gene regulation occur through the sensing of one-carbon metabolism. *Cell Metab.* 22 (5), 861–873. doi:10.1016/j.cmet.2015.08.024

Moore, L. D., Le, T., and Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology* 38 (1), 23–38. doi:10.1038/npp.2012.112

Movahhed, P., Saberiyan, M., Safi, A., Arshadi, Z., Kazerouni, F., and Teimori, H. (2022). The impact of DAPK1 and mTORC1 signaling association on autophagy in cancer. *Mol. Biol. Rep.* 49, 4959–4964. doi:10.1007/s11033-022-07154-1

Musialik, E., Bujko, M., Kober, P., Wypych, A., Gawle-Krawczyk, K., Matysiak, M., et al. (2015). Promoter methylation and expression levels of selected hematopoietic genes in pediatric B-cell acute lymphoblastic leukemia. *Blood Res.* 50 (1), 26–32. doi:10.5045/br.2015.50.1.26

Newton, T. P., Cummings, C. T., Graham, D. K., and Bernt, K. M. (2014). Epigenetics and chemoresistance in childhood acute lymphoblastic leukemia. *Int. J. Hematol. Oncol.* 3 (1), 19–30. doi:10.2217/ijh.13.68

Nguyen, C. H., Glüxam, T., Schlerka, A., Bauer, K., Grandits, A. M., Hackl, H., et al. (2019). SOCS2 is part of a highly prognostic 4-gene signature in AML and promotes disease aggressiveness. *Sci. Rep.* 9 (1), 9139. doi:10.1038/s41598-019-45579-0

Nordlund, J., and Syvänen, A. C. (2018). Epigenetics in pediatric acute lymphoblastic leukemia. *Semin. Cancer Biol.* 51, 129–138. doi:10.1016/j.semcancer.2017.09.001

Ogawa, S., Tsuchida, M., Kaizu, K., Manabe, A., Sato, Y., Aoki, T., et al. (2016). Long-term outcome of 6-month maintenance chemotherapy for acute lymphoblastic leukemia in children. *Leukemia* 31 (3), 580–584. doi:10.1038/leu.2016.274

Ouyang, W., Ren, L., Liu, G., Chi, X., and Wei, H. (2019). Lncrna mir4435-2hg predicts poor prognosis in patients with colorectal cancer. *PeerJ* 2019 (4). doi:10.7717/peerj.6683

Patel, S. A., and Vanharanta, S. (2017). Epigenetic determinants of metastasis. *Mol. Oncol.* 11 (1), 79–96. doi:10.1016/j.molonc.2016.09.008

Poli, V., Fagnocchi, L., and Zippo, A. (2018). Tumorigenic cell reprogramming and cancer plasticity: interplay between signaling, microenvironment, and epigenetics. *Stem Cells Int.* 2018, 4598195. doi:10.1155/2018/4598195

Potter, C., Moorman, A. V., Relton, C. L., Ford, D., Mathers, J. C., Strathdee, G., et al. (2018). Maternal red blood cell folate and infant vitamin B12 status influence methylation of genes associated with childhood acute lymphoblastic leukemia. *Mol. Nutr. Food Res.* 62 (22), 1800411. doi:10.1002/mnfr.201800411

Pui, C.-H., Leslie, L., and Robison, A. T. L. (2008). Acute lymphoblastic leukaemia. *Lancet* 371, 1030–1043. doi:10.1016/s0140-6736(08)60457-2

Qin, Y., Ye, G. X., Wu, C. J., Wang, S., Pan, D. B., Jiang, J. Y., et al. (2014). Effect of DAPK1 gene on proliferation, migration, and invasion of carcinoma of pancreas BxPC-3 cell line. *Int. J. Clin. Exp. Pathol.* 7, 7536–7544.

Roman-Gomez, J., Jimenez-Velasco, A., Barrios, M., Prosper, F., Heiniger, A., Torres, A., et al. (2007). Poor prognosis in acute lymphoblastic leukemia may relate to promoter hypermethylation of cancer-related genes. *Leuk. Lymphoma* 48 (7), 1269–1282. doi:10.1080/10428190701344899

Sandoval, J., Heyn, H., Méndez-González, J., Gomez, A., Moran, S., Baiget, M., et al. (2012). Genome-wide DNA methylation profiling predicts relapse in childhood B-cell acute lymphoblastic leukaemia. *Br. J. Haematol.* 160, 406–409. doi:10.1111/bjh.12113

San Jose-Eneriz, E., Agirre, X., Rodriguez-Otero, P., and Prosper, F. (2013). Epigenetic regulation of cell signaling pathways in acute lymphoblastic leukemia. *Epigenomics* 5 (5), 525–538. doi:10.2217/epi.13.56

Schulze, I., Rohde, C., Scheller-Wendorff, M., Bäumer, N., Krause, A., Herbst, F., et al. (2016). Increased DNA methylation of Dnmt3b targets impairs leukemogenesis. *Blood* 127 (12), 1575–1586. doi:10.1182/blood-2015-07-655928

Shen, H., Sun, B., Yang, Y., Cai, X., Bi, L., Deng, L., et al. (2020). MIR4435-2HG regulates cancer cell behaviors in oral squamous cell carcinoma cell growth by upregulating TGF-β1. *Odontology* 108 (4), 553–559. doi:10.1007/s10266-020-00488-x

Sial, N., Ahmad, M., Hussain, M. S., Iqbal, M. J., Hameed, Y., Khan, M., et al. (2021). CTHRC1 expression is a novel shared diagnostic and prognostic biomarker of survival in six different human cancer subtypes. *Sci. Rep.* 11 (1), 19873. doi:10.1038/s41598-021-99321-w

Singh, P., Ravanan, P., and Talwar, P. (2016). Death associated protein kinase 1 (DAPK1): a regulator of apoptosis and autophagy. *Front. Mol. Neurosci.* 9, 46. doi:10.3389/fnmol.2016.00046

Sok, P., Brown, A. L., Taylor, O. A., Bernhardt, M. B., Bernini, J. C., Erana, R. A., et al. (2022). Abstract 3633: disparities in relapse among a large multi-ethnic population of children diagnosed with acute lymphoblastic leukemia (ALL): a report from the Reducing Ethnic Disparities in Acute Leukemia (REDIAL) Consortium. *Cancer Res.* 82 (12), 3633. doi:10.1158/1538-7445.AM2022-3633

Song, S., Zhang, J., Su, Q., Zhang, W., Jiang, Y., Fan, G., et al. (2021). Downregulation of ITGA6 confers to the invasion of multiple myeloma and promotes progression to plasma cell leukaemia. *Br. J. Cancer* 124 (11), 1843–1853. doi:10.1038/s41416-021-01362-5

Sørensen, K. D., Wild, P. J., Mortezavi, A., Adolf, K., Tørring, N., Heebøll, S., et al. (2009). Genetic and epigenetic SLC18A2 silencing in prostate cancer is an independent adverse predictor of biochemical recurrence after radical prostatectomy. *Clin. Cancer Res.* 15 (4), 1400–1410. doi:10.1158/1078-0432.CCR-08-2268

Stary, J., Zimmermann, M., Campbell, M., Castillo, L., Dibar, E., Donska, S., et al. (2014). Intensive chemotherapy for childhood acute lymphoblastic leukemia: results of the randomized intercontinental trial ALL IC-BFM 2002. *J. Clin. Oncol.* 32 (3), 174–184. doi:10.1200/JCO.2013.48.6522

Stumpel, D. J. P. M., Schneider, P., van Roon, E. H. J., Boer, J. M., de Lorenzo, P., Valsecchi, M. G., et al. (2009). Specific promoter methylation identifies different subgroups of MLL-rearranged infant acute lymphoblastic leukemia, influences clinical outcome, and provides therapeutic options. *Blood* 114 (27), 5490–5498. doi:10.1182/blood-2009-06-227660

Tong, C. W., Wang, J. L., Jiang, M. S., Hsu, C. H., Chang, W. T., and Huang, A. M. (2013). Novel genes that mediate nuclear respiratory factor 1-regualted neurite outgrowth in neuroblastoma IMR-32 cells. *Gene* 515 (1), 62–70. doi:10.1016/j.gene.2012.11.026

Tsellou, E., Troungos, C., Moschovi, M., Athanasiadou-Piperopoulou, F., Polychronopoulou, S., Kosmidis, H., et al. (2005). Hypermethylation of CpG islands in the promoter region of the p15INK4B gene in childhood acute leukaemia. *Eur. J. Cancer* 41 (4), 584–589. doi:10.1016/j.ejca.2004.12.010

Van Dongen Jjm, Lhermitte, L., Böttcher, S., Almeida, J., Van Der Velden, V. H. J., Flores-Montero, J., et al. (2012). EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. *Leukemia* 26 (9), 1908–1975. doi:10.1038/leu.2012.120

Van Dongen Jjm, Van Der Velden, V. H. J., Brüggemann, M., and Orfao, A. (2015). Minimal residual disease diagnostics in acute lymphoblastic leukemia: need for sensitive, fast, and standardized technologies. *Blood* 125 (26), 3996–4009. doi:10.1182/blood-2015-03-580027

van Dongen Jacques, J. M., Taku, S., Panzer-Grümayer, E. R., Andrea, B., Pongers-Willemse Marja, J., Lilly, C., et al. (1998). Prognostic value of minimal residual disease in acutelymphoblastic leukaemia in childhood. *Lancet* 352, 1731–1738. doi:10.1016/S0140-6736(98)04058-6

Vera, A. M., Pardo, C., Duarte, M. C., and Suárez, A. (2012). Experiencia en el análisis de la mortalidad por leucemia aguda pediátrica en el Instituto Nacional de Cancerología. *Biomédica* 32 (3), 355–364. doi:10.7705/biomedica.v32i3.691

Wainwright, E. N., and Scaffidi, P. (2017). Epigenetics and cancer stem cells: unleashing, hijacking, and restricting cellular plasticity. *Trends Cancer* 3 (5), 372–386. doi:10.1016/j.trecan.2017.04.004

Walsh, K. M., Chokkalingam, A. P., Hsu, L. I., Metayer, C., De Smith, A. J., Jacobs, D. I., et al. (2013). Associations between genome-wide Native American ancestry, known risk alleles and B-cell ALL risk in Hispanic children. *Leukemia* 27, 2416–2419. doi:10.1038/leu.2013.130

Walsh, K. M., de Smith, A. J., Welch, T. C., Smirnov, I., Cunningham, M. J., Ma, X., et al. (2014). Genomic ancestry and somatic alterations correlate with age at diagnosis in Hispanic children with B-cell acute lymphoblastic leukemia. *Am. J. Hematol.* 89 (7), 721–725. doi:10.1002/ajh.23727

Wang, L. Q., Kwong, Y. L., Wong, K. F., Kho, C. S. B., Jin, D. Y., Tse, E., et al. (2014). Epigenetic inactivation of mir-34b/c in addition to mir-34a and DAPK1 in chronic lymphocytic leukemia. *J. Transl. Med.* 12 (1), 52. doi:10.1186/1479-5876-12-52

Wei, J., Xie, Q., Liu, X., Wan, C., Wu, W., Fang, K., et al. (2020). Identification the prognostic value of glutathione peroxidases expression levels in acute myeloid leukemia. *Ann. Transl. Med.* 8 (11), 678. doi:10.21037/atm-20-3296

Wu, L., Sun, S., Qu, F., Liu, X., Sun, M., Pan, Y., et al. (2022). ASCL2 affects the efficacy of immunotherapy in colon adenocarcinoma based on single-cell RNA sequencing analysis. *Front. Immunol.* 13, 829640. doi:10.3389/fimmu.2022.829640

Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* 18 (9), 517–534. doi:10.1038/nrg.2017.33

Yamakawa, N., Kaneda, K., Saito, Y., Ichihara, E., and Morishita, K. (2012). The increased expression of integrin α6 (itga6) enhances drug resistance in evi1 high leukemia. *PLoS One* 7 (1), e30706. doi:10.1371/journal.pone.0030706

Younesian, S., Shahkarami, S., Ghaffari, P., Alizadeh, S., Mehrasa, R., and Ghaffari, S. H. (2019). Residual methylation of tumor suppressor gene promoters, RASSF6 and RASSF10, as novel biomarkers for minimal residual disease detection in adult acute lymphoblastic leukemia. *Ann. Hematol.* 98 (12), 2719–2727. doi:10.1007/s00277-019-03775-y

Yuan, W., Chen, J., Shu, Y., Liu, S., Wu, L., Ji, J., et al. (2017). Correlation of DAPK1 methylation and the risk of gastrointestinal cancer: a systematic review and meta-analysis. *PLoS One* 12 (9), e0184959. doi:10.1371/journal.pone.0184959

Zhong, C., Xie, Z., Zeng, L. H., Yuan, C., and Duan, S. (2022). MIR4435-2HG is a potential pan-cancer biomarker for diagnosis and prognosis. *Front. Immunol.* 13, 855078. doi:10.3389/fimmu.2022.855078

Zhu, L., Wang, A., Gao, M., Duan, X., and Li, Z. (2020). LncRNA MIR4435-2HG triggers ovarian cancer progression by regulating miR-128-3p/CKD14 axis. *Cancer Cell Int.* 20 (1), 145. doi:10.1186/s12935-020-01227-6

Zuo, Q., Wang, J., Chen, C., Zhang, Y., Feng, D. X., Zhao, R., et al. (2018). ASCL2 expression contributes to gastric tumor migration and invasion by downregulating miR223 and inducing EMT. *Mol. Med. Rep.* 18 (4), 3751–3759. doi:10.3892/mmr.2018.9363

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 01 May 2024
DOI 10.3389/fmolb.2024.1394585

Check for updates

OPEN ACCESS

EDITED BY
Matteo Becatti,
University of Firenze, Italy

REVIEWED BY
Zheng He,
Qilu Hospital of Shandong University, China
Junwei Cui,
Peking University, China

*CORRESPONDENCE
Lixue Xuan,
xuanlixuepumc@126.com

RECEIVED 01 March 2024
ACCEPTED 18 April 2024
PUBLISHED 01 May 2024

CITATION
Lei C, Li Y, Yang H, Zhang K, Lu W, Wang N and Xuan L (2024), Unraveling breast cancer prognosis: a novel model based on coagulation-related genes.
Front. Mol. Biosci. 11:1394585.
doi: 10.3389/fmolb.2024.1394585

COPYRIGHT
© 2024 Lei, Li, Yang, Zhang, Lu, Wang and Xuan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Unraveling breast cancer prognosis: a novel model based on coagulation-related genes

Chuqi Lei[1,2], Yuan Li[1,2], Huaiyu Yang[1,2], Ke Zhang[1,2], Wei Lu[1,2], Nianchang Wang[1,2] and Lixue Xuan[1,2]*

[1]Department of Breast Surgical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Beijing, China, [2]Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Objective: Breast cancer is highly heterogeneous, presenting challenges in prognostic assessment. Developing a universally applicable prognostic model could simplify clinical decision-making. This study aims to develop and validate a novel breast cancer prognosis model using coagulation-related genes with broad clinical applicability.

Methods: A total of 203 genes related to coagulation were obtained from the KEGG database, and the mRNA data of 1,099 tumor tissue samples and 572 samples of normal tissue were retrieved from the TCGA-BRCA cohort and GTEx databases. The R package "limma" was utilized to detect variations in gene expression related to coagulation between the malignancies and normal tissue. A model was constructed in the TCGA cohort through a multivariable Cox regression analysis, followed by validation using the GSE42568 dataset as the testing set. Constructing a nomogram incorporating clinical factors to enhance the predictive capacity of the model. Utilizing the ESTIMATE algorithm to investigate the immune infiltration levels in groups with deferent risk. Performing drug sensitivity analysis using the "oncoPredict" package.

Results: A risk model consisting of six coagulation-associated genes (SERPINA1, SERPINF2, C1S, CFB, RASGRP1, and TLN2) was created and successfully tested for validation. Identified were 6 genes that serve as protective factors in the model's development. Kaplan-Meier curves revealed a worse prognosis in the high-risk group compared to the low-risk group. The ROC analysis showed that the model accurately forecasted the overall survival (OS) of breast cancer patients at 1, 3, and 5 years. Nomogram accompanied by calibration curves can also provide better guidance for clinical decision-making. The low-risk group is more likely to respond well to immunotherapy, whereas the high-risk group may show improved responses to Gemcitabine treatment. Furthermore, individuals in distinct risk categories displayed different responses to various medications within the identical therapeutic category.

Conclusion: We established a breast cancer prognostic model incorporating six coagulation-associated genes and explored its clinical utility. This model offers valuable insights for clinical decision-making and drug selection in breast cancer patients, contributing to personalized and precise treatment advancements.

KEYWORDS

breast cancer, coagulation, prognostic model, drug sensitive, immune response

1 Introduction

Breast cancer is a global health problem on the rise that affects women of different ages in various countries. The most recent global cancer statistics from the World Health Organization (WHO) show that breast cancer has a prevalence of 11.6% worldwide, placing it in second position after lung cancer. Breast cancer has a mortality rate of 15.4% in women, leading to the highest number of cancer-related deaths among women (Siegel et al., 2024). Breast cancer exhibits diversity at the biological level, with prognosis being impacted by a range of factors such as molecular subtypes, age at diagnosis, and histopathological features (Brown et al., 2023). Recent advancements in early detection methods for breast cancer and the ongoing progress in treatment options have greatly enhanced the outlook for patients with this disease (Will et al., 2023; Brown et al., 2024). Especially as people's understanding of the immune microenvironment deepens, the utility of immunotherapy in breast cancer treatment continues to ascend. However, due to the substantial heterogeneity of the disease and intrinsic or acquired drug resistance, current clinical treatments still face significant challenges.

The tumor microenvironment (TME), serving as both a driving force and a regulatory factor in cancer development, facilitates tumor proliferation, migration, and treatment resistance (Gao et al., 2022; Shahzad et al., 2022; Xing et al., 2022). New evidence suggests that there is a close association between the coagulation process and the tumor microenvironment (Xu et al., 2022; Wei et al., 2023). Malignant solid tumors typically activate coagulation directly by releasing procoagulants like tissue factor (TF, encoded by the F3 gene). This establishment of a hypercoagulable state can lead to venous thromboembolism, resulting in local hypoxia and necrosis, thus reshaping the tumor microenvironment. This reshaping includes the accumulation of immunosuppressive immune cells, microvascular proliferation, and tumor cell migration, all of which promote tumor growth and metastasis (Riedl et al., 2017; Gofrit and Shavit-Stein, 2019; Feinauer et al., 2021). A comprehensive analysis across multiple types of cancer revealed a strong correlation between elevated levels of the fibrinolysis gene cluster and characteristics of the tumor microenvironment (TME), including the presence of monocytes and increased expression of immune checkpoint markers (Saidak et al., 2021). Additionally, research has shown that platelets have the ability to hinder the function of immune cells (such as by dampening the cytotoxic effects of Nature Killer cells and T cells) through the release of growth factors, cytokines, and coagulation factors, ultimately facilitating immune escape during tumor progression (Dann et al., 2018). More than just platelets, Graf's research indicates that FX synthesized by myeloid cells can also promote immune evasion in tumors. The group discovered that the utilization of a coagulation factor FX inhibitor could boost the presence of dendritic cells (DCs) and cytotoxic T cells in the tumor location (Graf et al., 2019). They also discovered that combining the FX inhibitor with anti-PD-L1 inhibitor significantly improved anti-tumor immunity. The investigations above demonstrate a tight interplay between the coagulation process and TME, this connection has a constantly impact on tumor initiation, progression, and regulating anti-tumor immune responses.

Different types of cells interact harmoniously in the tumor microenvironment (Pitt et al., 2016). Local recruitment of leukocytes and activation of inflammation in the TME intricately regulate coagulation and fibrin formation. The tumor coagulome, a molecular effector network favored by cancer, contributes to thrombosis or bleeding, has emerged as a hot topic in cancer research (Wahab et al., 2023). Recent studies (Tinholt et al., 2024) have revealed the crucial role of coagulation-related genes in the tumor microenvironment of breast cancer, particularly in predicting patient prognosis and response to chemotherapy. These findings suggest that targeted therapy strategies against coagulation group genes have the potential to enhance the efficacy of immunotherapy and reduce the risk of thrombosis, thereby opening up new avenues for breast cancer treatment. Based on the findings, it is essential to further explore the prognostic potential of coagulation group genes in breast cancer and their impact on clinical treatment decisions. This will facilitate a more comprehensive understanding and utilization of these genes as potential biomarkers and therapeutic targets.

With the benefit of latest progress in bioinformatics and genomic information, we are able to investigate the connection between tumor coagulome and TME in breast cancer with greater accuracy and thoroughness. By using the COX regression analyses, we identified important prognostic genes from coagulation-related genes (CRGs) in breast cancer, creating a detailed prognostic model. This model includes 6 key genes and integrates clinical pathological features. Through the model, we identified differences in prognosis, immune microenvironment, and drug sensitivity among different risk groups. This provides personalized recommendations for clinical treatment of breast cancer and guidance for identifying beneficiaries of immunotherapy in breast cancer.

2 Materials and methods

2.1 Patients and mRNA sequences data acquisition

Acquired cancer tissues mRNA-seq information in TPM form from The Cancer Genome Atlas (TCGA) and the mRNA data of normal tissues from Genotype-Tissue Expression (GTEx) database. Extracted data corresponding to 1,099 cases of invasive breast cancer from TCGA, along with 113 adjacent normal tissue samples, and 459 normal tissues data from GTEx. Retrieved the microarray data set GSE42568 from the Gene Expression Omnibus (GEO) database. All samples were included in the analysis of gene expression differences and correlations. Cases with complete clinical and pathological data were used for clinical correlation and prognosis analysis.

2.2 Selection and analysis of genes related to coagulation for differential expression

To identify a subset of coagulation-related genes from a vast pool of candidates, we leveraged the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. KEGG is a comprehensive repository housing detailed functional information on genes, genomes, chemical molecules, and cellular processes. By delineating pathways and networks involving genes and their

products, KEGG facilitates researchers in comprehending cellular functions and disease mechanisms. Employing pathway data from KEGG allowed us to pinpoint genes already implicated in the coagulation cascade, thereby streamlining our investigation and directing focus towards genes most pertinent to our research inquiry. The tumor tissues from the TCGA-BRCA cohort were used as experimental group samples, while the control group consisted of 113 adjacent-tumor samples and 459 normal tissue samples from GTEx. The "limma" package was utilized to generate a matrix of gene expression differences, with a threshold set at adjusted p-value <0.05 and |LogFC |>1 to identify differentially expressed mRNAs. Visualizing differential gene expression using the "pheatmap" package.

2.3 Analysis of functional enrichment

Functional enrichment analysis was performed on data to confirm the intrinsic functionality of differentially expressed genes. The Gene Ontology (GO) serves as a prevalent method for gene annotation, encompassing molecular function (MF), biological pathways (BP), and cellular components (CC). After the initial KEGG analysis, we identified coagulation-related genes. Subsequently, we conducted another enrichment analysis using the KEGG database to further explore the functional validation and mechanisms of the differentially expressed coagulation genes. This step represents the validation of differential coagulation gene functions and the exploration of their mechanisms. It aids in uncovering the potential roles of these genes in the occurrence and development of breast cancer, as well as their interactions and impact on the disease progression.

The KEGG database is a useful resource for gaining a deep understanding of genome-wide functionality. By performing Gene Set Enrichment Analysis (GSEA), we identified pathways that were enriched in the differential genes between the high-risk and low-risk groups. To fully comprehend the target mRNA carcinogenesis, we employed the "ClusterProfiler" to examine GO function, enrich KEGG pathway, and conduct GSEA. Visualizations of enrichment analysis results were generated using the ggplot2 package.

2.4 Analysis of protein-protein interactions (PPI) networks and identification of central genes

The CRGs that showed differential expression were analyzed for protein-protein interactions using the STRING database, with a minimum interaction score of 0.7. Central genes were extracted with the degree centrality algorithm in Cytoscape software (version 3.8.2) with the assistance of the "cytoHubba" plugin for network visualization. According to this algorithm, the degree of nodes in the network can be calculated, which refers to the number of edges directly connected to a node. The higher the degree of a node, the greater its importance in the network. Node degree ranking is arranged in descending order according to the degree of nodes, with higher-ranking nodes having greater influence. Subnetworks were identified using the MCODE plugin.

2.5 Development and validation of the prognostic model

The prognostic significance of 59 differentially expressed CRGs was investigated using univariate Cox regression. With the aid of multivariate Cox regression analysis, a prognostic model was developed based on those prognostic related CRGs. The risk score was calculated using the regression coefficients and the values of gene expression with the following formula:

$$Risk\ score = \sum_i Coefficient\ of\ (i) \times Expression\ of\ gene\ (i)$$

Using the patients from the TCGA-BRCA cohort as the training set, the risk score for each patient was calculated using the above formula. Patients were stratified into high-risk and low-risk categories based on the medium cut-off value of risk scores. Kaplan-Meier (K-M) survival analysis was conducted using the Log-rank test. The "pROC" package was employed to execute receiver operating characteristic (ROC) analysis. The model's accuracy on predicting OS was evaluated using AUC values from the ROC curve. The "timeROC" in R software was utilized to generate ROC curves for 1-year, 3-year, and 5-year periods, and Kaplan-Meier survival analysis was performed using the calculated risk scores. To enhance the credibility of the model, we validated it using the dataset GSE42568. After normalizing the expression of differential CRGs, the risk score was calculated. Likewise, individuals were categorized into different risk categories followed by the Kaplan-Meier survival analysis.

2.6 Construction of prognosis nomogram and establishment of calibration curve

Using the "survival" package for proportional hazards analysis. Using clinical factors including age, clinical T stage, clinical N stage, pathological stage, and risk score in both univariate and multivariate Cox independent prognosis analyses and presenting the results with a forest plot generated by "forestplot" package. Utilizing the outcomes of proportional hazards (PH) analysis, create a nomogram with the assistance of the "rms" to estimate the OS at each period. Creating calibration graphs for 1-, 3-, and 5-year endpoints to evaluate the agreement between endpoint occurrences and observed outcomes.

2.7 Immune checkpoint and immune infiltration

Evaluating the relationship between risk score and different immune checkpoints and various types of immune cells using Spearman correlation analysis. Using ssGSEA to assess immune infiltration, applying the GSVA algorithm in R software, and determining the immune infiltration condition of high-risk and low-risk groups using markers of 24 immune cell types (Bindea et al., 2013). Utilizing the ESTIMATE algorithm to calculate immune cell scores for various risk categories, thus deducing the composition of immune cells. Employing the "stats" package along with the "car" package to conduct Wilcoxon rank sum tests and utilizing "ggplot2" for visualizing the results.

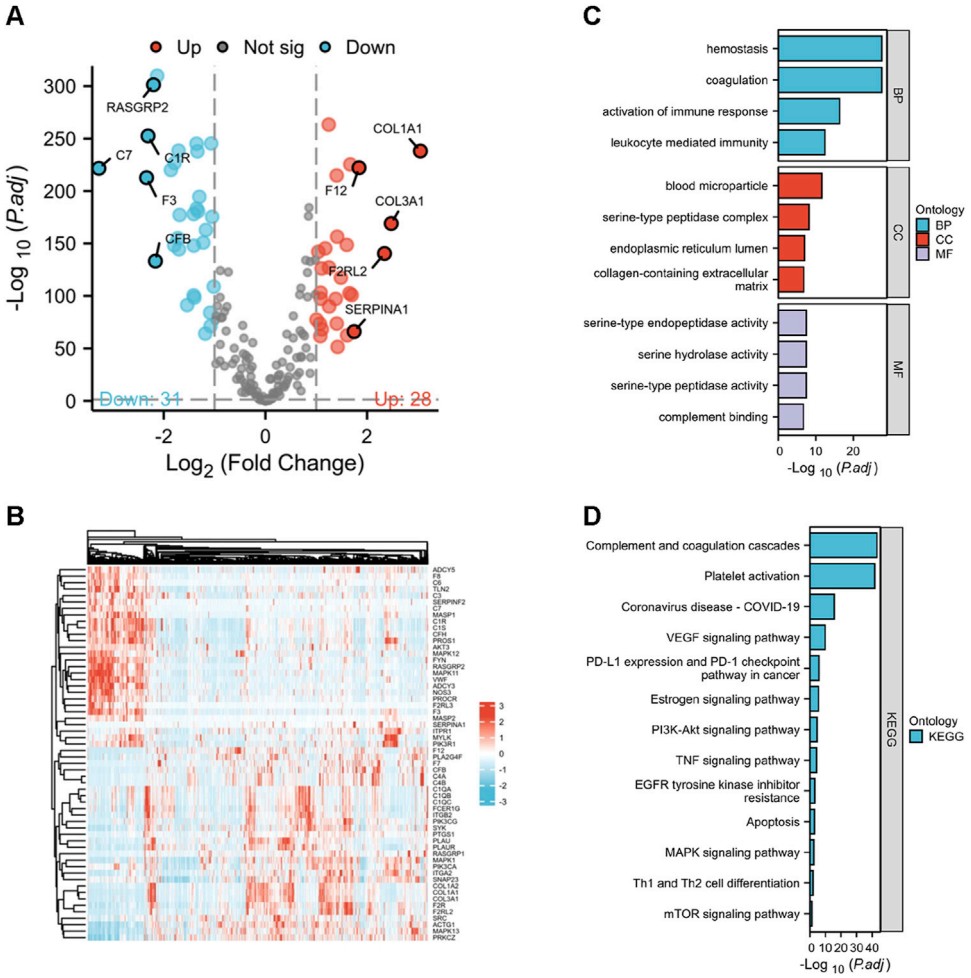


FIGURE 1
The outcomes of differentially analysis of CRGs and the analysis of functional enrichment. **(A,B)** the Volcano plot and heatmap of determined differential CRGs. **(C,D)** the GO and KEGG enrichment analysis outcomes of differential CRGs.

2.8 Drug sensitivity analysis

"OncoPredict" package was employed to forecast drug response in cancer patients (Maeser et al., 2021). Predicting drug responses using the IC50 values of 198 compounds tested on 809 cell lines from the GDSC database version 2 as the training dataset. Through computation, the OncoPredict package can generate sensitivity scores for individual drugs, which exhibit a positive correlation with IC50 values. Pearson correlation analysis was utilized to evaluate the relationship between risk scores and different drug sensitivities. The Mann-Whitney U test was employed to access the variations in drug responses of popular breast cancer treatments among the risk categories.

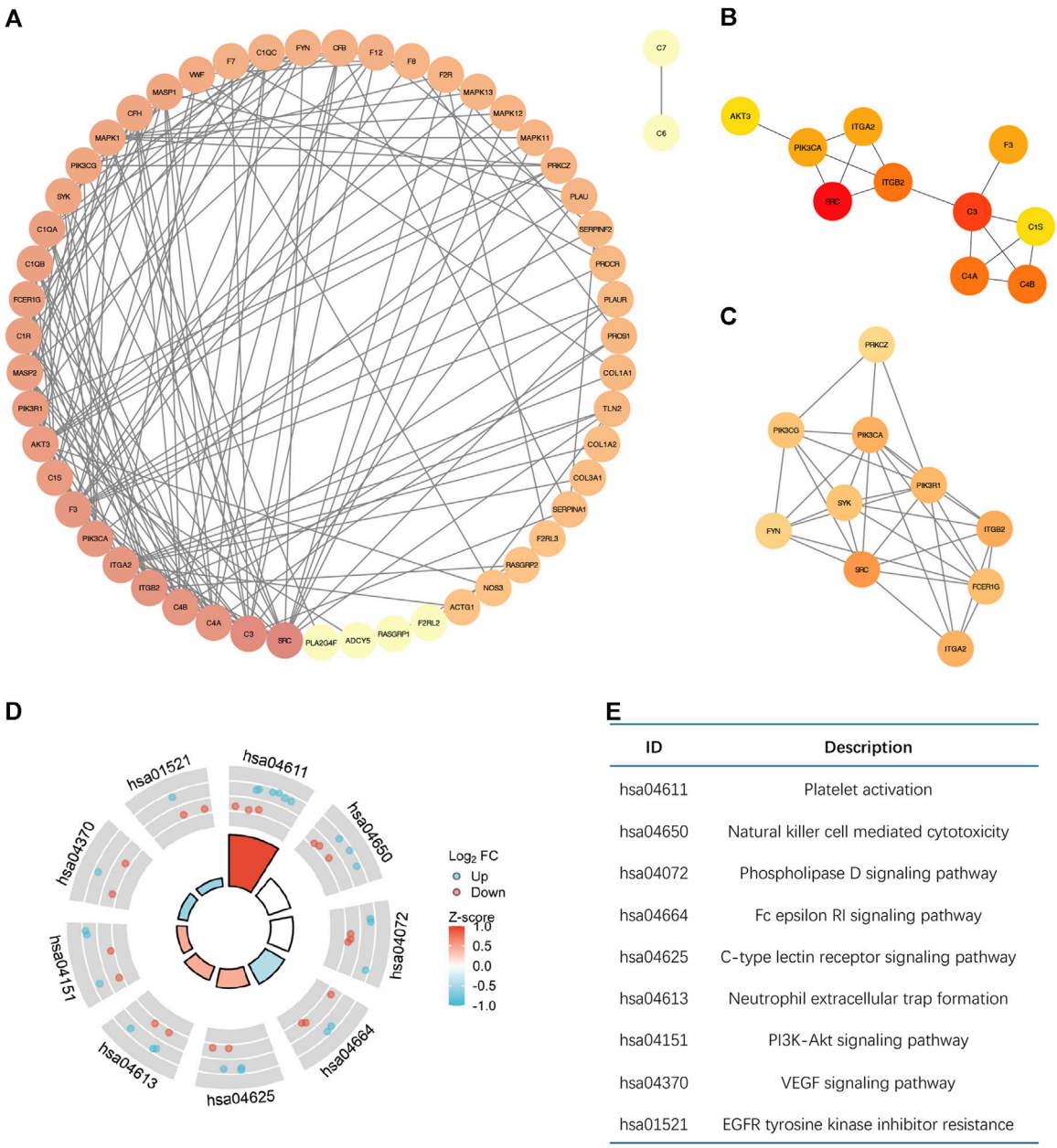
2.9 Statistical analysis

R software (version 4.2.2) was used for data analysis. The Mann-Whitney U test was utilized to assess variations between two sets of continuous variables that were not normally distributed. The Pearson test or Spearman test was used to examine the correlation between continuous variables. Chi-square and Fisher's exact test were employed to compare the differences between categorical variables. Statistics were analyzed using two-sided *p* values with a *p*-value of 0.05 defining statistical significance.

3 Results

3.1 Determination of differentially expressed CRGs and functional enrichment analysis

Totally 203 genes were collected from the KEGG database, derived separately from hsa04610 (complement and coagulation cascades) and hsa04611 (platelet activation). Define these genes as coagulation-related genes (Supplementary Table S1). The mRNA sequences data were gathered from the TCGA-BRCA cohort (including 1,099 cancer samples and 113 adjacent normal tissue samples) and GTEx (459 normal tissues samples). In breast cancer tissues, a total of 59 coagulation-related genes were found to have differential expression, with 31 showing decreased levels and 28 showing increased levels (see Supplementary Table S2 for



ID	Description
hsa04611	Platelet activation
hsa04650	Natural killer cell mediated cytotoxicity
hsa04072	Phospholipase D signaling pathway
hsa04664	Fc epsilon RI signaling pathway
hsa04625	C-type lectin receptor signaling pathway
hsa04613	Neutrophil extracellular trap formation
hsa04151	PI3K-Akt signaling pathway
hsa04370	VEGF signaling pathway
hsa01521	EGFR tyrosine kinase inhibitor resistance

FIGURE 2
The Molecular interactions analysis of differential CRGs based on PPI. **(A)** The PPI network diagram of 59 differential CRGs. The degree of nodes is sorted by color depth. This PPI network contains 58 nodes and 166 edges. **(B)** The top ten differentially expressed genes selected by the degree centrality algorithm are SRC, C3, C4B, C4A, ITGB2, PIK3CA, F3, ITGA2, C1S, and AKT3. **(C)** The subnetwork consists of 10 nodes and 33 edges selected by the MCODE plugin. **(D,E)** Enrichment Analysis of the genes in the subnetwork.

details). Figure 1A displayed the top ten genes with the most significant differential expression based on logFC, while Figure 1B showed the expression heatmap of differentially expressed CRGs.

The analysis of the 59 differentially expressed CRGs using Gene Ontology (GO) showed that aside from hemostasis and blood coagulation, these genes are also involved in biological processes like immune response activation and leukocyte-mediated immunity. CRGs are abundant in blood microparticles, serine-type peptidase complex, endoplasmic reticulum lumen, and extracellular matrix containing collagen when it comes to cellular components. Serine-type endopeptidase/peptidase activity, serine hydrolase activity, and complement binding are all enriched in molecular functions, as shown in Figure 1C. The outcomes of KEGG pathways enrichment conveyed that the distinct genes are enriched not only in Complement and coagulation cascades and Platelet activation, but also in pathways like VEGF signaling, Estrogen signaling, PI3K-Akt signaling, TNF signaling, EGFR tyrosine kinase inhibitor resistance, Apoptosis, and MAPK signaling. These pathways control different physiological processes such as cell growth and cell death. Furthermore, it was noted that these

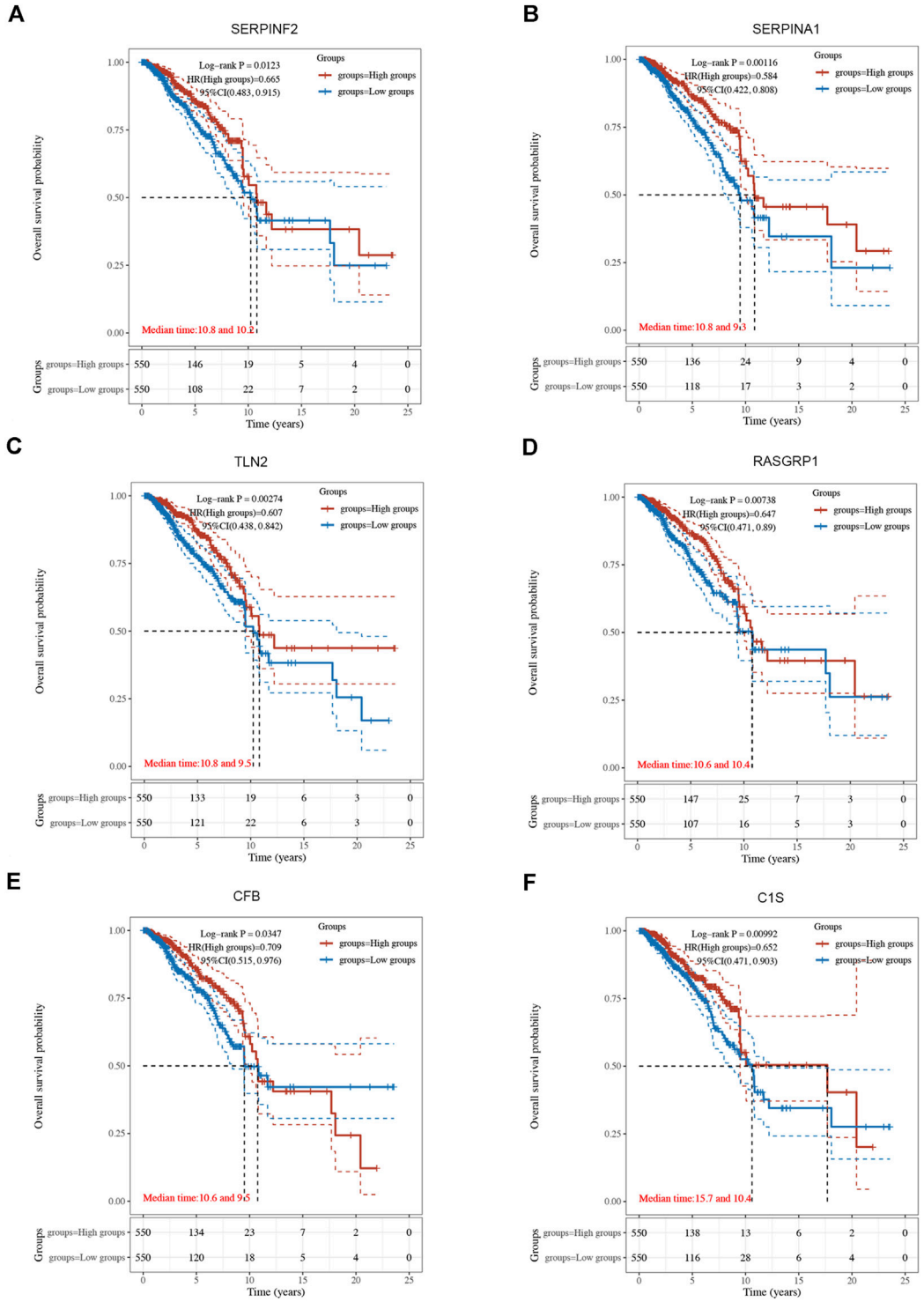


FIGURE 3
The Kaplan-Meier survival curves of 6 integrated genes from TCGA-BRCA cohort. **(A)** Curve of SERPINF2 (p = 0.012, HR = 0.665, 95%CI (0.483, 0.915). **(B)** Curve of SERPINA1(p = 0.001, HR = 0.584, 95%CI (0.422, 0.808). **(C)** Curve of TLN2 (p = 0.003, HR = 0.607, 95%CI (0.438, 0.842). **(D)** Curve of RASGRP1 (p = 0.007, HR = 0.647, 95%CI (0.471, 0.890). **(E)** Curve of CFB (p = 0.034, HR = 0.709, 95%CI (0.515, 0.976). **(F)** Curve of C1s (p = 0.010, HR = 0.652, 95%CI (0.471, 0.903).

distinct genes play a role in the expression of PD-L1/PD-1 checkpoint pathway, as well as in controlling the differentiation of Th1 and Th2 immune lymphocytes (Figure 1D). The results above indicated the involvement of differential coagulation genes in signaling pathways and the regulation of the immune microenvironment.

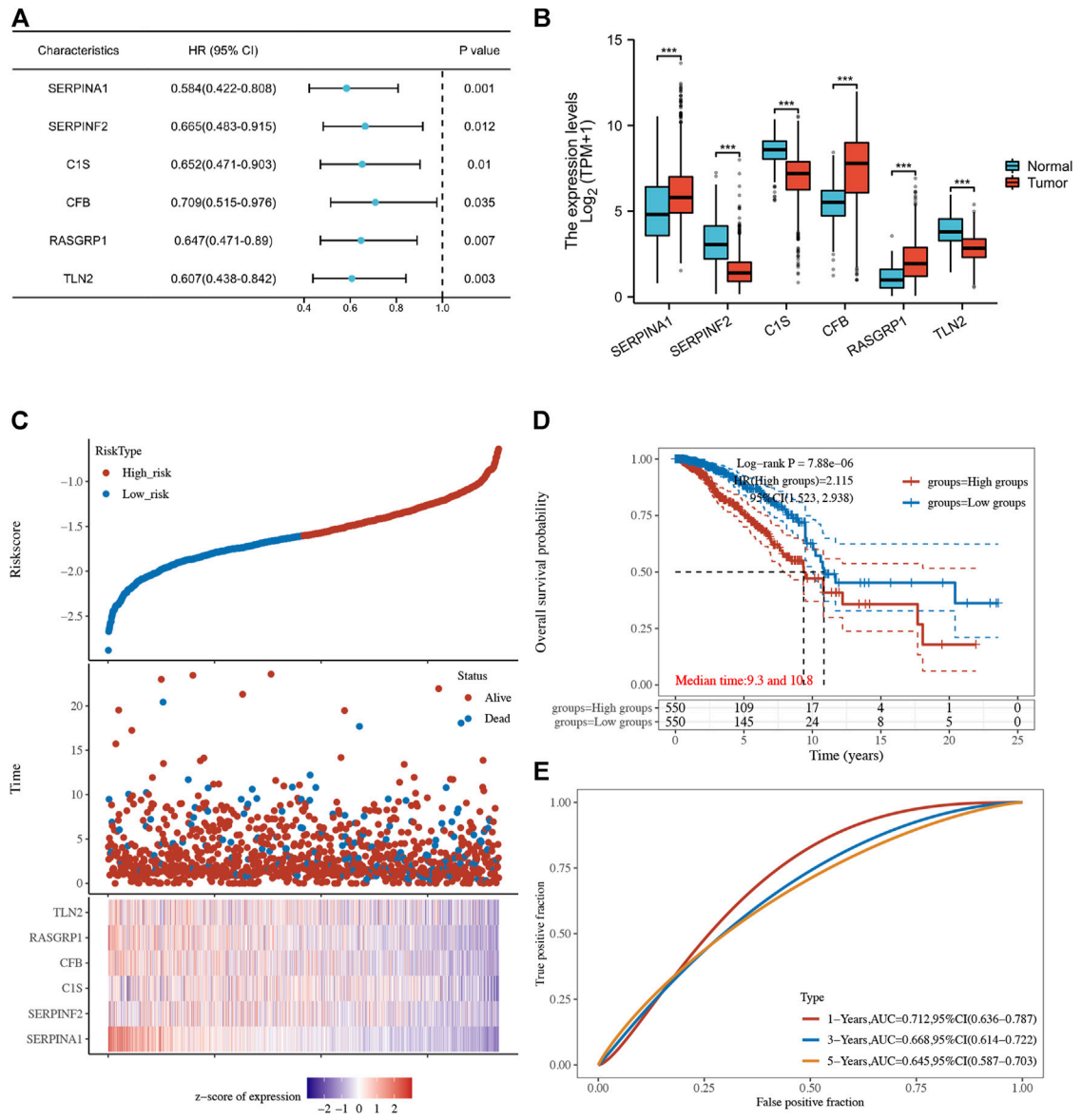


FIGURE 4
(A) Forest plot of 6 integrated genes. **(B)** The expression of 6 genes in TCGA-BRCA cohort and GTEx database. The red box represents the expression of genes in the tumor group, whereas the blue box indicates the expression of genes in normal control. **(C,D)** The risk predictor plot and the Kaplan-Meier survival curve of the model. **(E)** the ROC curve for model's predictive precision. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

3.2 Molecular interactions analysis of differential CRGs

The PPI network that was produced contains 58 nodes and 166 edges, as shown in Figure 2A. Through node analysis, we identified SRC, C3, C4A, C4B, ITGB2, ITGA2, PIK3CA, F3, C1S, and AKT3 as the top 10 genes interacting most with other differentially expressed genes. By utilizing Cytoscape, we implemented the degree centrality algorithm to identify hub genes, choosing the top 10 genes based on their node degree ranking. These genes are SRC, C3, C4B, C4A, ITGB2, PIK3CA, F3, ITGA2, C1S, and AKT3, which are consistent with the top genes identified in the node analysis (Figure 2B). The PPI network revealed intricate interactions among the differentially expressed CRGs in breast cancer. Next, we identified a subnetwork using the MCODE plugin, which consists of 10 nodes and 33 edges (Figure 2C). Analysis of the genes in the subnetwork showed enrichment in various signaling pathways that control cell growth and apoptosis, as depicted in Figures 2D, E. This suggests that hub genes associated with coagulation may involve in controlling the onset and progression of breast cancer.

3.3 Establishment of breast cancer prognostic model based on CRGs

Using univariate Cox regression analysis, 6 genes were identified as prognostically significant out of a pool of 59 differential genes.

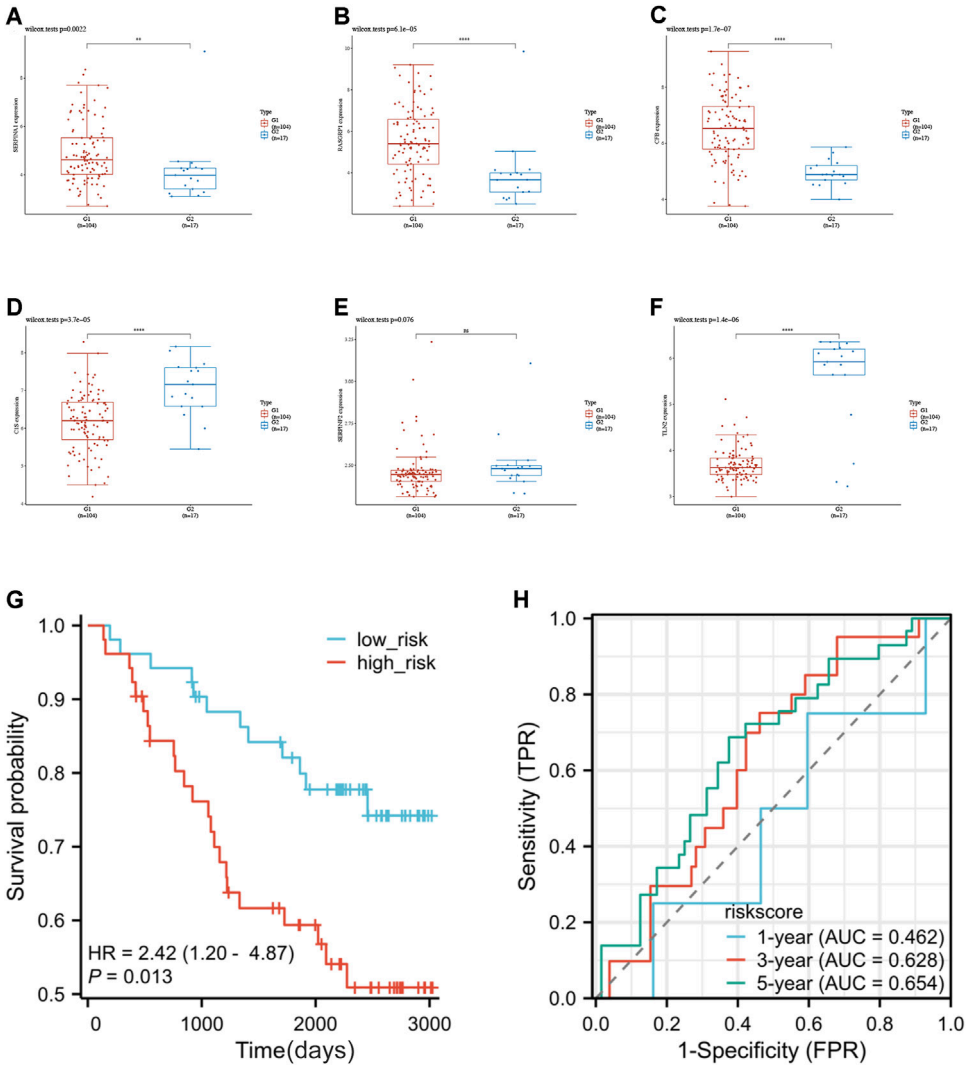


FIGURE 5
(A–F) The Wilcoxon rank sum tests results to illustrate the expression of 6 genes in the test set. The red box represents for the tumor group, whereas the blue box indicates for normal control. [**(A)** SERPINA1. **(B)** RASGRP1. **(C)** CFB. **(D)** C1S. **(E)** SERPINF2. **(F)** TLN2]. **(G)** The K-M curves of the model in test set. **(H)** the ROC curve for model's predictive precision in test set. (*p < 0.05, **p < 0.01, ***p < 0.001, ns: no significance).

These genes include SERPINA1, SERPINF2, C1S, CFB, RASGRP1, and TLN2. Prognostic analyses were conducted on each of these 6 genes to observe their efficacy as prognostic biomarkers in breast cancer. Based on the median gene expression, patients from the TCGA-BRCA cohort were categorized into groups with high or low gene expression, and then Kaplan-Meier curves were generated. Kaplan-Meier curves indicate that increased levels of the six genes are linked to improved prognosis (HR < 1, Log-rank p < 0.05) (Figures 3, 4A). Indicated that all six genes are protective factors for breast cancer. The AUC values for the 1-, 3-, and 5-year overall survival (OS) of breast cancer patients are all greater than 0.5 for the six specified genes. Among them, SERPINA1, RASGRP1, and CFB demonstrated the highest prognostic efficacy. Additional analysis of the 6 genes in breast cancer showed notable variations in expression levels between tumor and healthy tissues. Specifically, tumor tissues show increased expression of SERPINA1, RASGRP1, and CFB compared to normal tissues, whereas C1S, SERPINF2, and TLN2 display the opposite pattern (Figure 4B). The prognostic model was established with the regression coefficient of the 6 genes. The model yielded an AIC (Akaike information criterion) of 1733.7347. The final formula settled as follows: Risk score = (−0.1377) * SERPINA1 + (−0.0524) * SERPINF2 + (−0.0014) *C1S + (−0.0565) * CFB + (−0.108) * RASGRP1 + (−0.0724) * TLN2.

Following this, patients from the TCGA-BRCA cohort were stratified into high- and low-risk categories according to the median cut-off value of the risk score. Those classified in the high-risk category showed decreased overall survival rates in comparison to those in the low-risk category, as demonstrated by the Kaplan-Meier curve (p = 7.88*10-6, HR = 2.115, 95% CI 1.523–2.938) (Figures 4C, D). ROC analysis over time showed that the model's predictive precision was 0.712 (0.636–0.787) for 1-year OS, 0.668 (0.614–0.722) for 3-year OS, and 0.645 (0.587–0.703) for five-year OS (Figure 4E). Overall, our model demonstrated stable prognostic prediction accuracy in the training dataset.

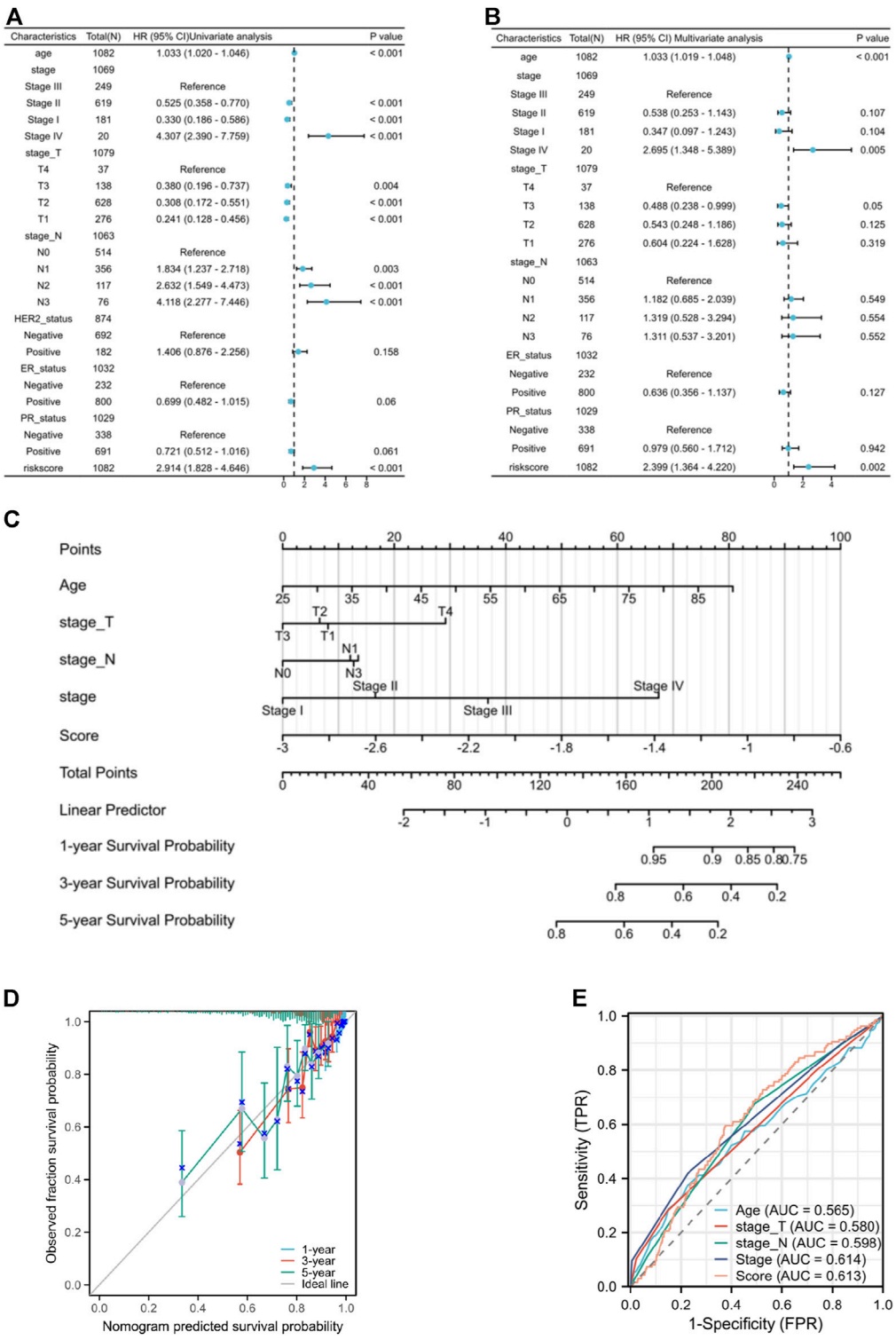


FIGURE 6
(A) the forest plot of the univariate independent prognostic analysis. **(B)** the forest plot of the multivariate independent prognostic analysis. **(C)** the Nomogram that integrating age, clinical T stage, clinical N stage, and risk score. **(D)** the calibration curves of the Nomogram. **(E)** ROC curves of Age, T stage, N stage, pathological stage, and risk score.

3.4 Validating the model in the test set

To validate the model's generalizability, we obtained data from 104 breast cancer patients in the GSE42568 as the test set. Risk scores were computed utilizing an identical formula, leading to the classification of patients into high-risk (52 cases) and low-risk (52 cases) groups. Initially, 6 genes' expression in the test group were examined and we observed that except for SERPINF2, other

five genes exhibited similar expression patterns as the training set. As shown in Figures 5A–F, it demonstrated a notably elevated expression of SERPINA1, RASGRP1, and CFB in tumor tissues, while C1S and TLN2 showed significantly higher expression in normal tissues. There were no notable variations in the SERPINF2 expression, possibly due to the limited sample size. The KM curves indicated extended survival in the low-risk individuals ($p = 0.013$) (Figure 5G). ROC analysis over time showed that the predictive precision was 0.46 (0.184–0.7403) for 1-year overall survival, 0.628 (0.5061–0.7505) for 3-year overall survival, and 0.654 (0.5367–0.7722) for five-year overall survival (Figure 5H). The findings indicate that our model has a degree of precision in forecasting the outcome of breast cancer survival rates at 3 and 5 years.

3.5 Coagulation-related model serve as an independent prognostic factor in breast cancer

The univariate independent prognostic analysis indicates that features associated with overall survival (OS) include age at diagnose, pathological stage after surgery, clinical T stage, clinical N stage and risk score (Figure 6A). When it comes to the multivariate independent prognostic analysis, it is evident that the risk score along with age, and pathological stage can all serve as independent prognostic indicators for breast cancer patients (Figure 6B). Nomogram has been developed by integrating age, clinical T stage, clinical N stage, and risk score to provide clinicians with a quantitative method of predicting BRCA patients' prognosis (Figure 6C) and the calibration curves show good survival prediction capability (Figure 6D). Furthermore, compared to traditional prognostic scoring systems, our risk model exhibits a higher AUC value (AUC = 0.613) (Figure 6E) which demonstrated that the risk score makes a substantial contribution to prognosis prediction.

3.6 Clinicopathological features and risk score

No variations in age and N stage were observed between high and low-risk individuals, as shown in Supplementary Table S3. Nevertheless, there are notable variances in T stage ($p = 0.004$), M stage ($p = 0.020$), pathological stage ($p = 0.016$), HER2 status ($p < 0.001$), ER status ($p < 0.001$), and PR status ($p < 0.001$) between the risk groups. The low-risk group have a higher proportion of individuals at T1 stage than the high-risk group. Conversely, the high-risk group has more patients with T2, T3, and T4 stage. Furthermore, the high-risk patients appear to be more prone to distant metastasis and individuals in this group typically exhibit elevated pathological stages. Low-risk patients typically show hormone receptor (HR)-positive status and low HER2 expression, whereas high-risk patients exhibit the opposite (HR-/HER2 high expression). The findings of the prognostic efficacy suggest that individuals classified as high-risk have worse overall survival rates in different subcategories (Figure 7). Nevertheless, the variation in operating systems within subcategories with M1 metastasis did not show a notable discrepancy, potentially as a result of the small sample size of M1 individuals, hindering statistical distinction.

3.7 Enrichment analysis of genes that differ between risk groups showed unique pathway enrichments

The GSEA of differentially expressed genes in risk groups identified unique pathway enrichments shown in Figures 8A, B. Significant enrichments in pathways such as ASCORBATE_AND_ALDARATE_METABOLISM, PENTOSE_AND_GLUCURONATE_INTERCONVERSIONS, OXIDATIVE_PHOSPHORYLATION, DNA_REPLICATION, MISMATCH_REPAIR, and CELL_CYCLE were observed within the low-risk group. In contrast, the group at high risk showed enhancements in pathways linked to T cell receptor signaling, Toll-like receptor signaling, JAK-STAT and MAPK signaling pathway, B cell receptor signaling, and apoptosis. These pathways are related to tumorigenesis, inflammation, immune responses, and cellular apoptosis. An association between the risk score and the tumor environment has been indicated.

3.8 Disparities in the TME compositions and immune infiltration across risk categories

It is a negative correlation between risk score and six types of immune cell. The expression levels of immune cells decrease along as the risk score increases (Figure 8F). The strongest correlation was found with Neutrophils (Spearman coefficient −0.32), with $CD4^+$ T cells, myeloid DCs, and $CD8^+$ T cells following closely behind. Following this, we performed ssGSEA on individual samples and identified notable variances in immune infiltration levels among high and low-risk categories. Immune cells, aside from B cells, Tgd cells, and Treg cells, did not display variations between risk groups, while the remaining showed significant statistical variances between the two risk categories. In the high-risk group, there was a notable increase in Th2 cell infiltration compared to the group of low risk, which had higher infiltration of other immune cells (Figure 8C). Next, the outcomes of ESTIMATE algorithm containing Immune score, Stromal score, and ESTIMATE score were considered for inferring the composition of immune cells. Since increased scores suggest greater presence of infiltrating elements within the tumor microenvironment, this study found that the individuals of low-risk group exhibited higher infiltration of both stromal ($p = 8.88*10\text{-}10$) and immune ($p = 1.49*10\text{-}11$) components compared to the high-risk group (Figure 8D).

3.9 Differences in immune response and drug sensitivity among risk groups

Examining the relationship between risk score and the activity of Immune checkpoint genes to forecast response to immunotherapy in various risk categories. The results indicate that not only the classic immune checkpoints (PDCD1, CTLA4, and PDCD1LG2) but also the emerging target genes (SIGLEC15, TIGIT, CD274, HAVCR2) are highly expressed in patients from the low-risk group (Figure 8E).

Correlations between the IC50 values of 4 drugs and the risk score were identified through with a threshold of $|cor| < 0.3$ and $p < 0.05$. These drugs are Gefitinib_1010, GSK2606414_1618, Ribociclib_1632, and Pyridostatin. A negative correlation was

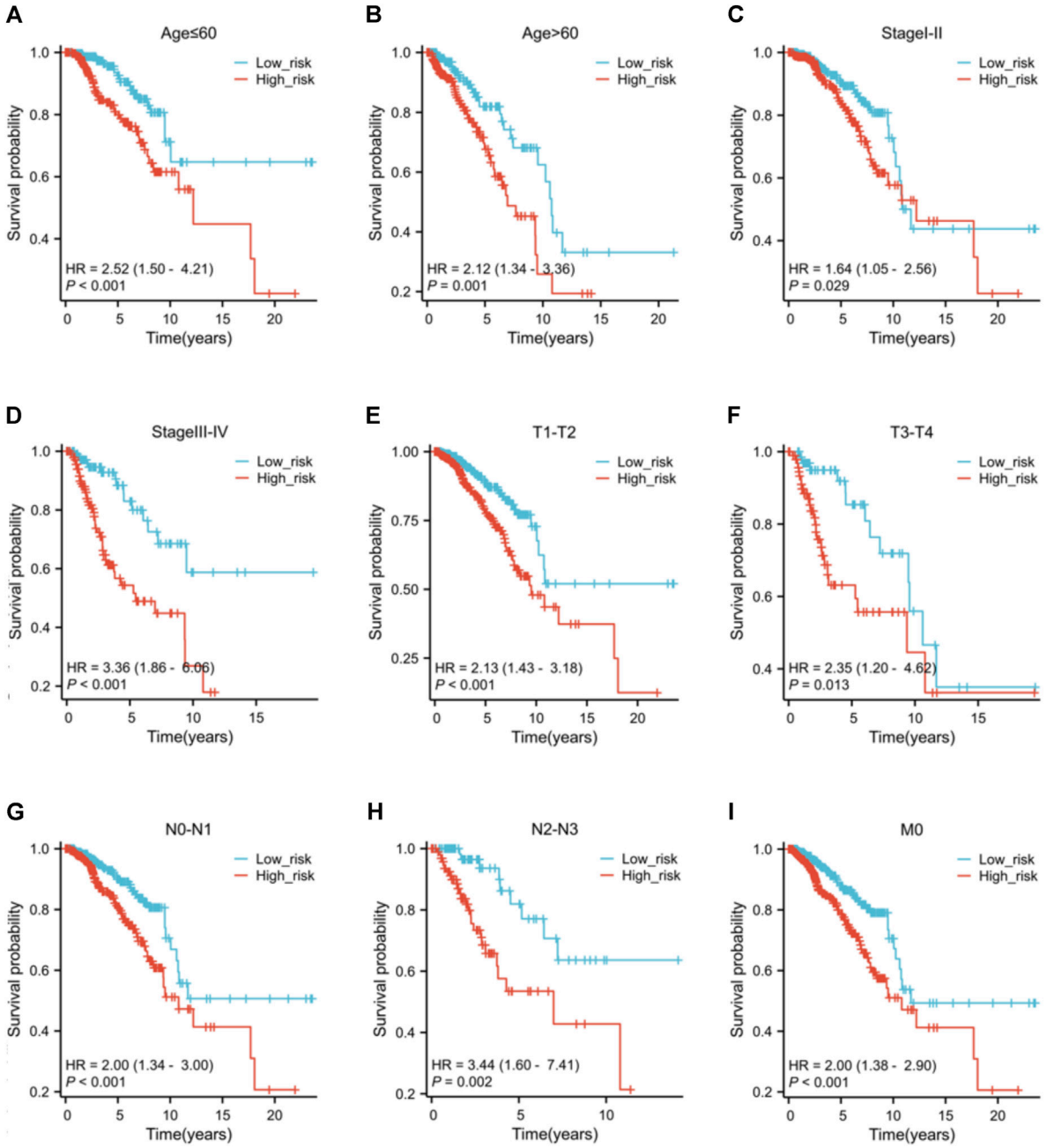


FIGURE 7
The K-M curves of the risk score across different subgroups. Subgroups of patients with **(A)** age ≤60 ; **(B)** age >60; **(C)** clinical pathological stage I-II; **(D)** clinical pathological stage III-IV; **(E)** tumor stage T1-T2; **(F)** tumor stage T3-T4; **(G)** nodal stage N0-N1; **(H)** nodal stage N2-N3; **(I)** metastasis stage M0.

observed between Gefitinib and the risk score, suggesting that higher risk scores are linked to lower drug sensitivity scores for Gefitinib. Significant variations in drug sensitivity were observed among the groups for the four drugs analyzed (Figure 9A). The sensitivity score calculation shows a positive correlation with the IC50 value of the drugs, indicating that patients classified as low-risk demonstrate increased resistance to Gefitinib and heightened sensitivity to the remaining three drugs.

Next, we explored the differences in sensitivity to several commonly used or hot-spot drugs in breast cancer patients. The low-risk individuals exhibited greater resilience to Cisplatin and Cyclophosphamide than those in the high-risk category when it came to chemotherapy medications. Conversely, the low-risk individuals exhibited greater sensitivity to Oxaliplatin and Docetaxel. No notable distinction was observed in the sensitivity to Paclitaxel, Gemcitabine, and Epirubicin between the high and low-risk groups (Figure 9B). In the realm of endocrine treatments, Tamoxifen was found to be more effective for high-risk patients, while Fulvestrant illustrated better results for low-risk group patients. Furthermore, individuals classified as low-risk exhibited increased responsiveness

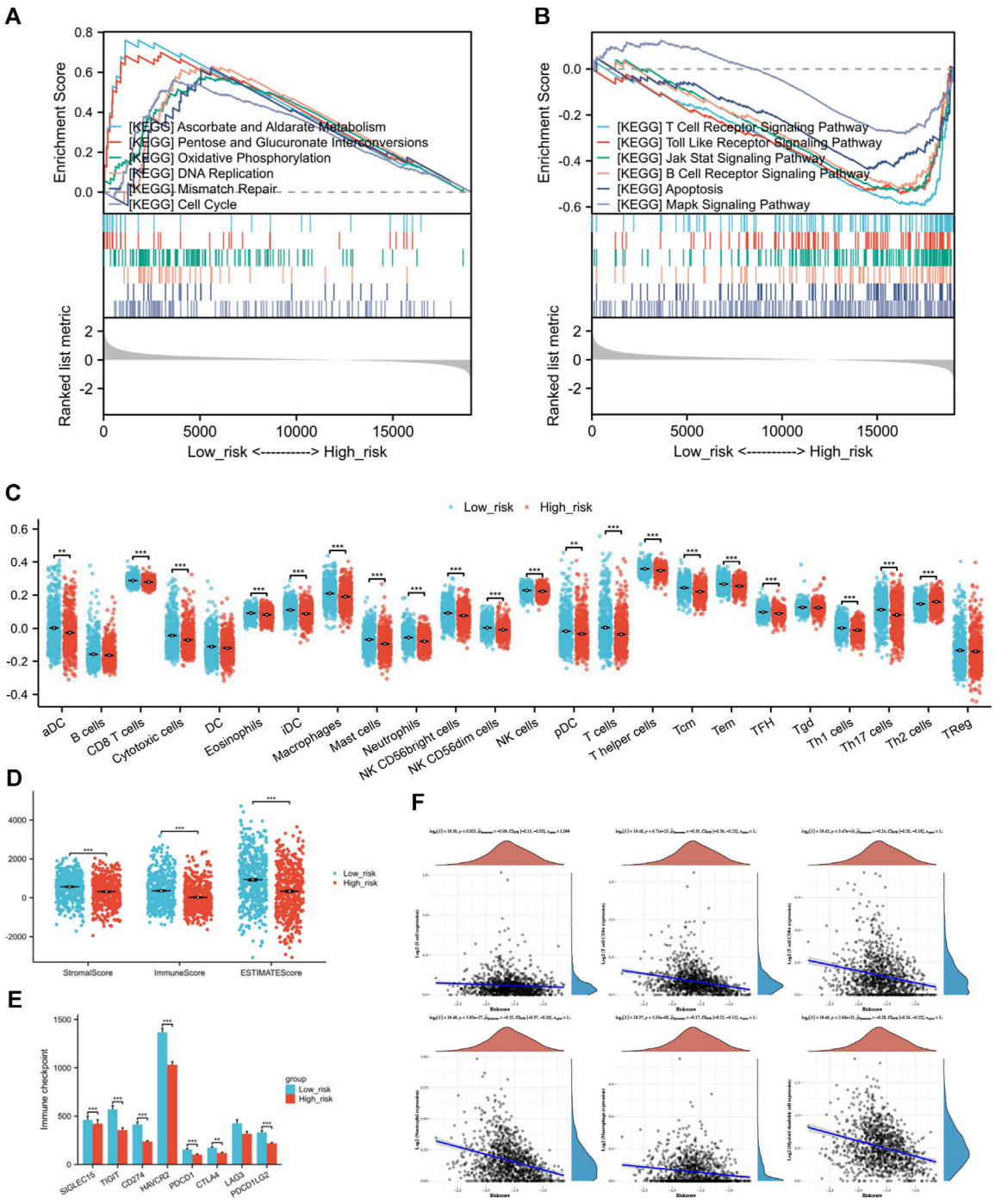


FIGURE 8
(A) GSEA of differentially expressed genes in low-risk individuals. **(B)** GSEA of differentially expressed genes in high-risk individuals. **(C)** The infiltration condition of 24 types of immune cells based on the specific immune markers among the risk group. **(D)** Immune score, Stromal score, and ESTIMATE score of ESTIMATE algorithm. **(E)** Comparison of immune checkpoint genes expression between high and low-risk groups **(F)** The correlation analysis between risk score and the counts of immune cells calculated by TIMER. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

to CDK4/6 inhibitors Palbociclib and Ribociclib in comparison to those in the high-risk category (Figure 9C). Individuals classified as low-risk showed notably higher resilience to the EGFR inhibitor Gefitinib in comparison to those categorized as high-risk. In Figure 9D, patients classified as low risk showed increased sensitivity to the PARP inhibitor Olaparib. Zoledronate is frequently prescribed for breast cancer patients who have bone metastases. Individuals classified as high-risk showed a notably higher response to Zoledronate in comparison to those categorized as low risk. Finally, we also investigated Mitoxantrone, a chemotherapy drug currently used as a tracer in sentinel lymph node biopsy during the breast cancer surgery. A notable variation in drug sensitivity to Mitoxantrone was also noted between the risk groups, with the low-risk group showing increased sensitivity and the high-risk group displaying greater resistance (Figure 9D).

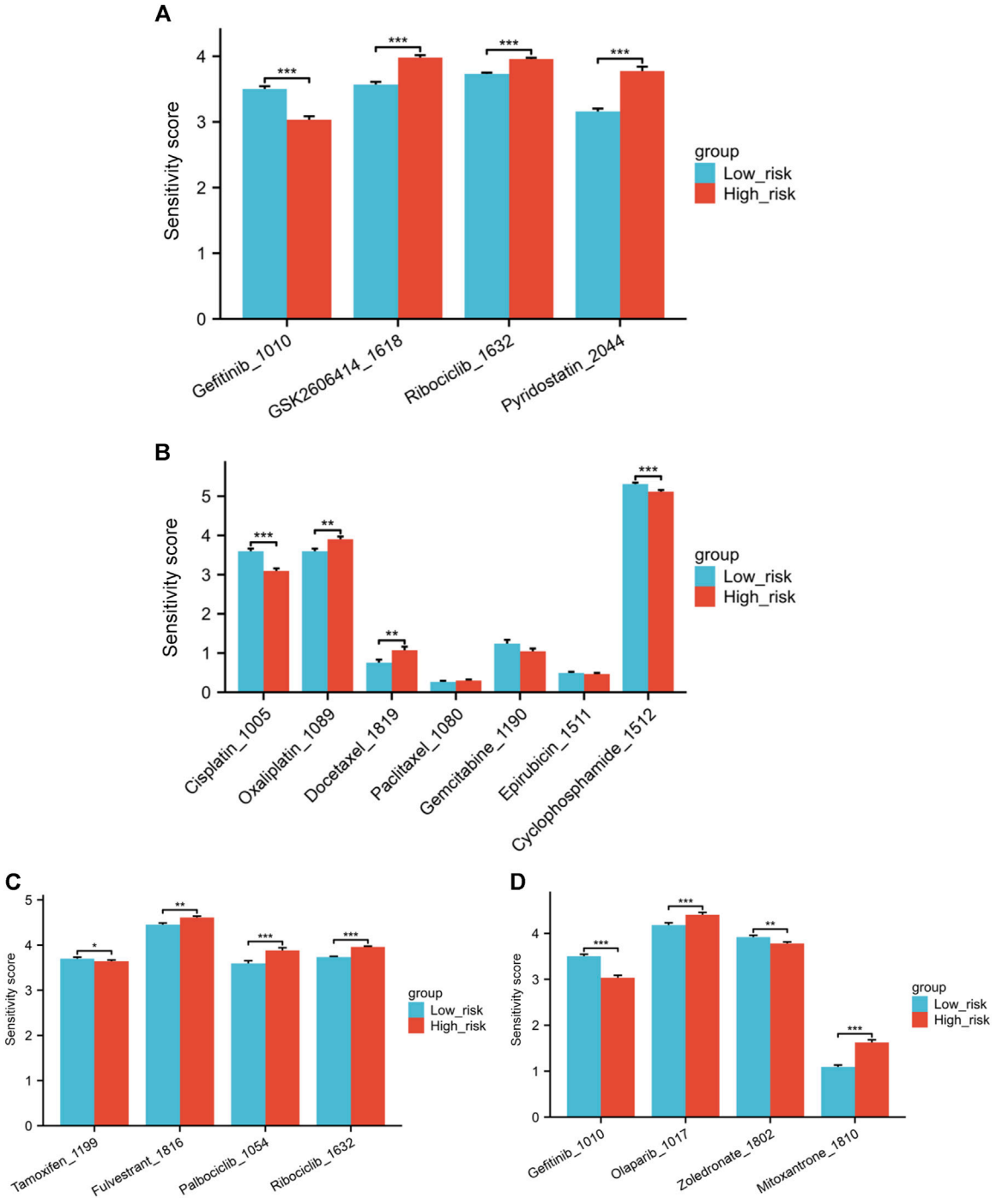


FIGURE 9
(A) Variations in drug sensitivity score of the top four drugs with the highest correlation between IC50 values and risk score. **(B)** Differences in sensitivity score to several chemotherapy medications. **(C)** Differences in sensitivity score to endocrine treatments and CDK4/6 inhibitors. **(D)** Differences in sensitivity score to Gefitinib, Olaparib, Zoledronate and Mitoxantrone (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

4 Discussion

The intimate relationship between malignant tumors and the coagulation process is widely observed. Patients with malignant tumors often experience a hypercoagulable state, which is typically attributed to an imbalance in the coagulation process caused by increased procoagulant activity and decreased anticoagulant factors (Falanga et al., 2015). The abnormal expression of genes related to coagulation is a key factor in this phenomenon. Several studies have demonstrated that coagulation-related genes play distinct roles in various types of cancers. In hepatocellular carcinoma, X. Ai et al. discovered that PIK3R1 is expressed abnormally, leading to heightened proliferation and invasion of tumor cells, as well as the suppression of apoptosis (Ai et al., 2018). L. Ma and others found

that high expression of ITGA2 can promote ovarian cancer cell proliferation and resistance to paclitaxel through the AKT/FOXO1 signaling axis (Ma et al., 2020). Additionally, studies have shown that the majority of thrombogenic tumors themselves express high levels of F3 mRNA, leading to tumor-associated thrombosis and increased mortality (Timp et al., 2013).

Nevertheless, there is insufficient research on the impact of coagulation-related genes on breast cancer cells. The goal of this study is to develop a predictive model focusing on genes attached to coagulation activity in patients with breast cancer, examining how they affect the advancement of tumors, evaluation of prognosis, changes in the immune environment, and responsiveness to treatments.

In this study, to encompass as many comprehensive coagulation-related genes as possible, we retrieved gene sets from the KEGG pathway database using keywords such as platelet, coagulation, and fibrinolysis. Specifically, we selected gene sets from hsa04610 (complement and coagulation cascades) and hsa04610 (platelet activation) for further investigation. Data from the TCGA-BRCA cohort and the GTEx databases revealed 59 differential genes related to coagulation, which were found to be enriched in various classical signaling pathways including the VEGF pathway, PI3K-Akt signaling pathway, and TNF pathway. Multiple studies have verified the participation of specific coagulation genes in controlling signaling pathways: TF can enhance VEGF levels and drive cancer advancement through PAR2 activation, while changes in PIK3R1 expression can lead to oncogenic transformations in different cancers via PI3K-Akt pathway activation (Hisada and Mackman, 2021; Chakraborty et al., 2022). The extensive and complex interplay between coagulation genes and signaling pathway regulation is evident (Taniguchi et al., 2010; Cizkova et al., 2013). Additional examination showed that distinct coagulation genes were involved in controlling the expression of PD-L1/PD-1 checkpoint pathway, along with the regulation of immune lymphocyte differentiation like Th1 and Th2, highlighting the complex connection between coagulation genes and the immune microenvironment of tumors. This study conducted GO and KEGG enrichment analyses on differentially expressed coagulation-related genes, revealing their multifaceted roles in breast cancer. It is noteworthy that our research involved two rounds of KEGG analysis, each with distinct objectives and emphases despite utilizing the same database resource. In the first stage, KEGG analysis was employed to filter a set of genes associated with coagulation, laying the foundation for our study and enabling us to focus on genes closely related to the research topic. The second stage aimed to functionally validate and explore the mechanisms of these filtered genes, unveiling their involvement not only in the coagulation process but also in regulating various signaling pathways and the immune microenvironment in breast cancer. This multi-tiered analytical approach provides us with a more comprehensive and in-depth understanding of the roles of coagulation genes in the occurrence and development of breast cancer, thereby aiding in the construction of more accurate prognostic models and the formulation of effective treatment strategies.

A predictive model was developed through the analysis of six genes (SERPINA1, SERPINF2, C1S, CFB, RASGRP1, and TLN2) using multivariate cox regression. Surprisingly, all the involving genes were prognostic protective factors.

SERPINA1, a key component of the serine protease inhibitor group, produces an anti-trypsin that is vital for regulating cell balance through the irreversible inhibition of different serine endopeptidases. Its prognostic value was observed across multiple cancer types, yet its role varies among different malignancies. A recent study by Kuai X. et al. found that increased SERPINA1 expression is linked to improved overall survival rates in BRCA, COAD, SARC, and SKCM, but worse survival rates in GBMLGG, HNSC, LGG, LIHC, and LUSC (Kuai et al., 2023). This research further confirmed its important function as a standalone predictor in forecasting the outcome of breast cancer. Moreover, in the context of breast cancer, current research has illuminated an association between reduced expression of SERPINA1 and more aggressive tumor phenotypes, poorer prognosis, and tumor metastasis. This hints at a potential tumor-suppressive function of SERPINA1 in breast cancer, whereby its diminished expression may facilitate tumor progression and metastasis (Chan et al., 2015; Zhao et al., 2018). Belonging to the same family as SERPINA1, SERPINF2 specifically enriches in hepatocytes (Hou, 2022; Desoteux et al., 2023). Previous studies have indicated a correlation between SERPINF2 with unfavorable outcomes in advanced serous ovarian cancer, and laboratory tests have confirmed its crucial involvement in tumor growth and spread (Huh et al., 2022). Although the prognostic efficacy is relatively low, our study found certain value of it in breast cancer. C1S participates in the formation of the C1 complex, which serves as the initiator of the classical complement activation pathway (Kim et al., 2013; Riihilä et al., 2020). C1S triggers complement activation and independently modulates tumor cell phenotype and tumor microenvironment, thereby promoting tumor progression (Daugan et al., 2021). Activation of the C1 receptor on monocytes can trigger a series of events that result in cell migration into tissues and transformation into macrophages or dendritic cells, ultimately promoting adaptive immunity and creating a tumor-promoting microenvironment. The high tumorigenicity of C1s has been noted in both renal cell carcinoma and cutaneous squamous cell carcinoma, highlighting its promise as a target for cancer therapy (Riihilä et al., 2020; Liu et al., 2021). Given that FDA has approved C1S antibodies for treatment in patients with cold agglutinin disease (a rare autoimmune disease), further clinical research is needed to confirm its efficacy in cancer therapy (Röth et al., 2021).

The complement factor B encoded by the CFB is a component of the alternative pathway of complement activation (Akhlaghpour et al., 2023). Its subunits cleaved by factor D associated with the proliferation and differentiation of pre-activated B lymphocytes, rapid expansion of peripheral blood monocytes, stimulation of lymphocyte follicle formation, and erythrocyte lysis. Mutations in the CFB gene cause reduced activation of B cells, resulting in changes in the tumor immune environment, potentially playing a role in its correlation with unfavorable outcomes in breast and lung cancer (He et al., 2021). RASGRP1 is also an important gene in this model. Belonging to the Ras superfamily guanine nucleotide exchange factor (GEF) gene family, this gene possesses the capacity to stimulate the Erk/MAP kinase cascade, regulate the proliferation, homeostasis, and differentiation of T cells and B cells, and exhibits significant potential as a therapeutic target for cancer. Cong Wang has unveiled, through experimentation and observation, the dual role of RASGRP1 in regulating acute inflammatory responses and inhibiting inflammation-related cancers and its promising prognostic value (Wang et al., 2022). The final gene in the

model, TLN2, encodes a protein associated with talin 1, which is a cell-skeletal protein playing a crucial role in the assembly of actin filaments and the dispersion and migration of various cells (including fibroblasts and osteoclasts). Prior research indicated that TLN2 suppressed kidney cancer by inhibiting the Wnt/β-catenin signaling pathway (Cai et al., 2022), Yet Fang's study contradicted this finding. Suggested that the presence of TLN2 was linked to the cancer-causing potential of liver cancer cells (Fang et al., 2016). Overall, the predictive significance of the six coagulation genes in breast cancer is significant, and additional experimental validation is needed to clarify their impact on the onset and progression of breast cancer. Notably, their direct role in immune microenvironment regulation holds the potential to become new targets for immunotherapy in the future.

In this study, we constructed an innovative breast cancer prognosis model based on six prognostic-related genes. The model demonstrated considerable accuracy in predicting overall survival (OS) of breast cancer patients. Through ROC analysis, we found that the model achieved AUC values of 0.712, 0.668, and 0.645 for 1-year, 3-year, and 5-year survival rates, respectively, in the training set. These results reveal the model's robust performance, particularly in short-term (especially 1-year and 3-year) predictions. However, we also acknowledge that the AUC values of the model did not reach our expected optimal level, which may indicate some limitations in predictive performance. These limitations could stem from the imbalance in sample distribution or the failure to incorporate all biologically relevant biomarkers closely associated with breast cancer prognosis during feature selection. To address these issues, we are actively exploring various strategies to optimize the model performance, including the application of more advanced feature selection techniques, experimentation with different machine learning algorithms, and fine-tuning of model parameters. Additionally, we plan to conduct more in-depth data mining to identify and integrate potential biomarkers that may have been overlooked by the current model. We also plan to validate the model in a broader and more diverse patient population to assess its applicability and limitations across different populations, which is crucial for the clinical translation of the model. While the AUC value is a key metric for evaluating the predictive ability of the model, we believe that the clinical utility of the model extends far beyond this. Our model not only provides risk stratification for breast cancer patients but also reveals differences in immune microenvironment and drug sensitivity among patients in different risk groups. In the following discussion, we will delve into these differences and their potential implications for personalized treatment strategies.

The treatment for breast cancer involves multidisciplinary collaboration, and the combination with immunotherapy promises to be a new therapeutic strategy. Several clinical studies are currently underway, however the number of patients benefiting from immunotherapy in breast cancer remains limited (Ye et al., 2023). In order to identify potential beneficiaries, we investigated the differences in the immune microenvironment between the risk groups. The findings showed that the primary distinction between the two groups is in how Th1 and Th2 cells are distributed. The group at high risk displayed increased Th2 cell infiltration, while low-risk group had more Th1 cell infiltration. Th1 and Th2 cell both have the ability to release cytokines that support their own growth while suppressing the growth of the other subset (Jia et al., 2021). Additionally, they participate in regulating the activation of helper B cells and contributing to humoral immunity. Typically, there is a harmonious equilibrium between Th1 and Th2 cells. However, when there are abnormalities in bodily functions, this balance may be disrupted, leading to a phenomenon known as "Th1/Th2 shift." Many cancer patients exhibit Th1/Th2 shift in the body, typically leaning towards Th2 dominance (Ruterbusch et al., 2020). The reason for this imbalance could be the capacity of Th2 cells to facilitate immune avoidance in tumors (Frafjord et al., 2021). Changes in TH1/TH2 cell cytokines were discovered to be linked to various molecular subtypes in breast cancer research (Hong et al., 2013). In TNBC, the secretion of cytokines from Th2 cells like IL-4, IL-5, and IL-10 increases, causing a change in the balance between TH1 and TH2 towards a higher ratio of TH2/TH1 cytokines. On the other hand, ER+ and other Luminal type breast cancer exhibit lower levels of Th2 cell cytokines and generally shift towards Th1 immune response. Regarding disease prognosis, the ratio of TH1/TH2 is associated with improved prognosis in ER+/PR + breast cancer, but worse OS in Basel like breast cancer. Thus, creating a shift towards anti-tumor TH1 responses may be a new treatment strategy aimed at improving the prognosis of tumor patients with high Th2 infiltration. It is essential to control the infiltration of Th2 cells to preserve the immune responses targeted against tumors. The research indicated that elevated Th2 cell infiltration in the high-risk group may contribute to immune evasion, while the heightened presence of Th1 and other immune cells with substantial infiltration in the low-risk group implies a heightened level of immune activation, potentially leading to a more favorable prognosis. The research also discovered that low-risk patients exhibited higher expression of various immune checkpoint genes, suggesting that those with low-risk profiles may respond more favorably to immune therapy.

To better facilitate clinical translation, we explored the potential of the model in assisting drug administration decisions and discovered the therapeutic benefits of Gefitinib for high-risk patients. Gefitinib is an orally administered targeted therapy drug, known as an inhibitor of the tyrosine kinase receptor for the epidermal growth factor (EGFR). Frequently utilized for the management of NSCLC, especially in individuals with activating mutations in the EGFR gene (Hosomi et al., 2020). Although there was literature reported EGFR as a potential therapeutic target in TNBC(Corkery et al., 2009), to date, there is still insufficient clinical trial evidence demonstrating significant efficacy of Gefitinib in breast cancer treatment. Several studies have also investigated the efficacy of Gefitinib when used in conjunction with other forms of treatment, including chemotherapy and hormonal therapy medications. Carine M. and her team covalently linked Gefitinib and Tamoxifen to develop a new anti-cancer drug conjugate, which showed promising effects in various types of breast cancer cells (Abdelmalek et al., 2022). Studies by other scientists have demonstrated that the joint use of Gefitinib and HER3 antibody can notably decrease the phosphorylation of HER3, EGFR, Akt, and ERK1/2 in TNBC cells, leading to successful growth suppression and cell death (Lyu et al., 2023). The research conducted on cells establishes a groundwork for the medical application of Gefitinib, confirming its therapeutic possibilities for breast cancer patients at high risk and setting the stage for upcoming clinical trials.

Breast cancer patients are mainly treated with chemotherapy, hormone therapy, targeted therapy, and radiation therapy, which are proven effective methods (Ben-Dror et al., 2022). Through drug sensitivity analysis in this study, we stratified breast cancer patients

using coagulation genes in a manner distinct from traditional molecular subtyping. Patients with diverse risk scores showed different responses to identical medications. This provides personalized recommendations for clinical drug administration, potentially reducing the occurrence of drug resistance and enhancing drug efficacy to some extent.

5 Conclusion and perspective

This study has utilized bioinformatics approaches to identify six coagulation-related genes with prognostic significance in breast cancer. By developing a risk-scoring model, we have demonstrated its potential in predicting breast cancer outcomes and its utility in assessing the immune environment, response to immunotherapies, and drug sensitivities. This model serves as a valuable tool for the personalized treatment of breast cancer patients. However, the study's findings are based on data from public sources, and the model's validation on real-world datasets remains a crucial next step. Looking ahead, several research directions and challenges present themselves to further advance the field: 1) Real-World Data Validation: Future research should focus on validating the prognostic model using real-world clinical data to confirm its applicability and generalizability in diverse patient populations. 2) Multi-Dataset Analysis: Expanding the model's validation to additional independent datasets, including those from different ethnicities and geographical regions, will enhance the robustness of the model and its predictive accuracy. 3) Functional Studies of Genes: In-depth experimental studies are needed to elucidate the biological functions of the identified coagulation-related genes in breast cancer progression. This includes investigating their roles in tumor growth, metastasis, and response to therapies. 4) Integration of Omics Data: Combining the current gene expression data with other omics data, such as proteomics and metabolomics, could provide a more comprehensive understanding of the molecular mechanisms underlying the prognostic value of coagulation-related genes. 5) Clinical Trial Design: Future clinical trials should consider incorporating the risk scores derived from our model to stratify patient populations and evaluate the efficacy of targeted therapies and personalized treatment strategies. 6) Technological Advancements: Keeping abreast of emerging technologies and bioinformatics tools will be essential for refining the model and incorporating new insights into the complex interplay between coagulation and cancer. By addressing these challenges and directions, we aim to contribute to the evolving landscape of precision medicine in breast cancer. We are committed to furthering our research to provide more targeted and effective treatment options for patients.

Data availability statement

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

Ethics statement

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

Author contributions

CL: Conceptualization, Data curation, Formal Analysis, Writing–original draft. YL: Conceptualization, Data curation, Writing–original draft. HY: Investigation, Software, Writing–review and editing. KZ: Formal Analysis, Project administration, Writing–original draft. WL: Methodology, Project administration, Writing–review and editing. NW: Resources, Supervision, Writing–review and editing. LX: Funding acquisition, Resources, Supervision, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Chinese Academy of Medical Sciences Clinical and Translational Medicine Research Innovation Fund Project (No. 2022-I2M-C&T-A-013), and the Chinese Academy of Medical Sciences Clinical and Translational Medicine Research Fund Project (No. 2019XK320067).

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/fmolb.2024.1394585/full#supplementary-material

References

Abdelmalek, C. M., Hu, Z., Kronenberger, T., Küblbeck, J., Kinnen, F. J. M., Hesse, S. S., et al. (2022). Gefitinib-tamoxifen hybrid ligands as potent agents against triple-negative breast cancer. *J. Med. Chem.* 65 (6), 4616–4632. doi:10.1021/acs.jmedchem.1c01646

Ai, X., Xiang, L., Huang, Z., Zhou, S., Zhang, S., Zhang, T., et al. (2018). Overexpression of PIK3R1 promotes hepatocellular carcinoma progression. *Biol. Res.* 51 (1), 52. doi:10.1186/s40659-018-0202-7

Akhlaghpour, M., Haritunians, T., More, S. K., Thomas, L. S., Stamps, D. T., Dube, S., et al. (2023). Genetic coding variant in complement factor B (CFB) is associated with increased risk for perianal Crohn's disease and leads to impaired CFB cleavage and phagocytosis. *Gut* 72 (11), 2068–2080. doi:10.1136/gutjnl-2023-329689

Ben-Dror, J., Shalamov, M., and Sonnenblick, A. (2022). The history of early breast cancer treatment. *Genes (Basel)* 13 (6), 960. doi:10.3390/genes13060960

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

Brown, R. B., Bigelow, P., Dubin, J. A., and Mielke, J. G. (2023). High dietary phosphorus is associated with increased breast cancer risk in a U.S. Cohort of middle-aged women. *Nutrients* 15 (17), 3735. doi:10.3390/nu15173735

Brown, R. B., Bigelow, P., Dubin, J. A., and Neiterman, E. (2024). Breast cancer, alcohol, and phosphate toxicity. *J. Appl. Toxicol.* 44 (1), 17–27. doi:10.1002/jat.4504

Cai, J., Huang, Z., Zhou, J., Wu, W., and Ye, Y. (2022). TLN2 functions as a tumor suppressor in clear cell renal cell carcinoma via inactivation of the Wnt/β-catenin signaling pathway. *Transl. Androl. Urol.* 11 (1), 39–52. doi:10.21037/tau-21-914

Chakraborty, G., Nandakumar, S., Hirani, R., Nguyen, B., Stopsack, K. H., Kreitzer, C., et al. (2022). The impact of PIK3R1 mutations and insulin-PI3K-glycolytic pathway regulation in prostate cancer. *Clin. Cancer Res.* 28 (16), 3603–3617. doi:10.1158/1078-0432.Ccr-21-4272

Chan, H. J., Li, H., Liu, Z., Yuan, Y. C., Mortimer, J., and Chen, S. (2015). SERPINA1 is a direct estrogen receptor target gene and a predictor of survival in breast cancer patients. *Oncotarget* 6 (28), 25815–25827. doi:10.18632/oncotarget.4441

Cizkova, M., Vacher, S., Meseure, D., Trassard, M., Susini, A., Mlcuchova, D., et al. (2013). PIK3R1 underexpression is an independent prognostic marker in breast cancer. *BMC Cancer* 13, 545. doi:10.1186/1471-2407-13-545

Corkery, B., Crown, J., Clynes, M., and O'Donovan, N. (2009). Epidermal growth factor receptor as a potential therapeutic target in triple-negative breast cancer. *Ann. Oncol.* 20 (5), 862–867. doi:10.1093/annonc/mdn710

Dann, R., Hadi, T., Montenont, E., Boytard, L., Alebrahim, D., Feinstein, J., et al. (2018). Platelet-derived MRP-14 induces monocyte activation in patients with symptomatic peripheral artery disease. *J. Am. Coll. Cardiol.* 71 (1), 53–65. doi:10.1016/j.jacc.2017.10.072

Daugan, M. V., Revel, M., Russick, J., Dragon-Durey, M. A., Gaboriaud, C., Robe-Rybkine, T., et al. (2021). Complement C1s and C4d as prognostic biomarkers in renal cancer: emergence of noncanonical functions of C1s. *Cancer Immunol. Res.* 9 (8), 891–908. doi:10.1158/2326-6066.Cir-20-0532

Desoteux, M., Maillot, B., Bévant, K., Ferlier, T., Leroux, R., Angenard, G., et al. (2023). Transcriptomic evidence for tumor-specific beneficial or adverse effects of TGFβ pathway inhibition on the prognosis of patients with liver cancer. *FEBS Open Bio* 13 (7), 1278–1290. doi:10.1002/2211-5463.13647

Falanga, A., Schieppati, F., and Russo, D. (2015). Cancer tissue procoagulant mechanisms and the hypercoagulable state of patients with cancer. *Semin. Thromb. Hemost.* 41 (7), 756–764. doi:10.1055/s-0035-1564040

Fang, K. P., Dai, W., Ren, Y. H., Xu, Y. C., Zhang, S. M., and Qian, Y. B. (2016). Both Talin-1 and Talin-2 correlate with malignancy potential of the human hepatocellular carcinoma MHCC-97 L cell. *BMC Cancer* 16, 45. doi:10.1186/s12885-016-2076-9

Feinauer, M. J., Schneider, S. W., Berghoff, A. S., Robador, J. R., Tehranian, C., Karreman, M. A., et al. (2021). Local blood coagulation drives cancer cell arrest and brain metastasis in a mouse model. *Blood* 137 (9), 1219–1232. doi:10.1182/blood.2020005710

Frafjord, A., Buer, L., Hammarström, C., Aamodt, H., Woldbæk, P. R., Brustugun, O. T., et al. (2021). The immune landscape of human primary lung tumors is Th2 skewed. *Front. Immunol.* 12, 764596. doi:10.3389/fimmu.2021.764596

Gao, Y., Zhou, H., Liu, G., Wu, J., Yuan, Y., and Shang, A. (2022). Tumor microenvironment: lactic acid promotes tumor development. *J. Immunol. Res.* 2022, 3119375. doi:10.1155/2022/3119375

Gofrit, S. G., and Shavit-Stein, E. (2019). The neuro-glial coagulonome: the thrombin receptor and coagulation pathways as major players in neurological diseases. *Neural Regen. Res.* 14 (12), 2043–2053. doi:10.4103/1673-5374.262568

Graf, C., Wilgenbus, P., Pagel, S., Pott, J., Marini, F., Reyda, S., et al. (2019). Myeloid cell-synthesized coagulation factor X dampens antitumor immunity. *Sci. Immunol.* 4 (39), eaaw8405. doi:10.1126/sciimmunol.aaw8405

He, C., Li, Y., Zhang, R., Chen, J., Feng, X., and Duan, Y. (2021). Low CFB expression is independently associated with poor overall and disease-free survival in patients with lung adenocarcinoma. *Oncol. Lett.* 21 (6), 478. doi:10.3892/ol.2021.12739

Hisada, Y., and Mackman, N. (2021). Tissue factor and extracellular vesicles: activation of coagulation and impact on survival in cancer. *Cancers (Basel)* 13 (15), 3839. doi:10.3390/cancers13153839

Hong, C. C., Yao, S., McCann, S. E., Dolnick, R. Y., Wallace, P. K., Gong, Z., et al. (2013). Pretreatment levels of circulating Th1 and Th2 cytokines, and their ratios, are associated with ER-negative and triple negative breast cancers. *Breast Cancer Res. Treat.* 139 (2), 477–488. doi:10.1007/s10549-013-2549-3

Hosomi, Y., Morita, S., Sugawara, S., Kato, T., Fukuhara, T., Gemma, A., et al. (2020). Gefitinib alone versus gefitinib plus chemotherapy for non-small-cell lung cancer with mutated epidermal growth factor receptor: NEJ009 study. *J. Clin. Oncol.* 38 (2), 115–123. doi:10.1200/jco.19.01488

Hou, M. (2022). Exploring novel independent prognostic biomarkers for hepatocellular carcinoma based on TCGA and GEO databases. *Med. Baltim.* 101 (43), e31376. doi:10.1097/md.0000000000031376

Huh, S., Kang, C., Park, J. E., Nam, D., Kim, S. I., Seol, A., et al. (2022). Novel diagnostic biomarkers for high-grade serous ovarian cancer uncovered by data-independent acquisition mass spectrometry. *J. Proteome Res.* 21 (9), 2146–2159. doi:10.1021/acs.jproteome.2c00218

Jia, Y., Kodumudi, K. N., Ramamoorthi, G., Basu, A., Snyder, C., Wiener, D., et al. (2021). Th1 cytokine interferon gamma improves response in HER2 breast cancer by modulating the ubiquitin proteasomal pathway. *Mol. Ther.* 29 (4), 1541–1556. doi:10.1016/j.ymthe.2020.12.037

Kim, M. K., Breitbach, C. J., Moon, A., Heo, J., Lee, Y. K., Cho, M., et al. (2013). Oncolytic and immunotherapeutic vaccinia induces antibody-mediated complement-dependent cancer cell lysis in humans. *Sci. Transl. Med.* 5 (185), 185ra63. doi:10.1126/scitranslmed.3005361

Kuai, X., Lv, J., Zhang, J., Xu, M., and Ji, J. (2023). Serpin family A member 1 is prognostic and involved in immunological regulation in human cancers. *Int. J. Mol. Sci.* 24 (14), 11566. doi:10.3390/ijms241411566

Liu, L., Du, X., Fang, J., Zhao, J., Guo, Y., Zhao, Y., et al. (2021). Development of an interferon gamma response-related signature for prediction of survival in clear cell renal cell carcinoma. *J. Inflamm. Res.* 14, 4969–4985. doi:10.2147/jir.S334041

Lyu, H., Shen, F., Ruan, S., Tan, C., Zhou, J., Thor, A. D., et al. (2023). HER3 functions as an effective therapeutic target in triple negative breast cancer to potentiate the antitumor activity of gefitinib and paclitaxel. *Cancer Cell Int.* 23 (1), 204. doi:10.1186/s12935-023-03055-w

Ma, L., Sun, Y., Li, D., Li, H., Jin, X., and Ren, D. (2020). Overexpressed ITGA2 contributes to paclitaxel resistance by ovarian cancer cells through the activation of the AKT/FoxO1 pathway. *Aging (Albany NY)* 12 (6), 5336–5351. doi:10.18632/aging.102954

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. *Brief. Bioinform* 22 (6), bbab260. doi:10.1093/bib/bbab260

Pitt, J. M., Marabelle, A., Eggermont, A., Soria, J. C., Kroemer, G., and Zitvogel, L. (2016). Targeting the tumor microenvironment: removing obstruction to anticancer immune responses and immunotherapy. *Ann. Oncol.* 27 (8), 1482–1492. doi:10.1093/annonc/mdw168

Riedl, J., Preusser, M., Nazari, P. M., Posch, F., Panzer, S., Marosi, C., et al. (2017). Podoplanin expression in primary brain tumors induces platelet aggregation and increases risk of venous thromboembolism. *Blood* 129 (13), 1831–1839. doi:10.1182/blood-2016-06-720714

Riihilä, P., Viiklepp, K., Nissinen, L., Farshchian, M., Kallajoki, M., Kivisaari, A., et al. (2020). Tumour-cell-derived complement components C1r and C1s promote growth of cutaneous squamous cell carcinoma. *Br. J. Dermatol* 182 (3), 658–670. doi:10.1111/bjd.18095

Röth, A., Barcellini, W., D'Sa, S., Miyakawa, Y., Broome, C. M., Michel, M., et al. (2021). Sutimlimab in cold agglutinin disease. *N. Engl. J. Med.* 384 (14), 1323–1334. doi:10.1056/NEJMoa2027760

Ruterbusch, M., Pruner, K. B., Shehata, L., and Pepper, M. (2020). *In vivo* CD4(+) T cell differentiation and function: revisiting the Th1/Th2 paradigm. *Annu. Rev. Immunol.* 38, 705–725. doi:10.1146/annurev-immunol-103019-085803

Saidak, Z., Soudet, S., Lottin, M., Salle, V., Sevestre, M. A., Clatot, F., et al. (2021). A pan-cancer analysis of the human tumor coagulome and its link to the tumor immune microenvironment. *Cancer Immunol. Immunother.* 70 (4), 923–933. doi:10.1007/s00262-020-02739-w

Shahzad, M. H., Feng, L., Su, X., Brassard, A., Dhoparee-Doomah, I., Ferri, L. E., et al. (2022). Neutrophil extracellular traps in cancer therapy resistance. *Cancers (Basel)* 14 (5), 1359. doi:10.3390/cancers14051359

Siegel, R. L., Giaquinto, A. N., and Jemal, A. (2024). Cancer statistics, 2024. *CA Cancer J. Clin.* 74 (1), 12–49. doi:10.3322/caac.21820

Taniguchi, C. M., Winnay, J., Kondo, T., Bronson, R. T., Guimaraes, A. R., Alemán, J. O., et al. (2010). The phosphoinositide 3-kinase regulatory subunit p85alpha can exert tumor suppressor properties through negative regulation of growth factor signaling. *Cancer Res.* 70 (13), 5305–5315. doi:10.1158/0008-5472.Can-09-3399

Timp, J. F., Braekkan, S. K., Versteeg, H. H., and Cannegieter, S. C. (2013). Epidemiology of cancer-associated venous thrombosis. *Blood* 122 (10), 1712–1723. doi:10.1182/blood-2013-04-460121

Tinholt, M., Tekpli, X., Torland, L. A., Tahiri, A., Geisler, J., Kristensen, V., et al. (2024). The breast cancer coagulome in the tumor microenvironment and its role in prognosis and treatment response to chemotherapy. *J. Thromb. Haemost.* doi:10.1016/j.jtha.2024.01.003

Wahab, R., Hasan, M. M., Azam, Z., Grippo, P. J., and Al-Hilal, T. A. (2023). The role of coagulome in the tumor immune microenvironment. *Adv. Drug Deliv. Rev.* 200, 115027. doi:10.1016/j.addr.2023.115027

Wang, C., Li, X., Xue, B., Yu, C., Wang, L., Deng, R., et al. (2022). RasGRP1 promotes the acute inflammatory response and restricts inflammation-associated cancer cell growth. *Nat. Commun.* 13 (1), 7001. doi:10.1038/s41467-022-34659-x

Wei, F., Su, Y., Quan, Y., Li, X., Zou, Q., Zhang, L., et al. (2023). Anticoagulants enhance molecular and cellular immunotherapy of cancer by improving tumor microcirculation structure and function and redistributing tumor infiltrates. *Clin. Cancer Res.* 29 (13), 2525–2539. doi:10.1158/1078-0432.Ccr-22-2757

Will, M., Liang, J., Metcalfe, C., and Chandarlapaty, S. (2023). Therapeutic resistance to anti-oestrogen therapy in breast cancer. *Nat. Rev. Cancer* 23 (10), 673–685. doi:10.1038/s41568-023-00604-3

Xing, X., Hu, Y. H., Wang, Y., Shao, Y., and Zou, M. (2022). No effect on tumorigenesis in MG63 cells induced by Co-cultured mesenchymal stem cells. *J. Oncol.* 2022, 4202439. doi:10.1155/2022/4202439

Xu, L., Li, X., Li, X., Wang, X., Ma, Q., She, D., et al. (2022). RNA profiling of blood platelets noninvasively differentiates colorectal cancer from healthy donors and noncancerous intestinal diseases: a retrospective cohort study. *Genome Med.* 14 (1), 26. doi:10.1186/s13073-022-01033-x

Ye, F., Dewanjee, S., Li, Y., Jha, N. K., Chen, Z. S., Kumar, A., et al. (2023). Advancements in clinical aspects of targeted therapy and immunotherapy in breast cancer. *Mol. Cancer* 22 (1), 105. doi:10.1186/s12943-023-01805-y

Zhao, Z., Ma, J., Mao, Y., Dong, L., Li, S., and Zhang, Y. (2018). Silence of α1-antitrypsin inhibits migration and proliferation of triple negative breast cancer cells. *Med. Sci. Monit.* 24, 6851–6860. doi:10.12659/msm.910665

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 02 May 2024
DOI 10.3389/fmolb.2024.1365655

Check for updates

OPEN ACCESS

EDITED BY
Yi Chieh Lim,
Danish Cancer Society Research Center (DCRC), Denmark

REVIEWED BY
Rajeev Nema,
Manipal University Jaipur, India
Hassan Dastsooz,
University of Turin, Italy

*CORRESPONDENCE
Ying Zhang,
zy1122812@163.com
Xi Cao,
caoxi@ahmu.edu.cn

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

RECEIVED 04 January 2024
ACCEPTED 12 April 2024
PUBLISHED 02 May 2024

CITATION
Zhu Z, Zhang W, Huo S, Huang T, Cao X and Zhang Y (2024), TUBB, a robust biomarker with satisfying abilities in diagnosis, prognosis, and immune regulation via a comprehensive pan-cancer analysis.
Front. Mol. Biosci. 11:1365655.
doi: 10.3389/fmolb.2024.1365655

COPYRIGHT
© 2024 Zhu, Zhang, Huo, Huang, Cao and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

TUBB, a robust biomarker with satisfying abilities in diagnosis, prognosis, and immune regulation via a comprehensive pan-cancer analysis

Zaifu Zhu[1†], Wei Zhang[2†], Shaohu Huo[1], Tiantuo Huang[1], Xi Cao[3,4]* and Ying Zhang[5,6]*

[1]Department of Pediatrics, First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China, [2]Department of Rehabilitation Medicine, First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China, [3]Department of Pharmacy, First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China, [4]The Grade 3 Pharmaceutical Chemistry Laboratory of State Administration of Traditional Chinese Medicine, Hefei, Anhui, China, [5]Department of Pathology, First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China, [6]Pathology Center, Anhui Medical University, Hefei, Anhui, China

Purpose: TUBB can encode a beta-tubulin protein. At present, the role of TUBB has not been ascertained in cancers. Hence, the importance of further systematic pan-cancer analyses is stressed to explore its value in the diagnosis, prognosis, and immune function of cancers.

Methods: By collecting and handling integrative data from the TCGA, Firehose, UCSC Xena, cBioPortal, GEO, CPTAC, TIMER2.0, TISCH, CellMiner, GDSC, and CTRP databases, we explored the potential diagnostic and prognostic roles of TUBB in pan-cancers from multiple angles. Moreover, the GSEA analysis was conducted to excavate the biological functions of TUBB in pan-cancers. In addition, survival profiles were described, and the differential expressions of TUBB in different molecular subtypes were discussed. Also, we utilized the cMAP function to search drugs or micro-molecules that have an impact on TUBB expressions.

Results: Based on the TCGA data, we found that TUBB was differentially expressed in a variety of tumors and showed an early-diagnostic value. Mutations, somatic copy number alterations, and DNA methylation would lead to its abnormal expression. TUBB expressions had relations with many clinical features. What's more, TUBB expressions were validated to be related to many metabolism-related, metastasis-related, and immune-related pathways. High TUBB expressions were proved to have a great impact on the prognosis of various types of cancers and would affect the sensitivity of some drugs. We also demonstrated that the expression of TUBB was significantly correlated to immunoregulator molecules and biomarkers of lymphocyte subpopulation infiltration.

Conclusion: TUBB and its regulatory genes were systemically analyzed in this study, showing that TUBB had satisfying performances in disease diagnosing and

prognosis predicting of multiple cancers. It could remodel the tumor microenvironment and play an integral role in guiding cancer therapies and forecasting responses to chemotherapy.

KEYWORDS

TUBB, diagnosis, pan-cancer, prognosis, TME

Introduction

Many people in the world are battling cancer (Dolgin, 2021). The explicit pathogenesis of cancer has always been a hot topic in cancer research. Multiform therapies like chemotherapy, radiotherapy, targeted therapy, and immunotherapy are separately or compositely used clinically. Chemotherapy is a basic approach to treating cancer. Immunotherapy has greatly revolutionized both the research and treatment of cancer (Szeto and Finley, 2019). However, due to the complexity of carcinogenesis, the efficacy of cancer therapies cannot be guaranteed. Unfortunately, they are not able to enhance the long-term prognosis of patients. Therefore, it is of great significance to analyze a potential biomarker gene at the pan-cancer level and explore its connections to the clinical outcome of cancer patients. Currently, the largest database of cancer genes is the Cancer Genome Atlas (TCGA) (https://www.cancer.gov/ccg/research/genome-sequencing/tcga). More than 11,000 tumors from 33 of the most prevalent forms of cancer have been analyzed, constituting a uniquely comprehensive, in-depth pan-cancer atlas that serves as an essential resource for the development of new therapies in the pursuit of precision medicine (Hoadley et al., 2018).

TUBB (β-tubulin) is a protein-coding gene, responsible for forming a heterodimer with α-tubulin and acting as a structural component of microtubules (Miller et al., 2010), which are a kind of long hollow polymers that are 25 nm wide and range in length from <1 μm to >100 μm (Goodson and Jonasson, 2018). Microtubules play a major role in controlling different aspects of cell architecture and function (Wu and Akhmanova, 2017), they are essential for motor-driven intracellular transport, interact with accessory proteins to assemble into larger structures, and coordinate with other types of cells as a mature network (Goodson and Jonasson, 2018). As a result, compounds that target microtubules can interfere with multiple vital cellular processes (Wordeman and Vicente, 2021), for example, they inhibit microtubule polymerization, destroy the dynamic imbalance of microtubules, damage spindles, block cell cycle, and cause tumor cell death. Such anti-tumor drugs are collectively called microtubule-targeting agents (MTAs), served as cancer therapy for many years, the first being paclitaxel, introduced in 1994 (Field et al., 2014). What's more, recent studies report that alterations in the expression of certain tubulin isotypes and associated post-translational modifications (PTMs) have been observed in human cancers, however, the exact implications of whether and how the tubulin code can mediate the biological progress of cancer cells are not yet clear (Lopes and Maiato, 2020). Therefore, TUBB as a mediator, has been studied in the cancer research area. In osteosarcoma, TUBB has been identified as the significant survival-predicting factor (Shao et al., 2022). Alhammad R. found that, the overexpression of TUBB would lead to a worse prognosis in ERα-positive and better prognosis in ERα-negative breast cancer (Alhammad, 2022). TUBB plays an important role in cancer progression and targeting TUBB may provide significant clues in cancer treatments. However, how TUBB modulates cancer initiation and advancement in pan-cancers remains controversial. Hence, it is necessary to explore TUBB profiles from the perspectives of multiple cancer cells. Thorough analyses are needed to understand the intrinsic role of TUBB in tumor immunity.

Considering the lack of pan-cancer analysis of TUBB, this study aims to conduct a comprehensive exploration of the potential roles of TUBB in malignant tumor cells and its underlying mechanisms in the prediction of clinical prognosis. The TUBB profiles, including their expressions, mutations, relations with aggressive tumor traits, and contributions to the survival of cancer patients, have been depicted. With the aid of several famous public databases, the analysis was performed based on the web tools and "R" software. We found that TUBB was significantly correlated with various cancer characteristics, tumor immune microenvironment (TIME), drug resistance, and survival states. All the results highlighted the critical roles of TUBB in cancer and they contribute to further studies of TUBB-related molecular mechanisms and therapeutic development.

Materials and methods

Acquisition and organization of public data from different databases

Firstly, the flow chart of this study is shown in Figure 1. Then, the mRNA expression data, copy number alteration threshold data, masked copy number segmentation data, and DNA methylation 450K data of both tumor and normal tissues in the TCGA pan-cancer cohort were obtained from the Firehose database (http://gdac.broadinstitute.org) (Deng et al., 2017). TCPA, mutation, and clinical data were acquired from the UCSC Xena database (https://xenabrowser.net/datapages/) (Wang et al., 2022). The mutation frequency of TUBB in the TCGA cohort was calculated using the cBioPortal database (https://www.cbioportal.org/) (Cerami et al., 2012). Based on the data from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) (Barrett et al., 2013), the TUBB expression at the transcript level was validated. We analyzed and estimated its expression at the protein level based on the CPTAC database (https://proteomics.cancer.gov/programs/cptac) (Zhang et al., 2014). Also, various immune infiltrating algorithms from the TIMER2.0 database (http://timer.cistrome.org) (Li et al., 2020) were adopted to depict the correlations between TUBB expression and tumor immune microenvironment (TIME). Via 99 single-cell datasets from the TISCH database (http://tisch.comp-genomics.org/home) (Sun et al., 2021), immune infiltration

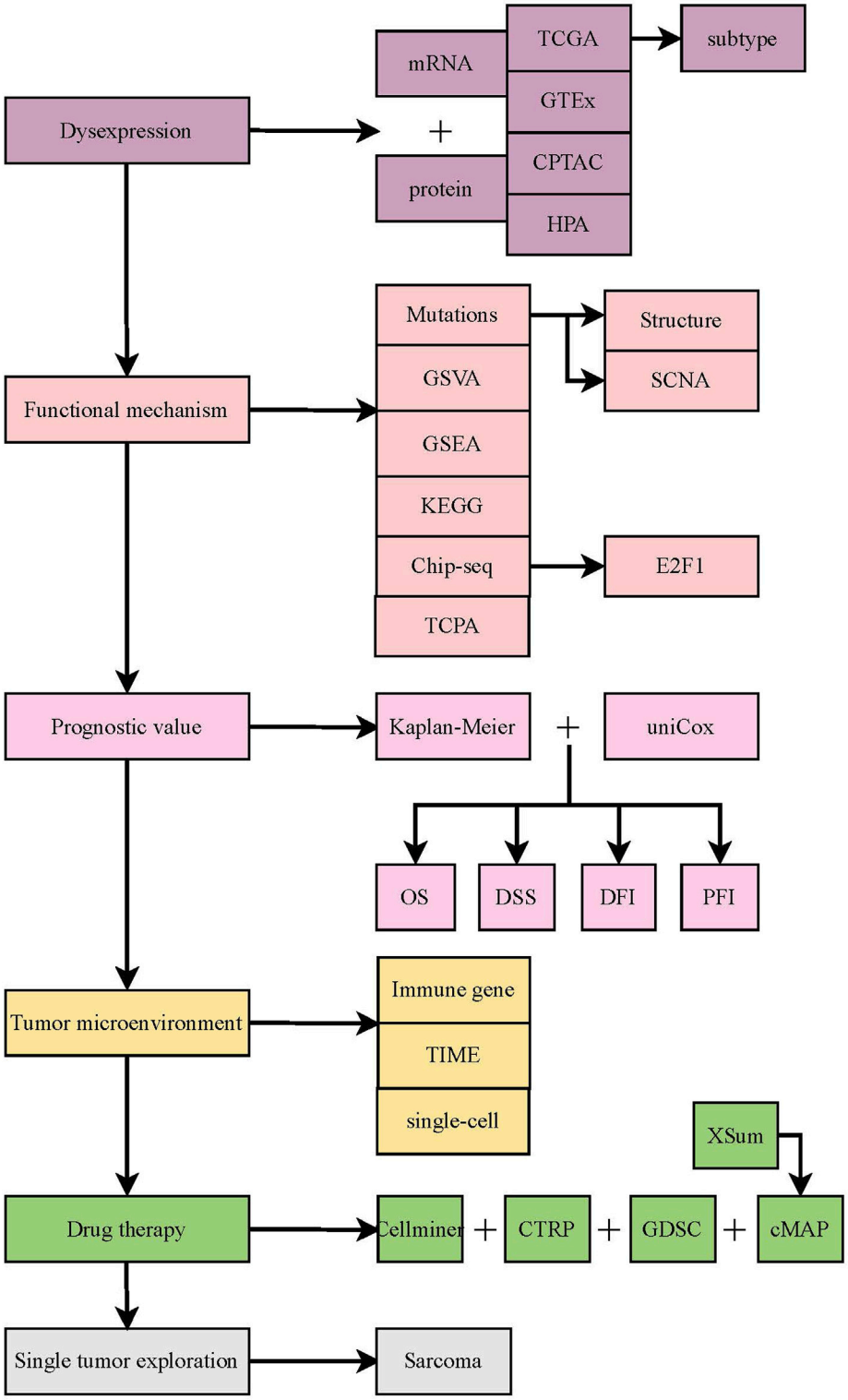


FIGURE 1
The flow chart of this study.

results by whole transcriptome analysis were verified. From the CellMiner (https://discover.nci.nih.gov/cellminer/CellMiner) (Shankavaram et al., 2009), GDSC (https://www.cancerrxgene.org/) (Yang et al., 2013), and CTRP (http://portals.broadinstitute.org/ctrp/) databases (Rees et al., 2016), we gained relevant chemotherapy data to illustrate how TUBB expressions interact with drug sensitivity. It should be noted that these public databases are free and open. The study strictly follows the data extraction policy of the databases and does not require the ethical review and approval of the ethics committee.

Differential expression analyses at multi-omics level

To observe whether imbalanced expressions of TUBB exist between tumor and normal tissues, differential analyses were conducted in terms of three dimensions.

First of all, we combined the data from the Genotype-Tissue Expression (GTEx) project because the TCGA-data was inadequate, to enhance the confidence by expanding the sample sizes. The "wilcox" test was applied to detect the variability ($p < 0.05$ was considered significant). The "gganatogram" R package was adopted to visualize the expressions in human's different organs. Then, the "wilcox" analysis was performed to compare the differential mRNA expressions of TUBB between tumor and normal tissues based on TCGA data. Moreover, we performed the "wilcox" test on paired samples from the TCGA cancer groups ($p < 0.05$ was considered significant, $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$). Using the "pROC" R package, the importance of TUBB in the diagnosis of pan-cancers was addressed. The area under curve (AUC) values were calculated (between 0.5 and 1, the closer the AUC value came to one, the better diagnostic performance TUBB would have). Usually, AUC values of 0.5–0.7, 0.7–0.9, and >0.9 indicate low accuracy, certainty accuracy, and high accuracy, respectively.

Based on the GEO database, we conducted the external validation at the transcriptional level. At the protein level, the protein expressions of TUBB were validated via the CPTAC database. What's more, the immunohistochemical (IHC) staining of tissues was collected from the HPA database. The "Kruskal" test was used to identify the expressions of TUBB in different stages or molecular subtypes.

Analyzing the survival status and clinical outcomes in pan-cancer

Survival data was obtained from the TCGA database, the "survival"and "survminer" R packages were utilized to describe the relations between the expression of TUBB and prognostic indexes of TUBB (including overall survival (OS), disease-specific survival (DSS), progression-free interval (PFI), and disease free interval (DFI)). Integrating the "Kaplan-Meier (KM)" analysis and the univariate COX analysis, we evaluated whether the TUBB was a protective or risky factor, and finally, the survival landscape of TUBB was drawn with a high level of confidence. Noticeably, the optimal cut-off between the high- and low-TUBB mRNA expression groups was determined via the "survminer" R package when the KM analysis was conducted. The "survfit" function was applied to conduct the "log rank" test to estimate the significance of the high- and low-TUBB expression groups. At last, the "forestplot" R package was used to visualize the results of survival data by the COX analysis.

Somatic copy number alteration (SCNA), mutation, and DNA methylation analysis

The cBioPortal (http://www.cbioportal.org) website is a powerful tool for retrieving, downloading, analyzing, and visualizing cancer genomic data from various types of genomic data such as somatic mutations, DNA copy-number alterations (CNAs), and DNA methylation. It can undertake multiple analyses, including mutation analysis and its visualization (Cerami et al., 2012). On cBioPortal, "TCGA Pan Cancer Atlas Study" was selected in the "Quick Search" section (Gao et al., 2013), and "TUBB" was typed to search for its genetic altering traits. The mutation frequency, types, and CNA data were collected from the "Cancer Type Summary" part. The mutation locations of TUBB were shown in a two-dimensional (2D) diagram of the protein structure using the "mutation" function. Somatic copy number alterations (SCNAs) and mutations can increase the CNA of the gene by amplifying and deleting the heterozygosity and homozygosity. Generally, high-frequency SCNA is defined with >5% mutation frequency. Next, the "Spearman" correlations between the expression level of TUBB and CNA scores were calculated to estimate the relation between SCNA and TUBB expression. The "IlluminaHumanMmethylation-450kanno.ilmn12.hg19" R package from "Bioconductor" was conducted to annotate the methylation probe of the TUBB promoter. Through the "Wilcoxon" rank test was performed to detect the differential methylation of TUBB between tumor and normal tissues. A p-value cutoff of 0.05 was used to identify cancers that were significantly hypomethylated or hypermethylated. The Spearman correlation between TUBB expressions and promoter DNA methylation Beta was calculated (p-value < 0.05 was considered significant).

Exploring pathways and functional mechanisms

To understand TUBB-related pathways, we divided tumor samples of each type according to TUBB expressions (top 30% and bottom 30%). Then the Gene Set Enrichment Analysis (GSEA) was carried out to compare differential activation or inhibition conditions of 50 hallmark gene sets and 83 metabolic gene sets in different tumors between the high- and low-TUBB group. Yuan H et al. have redefined 14 functional states of malignant tumoral features (Yuan et al., 2019). And, the "z-score" algorithm was proposed by Lee et al. (2008). It can reflect the activity of a given pathway by integrating expressions of characteristic genes. Using the "GSVA" R package, we conducted the "z-score" algorithm on the 14 functional state gene sets. The values were set as the z-score of each gene set. Then, the "Pearson" correlation analysis was performed to calculate statistical relations between TUBB expression and the z-score of each gene set. Additionally, we identified the differential genes between the high- and low-TUBB group. Moreover, to search for transcription factors that may affect the expression of TUBB, we used the CistromeDB database (http://dbtoolkit.cistrome.org/) to identify potential regulatory upstream factors for TUBB. As known, protein-protein interacting data usually contains those unlikely biological interactions that are impossible to happen in living cells. Therefore, we used the ComPPI database (https://comppi.linkgroup.hu/) which introduces two novel quantitative scores, the Localization Score and the Interaction Score, describing the calculated probability of the data correctness, to gain the proteins that interact with TUBB.

TUBB has been identified as a risky factor according to four survival indexes in sarcoma (SARC). Hence, we conducted the GSEA on multiple gene sets to analyze possible pathways that SARC may participate in. It was believed that if one gene had differential expression in more than five cancers, its functional gene would be thought to be related to TUBB and undertake a KEGG enrichment analysis to recognize the conservative functions or pathways that TUBB takes part in pan-cancers. Subsequently, we adopted the permutation test to identify the mutation that was statistically related to TUBB expressions in SARC. Then, using Fischer's exact test, we estimated the relationship between E2F1 and TUBB expressions. In the end, we conducted the Spearman correlation analysis to systemically identify BEND3-related proteins.

Exploring the tumor microenvironment and performing a single-cell analysis

The continuous interactions between tumor cells and the tumor microenvironment (TME) play a decisive role in tumor initiation, progression, metastasis, and responses to therapies. As a consequence, we confirmed correlations between TUBB expressions and immune-related genes (immunostimulatory genes, immunoinhibitory genes, chemokine genes, chemokine receptor genes, and MHC genes). We also adopted the same approach to conduct the differential analysis on TIP scores. Next, from the TIMER2.0 databases, seven most advanced algorithms were applied (CIBERSORT, CIBERSORT_ABS EPIC, MCPCOUNTER, QUANTISEQ, TIMER, XCELL) to estimate the immune filtration profiles of TCGA cancers (Li et al., 2020). At last, we used the TISCH database to download the expression landscape of TUBB in 99 single-cell datasets of 38 tumors, which helped us verify the TME analytic result at the single-cell level. In summary, we provided a comprehensive analysis and visualized the TME and immune infiltration profiles of TUBB in a pan-cancer cohort. It is worth noting that, we analyzed the anti-tumor immunity in seven steps of the cancer-immunity cycle (release of cancer cell antigens, cancer antigen presentation, priming and activation, trafficking of T cells to tumors, infiltration of T cells into tumors, recognition of cancer cells by T cells, killing of cancer cells (Chen and Mellman, 2013) and conducted the Gene Set Variation Analysis (GSVA) on each step, and then we compared their differences in the high- and low-TUBB groups.

Identification of chemical substances interacting with TUBB

Based on the GSCA database (http://bioinfo.life.hust.edu.cn/web/GSCALite/), we analyzed the relations between TUBB expressions and drug sensitivity (Liu et al., 2023). GSCA website contains 750 small molecule drugs from GDSC and CTRP databases. Besides, the gene expression data is used to mine the valuable small molecule drugs related to it. In addition, the National Cancer Institute (NCI) database was utilized. NCI serves as a trusted source of cancer information, and the platform of cancer cell lines from NCI has been widely used to screen out drugs that are related to a certain gene's expression. NCI-60 is a collection of 60 cancer cell lines from nine different kinds of cancers (leukemia, colon, lung, central nervous system, kidney, melanoma, ovarian, breast, and prostate cancers). Its data comes from the "CellMiner" database (Reinhold et al., 2023), for us to analyze the Spearman correlations between TUBB mRNA expressions and the z-score of drug sensitivity. Moreover, differentially expressed genes between the high- and low-TUBB group in different cancers were identified. We collected the top 150 upregulated and downregulated genes as TUBB-related biomarkers. The CMAP_gene_signatures.RData document containing 1,288 compound-related characteristics was downloaded from the BHKLAB database (https://www.pmgenomics.ca/bhklab/sites/default/files/downloads) and used to calculate matching scores. All analyzing processes kept to the methods from the previous literature (Malta et al., 2018). The R software was used to summarize and display the top five results of 32 kinds of cancers.

Statistical analysis

All data was processed using the web tools and R software (V.4.3.0, Institute of Statistics and Mathematics, Vienna, Austria). The Pearson correlation analysis was used on normally distributed data, otherwise the Spearman correlation analysis was used. The Kruskal–Wallis rank sum test, Wilcoxon rank sum and Signed rank tests were used to detect the differences between multiple variables or two variables, respectively. Using the "survival" R package, the COX and KM survival analyses were conducted. KM method adopted the log-rank test to detect significance. The "survminer" R package was used to visualize results from KM analysis. Hazard ratio (HR) and 95% confidence interval (CI) were used to describe relative risks. The "pROC" R package was utilized to perform ROC analysis to estimate the diagnostic ability of TUBB. All statistical tests were two-tailed. p-value < 0.05 was deemed as statistically significant. p-value < 0.0001 was deemed as greatly statistically significant ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$).

Results

Aberrant expressions of TUBB among cancers

To identify patterns of TUBB regulations in cancers, we combined TCGA and GTEx data to expand sample sizes and gained a boxplot describing TUBB expressions in pan-cancer (Figure 2A). Figure 2B showed TUBB expressions in different organs. TUBB had a differential expression in most cancers. It was greatly upregulated across cancers. Then, the differential analysis based on TCGA samples and paired samples were shown in Figures 2C, D. With the aid of logistics regression analysis using TCGA and TCGA-GTEx data, above results were validated (Figure 2E). An external validation of TUBB mRNA expressions was based on GEO database (Supplementary Figure S1). From CPTAC database, the validation at the protein level was

10.3389/fmolb.2024.1365655

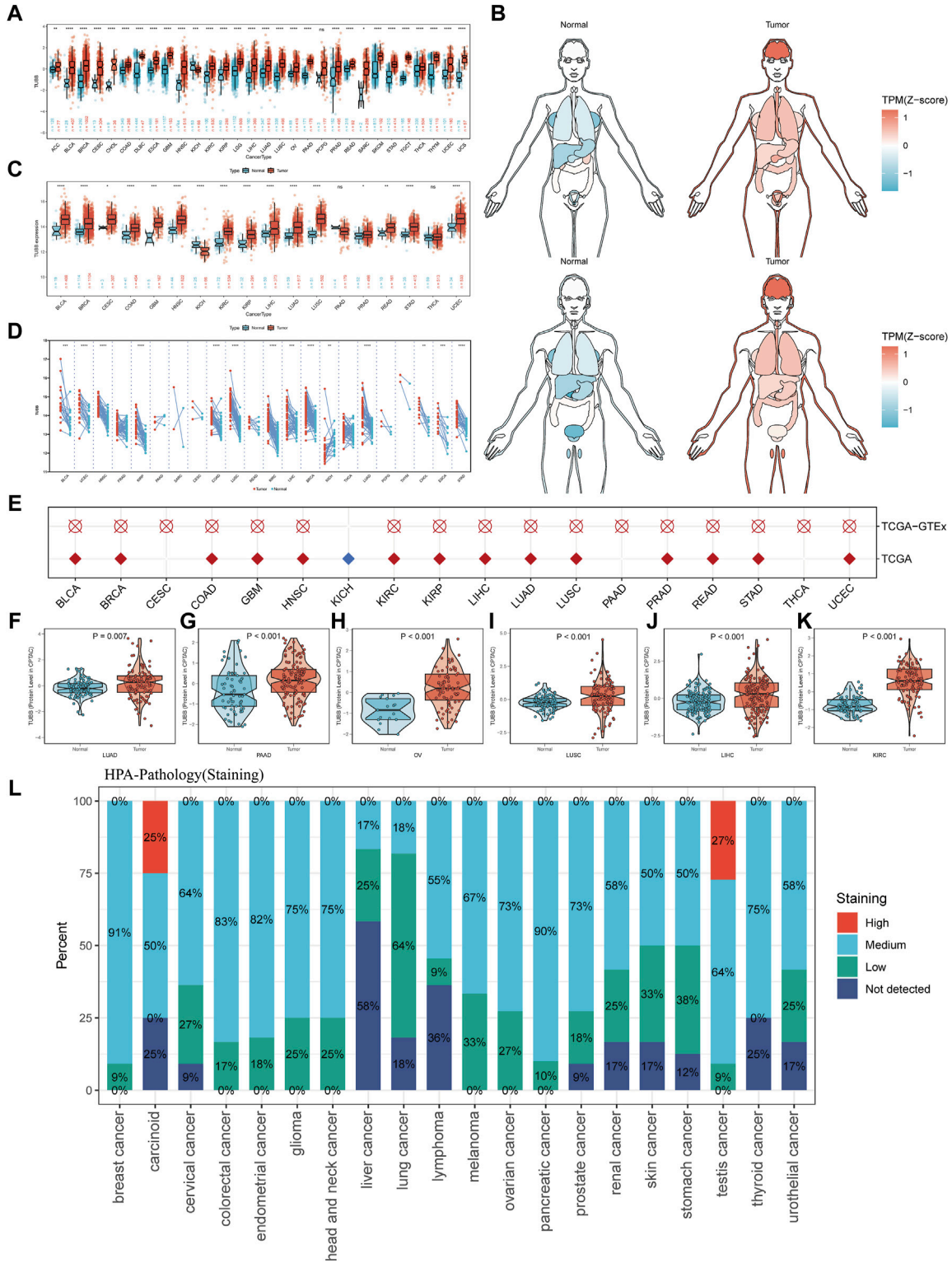


FIGURE 2
The expression profiles of TUBB in pan-cancer. **(A)** TUBB expressions in between tumor and normal tissues in various cancers using TCGA and GTEx data. **(B)** Expressions and distributions of TUBB between tumor and normal tissues in various organs. **(C)** TUBB mRNA expressions in the TCGA between tumor and normal tissues in TCGA. **(D)** Similar to **(C)**, but in paired samples grouped by cancer from TCGA. Each point representing one sample. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$) **(E)** Logistic regression analysis of TCGA, TCGA-GTEx. Red means OR>1, blue represents an OR value between 0 and 1. White circle means no significance. **(F–K)** Differential protein levels between tumor and normal tissues in LUAD, PAAD, OV, LUSC, LIHC, and KIRC based on the CPATC database. **(L)** IHC results indicate that TUBB showed weak to moderate cytoplasmic positivity in most malignant cells. Different colors represent different staining indicators.

10.3389/fmolb.2024.1365655

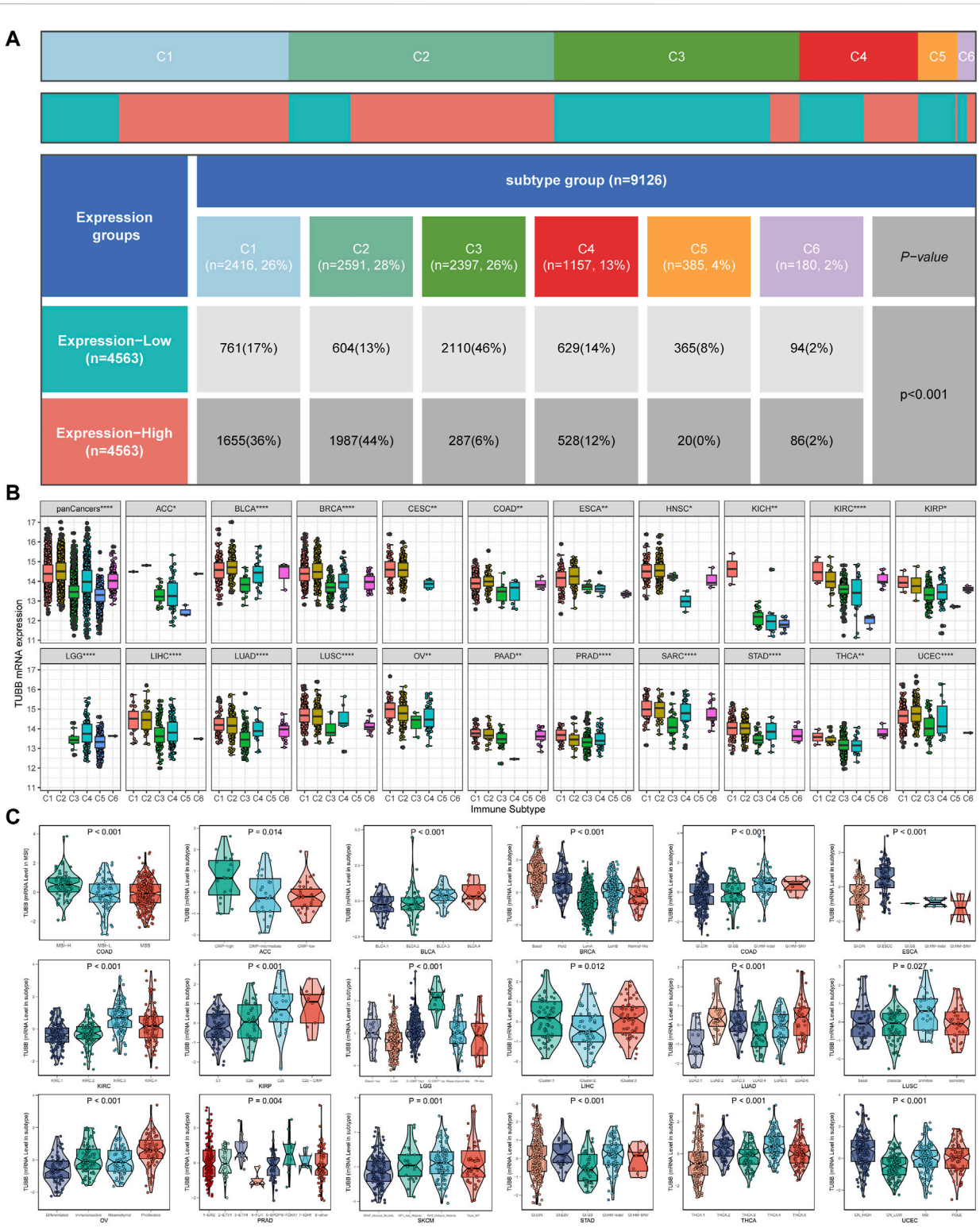


FIGURE 3
Analysis of clinical variables and molecular subtypes. **(A)** More patients with C1 and C2 immune subtypes in the TUBB high-expression group and more patients with C3 subtypes in the TUBB low-expression group. **(B)** Expression levels of TUBB in different immune subtypes. C1 (wound healing); C2 (IFN-gamma dominant); C3 (inflammatory); C4 (lymphocyte depleted); C5 (immunologically quiet); C6 (TGF-b dominant). The Kruskal test detects differences between six immune subtypes. **(C)** Differences in TUBB expression in different molecular subtypes.

conducted (Figures 2F–K). It was observed that results from different angles and databases had good consistency. IHC results showed that, most cancer cells displayed weak to moderate cytoplasmic immunoreactivity. A few seminomas and carcinoids were strongly stained. Most hepatocellular carcinomas were negative (Figure 2L). Moreover, TUBB expressions were correlated with

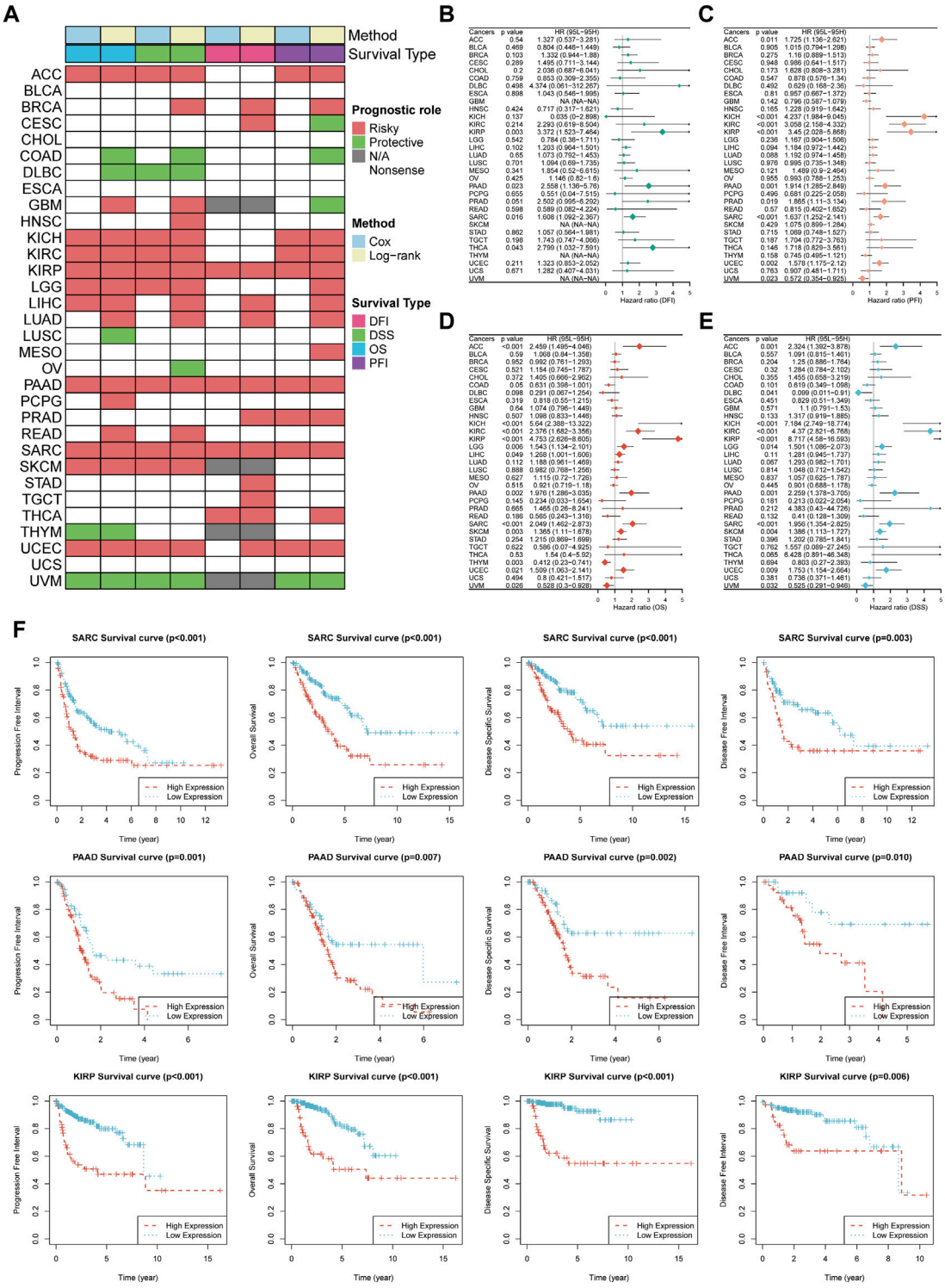


FIGURE 4
Survival profiles of TUBB in pan-cancers. **(A)** Correlations between TUBB expressions with overall survival (OS), disease-specific survival (DSS), disease free interval (DFI) and progression-free interval (PFI) based on the univariate Cox regression and Kaplan-Meier models. Red indicates that TUBB is a risk factor affecting the prognosis of cancer patients, and green represents a protective factor. Only *p* values <0.05 are shown. **(B–E)** Forest plots exhibit the prognostic role of TUBB in cancers by univariate Cox regression method. **(F)** Kaplan-Meier survival analysis and log-rank test were performed using "survival" and "survminer" packages.

tumor stages in seven cancers (ACC, ESCA, KIRC, KIRP, LIHC, SKCM, STAD) (Supplementary Figure S2), implying that TUBB may have a relationship with the progression of some cancers. The estimated ROC curves showed that (Supplementary Figure S3), the TUBB mRNA expression level showed satisfactory sensitivity and specificity for the diagnosis of nine kinds of tumors (AUC>0.7).

10.3389/fmolb.2024.1365655

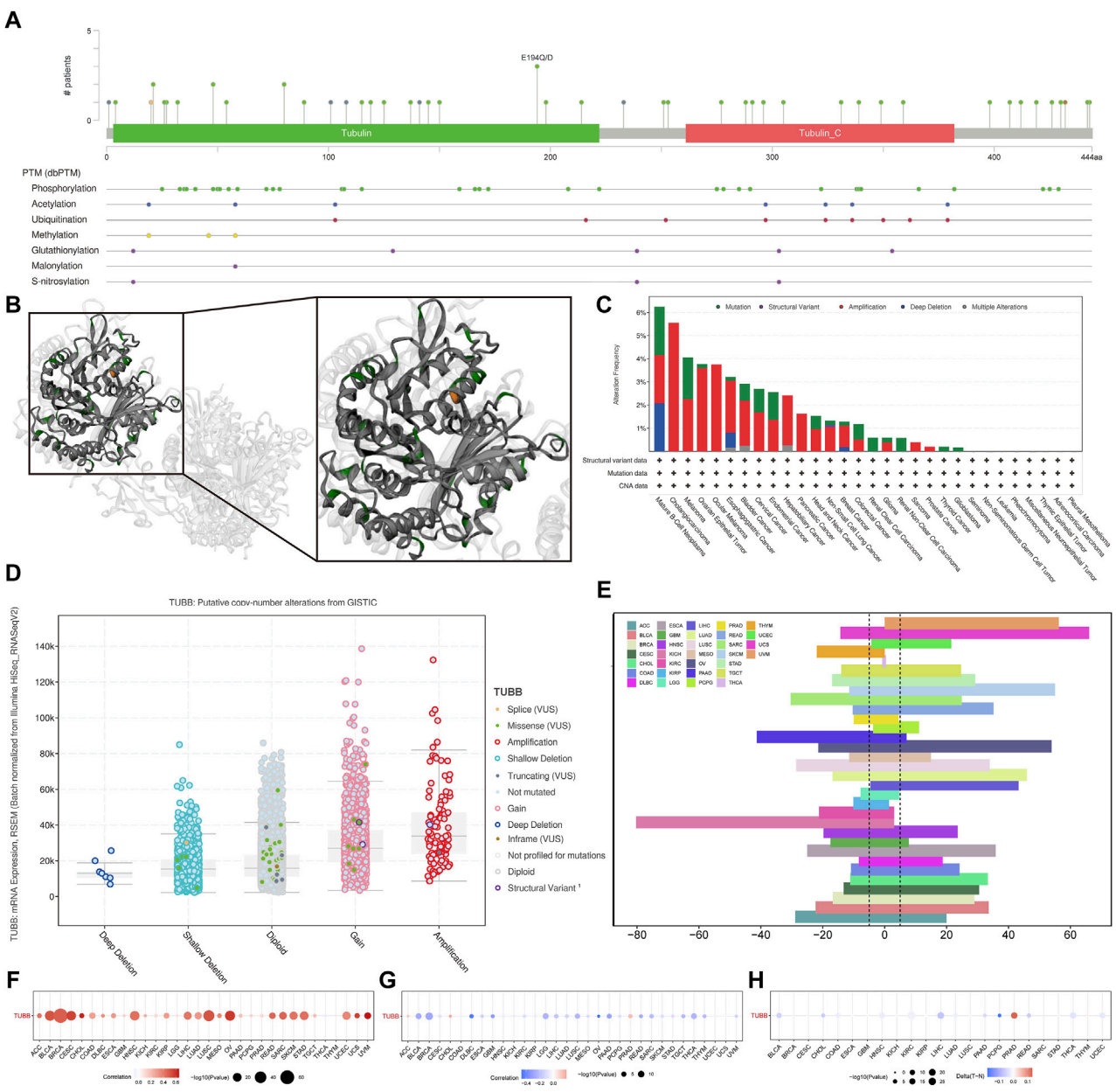


FIGURE 5
Genetic alterations of TUBB in cancers. **(A)** Sites and numbers of TUBB genetic alterations across cancers from cBioPortal. **(B)** 3D structure of TUBB mutation sites. **(C)** Frequency of TUBB mutations in different tumor types. **(D)** Relationship between TUBB mRNA expressions and genetic alterations. **(E)** Histogram shows the frequency of somatic copy number alterations for TUBB in each cancer type. **(F)** The Spearman's correlation between somatic copy number alterations and the expression of TUBB. **(G)** Spearman's correlation of TUBB between transcriptional expressions and promoter methylation. Red and blue represent positive and negative correlations, respectively. **(H)** Bubble Gram shows the differential methylation of TUBB in cancers; hypermethylated and hypomethylated TUBBs are marked in red and blue, respectively (Wilcoxon rank-sum test).

After expanding normal sample sizes with GTEx data, the results also were robust (Supplementary Figure S4). The repeatable and consistent results were demonstrated across multiple databases, multiple tumors, multiple methods, and multiple omics, suggesting that dysregulation of TUBB expression may play a crucial role in different cancers and was highly unlikely to be a false finding due to technical artifacts, chance, or bias in the eligibility criteria for TCGA samples. In addition, by the chi-square test, we found that C1 and C2 immune subtypes were more in the high-TUBB group (Figure 3A). The same results were obtained using the Kruskal test, that is, TUBB was expressed more in the C1 and C2 immune subtypes compared to other subtypes (Figure 3B). Interestingly, TUBB showed differences in a large number of molecular subtypes. For example, it had the lowest expression in LumA group, and in the basal group, its expression was the highest in breast cancer (BRCA), meaning that it had great values in the precision molecular stratification therapies and prognosis prediction (Figure 3C).

Clinical relevance of TUBB

To further dissect the clinical relevance of TUBB in cancers, the role of TUBB was analyzed. Survival profiles of pan-cancers showed that TUBB expressions were related to many kinds of survival indicators in pan-cancers (Figure 4A), and the relations had homogeneity. TUBB was identified as a risky factor in cancers, especially in KIRP, PAAD, and SARC because TUBB was risky weighed by all indicators DFI, DSS, OS, and PFI. In several cancers, TUBB was protective. For example, UVM patients with higher TUBB expressions would have better survival probability. Due to the highly heterogeneity across various cancers, in some cancers (BLCA, CHOL, ESCA, UCS), the tendency was vacant. Therefore, TUBB may play different roles in pan-cancers, suggesting further explorations of it should be addressed. Forest plots showed COX survival analysis results of four survival indicators. The hazard ratios of each cancer were listed, too (Figures 4B–E). KM curves were used to display the log-rank test results in KIRP, PAAD, and SARC (Figure 4F).

Genetic alterations of TUBB in cancers

We analyzed genomic data (genetic variation, somatic copy number alteration (SCNA), mRNA expression, and DNA methylation) of tumor and normal tissues from the TCGA cohort. Mutation sites of TUBB were visualized in 2D and 3D graphics (Figures 5A, B). TUBB had genetic changes in most cancers, and the most common alteration types were amplification and mutation (Figure 5C). Obviously, SCNA plays an essential role in regulating gene expression in cancers (Figure 5D). To study genetic changes of TUBB in cancers, we checked the percentage of SCNA. SCNA appeared frequently in most cancers (more than 5% among all samples). Only in a few types, the frequency was low (Figure 5E). Therefore, we estimated how SCNA affected TUBB mRNA expressions by calculating the Spearman correlation between TUBB expressions and masked copy number segment of TCGA. In most cancers, the mRNA expression of TUBB was significantly correlated with SNCA (Figure 5F). Therefore, it could be concluded that TUBB CNAs were common among cancers, and it could affect TUBB expressions. Besides, promoter methylation of TUBB can also regulate gene expressions. And, abnormal DNA methylation of promoters was associated with tumorigenesis. The methylation patterns of TUBB in pan-cancers were consistent with each other. From all results, it was observed that TUBB mRNA expressions were negatively correlated with DNA methylation (Figure 5G), and higher methylations existed in tumor tissues more than in normal tissues in most cancers (Figure 5H) because when TUBB is significantly overexpressed, the body increases promoter methylation to overcome this dysregulation to keep the balance. Great differences in the methylation patterns of TUBB suggested the complexity of TUBB regulation and the specificity of this process among different cancer species. In addition, we selected 10 transcription factors with the highest scores to annotate peaks and found that they were usually located in the promoter region of TUBB, showing that they may regulate the expression of TUBB.

Associations between TUBB and cancer-related pathways

Samples with top 30% and bottom 30% TUBB expressions were set as the high- and low-TUBB group. Based on two groups, we conducted the GSEA to explore relevant cell signaling in cancers. Furthermore, metabolism-related pathways were analyzed systematically, and the results had satisfying consistency across cancers, indicating the functions of TUBB were highly conservative. What's more, pathways related to cell-cycle were enriched in tumors with higher TUBB. Besides, TUBB may take part in metabolic disorder processes in cancers. Meanwhile, TUBB was believed to inhibit the course of drug metabolism, thus it may be related to responses to chemotherapy (Figure 6A). We analyzed the correlations between z-scores of 14 symbolic functional states of cancers (angiogenesis, apoptosis, cell cycle, differentiation, DNA damage, DNA repair, EMT, hypoxia, inflammation, invasion, metastasis, proliferation, quiescence, stemness) with the z-score of TUBB expressions. Among them, the z-score of the cell cycle had the highest R-value (0.51), showing a positive relation with TUBB expressions (Figure 6B). Then, in multiple cancers, we ran a Pearson correlation analysis between the z-score of TUBB expressions and the z-score of the cell cycle state. Learning from Figure 5C, in thymoma (THYM) the highest coefficient of 0.87 was detected (Figure 6C). CHIP-seq demonstrated that E2F1 might be an upstream transcription factor modulating the TUBB expression (Figure 6D). ComPPI helped identify the genes that may interact with TUBB (Figure 7A). The KEGG enrichment analysis was conducted on genes that were highly expressed in the high-TUBB group. Figure 7B showed that TUBB had relations with a lot of functions, especially with signal transduction, immune, and cancer-related pathways. As previously mentioned, TUBB was determined as a risky factor in terms of four survival indexes in SARC. Then we performed the GSEA on various gene sets to fully analyze the pathways that TUBB may be involved in (Figure 7C). From Figure 7D, the expression of TUBB probably was relevant to ATRX mutations. In SARC, E2F1 was statistically associated with TUBB expressions (Figure 7E). Besides, based on the TCPA database, some TUBB-related proteins were identified (Figures 7F–O). Among them, CYCLINB1 had the most significant relation with TUBB mRNA expressions with the highest coefficient of 0.57.

High TUBB expression correlates with immune infiltration in cancers

First of all, the heatmap displayed the landscape of immune infiltration of TUBB across cancers (Figure 8A). TUBB was significantly negatively correlated with immune-related genes (MHC, immune-inhibitor, immune-stimulator, chemokines). Notably, we scored seven steps and compared the differences between the high- and low-TUBB group, and the tendency was consistent in pan-cancers, namely the TIP scores were lower (Supplementary Figure S5). To illustrate cells regulated by TUBB in the TME, we used the TIMER2.0 database to explore correlations between TUBB mRNA expressions and immune infiltration and stroma cell abundance (Figure 8B). For example, in almost every cancer the TUBB mRNA expression level was positively related to the abundance of $CD4^+$ Th1 cells. In SARC, the TUBB mRNA

10.3389/fmolb.2024.1365655

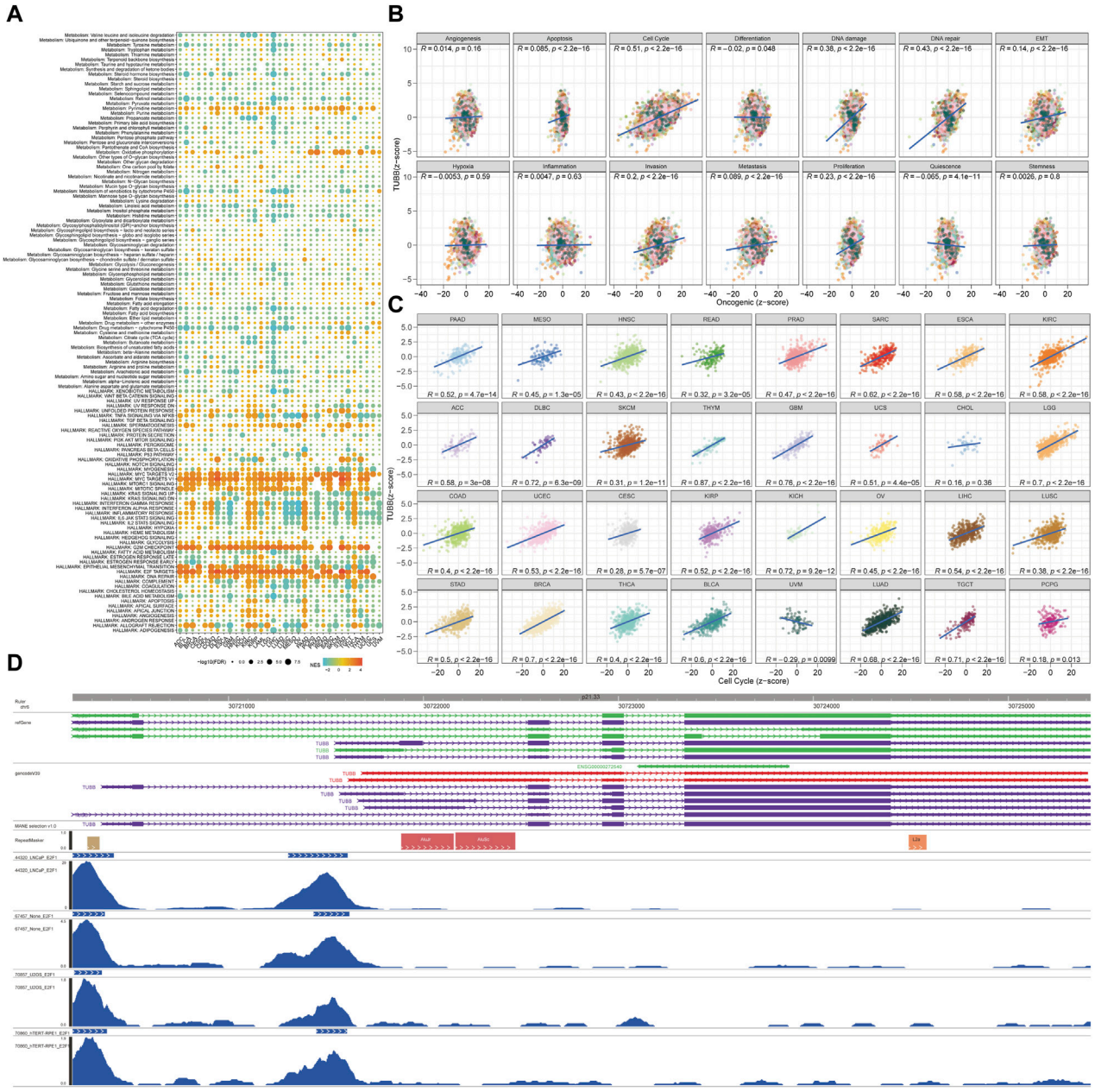


FIGURE 6
Analyses of pathways and underlying mechanisms. **(A)** Enrichment differences of TUBB in 50 HALLMARK and 83 metabolism gene sets. **(B)** The TUBB mRNA expression was highly correlated with 14 malignant features of all tumors. **(C)** The TUBB mRNA expression was highly correlated with cell cycle features of all tumors. **(D)** The 10 transcription factors with the highest scores were selected for peaks annotation, and it was found that they were usually located in the promoter region of TUBB.

expression level was negatively related to the abundance of B cells and CD8$^+$ T cells across different software, which could possibly explain why TUBB was risky in cancers. Different proportions of immune infiltration and unique TMEs in pan-cancers existed, so the correlations showed various change laws. However, we utilized seven different algorithms to conduct the analysis and made the results verify each other to ensure the accuracy of the study. Additionally, we could see from Figure 8C that the single-cell analysis showed that though TUBB was not expressed strikingly in most tumors, it mainly originated from malignant cells and proliferative T cells. This echoed previous results. In other words, TUBB might encourage the formation of immunological rejection or "immunological desert", and play a vital role in immunity-cancer crosstalk, especially in immune escape.

TUBB may influence responses to chemotherapy

Using Cellminer data, we found that TUBB was positively related to the sensitivity of a lot of drugs (Figure 9A), but negatively related to the half maximal inhibitory concentration

10.3389/fmolb.2024.1365655

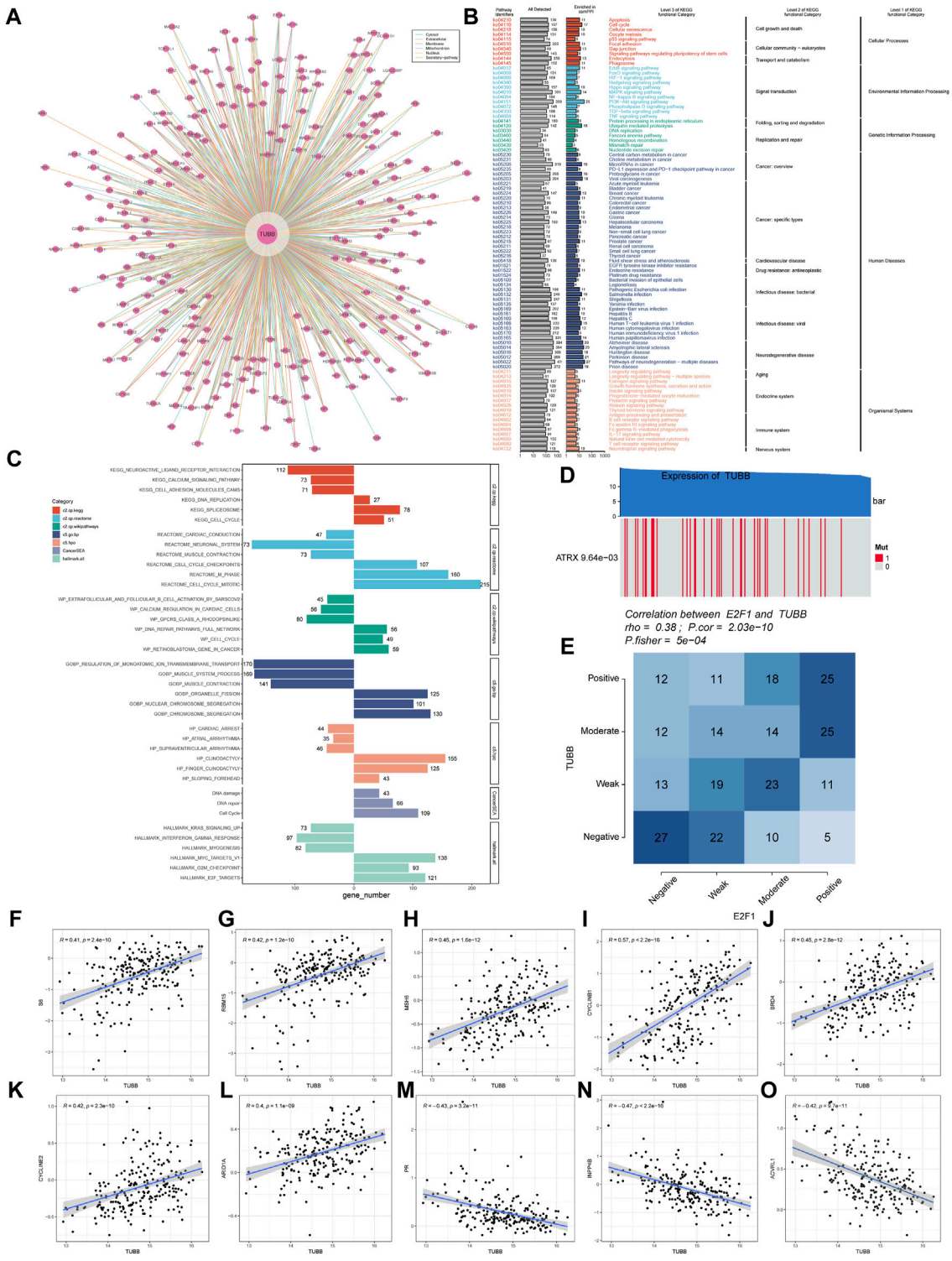


FIGURE 7
Deeper exploration of TUBB mainly in SARC. **(A)** Networks of genes interacting with TUBB. **(B)** KEGG enrichment analysis results showing highly expressed genes. **(C)** GSEA enrichment analysis results. Different colors represent different gene sets. Bar charts on the left means significant enrichment in the high-TUBB group, and *vice versa*, significant enrichment in the low-TUBB group. **(D)** The value adjacent to highly mutated gene is permutatio test *p*-value of gene expression between driver mutated (red) and not-mutated (gray) samples. **(E)** Fischer's precise test explores the correlation between E2F1 and TUBB. **(F–O)** TUBB related proteins of SARC in the TCPA database.

(IC50) values of many drugs based on CTRP, GDSC databases (Figures 9B, C). Therefore, we could deduce that TUBB was a potential chemotherapy-sensitive gene. To further explore underlying therapeutic regimens that offset the tumorigenic effect mediating by TUBB, CMap analysis was employed. Thus, a TUBB-related signature containing 150 significantly upregulated and

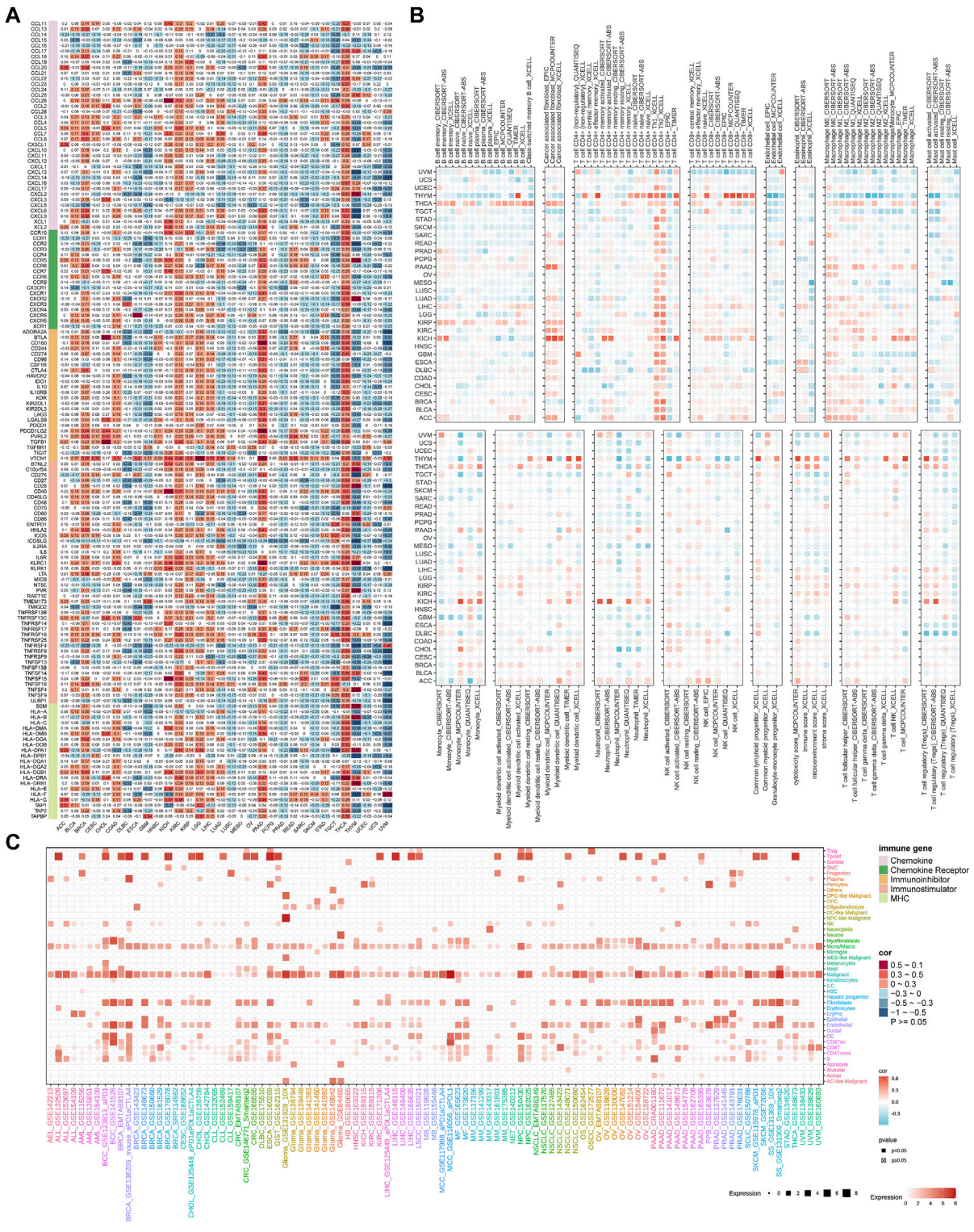


FIGURE 8
Association of TUBB expressions with immune infiltration. **(A)**. The heatmap shows correlations between TUBB mRNA expressions and expressions of chemokine, chemokine receptor, immune-inhibitor, immune-stimulatory, and MHC genes. **(B)** Seven software were used to evaluate the correlations between TUBB expression and cancer immune infiltration. **(C)** Cell sources of TUBB in pan-cancer at the single-cell level.

150 significantly downregulated genes that were selected by screening patients with high- or low-TUBB expressions in various cancers, was generated. Using the Extreme Summarization (XSum) method, TUBB-related traits were compared with CMap gene traits, and the similarity scores for 1,288 compounds were obtained. Fasudil, W.13, AH.6809 and

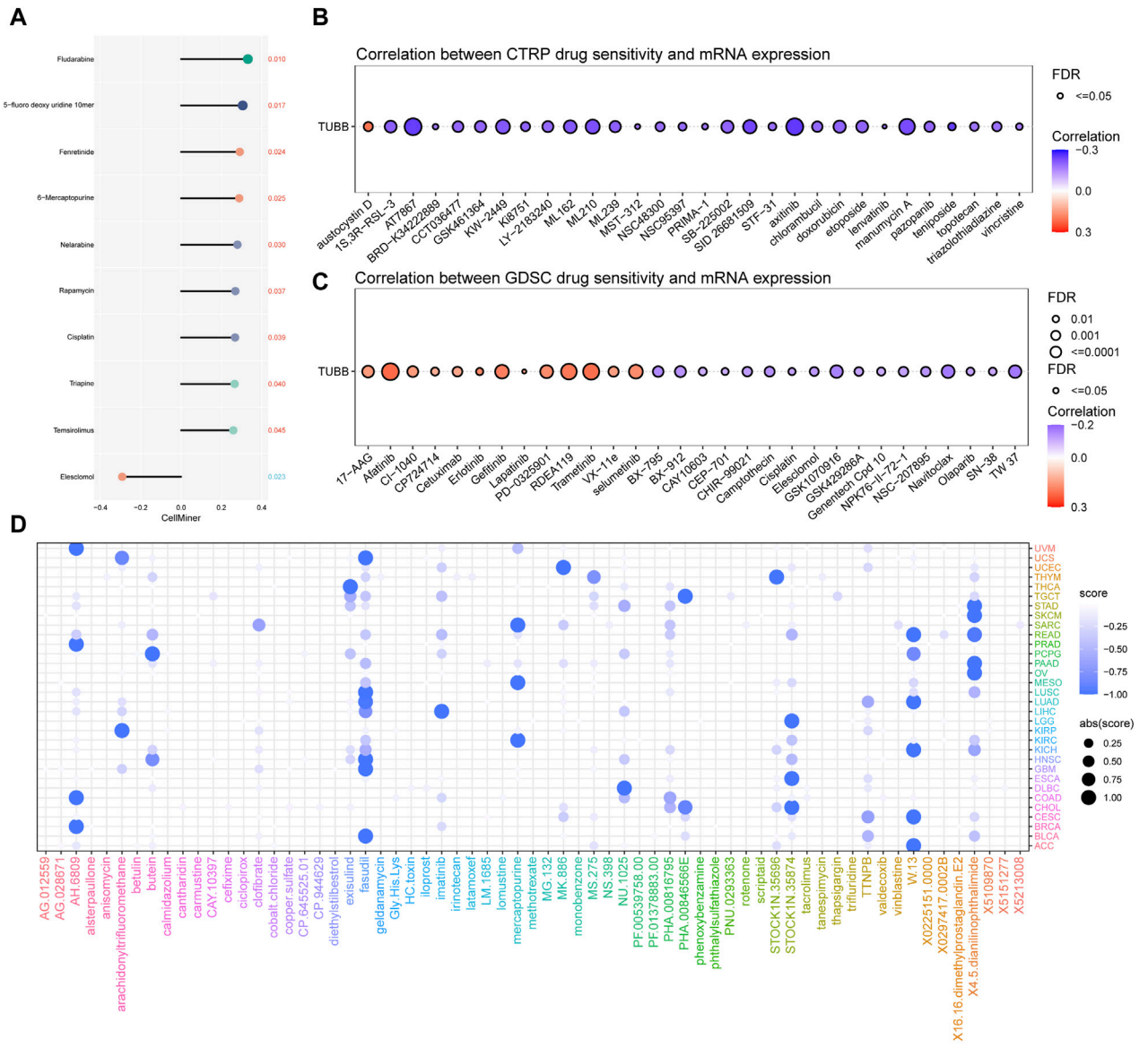


FIGURE 9
Chemotherapeutic drug resistance analysis. Correlations between TUBB expressions and drug sensitivity using the three different databases Cellminer **(A)**, CTRP **(B)**, GDSC **(C)** ($p < 0.05$ was considered statistically significant). **(D)** Prediction of potential compounds targeting TUBB.

X4.5. dianilinophthalimide showed relatively lower scores in most cancers, which demonstrated that they perhaps inhibit TUBB-mediated carcinogenic effects (Figure 9D). These findings provided substantial support for the validity of our predictions, although further researches are needed to elucidate the underlying mechanisms.

TUBB expressions in the tissue sample

To verify the expression of TUBB among the tissue samples of cancer patients, IHC staining was conducted. Figure 10A–C displayed the negative results of TUBB expression in the control subjects. TUBB exhibited strong diffuse staining in osteosarcoma tissues (Figures 10D, E). Positive staining was also found in chondrogenic sarcoma tissues (Figure 10F).

Discussion

Hitherto, cancer-related research has always been a main point and difficulty in the current domain. With data from multi-platforms, multi-omics approaches were used in this study. TUBB was identified as an important indicator across multiple cancers. TUBB was valuable in the early-detection, prognosis prediction, therapy selection in pan-cancers. Also, we found that TUBB was related to many vital biological pathways, indicating its indispensable roles in cancers. TUBB was validated as a risky factor

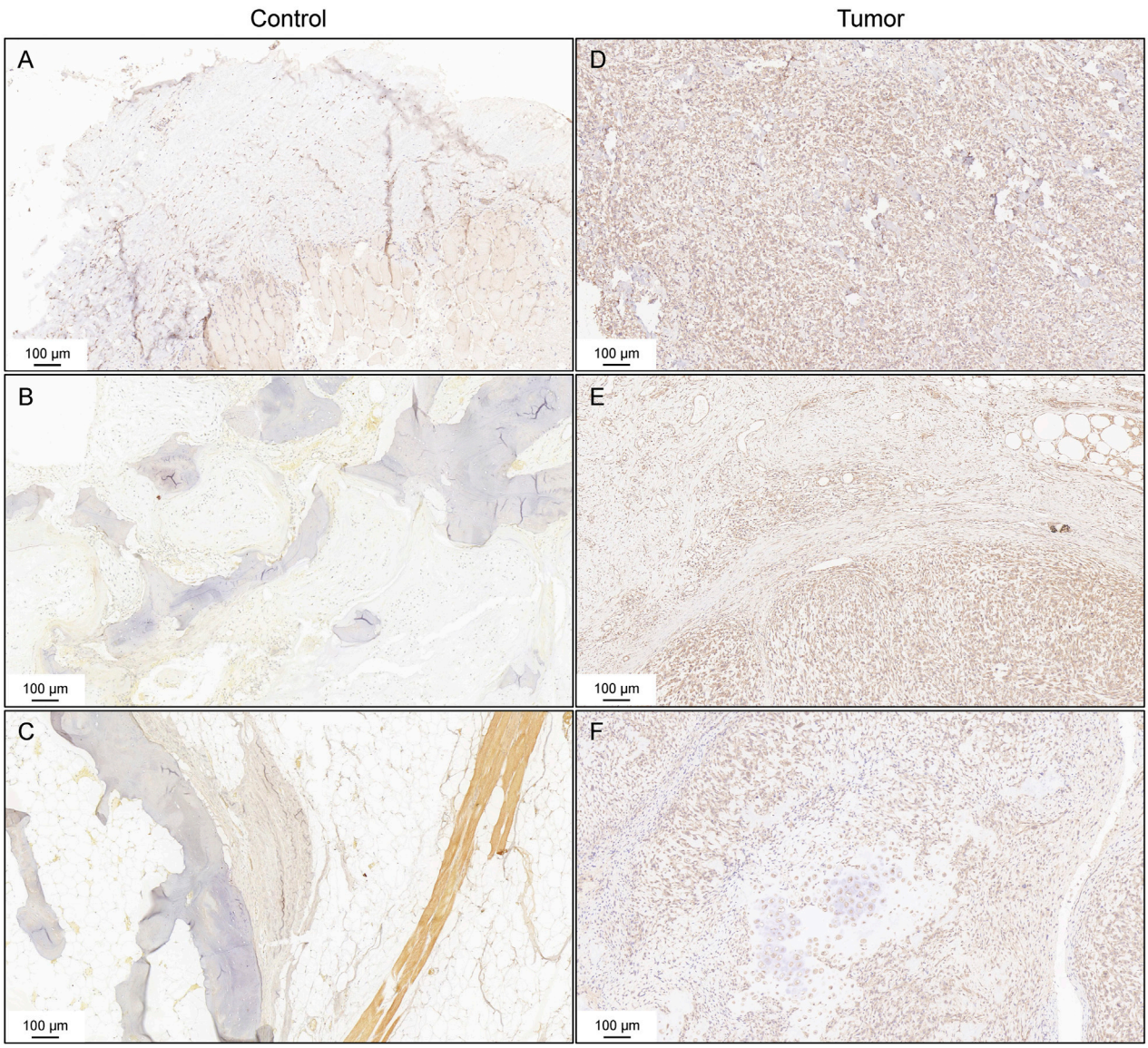


FIGURE 10
Immunohistochemistry staining for TUBB. Negative controls of the expression of TUBB **(A–C)**. IHC staining of osteosarcoma tissues **(D,E)**. IHC staining of chondrogenic sarcoma tissues **(F)**.

in some cancers, and immune infiltration analyses showed it may perhaps induce immune escape in tumor cells. Genes related to TUBB were identified too. Besides, we explored how TUBB could potentially affect responses to chemotherapies, in order to provide practical clues for clinical use. Compounds that can target and reverse the tumorigenic functions of TUBB. All results generated by multiple biological methods mutually pointed to the research value of TUBB in the future. And more experiments are wanted to verify our findings.

TUBB (Tubulin β class I gene) refers to a class of genes that can encode tubulin, a fundamental structure of microtubule cytoskeleton (Janakiraman et al., 2023). Microtubules are major components that dynamically control many vital functions of cells ranging from cell division to cell movement and vesicular transport (Gudimchuk and McIntosh, 2021). Nowadays, robust therapeutic targets of cancer treatments are widely and profoundly sought, among all kinds, microtubule-targeting agents (MTAs) are emerging as a time-proven anti-tumor chemical (Čermák et al., 2020). MTAs can be divided into two main categories: microtubule-stabilizing agents (MSAs) (Zhao et al., 2016) and microtubule-destabilizing agents (MDAs) (Borys et al., 2020), and they have been studied in many cancers (Karahalil et al., 2019; Khwaja et al., 2021; Anwar et al., 2022; Chen, 2023), epithelial ovarian cancer (Tymon-Rosario et al., 2021), non-small cell lung cancer (NSCLC) (Tagliamento et al., 2019), *etc.* Therefore, as the microtubule-encoding gene, TUBB's mediating functions in cancers should be noticed and clarified. MicroRNAs (miRNAs) are small endogenous non-coding RNAs that post-transcriptionally repress gene expression (Lu and Rothenberg, 2018). The miR-195 axis has been identified as a multipurpose bridge in regulating cancer processes including metastasis and chemoresistance. Yu et al. (2019) have demonstrated that TUBB is directly targeted by

miR-195, and participates in regulating the response of lung adenocarcinoma cells to MTAs. What's more, it has been experimentally verified that silencing LINC00665 contributes to melanoma cell viability decline, inhibited proliferation, migration, invasion, and cell cycle progression, and enhances apoptosis by regulating the miR-339-3p/TUBB axis (Liu Y. et al., 2022). However, the exact role of TUBB in pan-cancers has not been illustrated. Here we integrated multiple strategies and utilized pan-cancer data from public databases for further analysis. TUBB's underlying functions across cancers have been uncovered to some extent, providing new insights for cancer profiling in depth.

First of all, we determined the expression levels of TUBB mRNA and protein in tumor tissues compared to that in normal tissues across various cancers. Results showed that TUBB was found to be generally highly expressed in tumor cells more than normal cells. At transcriptional and protein level, we proved TUBB was much closer to be a risky factor for cancer. ROC curves showed high confidence to again confirm the results above. We conducted mutation analyses to depict the mutation profiles of mutation types, frequency, and sites in pan-cancer. Generally speaking, copy number variations were commonly seen in cancer, and these could in turn affect TUBB expressions. And, we exhibited TUBB mutation sites in a vivid way with 2D and 3D diagrams. Furthermore, we explained the importance of TUBB using the concept "z-score". We found that TUBB mRNA expressions were related to the z-score of 14 newly identified cancer-related function states. Among them, TUBB expressions was mostly positively related to the z-score of cell-cycle status, showing TUBB may participate in regulating the course of cell-cycle. Based on the z-scores of cell cycle, we analyzed their correlations with TUBB expressions in various cancers. Almost a positive tendency was observed in any cancer. In thymoma (THYM), the coefficient was the highest. Cell cycle is a highly controlled and regulated process enabling cell growth, duplication of genetic material, and cell division, aberrancy in its progression is one of the basic mechanisms underlying tumorigenesis, making regulators of the cell cycle machinery reasonable anticancer therapeutic targets (Suski et al., 2021; Liu J. et al., 2022; Matthews et al., 2022). Therefore, it was deduced that TUBB might be enrolled in modulating the process of cell cycle, and thus eventually have an effect on tumor cells. From immune infiltration analyses, we confirmed that TUBB plays an important role in regulating immunity across cancers. It was significantly negatively correlated with MHC, MHC, immune-inhibitor, immune-stimulator, and chemokines genes. Moreover, we utilized seven algorithms and found that in almost every cancer the TUBB mRNA expression level was positively related to the abundance of $CD4^+$ T helper 1 (Th1) cells, characterized by the production of proinflammatory cytokine interferon-gamma (IFN-γ), plays a central role in orchestrating cell-mediated immunity against tumor cells (Yang et al., 2020). They are responsible for generating effector and memory cytotoxic T lymphocytes (CTL) in facilitating immune responses, so $CD4^+$ Th1 cells are important targets in the field of tumor immunotherapy. In SARC, the TUBB mRNA expression level was negatively related to the abundance of B cells and $CD8^+$ T which are both strongly immune-related cells and pivotal targets for cancer therapies (Wennhold et al., 2019; Chow et al., 2022; Lundberg et al., 2022). From the single-cell analysis, TUBB was demonstrated to mainly originate from malignant cells and proliferative T cells. All results mutually indicated that TUBB was wicked and may induce immune escape in cancer, causing indifferent immune responses to cancer treatments, resulting in poor clinical outcomes at last. Based on it, we explored chemotherapeutic drug resistance on a deeper level to try to understand the rule behind lower responses to medications. Drug sensitivity data from three relevant databases were gleaned. From various angles, we analyzed correlations between TUBB expressions and drug sensitivity. TUBB played a vital role in responses to drugs. However, the exact role of TUBB remained unclear. Barrón-Gallardo et al. (2022) used to elucidate that in chemotherapy-resistant breast cancer patients, higher TUBB expressions were observed. As a consequence, more studies should be emphasized to clarify TUBB functions in medication guidance. Next, we found that Fasudil, W.13, AH.6809, and X4.5.dianilinophthalimide might reverse TUBB-mediated carcinogenic effects. These results may inspire relevant researchers. In addition, deeper explorations of TUBB mainly in SARC were conducted. We identified E2F transcription factor 1 (E2F1) as having a statistical relation with TUBB expressions in SARC. E2F1 has been reckoned as a tumor-promoting gene. From one pan-cancer analysis, the KAT2A/E2F1 complex promotes cell proliferation and metastasis by upregulating the UBE2C expression (Lin et al., 2022). An in SARC, the role of E2F1 has already been widely explored. As previously reported, E2F1 will promote Warburg effect and cancer progression via upregulating ENO2 expression in Ewing SARC (Jiang et al., 2022). What's more, MNK1 and MNK2 enforce expression of E2F1, FOXM1, and WEE1 to drive soft tissue sarcoma (Ke et al., 2021). In other types of cancer, like in hepatocellular carcinoma, scientists have proved that long non-coding RNA CDKN2B-AS1 could enhance tumor progression via the E2F1/G protein subunit alpha *Z*-axis (Tao et al., 2023). Targeting the E2F1/Rb/HDAC1 axis with the small molecule HR488B effectively inhibits colorectal cancer growth (Duan et al., 2023). Besides, CYCLINB1 protein had the most significant relation with TUBB mRNA expressions. CYCLINB1 has also been well-studied in multiple cancers. For example, Lv S et al. found that inhibiting CYCLINB1 resulted in suppressed proliferation, invasion, and epithelial mesenchymal transition of hepatocellular carcinoma cells and enhanced sensitivity to TRAIL-Induced apoptosis (Lv et al., 2020). Also, CYCLINB1 is a defined biomarker in esophageal squamous cell carcinoma (Li et al., 2023), penile squamous cell carcinoma (Tan et al., 2022), gastric cancer (Li et al., 2021), breast cancer (Liu et al., 2019), nasopharyngeal carcinoma (Xie et al., 2019), *etc.* As an important factor in so many cancers, our studies provided a new idea that TUBB may function in tumor formation by interacting with the CYCLINB1 protein.

In this study, we found that TUBB was generally differentially expressed between tumor tissues and normal tissues through a full-scale pan-cancer analysis, using a series of bioinformatics approaches. Correlation between TUBB expressions and clinical prognosis was uncovered in the heatmap, forest plots, and ROC curves, making our findings more reliable. TUBB was demonstrated as a potential independently working prognostic factor for most tumors. In particular, TUBB's risky role was emphasized. We confirmed that TUBB may be the reason why immune rejection and immune escape should be generated. Higher expressions of TUBB would result in less immune-related molecules in pan-cancer. Aiming at this, we further conducted drug resistance analysis and found that TUBB played an important role in regulating chemotherapy sensitivity across cancers. But this was controversial, reminding us to perform more and better research and development of drugs with safer and more effective clinical trials. At last, we selected SARC to study more. And we offered E2F1 as a

key gene interacting with TUBB. Moreover, CYCLINB1 was identified as another hub term in TUBB expressions. These results mutually provided novel insights in future studies on TUBB in cancers, and they are hoped to be helpful to further elucidate the explicit mechanisms of TUBB in cancer initiation and progression.

In a nutshell, using the multi-omics method, TUBB has been elucidated to be dysregulated in pan-cancers as a potential diagnostic marker. The article systematically describes the relationship between TUBB and clinical outcomes in pan-cancers. For the first time, the possibility of targeting TUBB with drugs and small molecules has been identified from multiple dimensions. Moreover, a large number of single-cell datasets were combined to identify cells expressing TUBB at high resolution. The pathway and metabolic disorders mediated by TUBB were validated, and the transcription factors regulating their expressions were identified using the CHIP-seq approach. TUBB expression was linked to multiple molecular subtypes, demonstrating the potential of stratified precision therapies.

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

ZZ: Data curation, Formal Analysis, Writing–original draft, Writing–review and editing. WZ: Formal Analysis, Writing–original draft, Resources, Visualization. SH: Investigation, Software, Writing–review and editing. TH: Data curation, Investigation, Writing–review and editing. XC: Conceptualization, Funding acquisition, Resources, Supervision, Validation, Writing–review and editing. YZ: Conceptualization, Investigation, Methodology, Project administration, Resources, 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 funded by the National Natural Science Foundation of China (Grant Number: 82204290), the Clinical Science Foundation of Anhui Medical University (Grant Number: 2021xkj129), and the Foundation by basic and clinical cooperative research promotion program of Anhui Medical University (Grant Number: 2022xkjT023).

Acknowledgments

The authors thank the reviewers for their valuable remarks.

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/fmolb.2024.1365655/full#supplementary-material

SUPPLEMENTARY FIGURE S1
TUBB mRNA levels between tumor and normal tissues using GEO data from different cancers.

SUPPLEMENTARY FIGURE S2
Differences in TUBB expressions in different tumor stages by Kruskal-Wallis Rank Sum test in ACC, ESCA, KIRC, KIRP, LIHC, SKCM, and STAD.

SUPPLEMENTARY FIGURE S3
ROC curves (AUC value > 0.7) demonstrated the diagnostic value of TUBB in distinguishing tumors based on the TCGA data.

SUPPLEMENTARY FIGURE S4
ROC curves (AUC value > 0.7) demonstrated the diagnostic value of TUBB in distinguishing tumors based on the TCGA-GTEx data.

SUPPLEMENTARY FIGURE S5
The difference of TIP scores between high and low expression groups of TUBB was calculated.

References

Alhammad, R. (2022). Bioinformatics identification of TUBB as potential prognostic biomarker for worse prognosis in erα-positive and better prognosis in erα-negative breast cancer. *Diagn. (Basel)* 12 (9), 2067. Published 2022 Aug 26. doi:10.3390/diagnostics12092067

Anwar, S., Shahwan, M., Hasan, G. M., Islam, A., and Hassan, M. I. (2022). Microtubule-affinity regulating kinase 4: a potential drug target for cancer therapy. *Cell. Signal* 99, 110434. doi:10.1016/j.cellsig.2022.110434

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 (Database issue), D991–D995. doi:10.1093/nar/gks1193

Barrón-Gallardo, C. A., Garcia-Chagollán, M., Morán-Mendoza, A. J., Delgadillo-Cristerna, R., Martínez-Silva, M. G., Aguilar-Lemarroy, A., et al. (2022). Transcriptomic analysis of breast cancer patients sensitive and resistant to chemotherapy: looking for overall survival and drug resistance biomarkers. *Technol. Cancer Res. Treat.* 21, 15330338211068965. doi:10.1177/15330338211068965

Borys, F., Joachimiak, E., Krawczyk, H., and Fabczak, H. (2020). Intrinsic and extrinsic factors affecting microtubule dynamics in normal and cancer cells. *Molecules* 25 (16), 3705. Published 2020 Aug 14. doi:10.3390/molecules25163705

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 (5), 401–404. doi:10.1158/2159-8290.CD-12-0095

Čermák, V., Dostál, V., Jelínek, M., Libusová, L., Kovář, J., Rösel, D., et al. (2020). Microtubule-targeting agents and their impact on cancer treatment. *Eur. J. Cell. Biol.* 99 (4), 151075. doi:10.1016/j.ejcb.2020.151075

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

Chen, Q. H. (2023). Crosstalk between microtubule stabilizing agents and prostate cancer. *Cancers (Basel)* 15 (13), 3308. Published 2023 Jun 23. doi:10.3390/cancers15133308

Chow, A., Perica, K., Klebanoff, C. A., and Wolchok, J. D. (2022). Clinical implications of T cell exhaustion for cancer immunotherapy. *Nat. Rev. Clin. Oncol.* 19 (12), 775–790. doi:10.1038/s41571-022-00689-z

Deng, M., Brägelmann, J., Kryukov, I., Saraiva-Agostinho, N., and Perner, S. (2017). FirebrowseR: an R client to the broad institute's Firehose pipeline. *Database (Oxford)* 2017, baw160. Published 2017 Jan 6. doi:10.1093/database/baw160

Dolgin, E. (2021). Cancer's new normal [published correction appears in Nat Cancer. 2022 Mar;3(3):372]. *Nat. Cancer* 2 (12), 1248–1250. doi:10.1038/s43018-021-00304-7

Duan, N., Hu, X., Qiu, H., Zhou, R., Li, Y., Lu, W., et al. (2023). Targeting the E2F1/Rb/HDAC1 axis with the small molecule HR488B effectively inhibits colorectal cancer growth. *Cell. Death Dis.* 14 (12), 801. Published 2023 Dec 7. doi:10.1038/s41419-023-06205-0

Field, J. J., Kanakkanthara, A., and Miller, J. H. (2014). Microtubule-targeting agents are clinically successful due to both mitotic and interphase impairment of microtubule function. *Bioorg Med. Chem.* 22 (18), 5050–5059. doi:10.1016/j.bmc.2014.02.035

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal* 6 (269), pl1. doi:10.1126/scisignal.2004088

Goodson, H. V., and Jonasson, E. M. (2018). Microtubules and microtubule-associated proteins. *Cold Spring Harb. Perspect. Biol.* 10 (6), a022608. Published 2018 Jun 1. doi:10.1101/cshperspect.a022608

Gudimchuk, N. B., and McIntosh, J. R. (2021). Regulation of microtubule dynamics, mechanics and function through the growing tip. *Nat. Rev. Mol. Cell. Biol.* 22 (12), 777–795. doi:10.1038/s41580-021-00399-x

Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., et al. (2018). Cell-of-Origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell.* 173 (2), 291–304.e6. doi:10.1016/j.cell.2018.03.022

Janakiraman, V., Sudhan, M., Alzahrani, K. J., Alshammeri, S., Ahmed, SSSJ, and Patil, S. (2023). Dynamics of TUBB protein with five majorly occurring natural variants: a risk of cortical dysplasia. *J. Mol. Model.* 29 (4), 100. Published 2023 Mar 16. doi:10.1007/s00894-023-05506-7

Jiang, X., Chen, Z., Zhu, J., Han, J., You, G., Li, Y., et al. (2022). E2F1 promotes Warburg effect and cancer progression via upregulating ENO2 expression in Ewing sarcoma. *Mol. Med. Rep.* 26 (1), 237. doi:10.3892/mmr.2022.12753

Karahalil, B., Yardım-Akaydin, S., and Nacak Baytas, S. (2019). An overview of microtubule targeting agents for cancer therapy. *Arh. Hig. Rada Toksikol.* 70 (3), 160–172. doi:10.2478/aiht-2019-70-3258

Ke, X. Y., Chen, Y., Tham, V. Y., Lin, R. Y. T., Dakle, P., Nacro, K., et al. (2021). MNK1 and MNK2 enforce expression of E2F1, FOXM1, and WEE1 to drive soft tissue sarcoma. *Oncogene* 40 (10), 1851–1867. doi:10.1038/s41388-021-01661-4

Khwaja, S., Kumar, K., Das, R., and Negi, A. S. (2021). Microtubule associated proteins as targets for anticancer drug development. *Bioorg Chem.* 116, 105320. doi:10.1016/j.bioorg.2021.105320

Lee, E., Chuang, H. Y., Kim, J. W., Ideker, T., and Lee, D. (2008). Inferring pathway activity toward precise disease classification. *PLoS Comput. Biol.* 4 (11), e1000217. doi:10.1371/journal.pcbi.1000217

Li, F. N., Zhang, Q. Y., Li, O., Liu, S. L., Yang, Z. Y., Pan, L. J., et al. (2021). ESRRA promotes gastric cancer development by regulating the CDC25C/CDK1/CyclinB1 pathway via DSN1. *Int. J. Biol. Sci.* 17 (8), 1909–1924. Published 2021 May 10. doi:10.7150/ijbs.57623

Li, H., Yang, C., Chen, K., and Sun, M. (2023). Expression significance of Emi1, UBCH10 and CyclinB1 in esophageal squamous cell carcinoma. *Pathol. Oncol. Res.* 29, 1611081. Published 2023 Apr 24. doi:10.3389/pore.2023.1611081

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 (W1), W509-W514–W514. doi:10.1093/nar/gkaa407

Lin, S., Qiu, L., Liang, K., Zhang, H., Xian, M., Chen, Z., et al. (2022). *KAT2A/E2F1* promotes cell proliferation and migration via upregulating the expression of *UBE2C* in pan-cancer. *Genes. (Basel)* 13 (10), 1817. Published 2022 Oct 8. doi:10.3390/genes13101817

Liu, B., Liu, Y., Wang, Y., Xie, C., Gan, M., Han, T., et al. (2019). CyclinB1 deubiquitination by USP14 regulates cell cycle progression in breast cancer. *Pathol. Res. Pract.* 215 (10), 152592. doi:10.1016/j.prp.2019.152592

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 (1), bbac558. doi:10.1093/bib/bbac558

Liu, J., Peng, Y., and Wei, W. (2022b). 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

Liu, Y., Ma, S., Ma, Q., and Zhu, H. (2022a). Silencing LINC00665 inhibits cutaneous melanoma *in vitro* progression and induces apoptosis via the miR-339-3p/TUBB. *J. Clin. Lab. Anal.* 36 (9), e24630. doi:10.1002/jcla.24630

Lopes, D., and Maiato, H. (2020). The tubulin code in mitosis and cancer. *Cells* 9 (11), 2356. Published 2020 Oct 26. doi:10.3390/cells9112356

Lu, T. X., and Rothenberg, M. E. (2018). MicroRNA. *J. Allergy Clin. Immunol.* 141 (4), 1202–1207. doi:10.1016/j.jaci.2017.08.034

Lundberg, A., Li, B., and Li, R. (2022). B cell-related gene signature and cancer immunotherapy response. *Br. J. Cancer* 126 (6), 899–906. doi:10.1038/s41416-021-01674-6

Lv, S., Ning, H., Li, Y., Wang, J., Jia, Q., and Wen, H. (2020). Inhibition of cyclinB1 suppressed the proliferation, invasion, and epithelial mesenchymal transition of hepatocellular carcinoma cells and enhanced the sensitivity to TRAIL-induced apoptosis. *Onco Targets Ther.* 13, 1119–1128. Published 2020 Feb 5. doi:10.2147/OTT.S225202

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

Matthews, H. K., Bertoli, C., and de Bruin, R. A. M. (2022). Cell cycle control in cancer. *Nat. Rev. Mol. Cell. Biol.* 23 (1), 74–88. doi:10.1038/s41580-021-00404-3

Miller, L. M., Xiao, H., Burd, B., Horwitz, S. B., Angeletti, R. H., and Verdier-Pinard, P. (2010). Methods in tubulin proteomics. *Methods Cell. Biol.* 95, 105–126. doi:10.1016/S0091-679X(10)95007-3

Rees, M. G., Seashore-Ludlow, B., Cheah, J. H., Adams, D. J., Price, E. V., Gill, S., et al. (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.* 12 (2), 109–116. doi:10.1038/nchembio.1986

Reinhold, W. C., Wilson, K., Elloumi, F., Bradwell, K. R., Ceribelli, M., Varma, S., et al. (2023). CellMinerCDB: NCATS is a web-based portal integrating public cancer cell line databases for pharmacogenomic explorations. *Cancer Res.* 83 (12), 1941–1952. doi:10.1158/0008-5472.CAN-22-2996

Shankavaram, U. T., Varma, S., Kane, D., Sunshine, M., Chary, K. K., Reinhold, W. C., et al. (2009). CellMiner: a relational database and query tool for the NCI-60 cancer cell lines. *BMC Genomics* 10, 277. Published 2009 Jun 23. doi:10.1186/1471-2164-10-277

Shao, H., Ge, M., Zhang, J., Zhao, T., and Zhang, S. (2022). Osteoclasts differential-related prognostic biomarker for osteosarcoma based on single cell, bulk cell and gene expression datasets. *BMC Cancer* 22 (1), 288. Published 2022 Mar 17. doi:10.1186/s12885-022-09380-z

Sun, D., Wang, J., Han, Y., Dong, X., Ge, J., Zheng, R., et al. (2021). TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. *Nucleic Acids Res.* 49 (D1), D1420–D1430. doi:10.1093/nar/gkaa1020

Suski, J. M., Braun, M., Strmiska, V., and Sicinski, P. (2021). Targeting cell-cycle machinery in cancer. *Cancer Cell.* 39 (6), 759–778. doi:10.1016/j.ccell.2021.03.010

Szeto, G. L., and Finley, S. D. (2019). Integrative approaches to cancer immunotherapy. *Trends Cancer* 5 (7), 400–410. doi:10.1016/j.trecan.2019.05.010

Tagliamento, M., Genova, C., Rossi, G., Coco, S., Rijavec, E., Dal Bello, M. G., et al. (2019). Microtubule-targeting agents in the treatment of non-small cell lung cancer: insights on new combination strategies and investigational compounds. *Expert Opin. Investig. Drugs* 28 (6), 513–523. doi:10.1080/13543784.2019.1627326

Tan, X., Yuan, G., Wang, Y., Zou, Y., Luo, S., Han, H., et al. (2022). RAB20 promotes proliferation via G2/M phase through the chk1/cdc25c/cdc2-cyclinB1 pathway in penile squamous cell carcinoma. *Cancers (Basel)* 14 (5), 1106. Published 2022 Feb 22. doi:10.3390/cancers14051106

Tao, Z. G., Yuan, Y. X., and Wang, G. W. (2023). Long non-coding RNA CDKN2B-AS1 promotes hepatocellular carcinoma progression *via* E2F transcription factor 1/G protein subunit alpha Z axis. *World J. Gastrointest. Oncol.* 15 (11), 1974–1987. doi:10.4251/wjgo.v15.i11.1974

Tymon-Rosario, J., Adjei, N. N., Roque, D. M., and Santin, A. D. (2021). Microtubule-interfering drugs: current and future roles in epithelial ovarian cancer treatment. *Cancers (Basel)* 13 (24), 6239. Published 2021 Dec 12. doi:10.3390/cancers13246239

Wang, S., Xiong, Y., Zhao, L., Gu, K., Li, Y., Zhao, F., et al. (2022). UCSCXenaShiny: an R/CRAN package for interactive analysis of UCSC Xena data. *Bioinformatics* 38 (2), 527–529. doi:10.1093/bioinformatics/btab561

Wennhold, K., Shimabukuro-Vornhagen, A., and von Bergwelt-Baildon, M. (2019). B cell-based cancer immunotherapy. *Transfus. Med. Hemother* 46 (1), 36–46. doi:10.1159/000496166

Wordeman, L., and Vicente, J. J. (2021). Microtubule targeting agents in disease: classic drugs, novel roles. *Cancers (Basel)* 13 (22), 5650. Published 2021 Nov 12. doi:10.3390/cancers13225650

Wu, J., and Akhmanova, A. (2017). Microtubule-organizing centers. *Annu. Rev. Cell. Dev. Biol.* 33, 51–75. doi:10.1146/annurev-cellbio-100616-060615

Xie, X., Lin, W., Zheng, W., Chen, T., Yang, H., Sun, L., et al. (2019). Downregulation of G2/mitotic-specific cyclinB1 triggers autophagy via AMPK-ULK1-dependent signal pathway in nasopharyngeal carcinoma cells. *Cell. Death Dis.* 10 (2), 94. Published 2019 Jan 30. doi:10.1038/s41419-019-1369-8

Yang, W., Chen, X., and Hu, H. (2020). $CD4^+$ T-cell differentiation *in vitro*. *Methods Mol. Biol.* 2111, 91–99. doi:10.1007/978-1-0716-0266-9_8

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 (Database issue), D955–D961. doi:10.1093/nar/gks1111

Yu, X., Zhang, Y., Wu, B., Kurie, J. M., Pertsemlidis, A., et al. (2019). The miR-195 Axis regulates chemoresistance through TUBB and lung cancer progression through BIRC5. *Mol. Ther. Oncolytics* 14, 288–298. Published 2019 Aug 6. doi:10.1016/j.omto.2019.07.004

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 (D1), D900–D908. doi:10.1093/nar/gky939

Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., et al. (2014). Proteogenomic characterization of human colon and rectal cancer. *Nature* 513 (7518), 382–387. doi:10.1038/nature13438

Zhao, Y., Mu, X., and Du, G. (2016). Microtubule-stabilizing agents: new drug discovery and cancer therapy. *Pharmacol. Ther.* 162, 134–143. doi:10.1016/j.pharmthera.2015.12.006

Glossary

TME	tumor microenvironment
TCGA	the Cancer Genome Atlas
TUBB	β-tubulin
PTMs	post-translational modifications
TIME	tumor immune microenvironment
GTEx	the Genotype-Tissue Expression project
AUC	area under curve
IHC	immunohistochemical
OS	overall survival
DSS	disease-specific survival
PFI	progression free interval
DFI	disease free interval
KM	Kaplan-Meier
SCNA	Somatic copy number alteration
CNAs	copy-number alterations
GSEA	Gene Set Enrichment Analysis
SARC	sarcoma
TME	tumor microenvironment
GSVA	Gene Set Variation Analysis
NCI	National Cancer Institute
HR	Hazard ratio
CI	confidence interval
THYM	thymoma
IC50	half maximal inhibitory concentration
XSum	Extreme Summarization
MTAs	microtubule-targeting agents
MSAs	microtubule-stabilizing agents
MDAs	microtubule-destabilizing agents
NSCLC	non-small cell lung cancer
miRNAs	microRNAs
IFN-γ	interferon-gamma
CTL	cytotoxic T lymphocyte

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 22 May 2024
DOI 10.3389/fmolb.2024.1380384

Check for updates

OPEN ACCESS

EDITED BY
Lin Zhang,
Monash University, Australia

REVIEWED BY
Yash Gupta,
Penn State Milton S. Hershey Medical Center, United States
Udayan Bhattacharya,
NewYork-Presbyterian, United States

*CORRESPONDENCE
Zhe Lu,
zhelu@hznu.edu.cn

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

RECEIVED 01 February 2024
ACCEPTED 22 April 2024
PUBLISHED 22 May 2024

CITATION
Xie B, Wu T, Hong D and Lu Z (2024), Comprehensive landscape of junctional genes and their association with overall survival of patients with lung adenocarcinoma.
Front. Mol. Biosci. 11:1380384.
doi: 10.3389/fmolb.2024.1380384

COPYRIGHT
© 2024 Xie, Wu, Hong and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Comprehensive landscape of junctional genes and their association with overall survival of patients with lung adenocarcinoma

Bin Xie[1†], Ting Wu[1†], Duiguo Hong[2] and Zhe Lu[3,4*]

[1]School of Information Science and Technology, Hangzhou Normal University, Hangzhou, China, [2]Jincheng Community Health Service Center, Hangzhou, China, [3]Key Laboratory of Aging and Cancer Biology of Zhejiang Province, Hangzhou Normal University, Hangzhou, China, [4]School of Basic Medicine, Hangzhou Normal University, Hangzhou, China

Objectives: Junctional proteins are involved in tumorigenesis. Therefore, this study aimed to investigate the association between junctional genes and the prognosis of patients with lung adenocarcinoma (LUAD).

Methods: Transcriptome, mutation, and clinical data were retrieved from The Cancer Genome Atlas (TCGA). "Limma" was used to screen differentially expressed genes. Moreover, Kaplan–Meier survival analysis was used to identify junctional genes associated with LUAD prognosis. The junctional gene-related risk score (JGRS) was generated based on multivariate Cox regression analysis. An overall survival (OS) prediction model combining the JGRS and clinicopathological properties was proposed using a nomogram and further validated in the Gene Expression Omnibus (GEO) LUAD cohort.

Results: To our knowledge, this study is the first to demonstrate the correlation between the mRNA levels of 14 junctional genes (*CDH15, CDH17, CDH24, CLDN6, CLDN12, CLDN18, CTNND2, DSG2, ITGA2, ITGA8, ITGA11, ITGAL, ITGB4,* and *PKP3*) and clinical outcomes of patients with LUAD. The JGRS was generated based on these 14 genes, and a higher JGRS was associated with older age, higher stage levels, and lower immune scores. Thus, a prognostic prediction nomogram was proposed based on the JGRS. Internal and external validation showed the good performance of the prediction model. Mechanistically, JGRS was associated with cell proliferation and immune regulatory pathways. Mutational analysis revealed that more somatic mutations occurred in the high-JGRS group than in the low-JGRS group.

Conclusion: The association between junctional genes and OS in patients with LUAD demonstrated by our "TCGA filtrating and GEO validating" model revealed a new function of junctional genes.

KEYWORDS

lung adenocarcinoma, junctional genes, risk score, prognosis, overall survival

1 Introduction

Lung cancer has the highest mortality rate of all cancers. There were 2, 200, 000 incidences and 1,800,000 deaths from lung cancer in 2020 worldwide, accounting for more deaths than liver and stomach cancers combined (Leiter et al., 2023). The overall 5-year survival rate of patients with lung cancer is approximately 15%, and patients with distant metastases have an even lower survival rate of 8% (Sung et al., 2021). Most lung cancers form in the epithelial cells lining the respiratory tract. Non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) are the two main types of lung carcinoma (Zhao et al., 2021). Lung adenocarcinoma (LUAD) is the most common type of NSCLC, accounting for more than 40% of all lung cancer cases (Denisenko et al., 2018). Moreover, approximately 70% of patients with NSCLC have inoperable local or metastatic tumors at the time of diagnosis (Katzel et al., 2009). Hence, the identification of reliable prognostic biomarkers for high-risk patients with LUAD is important for designing treatment strategies.

Most studies have identified biomarkers for diagnosis and/or prognosis. The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases have been widely used to predict cancer prognoses. For example, differentially methylated sites (DMSs) were selected from a TCGA-LUAD cohort, used to construct a robust DMS-based prognostic signature, and validated in a GEO cohort (Wang X. et al., 2021). Tumor mutation burden (TMB) can affect immune infiltrates and alter gene expression. A previous study stratified patients with LUAD into higher- and lower-TMB subgroups, screened nine immune genes, and used a prognostic signature based on these nine immune genes to predict patient prognoses (Zhao et al., 2021). Other studies have used similar strategies to screen biomarkers and construct LUAD prognostic nomograms based on either a group of specific genes, a gene family, or biological/physiological factors, such as lncRNA (Zeng et al., 2023), pyroptosis (Song et al., 2021), T-cell marker genes (Peng et al., 2024), tumor microenvironment-related genes (Li et al., 2023), integrin genes (Wang Y. et al., 2021; Zhang S. et al., 2023) and oxidative stress (Qian et al., 2023). Although most studies used the TCGA and/or GEO databases and incorporated the genomic profiles and clinical information to construct prediction models, some used databases such as SEER without incorporating genomic information (Zuo et al., 2021) whereas a few others were based on cohorts from local hospitals (Sun et al., 2021). Recent studies that have focused on predicting the prognosis of patients with LUAD are summarized in Supplementary Table S1. The accuracy and efficiency of some models in previous studies were compromised. Some studies only provided calibration curves, but without the area under the curve (AUC) or C-index values (Li et al., 2023; Peng et al., 2024). Similarly, others only identified independent prognostic factors but did not propose a prognostic prediction model (Zhang Z. et al., 2023). Furthermore, some studies only included a limited number of patient samples for the training and testing cohorts. External verification ensures the universality of the prognostic prediction model. However, some studies did not conduct external validation. Therefore, we aimed to develop a more accurate and reliable overall survival (OS) prediction model for patients with LUAD.

Lung cancer generally develops in cells lining air passages, predominantly originating from epithelial cells. Junctional proteins play major roles in these cells, mainly regulating cell-cell adhesion and adhering cells to the extracellular matrix. Thus, they can seal cellular sheets and control the paracellular flux of ions and solutes (Buckley and St Johnston, 2022; Kuo et al., 2022; Troyanovsky, 2023). Potential biomarkers of junctional genes have been identified to monitor diseases, including cancers (Wang D. W. et al., 2022; Hashimoto and Oshima, 2022; Parrish et al., 2022; Lin et al., 2023; Nehme et al., 2023). Altered expression of junctional genes can disrupt cell-cell adhesion, which is an initial step in cancer cell invasion. Consequently, defective cell-cell adhesion allows extra nutrients and growth factors to flow from the luminal fluid and facilitates aggressive tumor growth. During intravasation and extravasation, cancer cells disrupt cell-cell junctions and transverse through the paracellular pathways of endothelial cells that serve as a barrier for cancer cells (Wang and Liu, 2022; Wautier and Wautier, 2022). Therefore, junctional genes play important roles in tumorigenesis and cancer progression. However, the predictive value of junctional genes for the prognosis of patients with LUAD has not been studied. We used Limma and Kaplan-Meier plot analyses to screen prognosis-related junctional genes based on the genomic and clinical data from TCGA. Furthermore, we constructed a nomogram for predicting the OS of patients with LUAD and validated it in four GEO datasets.

2 Materials and methods

2.1 Retrieval of analytical data

For the TCGA-LUAD cohort, mRNA expression data and clinical information of patients with LUAD were acquired from https://portal.gdc.cancer.gov/. Originally, the data were obtained from 617 patients. Patients without overall survival data or events were subsequently excluded. We included 463 patients with LUAD and 59 healthy lung tissues as controls. The database contains 105 members of the junctional gene family, as listed in Supplementary Table S2. Four validation datasets were acquired from GEO (https://www.ncbi.nlm.nih.gov/geo/): GSE17538, GSE31210, GSE37745, and GSE72094. Each GEO dataset included 232, 226, 106, 398 patients, respectively. Among them, GSE17538 is a colon cancer cohort, which was used here as an extra external validation cohort. The rest three GSE cohorts are LUAD cohorts. Clinical information and mRNA expression data were extracted from these datasets.

2.2 Screening of differentially expressed genes (DEGs) using "Limma"

TCGA data comprised 24,987 genes with RNASeq data. The "Limma" package (version 3.52.4) in R (version 4.2.1, R Foundation for Statistical Computing, Vienna, Austria) was used to screen genes with differential expression in LUAD tissues compared to normal lung tissues. DEGs with |log 2-fold change (log2FC)| >1 and false discovery rate (FDR) <0.05 were considered significant, and volcano plots were constructed. Heatmaps were plotted using the "pheatmap" package in R (version 1.0.12). The "ggpubr" package (version 0.6.0) was used to draw a boxplot for observing the

differential expression of the selected DEGs between LUAD and normal lung tissues.

2.3 Kaplan–Meier survival analysis

Kaplan-Meier plots were constructed to identify predictive DEGs for the OS of patients with LUAD. First, the "survival" (version 3.5.7) and "survminer" (version 0.4.9) packages were used to define the optimal cutoff point and draw survival curves for each low or high gene expression group. Log-rank tests were subsequently conducted to assess the predictive potential of junctional genes for the survival probability of patients with LUAD. Genes with $p < 0.01$ were considered to have significant predictive value and subjected to subsequent analyses.

2.4 Construction of the junctional genes-related risk score (JGRS)

Genes with $p < 0.01$ were selected from Kaplan-Meier plots and incorporated into multivariate Cox regression to develop the JGRS. For each gene, expression higher and lower than the cut-off point was designated as 1 and 0, respectively. The following formula was used to calculate the risk score of each patient with LUAD:

$$\mathrm{JGRS} = \sum \beta i \times Expi$$

Where βi represents the coefficient of each gene and Expi represents the designated gene expression value, which was either 1 or 0. The association between JGRS and eight clinical characteristics, including sex, age, *p*-stage, tumor (T), node (N), metastasis (M) stage, and immune and stromal scores, was calculated. The "estimate" package (version 1.0.13) in R was used to obtain the immune and stromal scores. The Wilcoxon rank-sum test was used for paired comparisons, and $p < 0.05$ was considered significant. Patients were then stratified into high- and low-JGRS groups according to the JGRS cutoff point, and Kaplan-Meier curves were plotted to compare survival between the two groups.

2.5 Development and assessment of the nomogram

OS is the life span of patients upon pathological diagnosis until the day of death or the last follow-up. Here, patients were censored if they were alive or had no adverse events at the last follow-up visit. Nomograms are commonly employed to model cancer prognosis by combining all predictors. In the present study, age, sex, and *p*-stages were defined as continuous, binary, and multiple categorical variables, respectively. Each predictor contributed to a score, and then a final total point was obtained by summing all the contributors and scaling to the axis of the probabilities of survival to predict the 1-, 2-, and 3-year OS probabilities (Iasonos et al., 2008). The area under the curve (AUC) by "timeROC" package (version 0.4), C-index by "simplevis" package (version 7.0.0), calibration curves by "rms" package (version 6.5.0), and decision curve analysis (DCA) by "ggDCA" package (version 1.2) were used to determine the effectiveness of the nomogram. The nomogram was further validated using four GEO cohorts.

TABLE 1 Clinical characteristics of patients with LUAD in the TCGA cohort.

Characteristics	TCGA cohort (*N* = 463)	
	N	%
Sex		
Men	213	46
Women	250	54
Age		
Mean (SD)	65.0	10.1
Median [Min, Max]	65.0	[33.0, 88.0]
Na	10	2.2
T stage		
T1	162	35
T2	243	52.5
T3	40	8.65
T4	15	3.2
TX	3	0.65
N stage		
N1	320	69.1
NX	143	30.9
M stage		
M0	304	65.7
M1	20	4.3
MX	139	30
Pathological stage		
Stage I	259	56
Stage II	108	23.3
Stage III	68	14.7
Stage IV	21	4.5
Na	7	1.5

2.6 Gene set enrichment analysis (GSEA)

GSEA is an efficient analytical method that focuses on two opposing biological states to determine statistically significant differences in biological pathways (Subramanian et al., 2007). The "clusterProfiler" package (version 4.4.4) was used to conduct GSEA to identify functionally relevant pathways regulated by JGRS; $p < 0.05$ was considered statistically significant.

2.7 Analysis of the tumor immune microenvironment (TIME)

CIBERSORT and TIMER methods based on the "IOBR" R package (version 0.99.9) were used to evaluate the infiltration condition of immune cells. The results were visualized using the "ggplot2" package (version 3.4.2). Furthermore, the correlation between the expression of each gene input into the JGRS formula and abundance of immune cells

were analyzed through "xCell" package (version 1.1.0) and demonstrated using "ggplot2" package (version 3.4.2).

2.8 Analysis of the genetic mutation status in the low- and high-JGRS groups

Somatic mutations in the TCGA-LUAD cohort, including nonsynonymous and synonymous mutations, were downloaded from https://portal.gdc.cancer.gov/. Significantly differentially mutated genes ($p < 0.05$) between the low- and high-JGRS groups were screened, and the correlations between these mutated genes were analyzed using "maftools" (version 2.14.0). Only genes mutated more than 30 times in at least one group were considered. The statistical test for the frequency of mutations was evaluated through a one-sided z-test and two-sided Chi-square test; $p < 0.05$ was considered significant.

3 Results

3.1 Clinical information of LUAD cohorts and the analytical scheme of the study

Clinical and mRNA expression data were extracted from the TCGA database, including 59 normal and 463 LUAD samples. The median age of patients with LUAD in the TCGA cohort was 65 years, and all patients were between 33 and 88 years of age. Among them, 213 patients were men (46%) and 250 were women (54%). Information on the tumor (T), node (N), as well as metastasis (M) and *p*-stages is shown in Table 1. Clinical information of the four GEO cohorts is provided in Supplementary Table S3. We first screened differentially expressed junctional genes between normal lung and LUAD tissues. The genes associated with OS of patients were used to establish the JGRS. Subsequently, we proposed a nomogram incorporating the JGRS and corresponding clinical parameters and further validated it in four GEO cohorts. In addition, the characteristics of the biological pathways, immune infiltration, and mutational status between the low- and high-JGRS groups were analyzed. The workflow for screening the potential junctional gene prognostic panel is shown in Figure 1. DEGs, heat maps, Kaplan-Meier survival curves, nomogram interactive line diagrams, ROC curves, calibration curves, GSEA, and gene mutation status analyses were performed. The corresponding results were generated using RStudio (version 4.2.1).

3.2 Identification of DEGs

To identify the DEGs associated with OS of patients with LUAD, we first compared the mRNA expression profiles of normal lung and LUAD tissues in the TCGA cohort. Of the 24,987 genes identified, 105 were junctional genes (Supplementary Table S2). Moreover, 1716 and 1,667 genes whose expression was upregulated and downregulated, respectively, were identified; the volcano plot is shown in Figure 2A. A heat map of the expression of 105 junctional genes in normal lung and LUAD tissues was plotted (Figure 2B). Of the 3,383 genes with altered expression, 28 were junctional genes, as shown in Figure 2C. These junctional genes showed significant differences in expression between normal lung and LUAD tissues (Supplementary Figure S1). Of the 28 junctional genes identified, the expression of 17 (*CDH15, CLDN10, CDH17, CTNND2, ITGA11, CLDN9, CLDN3, CLDN4, CLDN6, CDH24, DSP, CDH3, ITGB4, PKP3, ITGA2, CLDN12,* and *DSG2*) was significantly upregulated in LUAD tissues, whereas that of 11 genes (*JAM3, CDH13, ITGA8, CDH5, JAM2, ITGAL, CLDN18, CLDN5, CDH19, DST*, and *ITGA10*) was significantly downregulated in LUAD tissues. The 28 junctional genes were then subjected to Kaplan-Meier survival analysis to screen for genes associated with the OS of patients with LUAD. Our analyses revealed 14 junctional genes, namely *CDH15, CDH17, CDH24, CLDN6, CLDN12, CLDN18, CTNND2, DSG2, ITGA2, ITGA8, ITGA11, ITGAL, ITGB4*, and *PKP3*, whose expression was significantly associated with OS (Figures 3A–N), indicating that they were valuable prognostic predictors. Additionally, we used four external databases and the samples from our local hospital to confirm the mRNA and protein expression of these 14 junctional genes. Among the four GSE validation cohorts, GSE31210 cohort contains the mRNA expression information of both LUAD tissues and normal lung tissues, so we compared the mRNA expression levels of 14 junctional genes in LUAD tissues with normal lung tissues. As shown in Supplementary Figure S2, the RNA expression changes of 14 junctional genes all coincided well with these in the TCGA cohort. Then, the clinical Proteomic Tumor Analysis Consortium (CPTAC) database was used to explore protein expression levels in LUAD tissues. The results showed that the protein expression levels of CDH15, DSG2, ITGA11, and PKP3 were significantly upregulated, whereas these of CLDN18, ITGA8, and ITGAL were significantly downregulated in LUAD tissues compared with these in normal lung tissues. The protein expression changes of the above seven junctional genes in LUAD tissues coincided well with their mRNA expression changes in the TCGA and GSE31210 cohorts. There were no significant protein expression changes of CDH17, CTNND2, ITGA2, and ITGB4 and no data of CDH24, CLDN6, and CLDN12 protein expression in the CPTAC database (Supplementary Figure S3). Furthermore, we compared the mRNA expression levels of 14 junctional genes in normal lung epithelial cells-ciliated cells with these in the lung cancer cells, using the data from The Human Protein Atlas database (HPA, www.proteinatlas.org). LUAD originates mainly from the epithelium of the bronchi, and ciliated cells are the predominate epithelial cells in the respiratory tract (Li et al., 2021). We used the average expression data of a total of 232 lung cancer cell lines available on the website. It was found that except for *CTNND2* and *ITGA2*, the mRNA expression changes of the rest 12 junctional genes all coincided well with their changes in TCGA and GSE31210 cohorts in that the mRNA expression levels of *CDH15*, *CDH17*, *CDH24*, *CLDN6*, *CLDN12*, *DSG2*, *ITGA11*, *ITGB4*, and *PKP3* were increased and the mRNA expression levels of *CLDN18*, *ITGA8*, and *ITGAL* were decreased in 232 lung cancer cell lines compared with these in ciliated cells (Supplementary Figure S4).

We also collected 9 pairs of LUAD tumor samples and their corresponding adjacent normal tissues from the affiliated hospital of Hangzhou Normal University with subjects' informed consent and tested the mRNA expression levels of the 14 junctional genes

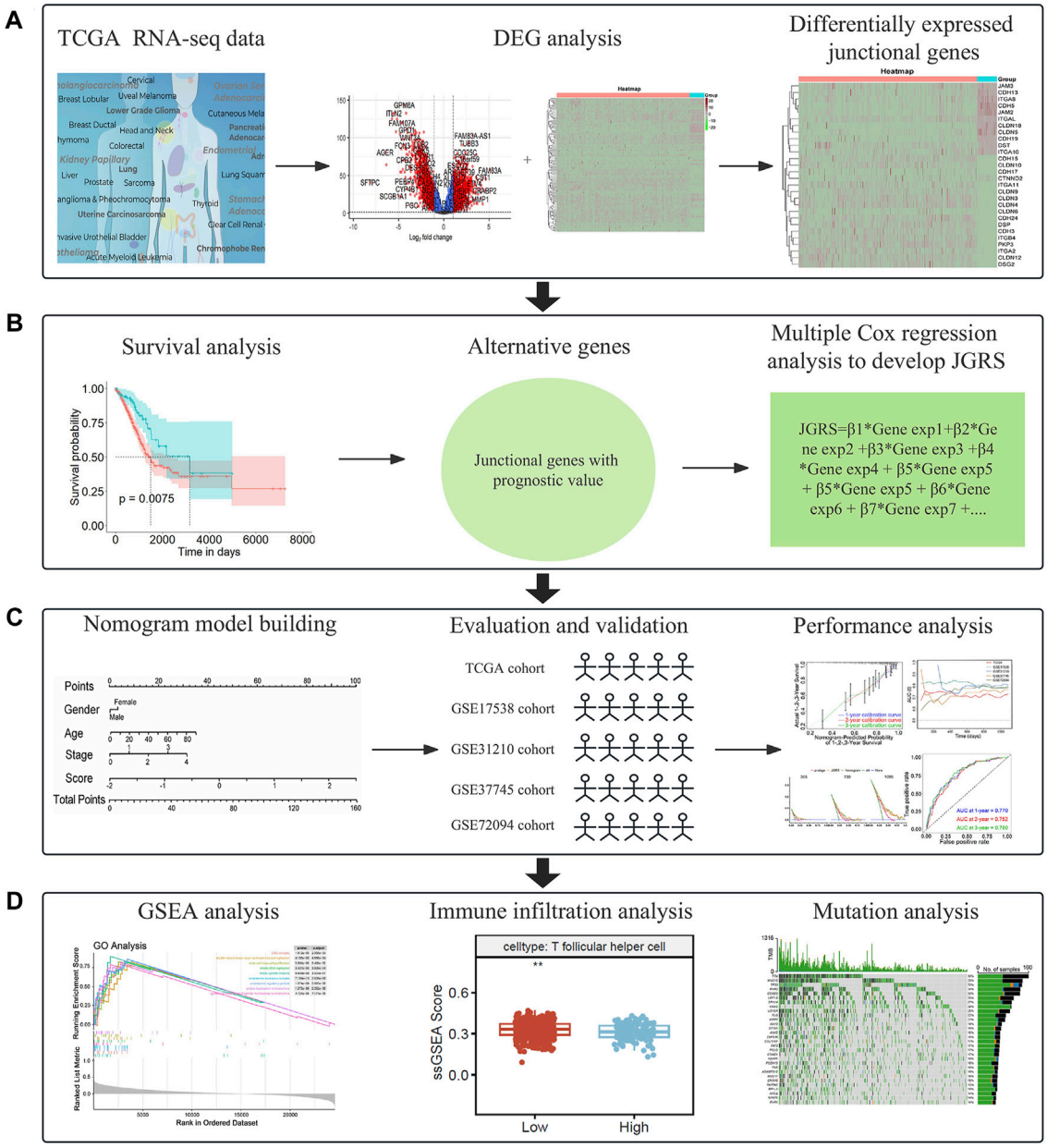


FIGURE 1
A flowchart of the study. **(A)** Screening of differentially expressed junctional genes between normal lung and LUAD tissues. **(B)** Development of a JGRS incorporating junctional genes with prognostic value. **(C)** Establishment, evaluation, and validation of the nomogram. **(D)** Functional analysis of the mechanisms behind the different JGRS.

(primers used were provided in Supplementary Table S4). As depicted in Supplementary Figure S5, consistent with the TCGA and GSE31210 cohorts, the mRNA expression levels of *CLDN12* and *ITGA11* were significantly increased and the mRNA expression levels of *CLDN18*, *CTNND2*, *ITGA8*, and *ITGAL* were significantly decreased in LUAD tissues compared with these in normal lung tissues. For *CDH15*, *CDH17*, *CLDN6*, *DSG2*, *ITGA2*, and *ITGB4*, although their mRNA expression changes did not reach to significant differences, the trends of changes were consistent with these in the TCGA and GSE31210 cohorts. It maybe because the number of our patient samples were not enough, and increasing the sample size may improve the results.

3.3 Development of the JGRS and its association with clinical characteristics

The prognostic risk score was computed as follows:

$$\mathrm{JGRS} = \sum Coef_i \times Exp_i$$

Multivariate Cox regression analysis incorporating the 14 screened junctional genes was used to generate regression coefficients. The final risk model was: JGRS = −0.39599 × *CDH15* + 0.44858 × *CDH17* + 0.20612 × *CDH24* + 0.58748 × *CLDN6* − 0.05628 × *CLDN18* − 0.46681 × *CTNND2* + 0.22880 ×

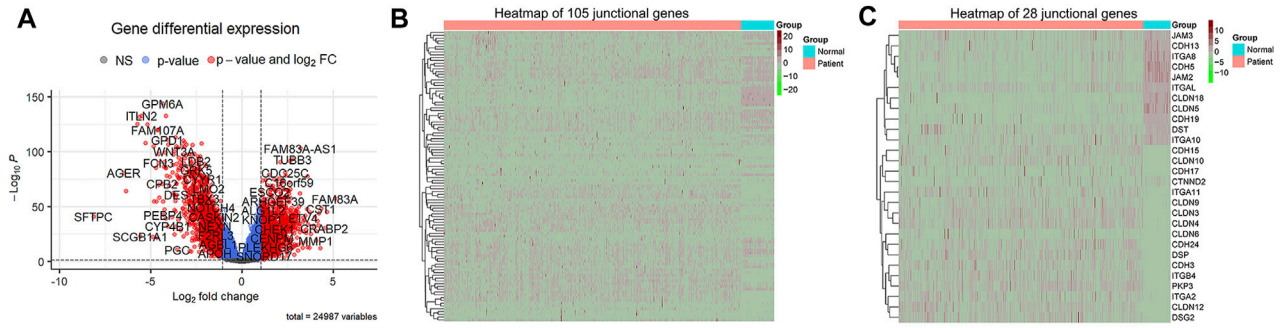


FIGURE 2
Identification of DEGs of TCGA datasets. adj. $p < 0.05$ and $|log2FC| > 1$ were used as the cut off criteria. **(A)** A volcano map of DEGs, which are denoted in red. **(B)** A heat map of the mRNA expression of 105 junctional genes between normal lung and LUAD tissues. Genes whose expression was upregulated are shown in red; the expression levels increase as the color darkens. Genes whose expression was downregulated are shown in green; the expression levels decrease as the color darkens. **(C)** A heat map of the mRNA expression of 24 differentially expressed junctional genes.

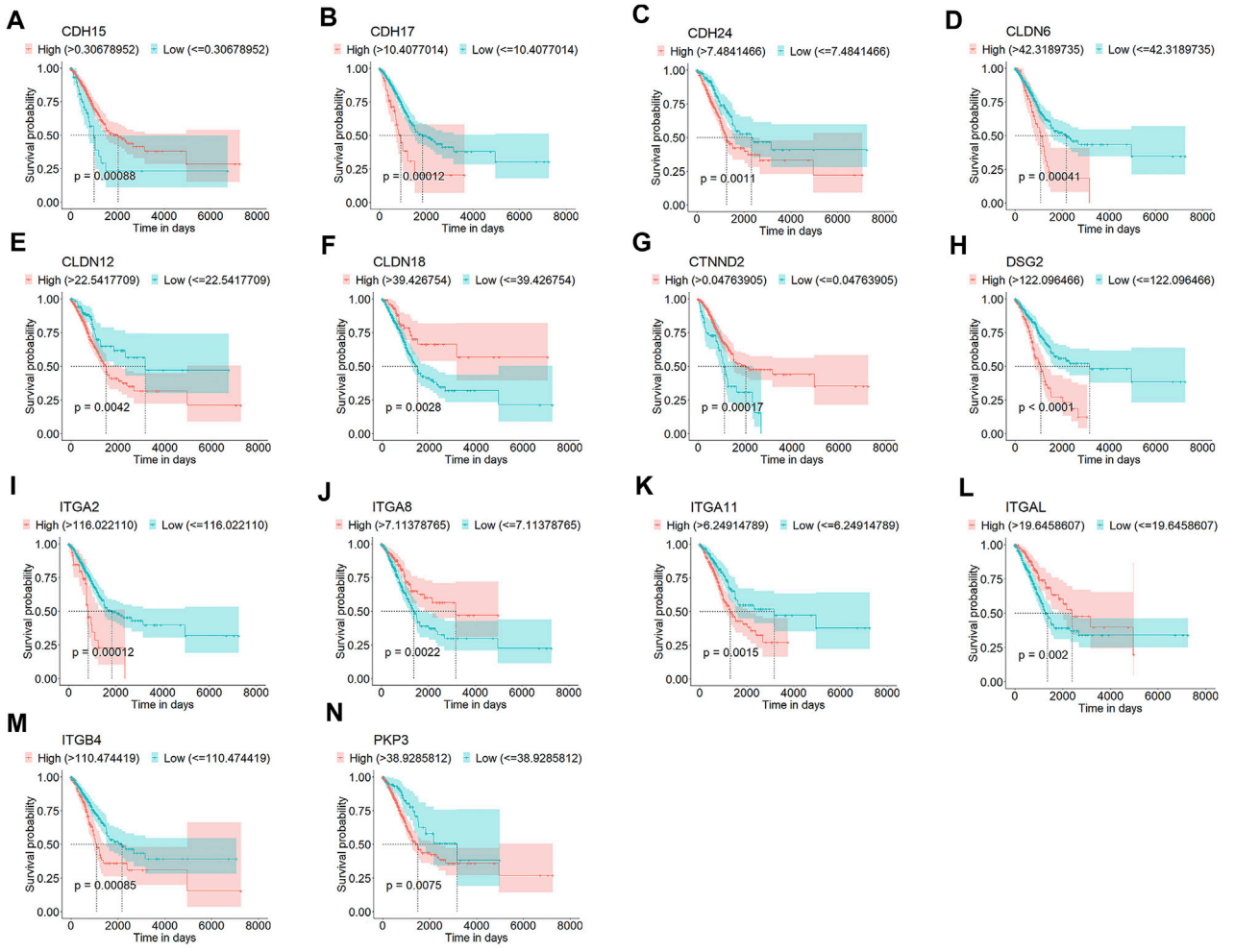


FIGURE 3
Kaplan-Meier survival curves for patients with LUAD grouped based on low and high expression levels of *CDH15* **(A)**, *CDH17* **(B)**, *CDH24* **(C)**, *CLDN6* **(D)**, *CLDN12* **(E)**, *CLDN18* **(F)**, *CTNND2* **(G)**, *DSG2* **(H)**, *ITGA2* **(I)**, *ITGA8* **(J)**, *ITGA11* **(K)**, *ITGAL* **(L)**, *ITGB4* **(M)**, and *PKP3* **(N)** in the TCGA cohort.

DSG2 + 0.55548 × *ITGA2* − 0.51734 × *ITGA8* − 0.34780 × *ITGAL* + 0.20949 × *ITGB4* + 0.32971 × *CLDN12* + 0.14356 × *PKP3* + 0.40963 × *ITGA11*. Patients were separated into low- and high-expression groups based on the cutoff value for each of the 14 genes, and hazard ratios (HRs) were calculated. *ITGAL*, *ITGA8*, *CTNND2*, *CLDN18*, and *CDH15* were protective factors for LUAD survival,

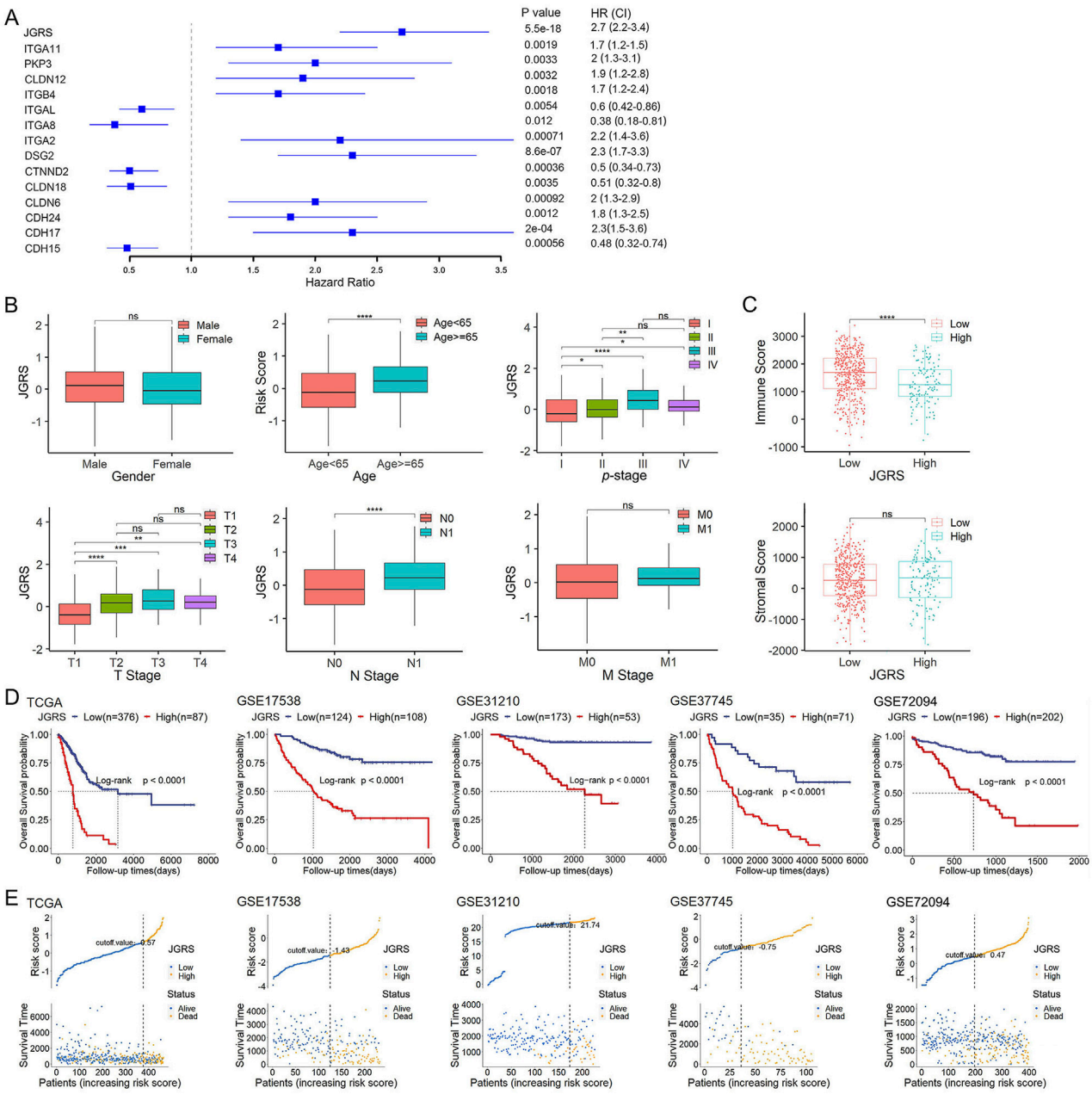


FIGURE 4
The JGRS can distinguish different clinicopathological features of LUAD. **(A)** A forest plot of the univariate Cox regression analysis of JGRS and 14 genes that were chosen for establishing a prognosis signature. **(B)** Different analyses of JGRS distribution based on sex, age, p-stage, as well as T, N, and M stages in TCGA cohort. **(C)** Distribution of immune and stromal scores between low- and high-JGRS groups in TCGA cohort. **(D)** Kaplan-Meier survival curves based on the JGRS in TCGA and four GEO cohorts. **(E)** JGRS distribution in TCGA and four GEO cohorts.

with HRs < 1. In contrast, *ITGA11, PKP3, CLDN12, ITGB4, ITGA2, DSG2, CLDN6, CDH24,* and *CDH17* were risk factors, with HRs >1. The JGRS had the higher HR than the individual genes (Figure 4A). We investigated the correlation between the JGRS and sex, age, TNM stage, and *p*-stage in patients with LUAD (Figure 4B). The JGRS was not correlated with sex but was significantly correlated with age. Patients with LUAD older than 65 years had significantly higher JGRS than younger patients. In addition, a higher JGRS score was significantly associated with higher T, N, and *p*-stage in patients with LUAD. From *p*-stages I to III, the JGRS increased with advancements in stage. The JGRS was significantly higher in the *p*-stage IV group than in the stage I group. However, there was no significant difference between the *p*-stage IV and II groups or between the *p*-stage IV and III groups. Similar results were observed in the T stage. The JGRSs of the T2/3/4 groups were all significantly higher than those of the T1 stage group. However, there were no significant differences between each pair of the T2/3/4 stage groups. These results can be attributed to preserved cell-cell and cell-matrix contacts at the early T stage, with these contacts lost when the tumor progressed to an advanced stage. For the N stage, the N1 stage

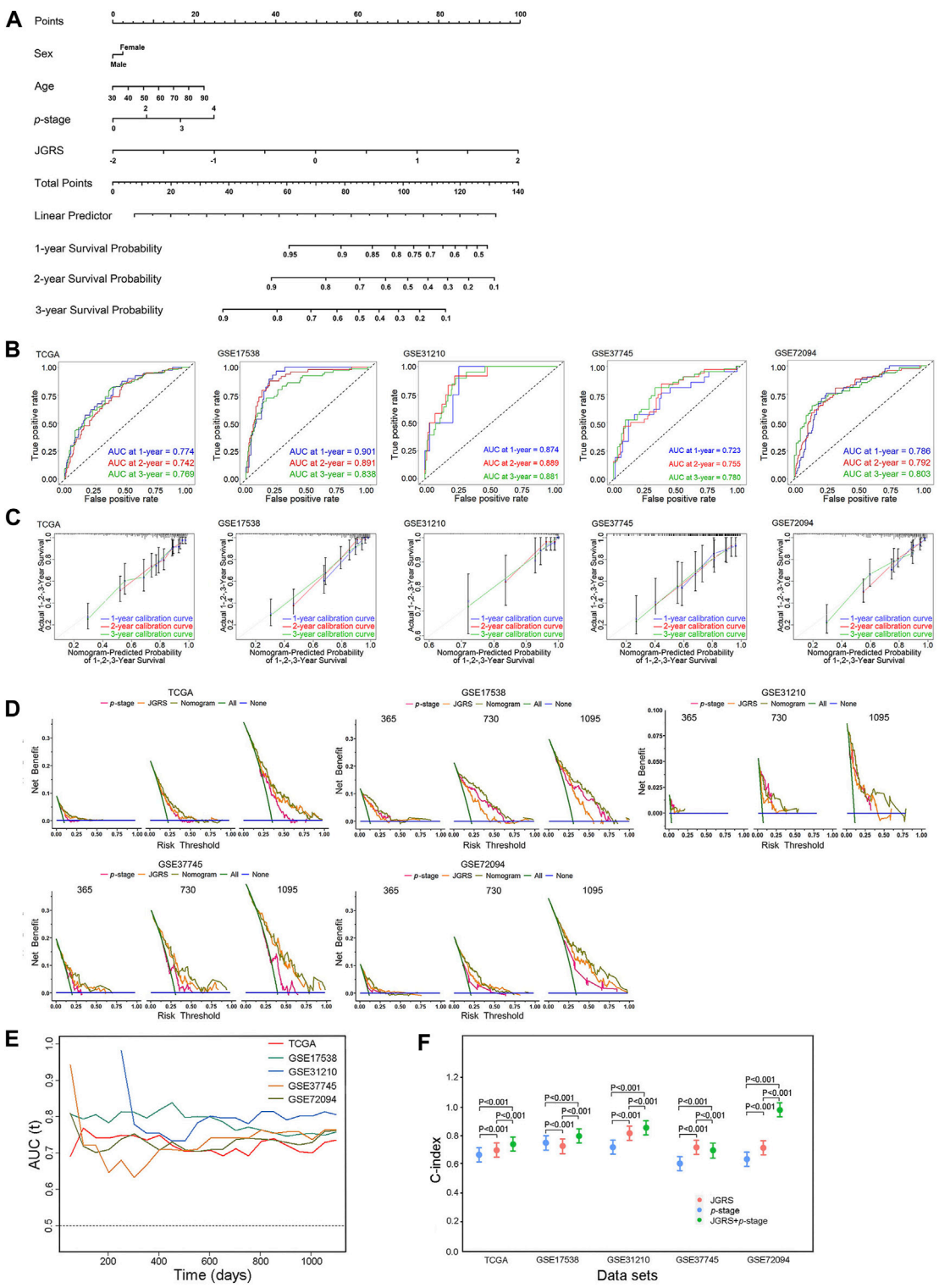


FIGURE 5
Establishment and assessment of the nomogram. **(A)** The nomogram plot was constructed based on sex, age, *p*-stage, and JGRS. ROC curves **(B)**, calibration curves **(C)**, and DCA **(D)** of the nomogram for 1-, 2-, and 3-year OS based on TCGA and four GEO cohorts. **(E)** Time-dependent AUC values in TCGA and four GEO cohorts. **(F)** C-indices of the *p*-stage, JGRS, and combined *p*-stage and JGRS in TCGA and four GEO cohorts.

group had significantly higher JGRS than the N0 stage group. In contrast, there was no significant difference in JGRS between the M0 and M1 groups in the M stage. This may be because any T or N stage was considered an M0 stage if it did not metastasize to a distant location. However, the loss of junctional genes occurs as the tumor grows from an early T or N stage to advanced stages.

In addition to the tumor cell-cell and cell-matrix interactions, stromal and immune cells can crosstalk with tumor cells and influence cancer growth and development. Therefore, we investigated whether the change in JGRS affects the immune and stromal scores. In this study, a higher JGRS score corresponded with a lower immune score. This negative correlation was statistically significant, whereas the stromal score was not significantly associated with JGRS (Figure 4C). We then stratified the patients in TCGA-LUAD, GSE17538, GSE31210, GSE37745, and GSE72094 cohorts into low- and high-JGRS groups based on the cutoff point in each cohort. Kaplan–Meier curves showed that patients in the low-JGRS group had significantly better OS rates than those in the high-JGRS group (Figure 4D). Figure 4E shows each of the patients on the x-axis from left to right based on their JGRS values and denotes the low-JGRS patients in blue and high-JGRS patients in yellow, analyzed by ggrisk (version 1.3). A higher percentage of deaths was observed in the high-JGRS group than in the low-JGRS group. These results indicate an association between JGRS and LUAD progression and OS in patients with LUAD.

3.4 Establishment and assessment of a prognostic nomogram for the OS of patients with LUAD

According to the DEG analysis results and Kaplan-Meier plots, a nomogram was constructed to predict the 1-, 2-, and 3-year OS probabilities for patients with LUAD based on the JGRS of 14 junctional genes. As sex, age, and *p*-stage of the cancer considerably affect the OS of patients, we included these three factors in the nomogram (Figure 5A). The general performance of this nomogram was assessed using four common evaluation methods: ROC curves, calibration curves, DCA curves, and the C-index. The ROC curve depicts both the sensitivity and specificity of the regression model. The AUC of the ROC curve is an effective method for assessing the overall diagnostic accuracy of a test. As shown in Figure 5B, the AUCs for 1-, 2-, and 3-year OS in the TCGA cohort were 0.774, 0.742, and 0.769, respectively. We further validated this nomogram using four GEO cohorts: GSE17538, GSE31210, GSE37745, and GSE72094. The AUCs for 1-, 2-, and 3-year OS in the GSE17538 cohort were 0.901, 0.891, and 0.838, respectively; those in the GSE37745 cohort were 0.723, 0.755, and 0.780, respectively; those in the GSE31210 cohort were 0.874, 0.889, and 0.881, respectively; and those in the GSE72094 cohort were 0.786, 0.792, and 0.803, respectively. These results indicate that the nomogram is both sensitive and specific for predicting OS in patients with LUAD. The calibration curve shows a consensus between the predicted value of the model and the observed value. The calibration curves had good consensus in the TCGA cohort and four GEO cohorts (Figure 5C), confirming the practicality of this nomogram in predicting patient OS (Figure 5C). The DCA is a statistical method used to evaluate the clinical consequences of models and tests. The DCA for this nomogram accurately predicted the 1-, 2-, and 3-year OS rates of patients with LUAD in TCGA and four GEO cohorts (Figure 5D). Furthermore, time-dependent AUC suggested that the nomogram accurately predicted the OS of patients with LUAD, with almost all AUC values above 0.7 overtime in all five cohorts (Figure 5E). Finally, the C-index reflects the predictive ability of a model. As demonstrated in Figure 5F, *p*-stage alone produced C-indices of 0.67, 0.755, 0.724, 0.609, and 0.639 in TCGA and the four GEO cohorts, respectively. The JGRS had significantly higher C-indices than the *p*-stage in all cohorts except GSE17538, with values of 0.702, 0.733, 0.823, 0.719, and 0.717. Furthermore, C-indices of the JGRS and *p*-stage combined were 0.737, 0.799, 0.852, 0.699, and 0.982. This significantly promoted C-indices in TCGA and the GSE17538, GSE31210, and GSE72094 GEO datasets. The C-index of the combined JGRS and *p*-stage was lower than that of JGRS alone in the GSE37745 cohort. However, the C-index of JGRS remained significantly higher than that of the *p*-stage, which verified the predictive value of JGRS.

3.5 JGRS-related biological pathways and genes

To explore the potential mechanisms underlying the differential OS outcomes of the JGRS determination, we performed GSEA. The top 10 differentially expressed gene ontology (GO) (Figure 6A) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Figure 6B) associated with the JGRS signature were identified at $p < 0.05$. Most GO and KEGG pathways enriched in the high-JGRS group were associated with cell replication (inner cell mass cell proliferation), mitotic processes (mitotic DNA replication/mitotic spindle midzone/protein localization to kinetochore/regulation of attachment of spindle microtubules to the kinetochore), and biosynthesis (aminoacyl-tRNA biosynthesis/biosynthesis of nucleotide sugars). Junctional genes primarily bind or unbind intercellular cells. Contact inhibition of proliferation occurs between cells under normal conditions, and unbinding of the cells is the precursor step of uncontrolled cell proliferation, which leads to precancerous cell development. Altered expression of junctional genes contributes to cell proliferation. For instance, high *CLDN1* expression and low *E-cadherin* expression promote cell proliferation and escape from their original sites (Bremnes et al., 2002; Wang D. W. et al., 2022). This explains the enrichment of cell proliferation and biosynthetic pathways in the high JGRS group. In addition, the top 10 GO terms for biological processes, such as immune receptor activity, B cell receptor signaling pathway, immunoglobulin-mediated immune response, and B cell-mediated immunity, were enriched in the high JGRS group (Figure 6C). KEGG enrichment analysis also revealed that immune- and cell proliferation-related categories, such as DNA replication, cell cycle, and IgA production, were enriched in the high-JGRS group (Figure 6D). This also confirmed the correlation between the immune score and JGRS (Figure 4C). In summary, the results obtained using either the GO or KEGG databases all pointed to the conclusion that the JGRS correlates with cell proliferation and immune-related processes, which may be the mechanism leading to the differences in the OS of patients with LUAD.

An xCell analysis that included 64 different types of immune and stromal cells was conducted to further elucidate the regulatory

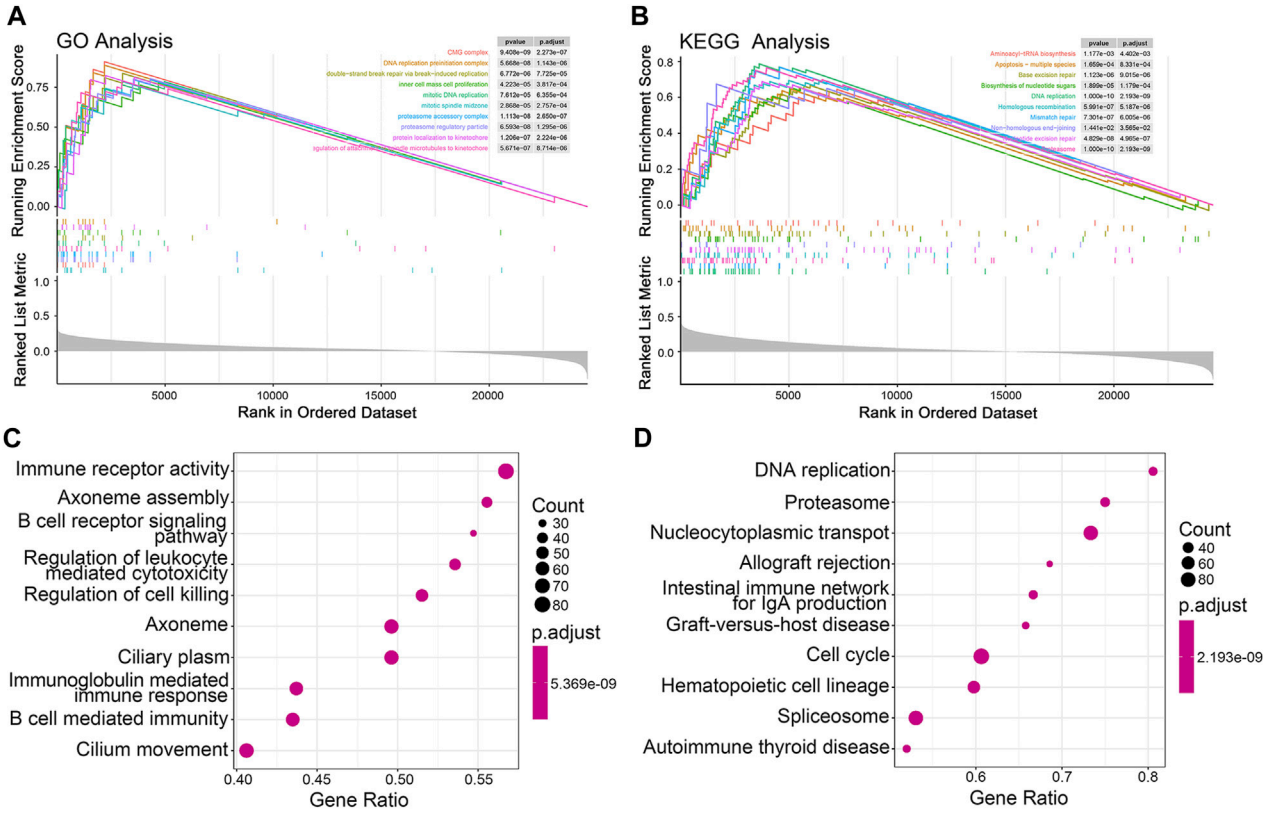


FIGURE 6
GSEA function enrichment analysis for JGRS. Curve graphs of the top 10 enriched pathways based on the GO **(A)** and KEGG **(B)** pathway databases, respectively. A dot plot of the top 10 enriched GO **(C)** and KEGG **(D)** terms, respectively.

immune cells involved in the tumor immune microenvironment (TIME). The abundance of infiltrating immune cells was determined based on the expression of marker genes. The correlations between each immune cell and the 14 JGRS signature genes are displayed as a heatmap in Supplementary Figure S6. Genes that were risk factors for OS, such as *CDH24* and *PKP3*, were negatively correlated with most immune cells, and *vice versa*. In contrast, genes that were protective factors for OS, such as *CLDN18*, *ITGA8*, and *ITGAL*, were positively correlated with most immune cells. Because the GO terms for biological processes were mainly enriched in B cell-related immunity, we investigated the infiltration of B cells into the TIME. CIBERSORT and TIMER methods were used to analyze the infiltration of different immune cell components into low- and high-JGRS tissues. As expected, both CIBERSORT (Figure 7A) and TIMER (Figure 7B) data showed a significantly higher infiltration of B cells in the low-JGRS group than in the high-JGRS group, suggesting that the poor OS rates of high-JGRS patients may be related to the reduced infiltration of B cells into the TIME. In addition to B cells, the CIBERSORT results showed a higher infiltration of dendritic cells and $CD4^+$T cells in the low-JGRS group. Furthermore, the high-JGRS group was associated with cell replication, mitotic processes (Figures 6A, B), and immune regulatory features. Therefore, we analyzed the mRNA expression of cell cycle control genes. Cyclins and cyclin-dependent kinases (CDKs) are important regulators that drive cell cycle progression (Basu et al., 2022). Figure 7C shows elevated levels of cyclins A (*CCNA2*), B (*CCNB1*), D (*CCND1*), and E (*CCNE1*), as well as those of CDK1, 2, 4, 6, and Cdc25A in the high-JGRS group. These results explain the expedited cell cycle in high-JGRS patients and are consistent with the results shown in Figure 6.

3.6 Analysis of mutation status between the low- and the high-JGRS groups

We analyzed the somatic mutations of patients in the TCGA cohort to further investigate the genetic mechanisms underlying the differential OS outcomes between the low- and high-JGRS groups. The 30 most frequently mutated genes in the low- and high-JGRS groups are shown in Figure 8A. The frequency of somatic mutations in the low-JGRS group was 90.23%, whereas that in the high-JGRS group was 97.07%. Further statistical analyses showed that the frequencies of total mutation counts, as well as non-synonymous and synonymous mutations, were significantly higher in the high-JGRS group than in the low-JGRS group (Figure 8B). Additionally, JGRS was significantly positively correlated with somatic, non-synonymous, and synonymous mutation counts ($p < 0.01$; Figure 8C). Comparison of the mutational frequencies of each gene revealed that 17 genes were significantly more frequently mutated in the high-JGRS group than in the low-JGRS group ($p < 0.01$), namely *USH2A*, *SORCS1*, *CPS1*, *APOB*, *DNAH8*, *LRP1B*, *TP53*, *COL6A3*, *AHNAK*, *TPTE*, *PCDH10*, *FAT4*,

LAMA1, NLRP3, RYR3, FBN2, and COL11A1 (Figure 8D). These genes were subjected to co-occurrence mutation analysis using maftools, and co-mutations were found among the 17 genes (Figure 8E).

3.7 Verification of prognostic DEGs using clinical tissue samples

To verify the reliability of the DEGs with prognostic values, we detected the protein expression of 14 genes in normal and LUAD tissues from the HPA website. Antibody-staining images except for CDH24 and CLDN6 were available on the HPA website. Of the rest 12 proteins, CDH15, CTNND2, and ITGA8, were all negatively expressed in both the normal and LUAD tissues. The expression of CDH17, CLDN12, DSG2, ITGB4, PKP3, ITGA11, and ITGA2 was all upregulated whereas the expression of ITGAL was downregulated in LUAD tissues compared to the normal lung tissues (Figure 9). These results coincide well with the JGRS formula in that *CDH17*, *CLDN12*, *DSG2*, *ITGB4*, *PKP3*, *ITGA11*, and *ITGA2* were risky factors whereas *ITGAL* was a protective factor for patient survival.

4 Discussion

Prediction of the prognosis in patients with cancer has garnered substantial research interest. Many researchers have screened for useful prognostic biomarkers using the genetic information of patients from TCGA and GEO databases or data from local hospitals combined with clinical properties to construct a nomogram that predicts patient OS. The ultimate goal was to develop an accurate and effective model. Some studies have focused on genome-wide screening by comparing cancer tissues with normal tissues, whereas others have focused on a specific gene group that targets common physiological processes or factors, such as apoptosis, cancer stem cells, tumor microenvironment, DNA methylation, gene mutation, and oxidative stress (Supplementary Table S1). In the present study, we first screened for differentially expressed junctional genes by comparing LUAD tissues with normal lung tissues. Of the 105 junctional genes identified, the expression of 28 was upregulated or downregulated in LUAD tissues. We constructed a Kaplan-Meier plot to select the genes that contributed to the OS of patients with LUAD. Fourteen of the 28 genes were selected to generate the JGRS, which was used to construct the nomogram. To the best of our knowledge, this is the first study to reveal the potential prognostic value of a panel of junctional genes.

The regulation of tumorigenesis is complex. Lung cancer often detected in the middle or late stages (Zhao et al., 2022). Approximately 40% of patients with lung cancer die from metastasis (Simonaggio et al., 2020). Migratory cells can escape from the primary site, invade normal tissues, travel through the lymphatic system or bloodstream, and spread to distant locations. Loss of cell connections is one of the initial hallmarks of epithelial cell migration (Prudkin et al., 2009). Junctional proteins mediate cell-cell and cell-matrix connections. Thus, we evaluated their role in OS prediction. Among the 14 junctional genes screened, *CLDN6, CLDN12*, and *CLDN18* belong to the tight junction family. In total, 27 claudins have been discovered to date (Osanai et al., 2017), with a high or low abundance of claudins described in diverse neoplastic

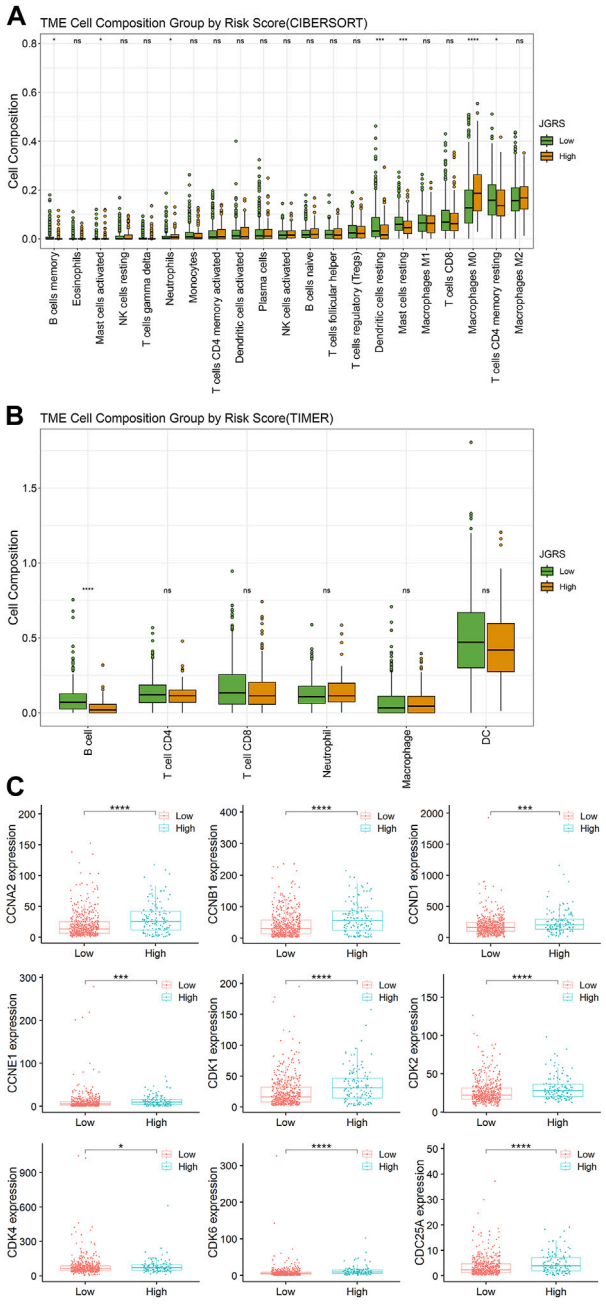


FIGURE 7
Infiltration of immune cells and the expression of cell cycle-related genes in the low- and high-JGRS groups of patients. Conditions of immune cell infiltration into the TIME based on the CIBERSORT **(A)** and TIMER **(B)** methods. **(C)** Expression of cell cycle-related genes in the low- and high-JGRS groups of patients.

tissues. For example, claudin-1 was decreased in pancreatic and ovarian cancers, as well as in LUAD (Osanai et al., 2017). Claudin-7 was reduced in LUAD (Lu et al., 2011), whereas claudin-3 and claudin-4 were increased in esophageal cancer (Osanai et al., 2017). Consistent with previous studies on the contributions of *CLDN6*, *CLDN12*, and *CLDN18* to LUAD, increased RNA expression of *CLDN6* and *CLDN12* in patients with LUAD increased the risk score and lowered the survival probability, whereas increased RNA expression of *CLDN18* had the opposite effect in the present

10.3389/fmolb.2024.1380384

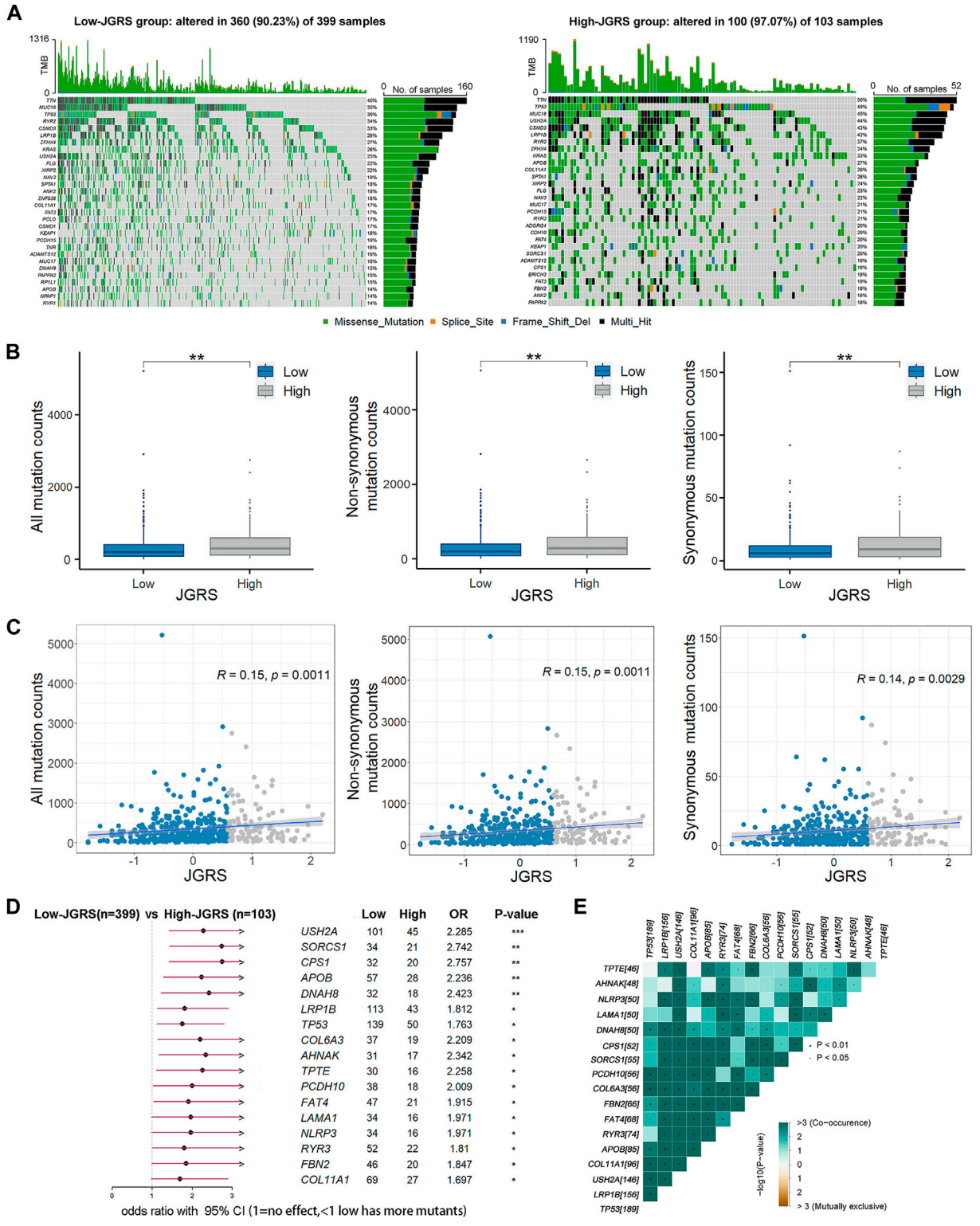


FIGURE 8
JGRS was related to tumor mutation status. **(A)** The top 30 most frequently mutated genes in the low- (left) and high-JGRS (right) groups. **(B)** The distribution of all, synonymous, and non-synonymous counts in the low- and high-JGRS groups. **(C)** Association between JGRS and all, non-synonymous, and synonymous mutation counts. **(D)** A forest plot of differentially mutated genes in the low- and high-JGRS groups, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. **(E)** Interaction effects between differentially mutated genes in the low-and the high-JGRS groups.

10.3389/fmolb.2024.1380384

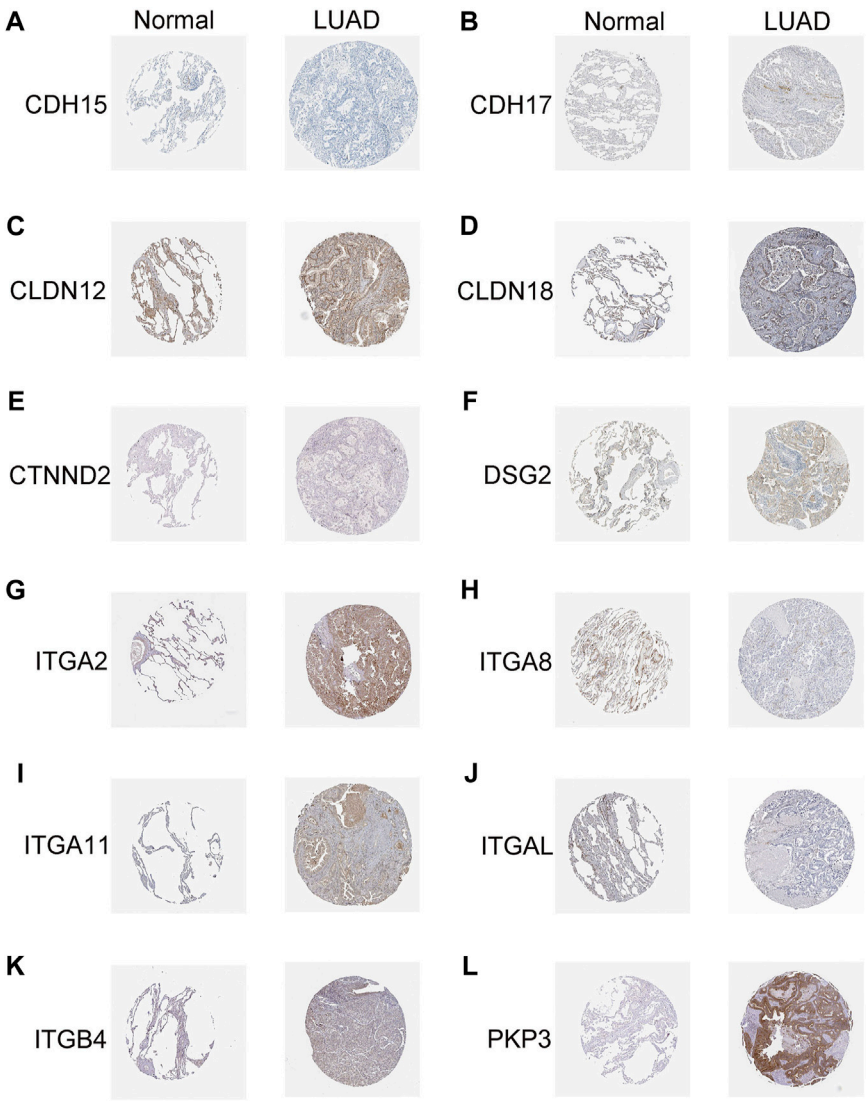


FIGURE 9
Immunohistochemical analysis of genes with prognostic values. **(A)** *CDH15*, **(B)** *CDH17*, **(C)** *CLDN12*, **(D)** *CLDN18*, **(E)** *CTNND2*, **(F)** *DSG2*, **(G)** *ITGA2*, **(H)** *ITGA8*, **(I)** *ITGA11*, **(J)** *ITGAL*, **(K)** *ITGB4*, and **(L)** *PKP3*.

study. Claudin-6 was significantly more frequent and most abundantly positive in adenocarcinoma (AC) than in squamous cell carcinoma (SCC), and was associated with poor prognosis in 164 patients with NSCLC from the University Hospital of Kuopio (Oini et al., 2022). This result was also confirmed in 196 patients with NSCLC at Uppsala University Hospital. These studies revealed distinct membrane-positive CLDN6 proteins in LUAD, with high CLDN6 expression associated with a worse prognosis (Micke et al., 2014). In addition, Kuner et al. conducted a global gene expression analysis of NSCLC subtypes and identified a striking presence of cell adhesion genes that were deregulated between SCC and AC subtypes. Among these, the expression of CLDN12 was upregulated in AC tissues compared to that in normal tissues (Kuner et al., 2009). Kuner et al. further detected the increased expression of other junctional genes, such as *DSG3, CLDN1, DSC2, CLDN3, CLDN7, CDH1,* and *CDH2* in the AC. However, these genes were not included in the nomogram developed in the present study. First, the genetic background of the patient tissues from their cohort (German Cancer Research Center) was, to some extent, different from that retrieved from TCGA. Second, although these genes were differentially expressed, they did not correlate with the OS of the patients; thus, they were not included in the nomogram. In addition, loss of *CLDN18* resulted in increased type 2 alveolar epithelial (AT2) cell proliferation and an increased frequency of LUAD in mice. Human LUAD, which originates from AT2 cells, also displayed reduced CLDN18 (Kotton, 2018). These findings indicate that *CLDN6, CLDN12,* and *CLDN18* markedly contribute to the prediction of OS in patients with LUAD.

Another large gene family that was incorporated into the nomogram was the integrin family. Of the selected 14 genes, five belonged to the integrin family. Integrins connect cells to the extracellular matrix. They comprise 18 α and 8 β subunits, different combinations of which can assemble into 24 complexes (Tvaroska et al., 2023). Smythe et al. discovered the downregulated

expression of the subunits α2, α3, α6, and β4 and the upregulated expression of the β6 subunit in NSCLC (Smythe et al., 1995). Integrins are expressed in a histological and type-specific manner. For example, α3 was strongly expressed in AC but was infrequent in SCLC.α4 was solely expressed in bronchioloalveolar carcinoma (Guo et al., 2009). Furthermore, αvβ3 was associated with tumorigenesis and metastasis steps, leading to poor survival in patients with LUAD (Kariya et al., 2021). Integrin α6 expression was significantly higher in LUAD tissues and positively correlated with the grade and T stage of LUAD, leading to poorer patient survival rates (Shen et al., 2019). In addition, Navab et al. revealed significantly impeded growth and metastasis of lung cancer cells in integrin α11-deficient severe combined immunodeficient (SCID) mice compared to wild-type SCID mice. This demonstrated the ability of integrin α11 to promote the growth and metastasis of NSCLC. Our results were consistent with this finding, as increased RNA expression of integrin α11 in patients with LUAD increased the JGRS and decreased survival probability. Four other integrin subunits, namely, integrins α2, α8, αL, and β4, were identified as prognostic-related biomarkers in the present study. Integrins α2 and β4 were positively correlated with the JGRS and negatively correlated with patient OS, whereas α8 and αL were the opposite. As integrins play crucial roles, effective inhibitors targeting integrin subunits have been discovered and are widely used clinically, including in the fields of cardiovascular and inflammatory bowel diseases. However, the clinical development of integrin inhibitors in cancer remains considerably challenging (Slack et al., 2022).

In addition to the claudin and integrin families, other junctional genes such as cadherins, δ-catenin *(CTNND2), DSG2*, and *PKP3* were correlated with the prognosis of patients with LUAD. As these junctional genes play pivotal roles in lung tumorigenesis, the JGRS based on the selected 14 genes can be an independent prognostic marker that best reflects the prognostic status of patients with LUAD. To some extent, previous studies have provided evidence of the mechanisms by which these genes affect lung tumorigenesis, such as cell proliferation, migration, and metastasis. Our GSEA also showed that high JGRS levels were mainly associated with cell proliferation and immune regulatory pathways. Further analysis showed more abundant infiltration of B cells and upregulated expression of cell cycle-related genes, such as cyclins and CDKs in the high-JGRS group (Figure 7).

Oncogenes and tumor suppressor genes are important "players" in cancer formation, and their mutations or expression changes can lead to neoplastic transformation in normal cells. In the present study, patients were separated into high- and low-JGRS groups, and their mutational status was analyzed. Seventeen genes were mutated significantly more frequently in the high-JGRS group than in the low-JGRS group. Of these, *COL6A3, AHNAK, CPS1, TPTE*, and *DNAH8* had twice as many mutations in the high-JGRS group than in the low-JGRS group. The involvement of these genes in tumorigenesis has been reported. *COL6A3* produces the alpha (α) 3 (VI) chain of type VI collagen, a component of the extracellular matrix. Dysregulated expression of *COL6A3* has been observed in several cancers, including cervical cancer and pancreatic adenocarcinoma (Annapurna et al., 2021; Wang H. L. et al., 2022). *AHNAK*, which encodes the giant protein desmoyokin, was originally identified as a nucleoprotein in neuroblastoma cells (Sundararaj et al., 2021). It was later identified as a tumor suppressor that can negatively regulate cell growth through TGFβ signaling (Lee et al., 2014). Carbamoyl phosphate synthetase 1 (CPS1) is a tumor promoter that either supports pyrimidine synthesis or prevents the buildup of intratumoral ammonia (Yao et al., 2020). Transmembrane phosphatase with tensin homology (TPTE) shares significant homology with the tumor suppressor protein PTEN. DNAH8 is a member of the dynein axonemal heavy chains (DNAHs), and its variants have been associated with heavy smoking (Wain et al., 2016). Failure to clear the toxins in the respiratory tract due to a variant change in DNAH8 may cause lung cancer. However, there are limited studies on the subject, and this hypothesis requires further validation. Owing to their association with various cancers, these mutated genes may directly or indirectly affect the prognosis of patients with LUAD.

The performance evaluation of a predictive model is crucial following its construction. The effectiveness of a model can be evaluated through the C-index, which is often used to measure how well a biomarker predicts the time to an event. It can also be determined through AUCs, which plot the rate of true positives against false positives. A calibration plot can also be used to assess the agreement between predicted and observed values. However, a calibration plot lacks an assessed value. Thus, effectiveness can only be determined by looking at the closeness between the prediction and diagonal lines. In addition, DCA is a statistical method that evaluates models and tests their clinical consequences. Papers predicting the prognosis of patients with LUAD published in the recent 5 years are summarized in Supplementary Table S1. In total, 47 studies were included. We have listed the predictive parameters of these studies, as well as the AUC and C-indices of the training and validation cohorts. Overall, the predictive values of the models were compromised. Some studies lacked external validation or were only validated in one or two cohorts (Peng et al., 2024). Others only focused on a subgroup of patients with LUAD, such as patients with early-stage LUAD or those with metastasis, which is not applicable to all patients with LUAD. The model with the highest C-index value (0.89) was constructed by Huang et al. However, this study neither provided AUC values nor externally validated the model (Huang et al., 2021). A prognostic model with DNA methylation profiling showed promising AUC values of 0.846, 0.900, and 0.909 for 1-, 3-, and 5-year predictions, respectively. However, it was only validated in one GSE cohort (Ma et al., 2020). Our nomogram has the advantage of robust performance, as it was validated using four GEO datasets. The AUCs for TCGA and the four validation GEO cohorts were all >0.72. For two GEO cohorts, GSE17538 and GSE72094, the values were >0.78. AUCs for the 1-, 2-, and 3-year OS rates were the highest in the GSE17538 cohort, reaching 0.901, 0.891, and 0.838, respectively. These results indicate that our nomogram is both sensitive and specific for predicting OS in patients with LUAD. The C-indices of JGRS and *p*-stage combined were all above 0.73 except for one cohort, GSE37745. The values in the GSE31210 and GSE72094 cohorts were high, reaching 0.852 and 0.982, respectively. This validated the predictive ability of the model. Furthermore, the calibration curves in all four GEO cohorts showed favorable consensus (Figure 5C), and DCA showed good clinical practicability of the model (Figure 5D). Hence, our results represent a small breakthrough compared to those of previous studies.

5 Conclusion

This study demonstrated a significant correlation between altered expression of junctional genes and the OS of patients with LUAD. Here, we constructed a junctional gene-related nomogram model to predict the OS of patients with LUAD. Although our model was multidimensionally validated, it has some limitations. The prognosis of LUAD depends on various factors, such as patient psychology, smoking status, surgical performance, and response to radiotherapy or chemotherapy. These factors were not taken into consideration in this study because of partial records. Additionally, although we evaluated the JGRSs in patients with different TNM stages and mutation characteristics of patients in the TCGA cohort, we did not perform them in the four GEO cohorts, owing to the lack of TNM stage information and mutation data. Moreover, future studies are required to explore the underlying molecular mechanisms and subsequently advance potential clinical applications, as these cell junctional genes may be valuable therapeutic targets.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: NIH GDC Data Portal (https://portal.gdc.cancer.gov/projects/TCGA-LUAD) and GEO repository, accession numbers: GSE17538, GSE31210, GSE37745 and GSE72094.

Ethics statement

Ethical approval was not required for the studies on humans in accordance with the local legislation and institutional requirements because the human samples used in this study were isolated as part of our previous study.

Author contributions

BX: Conceptualization, Investigation, Software, Supervision, Writing–review and editing, Methodology, Project administration, Validation. TW: Data curation, Formal Analysis, Methodology, Project administration, Software, Validation, Writing–original draft. DH: Data curation, Formal Analysis, Methodology, Writing–original draft. ZL: Conceptualization, Funding acquisition, Investigation, Validation, Writing–original draft, Writing–review and editing, Data curation, Project administration, Supervision.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by grants from the National Natural Science Foundation of China (grant number: 31200581); and Natural Science Foundation of Zhejiang Province (grant numbers: LY15H160068, LY23H010001).

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/fmolb.2024.1380384/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Expression of 28 differentially expressed junctional genes in normal and LUAD tissues.

SUPPLEMENTARY FIGURE S2
The RNA expression levels of 14 junctional genes in normal lung and LUAD tissues in the GSE31210 cohort.

SUPPLEMENTARY FIGURE S3
The protein expression levels of 14 junctional genes in normal lung and LUAD tissues by Clinical Proteomic Tumor Analysis Consortium.

SUPPLEMENTARY FIGURE S4
The RNA expression levels of 14 junctional genes in 232 lung cancer cell lines compared with normal human epithelial ciliated cells.

SUPPLEMENTARY FIGURE S5
Evaluation of the mRNA expression levels of 14 junctional genes between normal lung (n = 9) and LUAD specimens (n = 9) from our local hospital by PCR analysis.

SUPPLEMENTARY FIGURE S6
Correlation between the infiltration of 64 immune and stroma cell types with each signature genes in JGRS.

References

Annapurna, S. D., Pasumarthi, D., Pasha, A., Doneti, R., Sheela, B., Botlagunta, M., et al. (2021). Identification of differentially expressed genes in cervical cancer patients by comparative transcriptome analysis. *Biomed Res. Int.* 2021, 8810074. doi:10.1155/2021/8810074

Basu, S., Greenwood, J., Jones, A. W., and Nurse, P. (2022). Core control principles of the eukaryotic cell cycle. *Nature* 607 (7918), 381–386. doi:10.1038/s41586-022-04798-8

Bremnes, R. M., Veve, R., Hirsch, F. R., and Franklin, W. A. (2002). The E-cadherin cell-cell adhesion complex and lung cancer invasion, metastasis, and prognosis. *Lung Cancer* 36 (2), 115–124. doi:10.1016/S0169-5002(01)00471-8

Buckley, C. E., and St Johnston, D. (2022). Apical-basal polarity and the control of epithelial form and function. *Nat. Rev. Mol. Cell Biol.* 23 (8), 559–577. doi:10.1038/s41580-022-00465-y

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

Guo, L. L., Zhang, F., Cai, Y. Q., and Liu, T. F. (2009). Expression profiling of integrins in lung cancer cells. *Pathology Res. And Pract.* 205 (12), 847–853. doi:10.1016/j.prp.2009.07.005

Hashimoto, I., and Oshima, T. (2022). Claudins and gastric cancer: an overview. *Cancers* 14 (2), 290. doi:10.3390/cancers14020290

Huang, Q. J., Qu, T. Y., Qi, L. S., Liu, C. X., Guo, Y. H., Guo, Q. R., et al. (2021). A nomogram-based immune-serum scoring system predicts overall survival in patients with lung adenocarcinoma. *Cancer Biol. Med.* 18 (2), 517–529. doi:10.20892/j.issn.2095-3941.2020.0648

Iasonos, A., Schrag, D., Raj, G. V., and Panageas, K. S. (2008). How to build and interpret a nomogram for cancer prognosis. *J. Clin. Oncol.* 26 (8), 1364–1370. doi:10.1200/jco.2007.12.9791

Kariya, Y., Oyama, M., Suzuki, T., and Kariya, Y. (2021). αvβ3 Integrin induces partial EMT independent of TGF-β signaling. *Commun. Biol.* 4 (1), 490. doi:10.1038/s42003-021-02003-6

Katzel, J. A., Fanucchi, M. P., and Li, Z. (2009). Recent advances of novel targeted therapy in non-small cell lung cancer. *J. Hematol. Oncol.* 2, 2. doi:10.1186/1756-8722-2-2

Kotton, D. N. (2018). Claudin-18: unexpected regulator of lung alveolar epithelial cell proliferation. *J. Of Clin. Investigation* 128 (3), 903–905. doi:10.1172/Jci99799

Kuner, R., Muley, T., Meister, M., Ruschhaupt, M., Buness, A., Xu, E. C., et al. (2009). Global gene expression analysis reveals specific patterns of cell junctions in non-small cell lung cancer subtypes. *Lung Cancer* 63 (1), 32–38. doi:10.1016/j.lungcan.2008.03.033

Kuo, W. T., Odenwald, M. A., Turner, J. R., and Zuo, L. (2022). Tight junction proteins occludin and ZO-1 as regulators of epithelial proliferation and survival. *Ann. Of N. Y. Acad. Of Sci.* 1514 (1), 21–33. doi:10.1111/nyas.14798

Lee, I. H., Sohn, M., Lim, H. J., Yoon, S., Oh, H., Shin, S., et al. (2014). Ahnak functions as a tumor suppressor via modulation of TGFβ/Smad signaling pathway. *Oncogene* 33 (38), 4675–4684. doi:10.1038/onc.2014.69

Leiter, A., Veluswamy, R. R., and Wisnivesky, J. P. (2023). The global burden of lung cancer: current status and future trends. *Nat. Rev. Clin. Oncol.* 20 (9), 624–639. doi:10.1038/s41571-023-00798-3

Li, C., Yuan, Y., Jiang, X., and Wang, Q. (2023). Identification and validation of tumor microenvironment-related signature for predicting prognosis and immunotherapy response in patients with lung adenocarcinoma. *Sci. Rep.* 13 (1), 13568. doi:10.1038/s41598-023-40980-2

Li, Y., Shen, R., Wang, A., Zhao, J., Zhou, J., Zhang, W., et al. (2021). Construction of a prognostic immune-related LncRNA risk model for lung adenocarcinoma. *Front. Cell Dev. Biol.* 9, 648806. doi:10.3389/fcell.2021.648806

Lin, W. H., Cooper, L. M., and Anastasiadis, P. Z. (2023). Cadherins and catenins in cancer: connecting cancer pathways and tumor microenvironment. *Front. Cell And Dev. Biol.* 11, 1137013. doi:10.3389/fcell.2023.1137013

Lu, Z., Ding, L., Hong, H., Hoggard, J., Lu, Q., and Chen, Y. H. (2011). Claudin-7 inhibits human lung cancer cell migration and invasion through ERK/MAPK signaling pathway. *Exp. Cell Res.* 317 (13), 1935–1946. doi:10.1016/j.yexcr.2011.05.019

Ma, X. X., Cheng, J. C., Zhao, P., Li, L., Tao, K. X., and Chen, H. Y. (2020). DNA methylation profiling to predict recurrence risk in stage Iota lung adenocarcinoma: development and validation of a nomogram to clinical management. *J. Of Cell. And Mol. Med.* 24 (13), 7576–7589. doi:10.1111/jcmm.15393

Micke, P., Mattsson, J. S. M., Edlund, K., Lohr, M., Jirstrom, K., Berglund, A., et al. (2014). Aberrantly activated claudin 6 and 18.2 as potential therapy targets in non-small-cell lung cancer. *Int. J. Of Cancer* 135 (9), 2206–2214. doi:10.1002/ijc.28857

Nehme, Z., Roehlen, N., Dhawan, P., and Baumert, T. F. (2023). Tight junction protein signaling and cancer biology. *Cells* 12 (2), 243. doi:10.3390/cells12020243

Oini, Y., Pirinen, R., Takasawa, K., Osanai, M., and Takasawa, A. (2022). Claudin 6 is associated with a short survival and a short recurrent free interval in non-small cell lung carcinoma. *Pol. J. Of Pathology* 73 (1), 1–5. doi:10.5114/Pjp.2022.117171

Osanai, M., Takasawa, A., Murata, M., and Sawada, N. (2017). Claudins in cancer: bench to bedside. *Pflugers Archiv-European J. Of Physiology* 469 (1), 55–67. doi:10.1007/s00424-016-1877-7

Parrish, M. L., Broaddus, R. R., and Gladden, A. B. (2022). Mechanisms of mutant β-catenin in endometrial cancer progression. *Front. Oncol.* 12, 1009345. doi:10.3389/fonc.2022.1009345

Peng, Y., Dong, Y., Sun, Q., Zhang, Y., Zhou, X., Li, X., et al. (2024). Integrative analysis of single-cell and bulk RNA-sequencing data revealed T cell marker genes based molecular sub-types and a prognostic signature in lung adenocarcinoma. *Sci. Rep.* 14 (1), 964. doi:10.1038/s41598-023-50787-w

Prudkin, L., Liu, D. D., Ozburn, N. C., Sun, M. H., Behrens, C., Tang, X. M., et al. (2009). Epithelial-to-mesenchymal transition in the development and progression of adenocarcinoma and squamous cell carcinoma of the lung. *Mod. Pathol.* 22 (5), 668–678. doi:10.1038/modpathol.2009.19

Qian, J. Y., Hao, Y., Yu, H. H., Wu, L. L., Liu, Z. Y., Peng, Q., et al. (2023). A novel systematic oxidative stress score predicts the survival of patients with early-stage lung adenocarcinoma. *Cancers* 15 (6), 1718. doi:10.3390/cancers15061718

Shen, J. F., Xu, J. F., Chen, B. F., Ma, D. H., Chen, Z. X., Li, J. C., et al. (2019). Elevated integrin α6 expression is involved in the occurrence and development of lung adenocarcinoma, and predicts a poor prognosis: a study based on immunohistochemical analysis and bioinformatics. *J. Of Cancer Res. And Clin. Oncol.* 145 (7), 1681–1693. doi:10.1007/s00432-019-02907-1

Simonaggio, A., Elaidi, R., Fournier, L., Fabre, E., Ferrari, V., Borchiellini, D., et al. (2020). Variation in neutrophil to lymphocyte ratio (NLR) as predictor of outcomes in metastatic renal cell carcinoma (mRCC) and non-small cell lung cancer (mNSCLC) patients treated with nivolumab. *Cancer Immunol. Immunother.* 69 (12), 2513–2522. doi:10.1007/s00262-020-02637-1

Slack, R. J., Macdonald, S. J. F., Roper, J. A., Jenkins, R. G., and Hatley, R. J. D. (2022). Emerging therapeutic opportunities for integrin inhibitors. *Nat. Rev. Drug Discov.* 21 (1), 60–78. doi:10.1038/s41573-021-00284-4

Smythe, W. R., Lebel, E., Bavaria, J. E., Kaiser, L. R., and Albelda, S. M. (1995). Integrin expression in non-small-cell carcinoma of the lung. *Cancer And Metastasis Rev.* 14 (3), 229–239. doi:10.1007/Bf00690294

Song, J. H., Sun, Y. Y., Cao, H., Liu, Z. C., Xi, L., Dong, C. Q., et al. (2021). A novel pyroptosis-related lncRNA signature for prognostic prediction in patients with lung adenocarcinoma. *Bioengineered* 12 (1), 5932–5949. doi:10.1080/21655979.2021.1972078

Subramanian, A., Kuehn, H., Gould, J., Tamayo, P., and Mesirov, J. P. (2007). GSEA-P: a desktop application for gene set enrichment analysis. *Bioinformatics* 23 (23), 3251–3253. doi:10.1093/bioinformatics/btm369

Sun, D., Zhu, Y., Zhao, H., Bian, T., Li, T., Liu, K., et al. (2021). Loss of ARID1A expression promotes lung adenocarcinoma metastasis and predicts a poor prognosis. *Cell Oncol. (Dordr)* 44 (5), 1019–1034. doi:10.1007/s13402-021-00616-x

Sundararaj, S., Ravindran, A., and Casarotto, M. G. (2021). AHNAK: the quiet giant in calcium homeostasis. *Cell Calcium* 96, 102403. doi:10.1016/j.ceca.2021.102403

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

Troyanovsky, S. M. (2023). Adherens junction: the ensemble of specialized cadherin clusters. *Trends Cell Biol.* 33 (5), 374–387. doi:10.1016/j.tcb.2022.08.007

Tvaroska, I., Kozmon, S., and Kona, J. (2023). Molecular modeling insights into the structure and behavior of integrins: a review. *Cells* 12 (2), 324. doi:10.3390/cells12020324

Wain, L. V., Shrine, N., and Miller, S. (2016). Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank (vol 3, pg 769, 2015). *Lancet Respir. Med.* 4 (1), E4. doi:10.1016/S2213-2600(15)00283-0

Wang, D. W., Zhang, W. H., Danil, G., Yang, K., and Hu, J. K. (2022a). The role and mechanism of claudins in cancer. *Front. Oncol.* 12, 1051497. doi:10.3389/fonc.2022.1051497

Wang, H. L., Lu, L. Q., Liang, X. J., and Chen, Y. H. (2022b). Identification of prognostic genes in the pancreatic adenocarcinoma immune microenvironment by integrated bioinformatics analysis. *Cancer Immunol. Immunother.* 71 (7), 1757–1769. doi:10.1007/s00262-021-03110-3

Wang, J. Q., and Liu, H. (2022). The roles of junctional adhesion molecules (JAMs) in cell migration. *Front. Cell And Dev. Biol.* 10, 843671. doi:10.3389/fcell.2022.843671

Wang, X., Zhou, B., Xia, Y., Zuo, J., Liu, Y., Bi, X., et al. (2021a). A methylation-based nomogram for predicting survival in patients with lung adenocarcinoma. *BMC Cancer* 21 (1), 801. doi:10.1186/s12885-021-08539-4

Wang, Y., Hou, K., Jin, Y., Bao, B., Tang, S., Qi, J., et al. (2021b). Lung adenocarcinoma-specific three-integrin signature contributes to poor outcomes by metastasis and immune escape pathways. *J. Transl. Int. Med.* 9 (4), 249–263. doi:10.2478/jtim-2021-0046

Wautier, J. L., and Wautier, M. P. (2022). Vascular permeability in diseases. *Int. J. Of Mol. Sci.* 23 (7), 3645. doi:10.3390/ijms23073645

Yao, S. H., Nguyen, T. V., Rolfe, A., Agrawal, A. A., Ke, J. Y., Peng, S. Y., et al. (2020). Small molecule inhibition of CPS1 activity through an allosteric pocket. *Cell Chem. Biol.* 27 (3), 259–268. doi:10.1016/j.chembiol.2020.01.009

Zeng, J., Wu, Z. Y., Luo, M. J., Xu, X., Bai, W. J., Xie, G. J., et al. (2023). Development and validation of an endoplasmic reticulum stress long non-coding RNA signature for the prognosis and immune landscape prediction of patients with lung adenocarcinoma. *Front. Genet.* 14, 1024444. doi:10.3389/fgene.2023.1024444

Zhang, S., Liu, D., Ning, X., Zhang, X., Lu, Y., Zhang, Y., et al. (2023a). A signature constructed based on the integrin family predicts prognosis and correlates with the tumor microenvironment of patients with lung adenocarcinoma. *J. Environ. Pathol. Toxicol. Oncol.* 42 (2), 59–77. doi:10.1615/JEnvironPatholToxicolOncol.2022046232

Zhang, Z., Zhu, H., Wang, X., Lin, S., Ruan, C., and Wang, Q. (2023b). A novel basement membrane-related gene signature for prognosis of lung adenocarcinomas. *Comput. Biol. Med.* 154, 106597. doi:10.1016/j.compbiomed.2023.106597

Zhao, Y. L., Zhang, J. X., Yang, J. J., Wei, Y. B., Peng, J. F., Fu, C. J., et al. (2022). MiR-205-5p promotes lung cancer progression and is valuable for the diagnosis of lung cancer. *Thorac. Cancer* 13 (6), 832–843. doi:10.1111/1759-7714.14331

Zhao, Z., He, B., Cai, Q., Zhang, P., Peng, X., Zhang, Y., et al. (2021). Combination of tumor mutation burden and immune infiltrates for the prognosis of lung adenocarcinoma. *Int. Immunopharmacol.* 98, 107807. doi:10.1016/j.intimp.2021.107807

Zuo, Z. C., Zhang, G. C., Song, P., Yang, J., Li, S. T., Zhong, Z., et al. (2021). Survival nomogram for stage IB non-small-cell lung cancer patients, based on the SEER database and an external validation cohort. *Ann. Of Surg. Oncol.* 28 (7), 3941–3950. doi:10.1245/s10434-020-09362-0

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 24 May 2024
DOI 10.3389/fmolb.2024.1351888

Check for updates

OPEN ACCESS

EDITED BY
Georgina Gonzalez-Avila,
National Institute of Respiratory Diseases-Mexico (INER), Mexico

REVIEWED BY
Yun-Sheng Chen,
University of Illinois at Urbana-Champaign, United States
Sowmya Ramesh,
Johns Hopkins University, United States

*CORRESPONDENCE
Ibrahim Y. Hachim,
ihachim@sharjah.ac.ae

†These authors have contributed equally to this work

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

RECEIVED 07 December 2023
ACCEPTED 06 May 2024
PUBLISHED 24 May 2024

CITATION
Al Shareef Z, Hachim MY, Bouzid A, Talaat IM, Al-Rawi N, Hamoudi R and Hachim IY (2024), The prognostic value of Dickkopf-3 (Dkk3), TGFB1 and ECM-1 in prostate cancer. *Front. Mol. Biosci.* 11:1351888. doi: 10.3389/fmolb.2024.1351888

COPYRIGHT
© 2024 Al Shareef, Hachim, Bouzid, Talaat, Al-Rawi, Hamoudi and Hachim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The prognostic value of Dickkopf-3 (Dkk3), TGFB1 and ECM-1 in prostate cancer

Zainab Al Shareef[1,2], Mahmood Y. Hachim[3†], Amal Bouzid[1†], Iman M. Talaat[1,2,4], Natheer Al-Rawi[1,5], Rifat Hamoudi[1,2,6‡] and Ibrahim Y. Hachim[1,2*‡]

[1]Research Institute of Medical and Health Sciences, University of Sharjah, Sharjah, United Arab Emirates, [2]College of Medicine, University of Sharjah, Sharjah, United Arab Emirates, [3]College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates, [4]Faculty of Medicine, Alexandria University, Alexandria, Egypt, [5]College of Dental Medicine, University of Sharjah, Sharjah, United Arab Emirates, [6]Division of Surgery and Interventional Science, University College London, London, United Kingdom

Prostate cancer (PCa) is considered one of the most common cancers worldwide. Despite advances in patient diagnosis, management, and risk stratification, 10%–20% of patients progress to castration-resistant disease. Our previous report highlighted a protective role of Dickkopf-3 (DKK3) in PCa stroma. This role was proposed to be mediated through opposing extracellular matrix protein 1 (ECM-1) and TGF-β signalling activity. However, a detailed analysis of the prognostic value of DKK3, ECM-1 and members of the TGF-β signalling pathway in PCa was not thoroughly investigated. In this study, we explored the prognostic value of DKK3, ECM-1 and TGFB1 using a bioinformatical approach through analysis of large publicly available datasets from The Cancer Genome Atlas Program (TGCA) and Pan-Cancer Atlas databases. Our results showed a significant gradual loss of *DKK3* expression with PCa progression ($p < 0.0001$) associated with increased DNA methylation in its promoter region ($p < 1.63E-12$). In contrast, patients with metastatic lesions showed significantly higher levels of *TGFB1* expression compared to primary tumours ($p < 0.00001$). Our results also showed a marginal association between more advanced tumour stage presented as positive lymph node involvement and low *DKK3* mRNA expression ($p = 0.082$). However, while *ECM1* showed no association with tumour stage ($p = 0.773$), high *TGFB1* expression showed a significant association with more advanced stage presented as advanced T3 stage compared to patients with low *TGFB1* mRNA expression ($p < 0.001$). Interestingly, while *ECM1* showed no significant association with patient outcome, patients with high *DKK3* mRNA expression showed a significant association with favourable outcomes presented as prolonged disease-specific ($p = 0.0266$), progression-free survival ($p = 0.047$) and disease-free ($p = 0.05$). In contrast, high *TGFB1* mRNA expression showed a significant association with poor patient outcomes presented as shortened progression-free ($p = 0.00032$) and disease-free survival ($p = 0.0433$). Moreover, *DKK3, TGFB1* and *ECM1* have acted as immune-associated genes in the PCa tumour microenvironment. In conclusion, our findings showed a distinct prognostic value for this three-gene signature in PCa. While both DKK3 and

TGFB1 showed a potential role as a clinical marker for PCa stratification, ECM1 showed no significant association with the majority of clinicopathological parameters, which reduce its clinical significance as a reliable prognostic marker.

KEYWORDS

prostate cancer, Dickkopf-3, TGF-β signalling pathway, ECM-1, prognosis

Introduction

Prostate cancer (PCa) is still one of the most common male cancers (Greenberg et al., 2013) and one of the top five leading causes of death (Rawla, 2019; Barsouk et al., 2020). Moreover, many reports have highlighted an increase in the number of PCa cases diagnosed annually with the prediction that those numbers might increase in the near future (Mistry et al., 2011). An important element in determining the optimal management plan for patients with PCa is risk stratification (Greenberg et al., 2013; Yamazaki et al., 2021). Gleason score, PSA levels and clinical stage are considered important elements in risk stratification and prediction of tumour progression as well as recurrence (Yamazaki et al., 2021). The Gleason score is one of the main components of PCa prognosis and a risk stratification tool and is mainly based on the assessment of architectural features and glandular de-differentiation (Chen and Zhou, 2016; Tagai et al., 2019; van Leenders et al., 2020).

PCa patients may develop resistance to initial hormone therapy (10%–20%) and progress to a more advanced stage known as a castration-resistant disease (Vellky and Ricke, 2020; Harris et al., 2009; Yamazaki et al., 2021). This presents a clinical challenge, and there is a need to discover new prognostic biomarkers that can predict tumour progression and recurrence accurately (Frantzi et al., 2020).

Many reports highlighted the important role of cancer cell interactions with the reactive stroma in determining tumour behaviour, spread and progression. Some factors in the tumour microenvironment have tumour-inhibitory effects that might improve patient prognosis (Baghban et al., 2020), despite the stromal compartment in many cancers being pro-tumourigenic (Lopes-Coelho et al., 2018; Valkenburg et al., 2018). Recently, we revealed a protective role of stromal Dickkopf-3 (DKK3) in PCa (Al Shareef et al., 2018). Moreover, we also observed an inverse correlation between DKK3 and transforming growth Factor Beta Induced 1 (TGFB1) expression and that the activity of DKK3 was affected in different ways by TGFB1 and ECM-1 (Al Shareef et al., 2018; Niehrs, 2006; Kikuchi et al., 2021).

DKK3 is a member of the DKK family, secreted glycoproteins that negatively regulate Wnt signalling, although in the case of DKK3, this must be through a different mechanism (Veeck and Dahl, 2012). This might be attributed to the distinct amino acid sequence of DKK3, compared to other DKK family members, in the region of the LRP6-binding site (Niehrs, 2006; Kikuchi et al., 2021). Reports suggest that the tumour suppressive activity of DKK3 is related to its negative regulation of the β-catenin activity (Lee et al., 2020).

Previous reports highlighted a reduction in DKK3 expression in aggressive human cancer cells including basal breast cancer, melanoma, and hepatocellular carcinoma (HCC) through the promotor hypermethylation (Veeck and Dahl, 2012). Moreover, ectopic overexpression of DKK3 in some cancer cell lines inhibits their proliferation or induces apoptosis. Together, these studies suggest DKK3 has a tumour-suppressive function in various cancers (Hsieh et al., 2004; Kuphal et al., 2006; Lorsy et al., 2016). In contrast, some other reports highlight a tumour-promoting role of DKK3, for example, in oesophageal adenocarcinoma and oral squamous cell carcinoma (SCC), where it increases tumour cell proliferation and migration (Katase et al., 2013; Wang et al., 2015).

In the current study, we performed a comprehensive analysis of the prognostic value of *DKK3, TGFB1* and *ECM-1* as a multiple-gene prognostic signature in PCa using bioinformatics approaches.

Materials and methods

Gene expression evaluation in PCa nad normal samples

The publicly available application TNM plot (https://tnmplot.com/analysis/) was used to evaluate the mRNA expression levels of the genes of interest in normal (n = 106), tumour (n = 283) and metastatic (n = 6) samples (Bartha and Gyorffy, 2021). For statistical significance, the Kruskal–Wallis test was used to compare the gene expression levels among normal, tumour and metastatic tissue with $p < 0.01$ used as a cut-off value for statistical significance. In addition, using cBioPortal (cbioportal.org), we retrieved and analysed mRNA data (RNA SeqV2) of 493 patients with prostate adenocarcinoma from the Cancer Genome Atlas (TGCA) and Pan-Cancer Atlas. Log-rank test p-values were used to evaluate the statistical significance between the low and the high expression level (classified according to above and below mean mRNA expression of each biomarker. The Chi-squared test was used to evaluate the association between each biomarker and clinicopathological parameters including tumour stage as well as patient outcome. Moreover, the DNA methylation profiles retrieved from the TCGA database were assessed in order to evaluate the *DKK3* promoter DNA methylation levels using the UALCAN portal. DNA methylation analysis was carried out using normalized beta values from 502 PCa patients and 50 healthy controls considering beta value cut-offs of hyper-methylation [beta value: 0.7–0.5] or hypo-methylation [beta-value: 0.3–0.25].

Evaluation of the prognostic capacity

To evaluate the prognostic capacity of each marker of interest, we used the CANCERTOOL database, (http://web.bioinformatics.cicbiogune.es/CANCERTOOL), which is a comprehensive portal that aims to investigate different genes and their association with clinical data including disease progression, pathologic, and molecular characteristics, to evaluate the expression levels of our markers in normal *versus* malignant tissues.

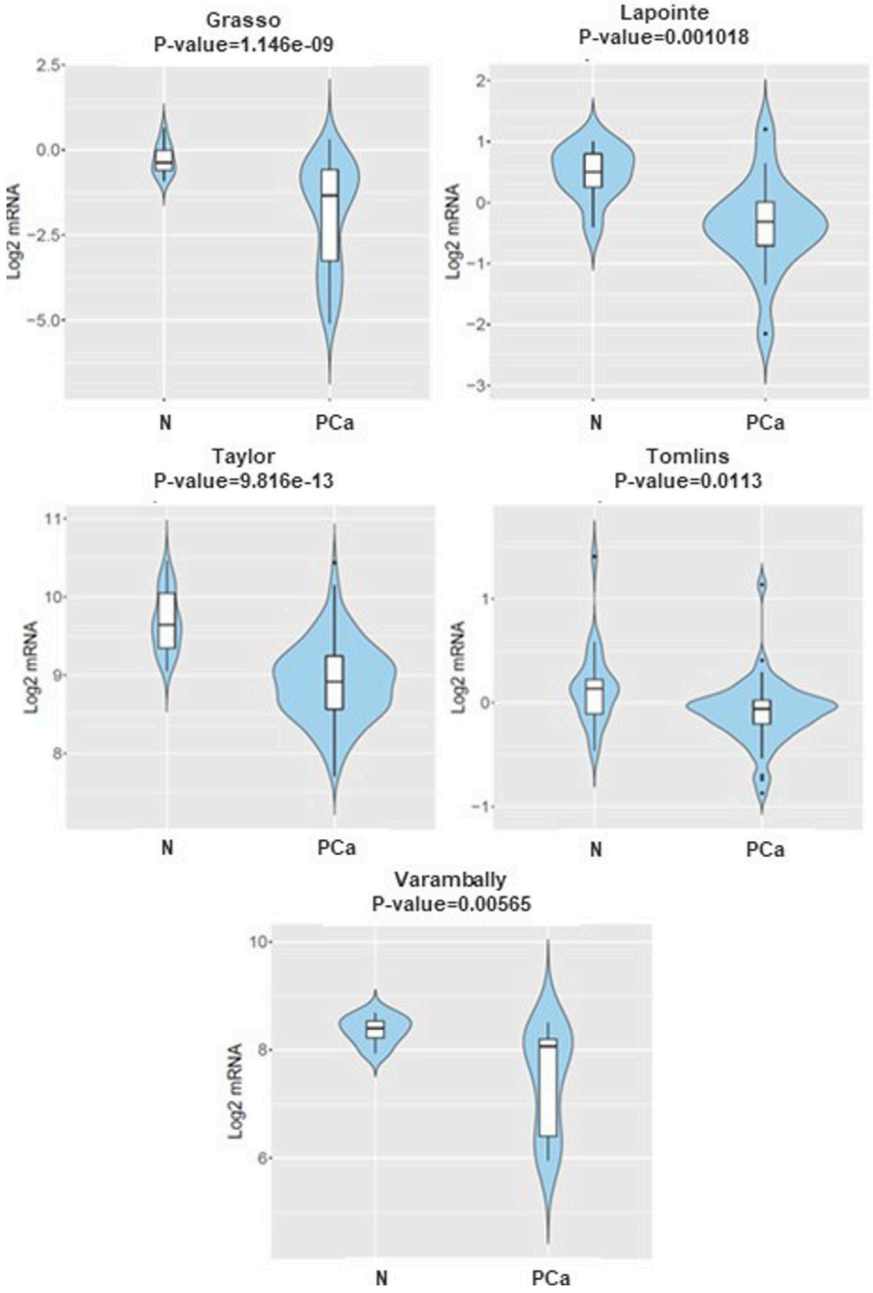


FIGURE 1
The mRNA expression of DKK3 in normal versus prostate cancer samples CANCERTOOL database. Violin plots show the expression of DKK3 in non-tumoral tissue (N) and primary tumours (PT) in various datasets of the CANCERTOOL database.

Correlation analysis of gene expression and immune infiltration in PCa tumour microenvironment

The relationship between *DKK3, TGFB1* and *ECM1* and immune cell infiltration in the in PCa microenvironment was explored using the Tumour IMmune Estimation Resource (TIMER) (https://cistrome.shinyapps.io/timer/), which consists of deconvolution and comprehensive analysis of tumour-infiltrating immune cells in various types of cancer. The association between *DKK3, TGFB1* and *ECM1* expression and the abundances of immune infiltrates including B cells, $CD4^+$ T cells, $CD8^+$ T cells, neutrophils, macrophages, and dendritic cells in the PCa tumour microenvironment was assessed, using Spearman's test ($p < 0.05$).

Construction of gene-gene interaction network and functional enrichment analysis

To assess the functional association between the studied genes *DKK3, TGFB1* and *ECM1* and their related connected genes, the gene-gene interaction network was constructed based on a large data

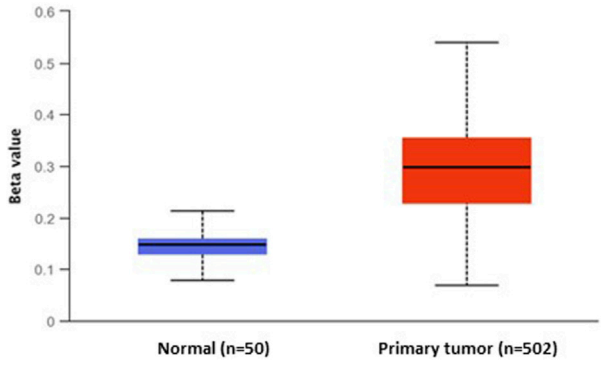


FIGURE 2
The differential DNA methylation levels of DKK3 promoter between normal and primary prostate tumours.

of functional-related data sets including genetic and protein interactions, physical interaction, co-expression, co-localization, and common pathway, using GeneMANIA bioinformatics tool with defaults parameters (Warde-Farley et al., 2010). Genes can be linked by the interacted network based on variable attributes. The network includes nodes that represent genes while edges represent connections. Besides, functional enrichment analysis on Gene Ontology terms including biological process (BP), cellular component (CC), and molecular function (MF) categories was conducted using the Enrichr database (Xie et al., 2021) to explore the biological significance and common pathways between the *DKK3, TGFB1* and *ECM1* genes. The hypergeometric distribution cut-off for the functional enrichment analyses was a p-value of <0.05.

Results

DKK3 is downregulated in PCa samples compared to normal tissue

To improve our understanding of the role of *DKK3, TGFB1* and *ECM-1* genes in tumourigenesis, we initially investigated the expression levels of *DKK3* in PCa samples compared to their normal counterparts using the CANCERTOOL database. Different datasets were used including Glinksy et al., Grasso et al., Lapointe et al., Taylor et al., TCGA (RNA-seq), Tomlins et al., and Varambally et al. (Figure 1). Our results showed a significant lower level of *DKK3* in PCa samples compared to samples obtained from their normal counterparts. Similarly, *TGFB1* showed also significant downregulation in PCa samples compared to normal counterparts. In contrast, ECM-1 levels showed no significant differences between the normal and PCa samples (Supplementary Figure S1).

Moreover, the evaluation of DNA methylation levels in the promoter region of *DKK3* using TCGA publicly available methylome data showed significant DNA methylation differences between primary PCa and normal samples (p < 1.62 E-12). Indeed, the *DKK3* promoter was highly hypermethylated among PCa cases (n = 502), while it was hypomethylated in the normal samples (n = 50). Our findings indicate that there is a notable decrease in the expression of DKK3 in PCa samples from various datasets, indicating that the promoter region of the DKK3 gene may be affected by DNA methylation, leading to its silencing (Figure 2).

The value of DKK3 expression as a marker of tumour progression in PCa samples

To improve our understanding of the role of *DKK3, ECM-1* and *TGF-β* in PCa tumourigenesis, we evaluated their expression in samples representing different stages of PCa progression (Figure 3). To achieve this, we analysed *DKK3, ECM-1* and *TGFB1* expression in normal prostate, primary tumours and samples obtained from more advanced diseases presented as patients with tumour metastases. Interestingly, our results showed a gradual and significant downregulation of *DKK3* expression during tumour progression. Indeed, the expression levels of *DKK3* showed significant downregulation in tumour samples compared to normal. Further downregulation was observed between tumour samples and samples obtained from patients with metastatic tumours (p < 0.0001). In contrast, while *TGFB1* expression was comparable between normal and tumour samples, metastatic lesions showed a significantly higher level of *TGFB1*, compared to both primary tumour and normal samples (p = 0.04). Moreover, while *ECM1* expression levels were higher in tumour samples compared to normal tissue, *ECM1* expression was significantly lower in metastatic samples compared to tumour, as well as normal samples (p < 0.0001). As a reference, we also investigated the expression levels of the classical marker *PSA* (*KLK3*) in the same samples. Our findings showed that while PSA levels initially increased in the tumour samples compared to normal, those levels were significantly lower in metastatic samples, compared to tumours (p < 0.0001). In summary, our results highlight the possible benefits of *DKK3* expression as a marker of PCa progression, owing to its gradual loss during tumourigenesis.

The association between *DKK3, ECM-1 and TGFB1* mRNA expression and tumour staging and grade in a large patient cohort

Next, we evaluated the mRNA expression of the three genes in association with the tumour stage using the prostate adenocarcinoma (TCGA, PanCancer Atlas) cohort (Figure 4). Our results showed a marginal significance (p = 0.082) between *DKK3* mRNA expression and LN involvement (Figure 4A). Patients with higher *DKK3* mRNA expression showed less chance of positive lymph node involvement (N1) compared to patients with lower *DKK3* mRNA expression. In comparison, while *ECM1* showed no association with tumour stage (p = 0.773), *TGFB1* showed a significant association with tumour stage (Figure 4B). Patients with high *TGFB1* mRNA expression presented with the more advanced T3A (tumour spread outside the prostate, but not invading seminal vesicles) and T3B (tumour spread outside prostate with seminal vesicles invasion) stages, compared to patients with low *TGFB1* mRNA expression (p < 0.001).

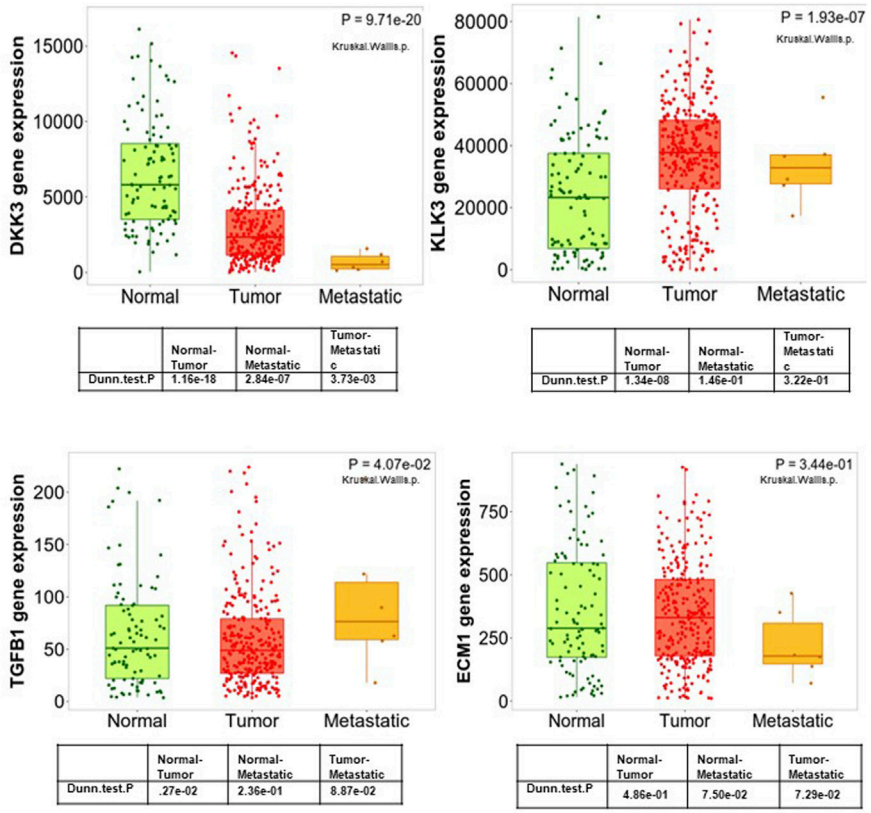


FIGURE 3
Boxplots of the mRNA expression of DKK3, ECM1, TGFB1 and PSA (KLK3) in a large patient cohort using the TNM plot database.

Interestingly, while patients with high *DKK3* mRNA expression showed marginally significant lower chance of having recurrence presented as new neoplastic events post initial therapy ($p = 0.082$), Patients with higher *TGFB1* mRNA expression showed a significant higher chance of recurrence presented as new neoplastic events post initial therapy compared with patients with low *TGFB1* mRNA expression ($p < 0.001$). In contrast, *ECM1* mRNA expression showed no significant association with new neoplastic events post initial therapy ($p = 0.932$) (Figure 4).

The analysis between the mRNA expression of the three markers with tumor grade showed that while both DKK3 and ECM1 showed no significant variation in samples from different Gleason grade, TGFB1 mRNA expression was significantly higher in patients presented with higher gleason score (G8,9&10) in the TCGA dataset ($p = 0.0001$) (Supplementary Figure S2).

The correlation between DKK3 mRNA expression and both TGFB1 and ECM1 in prostate cancer samples

Further analysis was performed to investigate the correlation between DKK3 expression and TGFB1 and ECM1 in prostate cancer samples. Indeed, this might better demonstrate the interplay between DKK3 and these markers. While our analysis showed no significant correlation between ECM1 and DKK3 mRNA levels (R = 0.07, $p = 0.23$), our results showed a significant and negative correlation between DKK3 mRNA expression levels and TGFB1 (R = -0.17, $p < 0.0001$). Additional analysis was performed to investigate the association between DKK3 and KLK3, which is most relevant marker for PCa diagnosis and management, and its expression is regulated by the upstream markers, such as AR. Our analysis showed no significant correlation between DKK3 mRNA expression levels and KLK3 expression (R = -0.09, $p = 0.14$) (Supplementary Figure S3).

The association between DKK3, ECM1 and TGFB1 and patient outcome

Next, we evaluated the association between our markers and patient outcome presented as progression-free survival (Figure 5). While ECM-1 showed no significant association with patient outcome progression-free survival ($p = 0.986$), disease specific ($p = 0.776$) and disease free survival ($p = 0.167$). Both DKK3 and TGFB1 showed a significant and contrasting association with patient outcomes. While patients with high DKK3 mRNA expression showed a significant association with favourable outcomes presented as prolonged progression-free survival ($p = 0.0470$), disease specific ($p = 0.0266$) and disease free survival ($p = 0.050$), high TGFB1 mRNA expression showed a significant association with poor patient outcome presented as shortened progression-free survival ($p = 3.267e-4$) and disease free survival ($p = 0.0433$).

10.3389/fmolb.2024.1351888

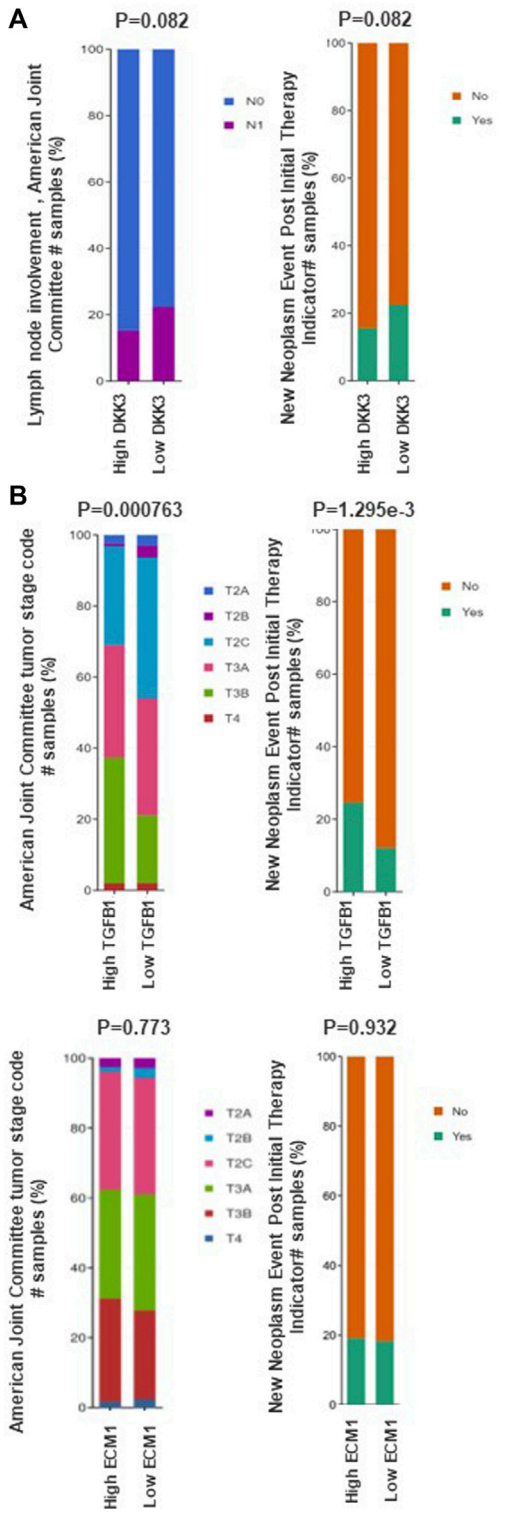


FIGURE 4
The association between DKK3, ECM1 and TGFB1 mRNA expression and clinicopathological features of prostate cancer. **(A)** The association between DKK3 and prostate cancer clinicopathological characteristics using data from prostate adenocarcinoma (TCGA) through cBioPortal tool. **(B)** The association between ECM1 and TGFB1 mRNA expression prostate cancer clinicopathological characteristics using data from prostate adenocarcinoma (TCGA) through cBioPortal tool.

DKK3, TGFB1 and *ECM1* act as the immune-associated genes in PCa

We were also interested to analyze the relationship between *DKK3, TGFB1* and *ECM*1 and immune cell infiltration in prostate cancer (Figure 6). The results showed that gene expression level against tumor purity is highly significant for the three biomarkers with negative Spearman's rho values (−0.5, −0.329 and −0.41), respectively in *DKK3, TGFB1* and *ECM1* suggesting that these biomarkers are highly expressed in the microenvironment of PCa (Figure 6). Moreover, all biomarkers have a positive significant correlation with B cells, $CD4^+$ T cells, $CD8^+$ T cells, Neutrophils, Macrophages, and Dendritic cells. Importantly, *DKK3* has the highest correlation with macrophage at 60%, while *ECM1* showed the highest correlation at about 54% with dendritic cells and *TGFB1* has the highest correlation with neutrophil and dendritic cells at 60% and 68% respectively. Importantly, all biomarkers had a better correlation with CD4 than CD8.

Biological functions and interconnections of the *DKK3, TGFB1 and ECM1* genes

To explore the potential mechanisms that involve *DKK3, TGFB1 and ECM1* in the carcinogenesis of PCa, we explored GeneMania to build up a gene–gene interaction network for these biomarkers. As shown in Figure 7, the network revealed that *DKK3, TGFB1* and *ECM1* share many genes within the same gene family or similar protein domain under a series of cooperators suggesting that there is a possibility of the interplay between the 3 markers through interaction with proteins from a similar family. For instance, *MMP9* presents a key marker in the network that showed physical interconnection with *TGFB1* and *ECM1*, and co-expression with *DKK3* suggesting the potential direct/indirect connections between these particular biomarkers. Knowing that *MMP9* codes for a matrix metalloproteinase protein that is reported to promote metastasis and angiogenesis through decomposition of the extracellular matrix in several tumours which further supports the possible involvement of *DKK3, TGFB1* and *ECM1* in the PCa progression.

Furthermore, gene ontologies including biological process and molecular function were performed to reveal the functional enrichment of *DKK3, TGFB1* and *ECM1* and their co-operators. The significant annotated pathways were represented in Figure 7. Interestingly, it identifies cytokine activity (GO:0005125, *p*-value = 0.02) and macrophage-related pathways such as negative regulation of macrophage cytokine production (GO:0010936, *p*-value = 7.5E-04) among the differentially regulated pathways involving mainly *TGFB1* as a key player which further supports the findings of immune cell infiltration in PCa (Figure 8). Additionally, the regulation of canonical Wnt signaling pathway (GO:0060828, *p*-value = 4.74E-04) was identified among the relevant significantly regulated biological processes involving *TGFB1* and *DKK3,* which could be of a pivotal role in PCa pathogenesis, proliferation and resistance to treatment. Taken together, these findings highlight the potential utility of *DKK3, TGFB1 and ECM1*as prognostic markers for PCa.

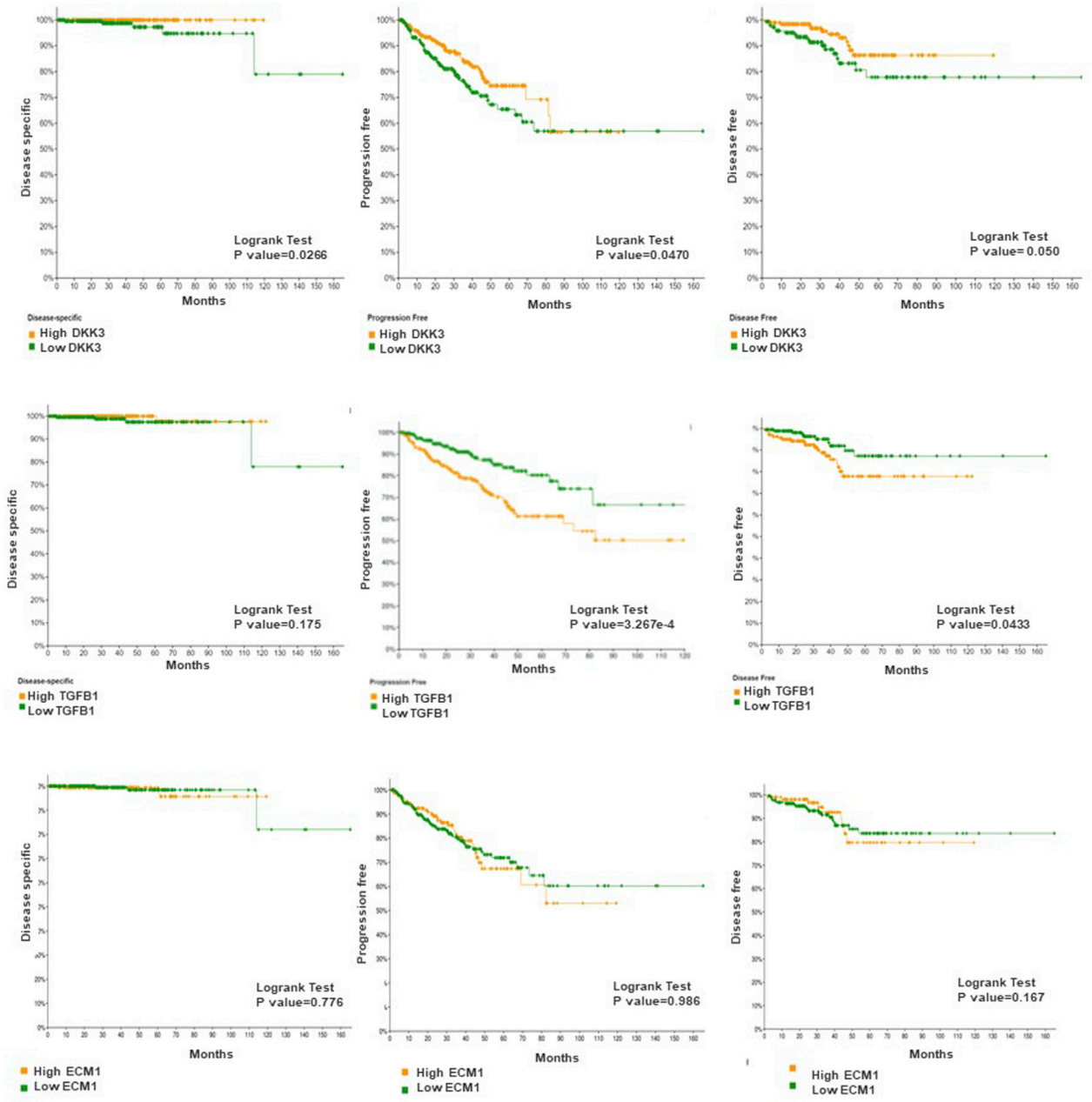


FIGURE 5
The association between DKK3, ECM1 and TGFB1 mRNA expression and patient survival. Log-rank test p-values were used to evaluate the statistical significance between the low and the high expression level.

Discussion

Previous reports highlighted the role of the Dickkopf (Dkk) gene family in determining cell fate during embryonic development (Hoang et al., 2004; Krupnik et al., 1999; Hoang et al., 2004; Niehrs, 2006). Moreover, the *DKK3* gene was found to encode essential intracellular regulators of cellular proliferation (Leonard et al., 2017). A tumour suppressor role of DKK3 was suggested since ectopic expression of Dkk3 was found to reduce cancer growth, in contrast, loss of Dkk3 expression induces hyperproliferation of cells (Veeck and Dahl, 2012; Fujii et al., 2014). In addition, other reports also found lower levels of DKK3 in various tumours including gastric, ovarian, lung, bladder and PCa (Kurose et al., 2004; Chim et al., 2008; Yue et al., 2008; Ueno et al., 2011).

Our earlier report proposed a tumour suppressor role of *DKK3* in PCa through opposing *TGFBI* and *ECM-1* function (Al Shareef et al., 2018). However, a comprehensive analysis of the prognostic role of *DKK3*, as well as its relationship with the expression of *TGFB1* and *ECM1* in a large sample size of PCa was not investigated. Here we have analysed the prognostic value of *DKK3, TGFB1* and *ECM1* and their association with well-known prognostic clinicopathological parameters in PCa using multiple cohorts in

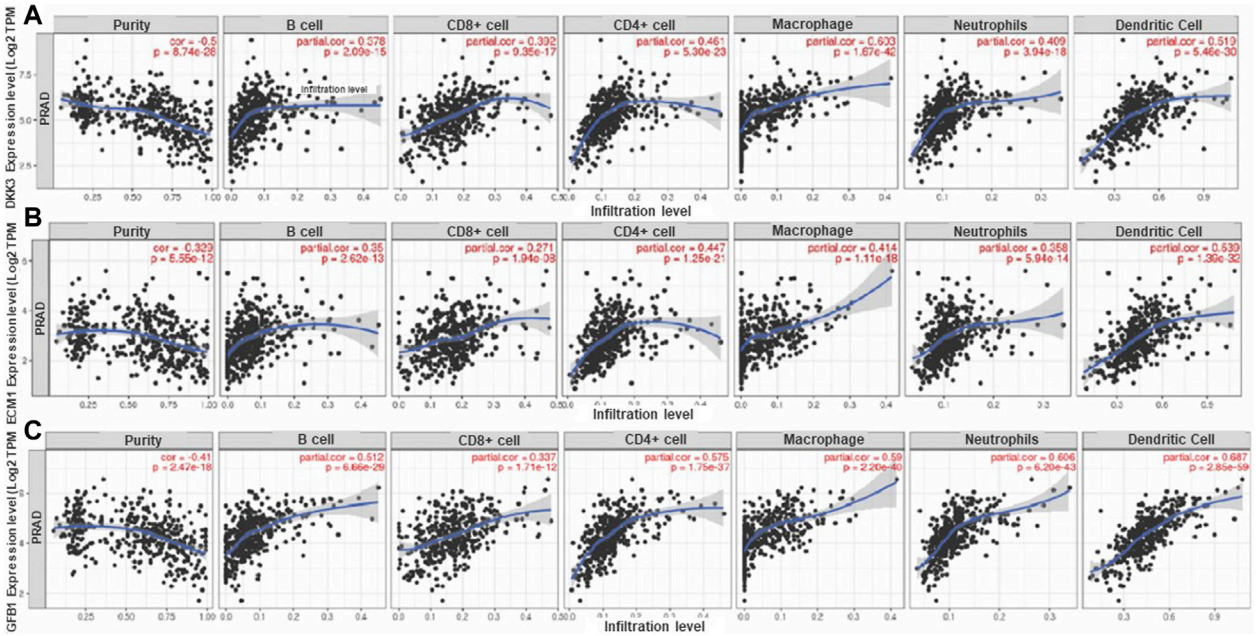


FIGURE 6
Correlation of **(A)** *DKK3*, **(B)** *ECM1* and **(C)** *TGFB1* gene expression level with immune cell infiltration in PCa. The scatterplots displayed Spearman's rho value correlation and statistical significance showing the potential interplay between *DKK3, TGFBI* and *ECMI* biomarkers and immune cells (B-Cell, CD8+ T Cell, CD4+ T Cell, Macrophage, Neutrophil and Dendritic Cell) in the prostate cancer tumor microenvironment. The log2 TPM gene expression values are presented on the y-axis, the average immune cell infiltration levels are presented on the x-axis. The blue curve and gray area in the figures show the general trend direction. TPM: transcripts per million. (Data generated from TIMER Webtool).

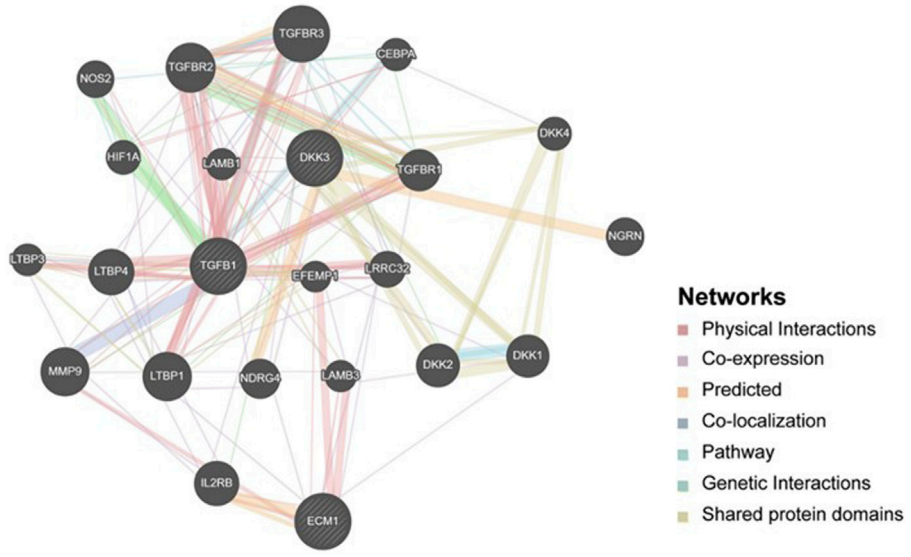


FIGURE 7
Gene-Gene interaction network including *DKK3*, *TGFB1* and *ECM1*. It shows interaction strength (edge thickness), interaction type (colour), multiple edges between nodes, and protein score (node size) defined using a stylesheet constructed with GeneMANIA. The interconnections between studied genes were evaluated based on physical interaction, co-expression, predicted, co-localization, common pathway, genetic interaction and shared protein domains.

various publicly available databases. Additionally, we conducted an investigation into the pathways involved in the interaction between DKK3, TGFB1, and ECM1.

This study reveals that *ECM1* and *TGFB1* are upregulated in cancerous tissues in comparison to normal tissues. Conversely, *DKK3* expression was found to be downregulated in malignant

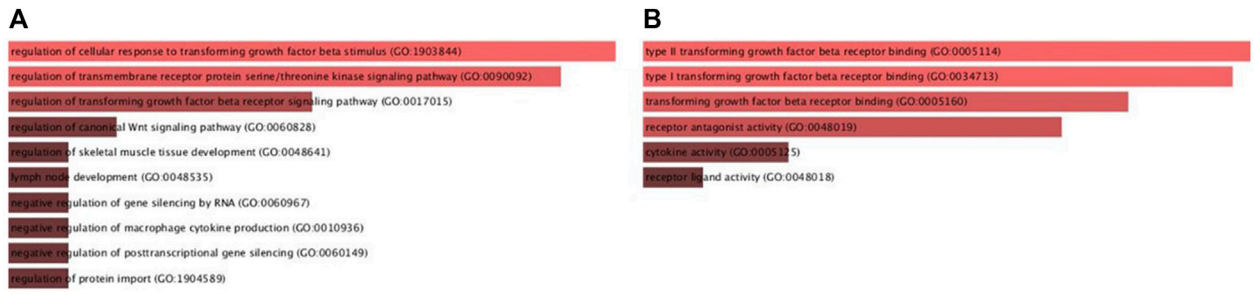


FIGURE 8
Gene ontologies enrichment analysis **(A)** Biological Process and **(B)** Molecular Functions of *DKK3, TGFBI* and *ECMI* using Enricher database.

tissue in comparison to normal tissue. These results align with prior research that has demonstrated decreased expression of *DKK3* in various types of cancer as compared to healthy tissue (Kurose et al., 2004; Chim et al., 2008; Yue et al., 2008; Ueno et al., 2011). This also supports our previous notion that DKK3 might exert a protective role against PCa (Al Shareef et al., 2018) and highlighted a possible role of its loss in PCa tumourigenesis.

Our results also showed a gradual loss of DKK3 mRNA expression between and during cancer progression with its least expression in samples obtained from patients with metastasis. Interestingly, our results showed that DKK3 expression was more able to predict tumour progression than KLK3 (coding for PSA). Indeed, PSA is considered one of the most relevant markers for PCa diagnosis and management (Bonk et al., 2019). Moreover, recent reports highlighted KLK3 as a possible predictive marker for molecular lymph-node staging in PCa patients (Lunger et al., 2021). However, our analysis showed that while KLK3 expression was significantly upregulated in primary cancer samples compared to normal, its expression was not significantly variable between metastatic and primary tumour samples. Indeed, this statement suggests that the clinical usefulness of PSA (KLK3) in assessing tumor progression is greatly restricted, and instead emphasizes the potential benefits of utilizing DKK3 expression as an indicator of cancer progression. In addition, our results also showed a significant association between low DKK3 expression and poor patient outcomes represented as shortened disease-specific, disease free and progression-free survival. While the association between DKK3 expression and patient outcome in PCa was not thoroughly investigated, recent reports showed a significant association between low DKK3 expression and reduced recurrence-free survival in breast cancer, cervical cancer, colorectal cancer, endometrial cancer and gastric cancer (Wang et al., 2012; Xu et al., 2012; Ryu et al., 2013; Lorsy et al., 2016).

Our results also highlighted a possible interplay between DKK3, TGFB1 and ECM1 in PCa through modulation of some biological processes and molecular functions including members of TGF-β signaling and the Wnt signaling pathway. Wnt Signaling was found to play an essential role in the modulation of PCa microenvironment to promote drug resistance as well as cancer stem cell expansion and self-renewal (Kypta and Waxman, 2012). Our analysis also identified pathways involved in the regulation of macrophage cytokine production as one of the enriched pathways in the regulation of DKK3, TGFB1 and ECM1 function in PCa. This highlighted a possible mechanism through which those genes modulate PCa microenviroment through regulation of macrophage function, which is now believed to play a role in PCa progression and metastatic cascades of PCa (Lo and Lynch, 2018). Further studies are still needed for comprehensive analysis for the role of DKK3, TGFB1 and ECM1 in PCa including the use of human single cell/spatial analysis of human Pca. Indeed this might improve our understanding to the prognostic role and their role in pathogenesis of prostate including progression and metastasis.

Conclusion

To summarize, our study utilized *in silico* analyses to investigate the potential utility of *DKK3*, *TGFB1* and *ECM1* as prognostic markers for PCa. Our findings suggest that while *DKK3*, *TGFB1* could serve as promising biomarkers for predicting disease progression and patient outcomes, *ECM1* expression showed no significant association with various clinicopathological parameters. For that reason, only *DKK3* and *TGFB1* can be used as a reliable prognostic markers in prostate cancer and the he incorporation of these biomarkers into clinical practice could enhance PCa patient stratification and facilitate the adoption of tailored therapeutic strategies for each individual. Furthermore, our analysis identified a group of key pathways that appear to be central in the interplay between DKK3, TGFB1 and ECM1, resulting in the modulation of the tumour microenvironment.

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

ZA: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Supervision, Writing–original draft, Writing–review and editing. MH: Data curation, Validation, Writing–original draft, Writing–review and editing, Conceptualization. AB: Data curation, Formal Analysis, Writing–original draft, Writing–review and editing. IT:

Validation, Writing–review and editing. NA-R: Writing–review and editing. RH: Data curation, Validation, Writing–review and editing. IH: Investigation, Methodology, Validation, Writing–original draft, Writing–review and editing, Conceptualization.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. ZA is funded by University of Sharjah, seed grant and targeted grant project number (1901090161). RH is funded by ASPIRE, the technology program management pillar of Abu Dhabi's Advanced Technology Research Council (ATRC), via the ASPIRE Precision Medicine Research Institute Abu Dhabi (VRI-20-10) grant.

Acknowledgments

We would like to thank Dr. Abdulqadir Alzarooni, Chief of ER and Surgical Services, Consultant & Chair of Urology for their support and contribution to this manuscript.

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/fmolb.2024.1351888/full#supplementary-material

References

Al Shareef, Z., Kardooni, H., Murillo-Garzon, V., Domenici, G., Stylianakis, E., Steel, J. H., et al. (2018). Protective effect of stromal Dickkopf-3 in prostate cancer: opposing roles for TGFBI and ECM-1. *Oncogene* 37 (39), 5305–5324. doi:10.1038/s41388-018-0294-0

Baghban, R., Roshangar, L., Jahanban-Esfahlan, R., Seidi, K., Ebrahimi-Kalan, A., Jaymand, M., et al. (2020). Tumor microenvironment complexity and therapeutic implications at a glance. *Cell. Commun. Signal. CCS* 18 (1), 59. doi:10.1186/s12964-020-0530-4

Barsouk, A., Padala, S. A., Vakiti, A., Mohammed, A., Saginala, K., Thandra, K. C., et al. (2020). Epidemiology, staging and management of prostate cancer. *Med. Sci. (Basel).* 8 (3), 28. doi:10.3390/medsci8030028

Bartha, A., and Gyorffy, B. (2021). TNMplot.com: a web tool for the comparison of gene expression in normal, tumor and metastatic tissues. *Int. J. Mol. Sci.* 22 (5), 2622. doi:10.3390/ijms22052622

Bonk, S., Kluth, M., Hube-Magg, C., Polonski, A., Soekeland, G., Makropidi-Fraune, G., et al. (2019). Prognostic and diagnostic role of PSA immunohistochemistry: a tissue microarray study on 21,000 normal and cancerous tissues. *Oncotarget* 10 (52), 5439–5453. doi:10.18632/oncotarget.27145

Chen, N., and Zhou, Q. (2016). The evolving Gleason grading system. *Chin. J. Cancer Res.* 28 (1), 58–64. doi:10.3978/j.issn.1000-9604.2016.02.04

Chim, C. S., Pang, R., and Liang, R. (2008). Epigenetic dysregulation of the Wnt signalling pathway in chronic lymphocytic leukaemia. *J. Clin. Pathol.* 61 (11), 1214–1219. doi:10.1136/jcp.2008.060152

Frantzi, M., Hupe, M. C., Merseburger, A. S., Schanstra, J. P., Mischak, H., and Latosinska, A. (2020). Omics derived biomarkers and novel drug targets for improved intervention in advanced prostate cancer. *Cancer. Diagn. (Basel)* 10 (9), 658. doi:10.3390/diagnostics10090658

Fujii, Y., Hoshino, T., and Kumon, H. (2014). Molecular simulation analysis of the structure complex of C2 domains of DKK family members and β-propeller domains of LRP5/6: explaining why DKK3 does not bind to LRP5/6. *Acta Med. Okayama* 68 (2), 63–78. doi:10.18926/AMO/52403

Greenberg, D. C., Wright, K. A., Lophathanon, A., Muir, K. R., and Gnanapragasam, V. J. (2013). Changing presentation of prostate cancer in a UK population--10 year trends in prostate cancer risk profiles in the East of England. *Br. J. Cancer* 109 (8), 2115–2120. doi:10.1038/bjc.2013.589

Harris, W. P., Mostaghel, E. A., Nelson, P. S., and Montgomery, B. (2009). Androgen deprivation therapy: progress in understanding mechanisms of resistance and optimizing androgen depletion. *Nat. Clin. Pract. Urol.* 6 (2), 76–85. doi:10.1038/ncpuro1296

Hoang, B. H., Kubo, T., Healey, J. H., Yang, R., Nathan, S. S., Kolb, E. A., et al. (2004). Dickkopf 3 inhibits invasion and motility of Saos-2 osteosarcoma cells by modulating the Wnt-beta-catenin pathway. *Cancer Res.* 64 (8), 2734–2739. doi:10.1158/0008-5472.can-03-1952

Hsieh, S. Y., Hsieh, P. S., Chiu, C. T., and Chen, W. Y. (2004). Dickkopf-3/REIC functions as a suppressor gene of tumor growth. *Oncogene* 23 (57), 9183–9189. doi:10.1038/sj.onc.1208138

Katase, N., Lefeuvre, M., Tsujigiwa, H., Fujii, M., Ito, S., Tamamura, R., et al. (2013). Knockdown of Dkk-3 decreases cancer cell migration and invasion independently of the Wnt pathways in oral squamous cell carcinomaderived cells. *Oncol. Rep.* 29 (4), 1349–1355. doi:10.3892/or.2013.2251

Kikuchi, A., Matsumoto, S., and Sada, R. (2021). Dickkopf signaling, beyond Wnt-mediated biology. *Semin. Cell. Dev. Biol.* 125, 55–65. doi:10.1016/j.semcdb.2021.11.003

Krupnik, V. E., Sharp, J. D., Jiang, C., Robison, K., Chickering, T. W., Amaravadi, L., et al. (1999). Functional and structural diversity of the human Dickkopf gene family. *Gene* 238 (2), 301–313. doi:10.1016/s0378-1119(99)00365-0

Kuphal, S., Lodermeyer, S., Bataille, F., Schuierer, M., Hoang, B. H., and Bosserhoff, A. K. (2006). Expression of Dickkopf genes is strongly reduced in malignant melanoma. *Oncogene* 25 (36), 5027–5036. doi:10.1038/sj.onc.1209508

Kurose, K., Sakaguchi, M., Nasu, Y., Ebara, S., Kaku, H., Kariyama, R., et al. (2004). Decreased expression of REIC/Dkk-3 in human renal clear cell carcinoma. *J. Urol.* 171 (3), 1314–1318. doi:10.1097/01.ju.0000101047.64379.d4

Kypta, R. M., and Waxman, J. (2012). Wnt/β-catenin signalling in prostate cancer. *Nat. Rev. Urol.* 9 (8), 418–428. doi:10.1038/nrurol.2012.116

Lee, E. J., Nguyen, Q. T. T., and Lee, M. (2020). Dickkopf-3 in human malignant tumours: a clinical viewpoint. *Anticancer Res.* 40 (11), 5969–5979. doi:10.21873/anticanres.14617

Leonard, J. L., Leonard, D. M., Wolfe, S. A., Liu, J., Rivera, J., Yang, M., et al. (2017). The Dkk3 gene encodes a vital intracellular regulator of cell proliferation. *PLoS One* 12 (7), e0181724. doi:10.1371/journal.pone.0181724

Lo, C. H., and Lynch, C. C. (2018). Multifaceted roles for macrophages in prostate cancer skeletal metastasis. *Front. Endocrinol. (Lausanne)* 9, 247. doi:10.3389/fendo.2018.00247

Lopes-Coelho, F., Gouveia-Fernandes, S., and Serpa, J. (2018). Metabolic cooperation between cancer and non-cancerous stromal cells is pivotal in cancer progression. *Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med.* 40 (2), 1010428318756203. doi:10.1177/1010428318756203

Lorsy, E., Topuz, A. S., Geisler, C., Stahl, S., Garczyk, S., von Stillfried, S., et al. (2016). Loss of Dickkopf 3 promotes the tumorigenesis of basal breast cancer. *PLoS One* 11 (7), e0160077. doi:10.1371/journal.pone.0160077

Lunger, L., Retz, M., Bandur, M., Souchay, M., Vitzthum, E., Jager, M., et al. (2021). KLK3 and TMPRSS2 for molecular lymph-node staging in prostate cancer patients

undergoing radical prostatectomy. *Prostate Cancer Prostatic Dis.* 24 (2), 362–369. doi:10.1038/s41391-020-00283-3

Mistry, M., Parkin, D. M., Ahmad, A. S., and Sasieni, P. (2011). Cancer incidence in the United Kingdom: projections to the year 2030. *Br. J. Cancer* 105 (11), 1795–1803. doi:10.1038/bjc.2011.430

Niehrs, C. (2006). Function and biological roles of the Dickkopf family of Wnt modulators. *Oncogene* 25 (57), 7469–7481. doi:10.1038/sj.onc.1210054

Rawla, P. (2019). Epidemiology of prostate cancer. *World J. Oncol.* 10 (2), 63–89. doi:10.14740/wjon1191

Ryu, S. W., Kim, J. H., Kim, M. K., Lee, Y. J., Park, J. S., Park, H. M., et al. (2013). Reduced expression of DKK3 is associated with adverse clinical outcomes of uterine cervical squamous cell carcinoma. *Int. J. Gynecol. Cancer* 23 (1), 134–140. doi:10.1097/IGC.0b013e3182754feb

Tagai, E. K., Miller, S. M., Kutikov, A., Diefenbach, M. A., Gor, R. A., Al-Saleem, T., et al. (2019). Prostate cancer patients' understanding of the gleason scoring system: implications for shared decision-making. *J. Cancer Educ.* 34 (3), 441–445. doi:10.1007/s13187-018-1320-1

Ueno, K., Hirata, H., Majid, S., Chen, Y., Zaman, M. S., Tabatabai, Z. L., et al. (2011). Wnt antagonist DICKKOPF-3 (Dkk-3) induces apoptosis in human renal cell carcinoma. *Mol. Carcinog.* 50 (6), 449–457. doi:10.1002/mc.20729

Valkenburg, K. C., de Groot, A. E., and Pienta, K. J. (2018). Targeting the tumour stroma to improve cancer therapy. *Nat. Rev. Clin. Oncol.* 15 (6), 366–381. doi:10.1038/s41571-018-0007-1

van Leenders, G., Verhoef, E. I., and Hollemans, E. (2020). Prostate cancer growth patterns beyond the Gleason score: entering a new era of comprehensive tumour grading. *Histopathology* 77 (6), 850–861. doi:10.1111/his.14214

Veeck, J., and Dahl, E. (2012). Targeting the Wnt pathway in cancer: the emerging role of Dickkopf-3. *Biochim. Biophys. Acta* 1825 (1), 18–28. doi:10.1016/j.bbcan.2011.09.003

Vellky, J. E., and Ricke, W. A. (2020). Development and prevalence of castration-resistant prostate cancer subtypes. *Neoplasia (New York, NY)* 22 (11), 566–575. doi:10.1016/j.neo.2020.09.002

Wang, W., Zhu, W., Xu, X. Y., Nie, X. C., Yang, X., Xing, Y. N., et al. (2012). The clinicopathological significance of REIC expression in colorectal carcinomas. *Histol. Histopathol.* 27 (6), 735–743. doi:10.14670/HH-27.735

Wang, Z., Lin, L., Thomas, D. G., Nadal, E., Chang, A. C., Beer, D. G., et al. (2015). The role of Dickkopf-3 overexpression in esophageal adenocarcinoma. *J. Thorac. Cardiovasc Surg.* 150 (2), 377–385. doi:10.1016/j.jtcvs.2015.05.006

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38 (Web Server issue), W214–W220. doi:10.1093/nar/gkq537

Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., Jenkins, S. L., et al. (2021). Gene set knowledge discovery with Enrichr. *Curr. Protoc.* 1 (3), e90. doi:10.1002/cpz1.90

Xu, X. Y., Xia, P., Yu, M., Nie, X. C., Yang, X., Xing, Y. N., et al. (2012). The roles of REIC gene and its encoding product in gastric carcinoma. *Cell. Cycle* 11 (7), 1414–1431. doi:10.4161/cc.19823

Yamazaki, H., Suzuki, G., Masui, K., Aibe, N., Shimizu, D., Kimoto, T., et al. (2021). Novel prognostic index of high-risk prostate cancer using simple summation of very high-risk factors. *Cancers (Basel)* 13 (14), 3486. doi:10.3390/cancers13143486

Yue, W., Sun, Q., Dacic, S., Landreneau, R. J., Siegfried, J. M., Yu, J., et al. (2008). Downregulation of Dkk3 activates beta-catenin/TCF-4 signaling in lung cancer. *Carcinogenesis* 29 (1), 84–92. doi:10.1093/carcin/bgm267

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 30 May 2024
DOI 10.3389/fmolb.2024.1395721

Check for updates

OPEN ACCESS

EDITED BY
Qingyu Luo,
Dana–Farber Cancer Institute, United States

REVIEWED BY
Lin Han,
Dana–Farber Cancer Institute, United States
Firoz Ahmed,
Jeddah University, Saudi Arabia

*CORRESPONDENCE
Gajendra P. S. Raghava,
raghava@iiitd.ac.in

RECEIVED 04 March 2024
ACCEPTED 13 May 2024
PUBLISHED 30 May 2024

CITATION
Jarwal A, Dhall A, Arora A, Patiyal S, Srivastava A and Raghava GPS (2024), A deep learning method for classification of HNSCC and HPV patients using single-cell transcriptomics. *Front. Mol. Biosci.* 11:1395721. doi: 10.3389/fmolb.2024.1395721

COPYRIGHT
© 2024 Jarwal, Dhall, Arora, Patiyal, Srivastava and Raghava. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A deep learning method for classification of HNSCC and HPV patients using single-cell transcriptomics

Akanksha Jarwal, Anjali Dhall, Akanksha Arora, Sumeet Patiyal, Aman Srivastava and Gajendra P. S. Raghava*

Department of Computational Biology, Indraprastha Institute of Information Technology, Delhi, India

Background: Head and Neck Squamous Cell Carcinoma (HNSCC) is the seventh most highly prevalent cancer type worldwide. Early detection of HNSCC is one of the important challenges in managing the treatment of the cancer patients. Existing techniques for detecting HNSCC are costly, expensive, and invasive in nature.

Methods: In this study, we aimed to address this issue by developing classification models using machine learning and deep learning techniques, focusing on single-cell transcriptomics to distinguish between HNSCC and normal samples. Furthermore, we built models to classify HNSCC samples into HPV-positive (HPV+) and HPV-negative (HPV–) categories. In this study, we have used GSE181919 dataset, we have extracted 20 primary cancer (HNSCC) samples, and 9 normal tissues samples. The primary cancer samples contained 13 HPV– and 7 HPV+ samples. The models developed in this study have been trained on 80% of the dataset and validated on the remaining 20%. To develop an efficient model, we performed feature selection using mRMR method to shortlist a small number of genes from a plethora of genes. We also performed Gene Ontology (GO) enrichment analysis on the 100 shortlisted genes.

Results: Artificial Neural Network based model trained on 100 genes outperformed the other classifiers with an AUROC of 0.91 for HNSCC classification for the validation set. The same algorithm achieved an AUROC of 0.83 for the classification of HPV+ and HPV– patients on the validation set. In GO enrichment analysis, it was found that most genes were involved in binding and catalytic activities.

Conclusion: A software package has been developed in Python which allows users to identify HNSCC in patients along with their HPV status. It is available at https://webs.iiitd.edu.in/raghava/hnscpred/.

KEYWORDS

HNSCC, gene biomarkers, single cell transcriptomics, machine learning, deep learning, classification models

1 Introduction

Head and neck cancer, encompasses a variety of malignancies that affect the respiratory tract and upper digestive tract. Head and Neck Squamous Cell Carcinoma (HNSCC) is the most typical kind among the head and neck cancer (Mody et al., 2021). In 2020, 562,328 people were diagnosed with head and neck cancer (HNC) worldwide, with a total count of 277,587 deaths due to the disease (Broutian et al., 2020). These carcinomas often develop in the salivary glands, larynx, oral cavity, throat, and sino-nasal tract epithelium. A number of head and neck malignancies are linked to the human papillomavirus (HPV) infection, notably HPV-16. However, some malignancies are also related to the other carcinogens like smoking, excessive alcohol, and other factors depending on the country or area. Hence, we can classify this cancer into two major categories—HPV-negative and HPV-positive. The median age of diagnosis for HPV associated HNSCC is about 66 years, whereas for HPV-associated oropharyngeal cancer the median age is ~53 years (Johnson et al., 2020).

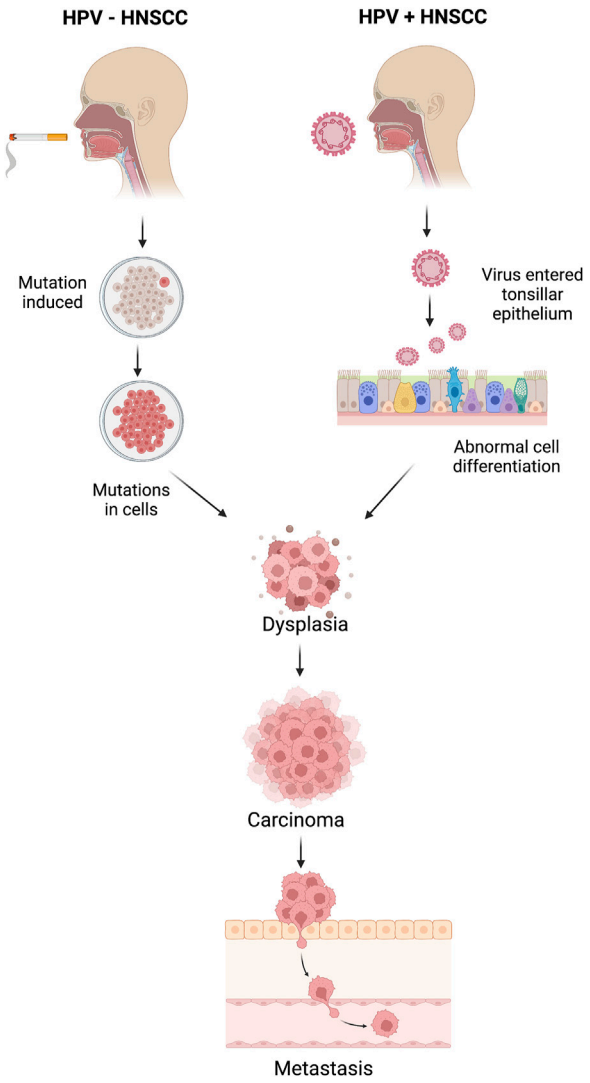


FIGURE 1
Mechanisms of head and neck squamous cell carcinoma (HNSCC) for HPV-positive and HPV-negative HNSCC patients.

Distinguishing between HPV-positive and HPV-negative Head and Neck Squamous Cell Carcinoma (HNSCC) samples holds profound significance in clinical practice as it unveils distinct molecular mechanisms underlying tumorigenesis and guides tailored therapeutic interventions. HPV-positive HNSCCs, primarily driven by high-risk human papillomavirus (HPV) infection, often manifest with activated cell cycle pathways, particularly the retinoblastoma protein (pRB) pathway, leading to enhanced cell proliferation (Leemans et al., 2018). Conversely, HPV-negative tumors frequently arise from genomic instability induced by environmental factors such as tobacco and alcohol exposure, resulting in diverse genetic alterations, such as mutations in tumor suppressor genes and oncogenes. Consequently, HPV-positive tumors exhibit heightened sensitivity to radiotherapy and chemotherapy due to their intact DNA repair mechanisms and increased expression of apoptosis-regulating proteins (Fakhry et al., 2008). Conversely, HPV-negative tumors, characterized by aberrant DNA repair pathways and resistance to apoptosis, necessitate more aggressive therapeutic strategies. Understanding the HPV status in HNSCC thus facilitates personalized treatment approaches, optimizing patient outcomes by targeting specific molecular vulnerabilities (Dok and Nuyts, 2016). Better understanding the HPV status of HNSCC tumors enables clinicians to tailor treatment strategies and provide accurate prognostic information, ultimately improving patient management and outcomes (Ang et al., 2010; Chaturvedi et al., 2011; Gillison et al., 2012). The mechanisms of HPV+ and HPV- associated HNSCC are explained in Figure 1.

Despite thorough and targeted treatment efforts, the chances of survival are reduced due to the majority of head and neck cancer cases being diagnosed at advanced stages. The traditional diagnosis of HNSCC is based on the physical examination, radiological investigation, and histological analysis of the tissue sections obtained from biopsies or surgical resections. These procedures can take a lot of time and are susceptible to mistakes in observation or interpretation, which can lead to discrepancies in cancer grading and prognostication (Mahmood et al., 2021). In addition to this, most of the HNSCC cancers are detected at a later stage. The reasons range from limited symptomatology in early-stage patients, swift progression from early to advanced stage, indistinctive diagnostic characteristics, and imprecise history information (Basheeth and Patil, 2019).

Identification of molecular biomarkers of HNSCC can lead to early diagnosis of this cancer and can also help in preventive management of HNSCC. The cancer biomarkers not only influence diagnosis but they also have the potential to improve the treatment outcomes using targeted therapy. The currently known biomarker of HNSCC is PD-L1 which is commonly used in treatment decision making in advanced stage of HNSCC. It has a moderate predictive value and has many limitations due to the lack of standardization and highly dynamic nature of PD-L1 expression. Currently, there are no any other FDA approved molecular biomarkers for HNSCC diagnosis or prognosis (Basheeth and Patil, 2019).

In this study, we made an attempt to identify biomarkers for HNSCC using single-cell sequencing data. On the basis of the 100 biomarkers identified in this study, we have developed a

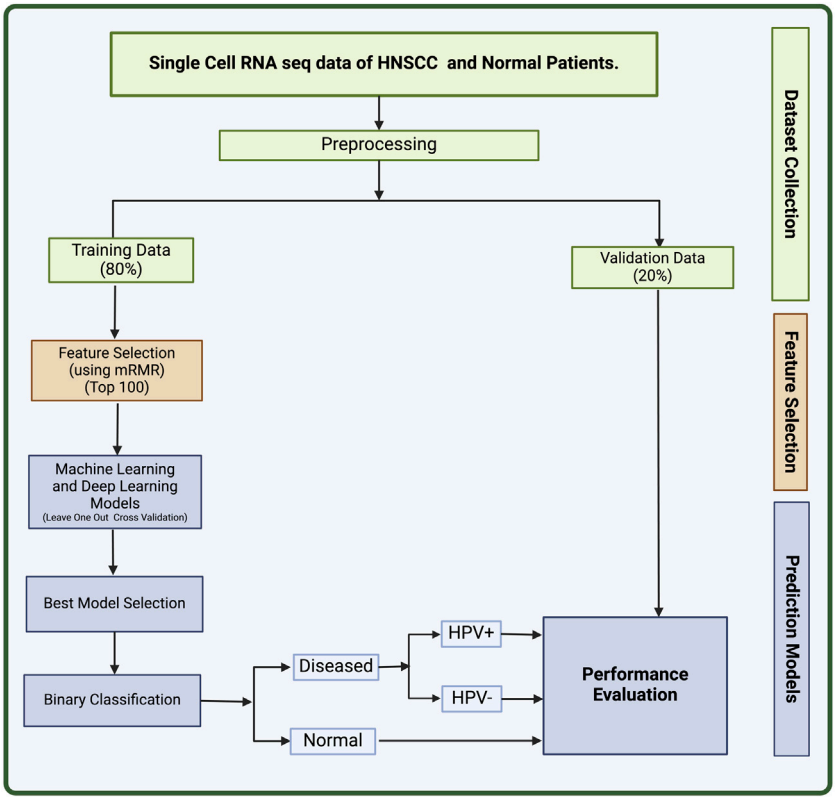


FIGURE 2
The full workflow of the study.

method that can predict the HNSCC cancer along with HPV+ or HPV− status. Single-cell data collected from individual cells using next-generation sequencing methods provides a better knowledge of the activity of a single cell in relation to its microenvironment (Eberwine et al., 2014). Cell-to-cell variation can be revealed by single-cell sequencing of RNA or epigenetic alterations, which may aid the populations in quickly adapting to new circumstances (Saliba et al., 2014). The significance of gene mosaicism, as well as intra-tumor genetic heterogeneity in the genesis of cancer or response to therapy, can be uncovered by single-cell precision (Gawad et al., 2016). Single-cell technology makes it possible to detect molecular alterations in individual cancer cells. This can increase the research of more specialized biomarkers with excellent resolution, leading to the development of a complete landscape of distinct cell types within tumors (Radpour and Forouharkhou, 2018). The full workflow of this study is described in Figure 2.

2 Materials and methods

2.1 Data collection

We retrieved the dataset used in this study (GSE181919) from National Centre for Biotechnology Information's (NCBI) Gene Expression Omnibus (GEO) (Clough and Barrett, 2016; Choi et al., 2023). The GSE181919 dataset comprise of 37 tissue specimens from 23 patients with Head and Neck Squamous Cell Carcinoma (HNSCC), covering a range of tissues, including normal tissues (n = 9), precancerous leukoplakia (n = 4), primary HNSCC (n = 20), and metastasized tumors (n = 4). Choi et al. methodology involved aligning sequencing data to the human reference genome (GRCh38) and processing it using CellRanger 2.1.1 by 10X Genomics. Subsequently, cell-level transcripts were clustered using the "Seurat"package's shared nearest neighbor method. To ensure the clarity in the dataset, we chose two distinct groups: normal tissue (n = 9) and primary HNSCC tissues (n = 20). Therefore, in this study, we have only taken 29 total samples comprising 20 primary cancer samples, and 9 normal samples. In addition, these cancer samples are divided into 13 HPV− and 7 HPV+ samples. The information on whether the samples were derived from HPV+ or HPV− patients was derived from the metadata provided on GEO. This dataset used Illumina HiSeq 4000 as the platform for scRNA sequencing. The 80% of this dataset was used to train machine learning (ML) and deep learning (DL) models and 20% was used as validation set.

2.2 Data pre-processing

After the retrieval of data from GEO, we processed the data using in-house python scripts. Firstly, we converted the sparse data

into a matrix and removed insignificant columns from our training data. The genes that had no mapped readings to more than 80% of the cells were eliminated, and cells containing zeroes were filtered leading to 2,604 genes. The sequencing depth affects the range of values for the features, which necessitates normalizing the count data before doing any sort of analysis. Hence, we performed counts per million (CPM) normalization and log transformation on the data using scanpy package in python (Wolf et al., 2018).

2.3 Feature selection

We applied feature selection to the set of 2,604 genes obtained after pre-processing to obtain a set of biomarkers for HNSCC. This was achieved using mRMR (Minimum Redundancy and Maximum Relevance) feature selection algorithm (Radovic et al., 2017). mRMR selects a subset of features that have the least correlation amongst themselves but high correlation with the output class. The advantage of using this method is that it provides with a small set of features with high predictive potential. The redundancy between genes is taken into account in this technique in addition to the relationship between samples and genes. The most relevant feature will be considered out of the numerous identical features. We used the value K = 100 for mRMR to extract 100 most relevant genes for the prediction of HNSCC (Zhao et al., 2019a). This strategy has been previously demonstrated to be useful and often utilized in single-cell RNA sequencing analysis (Ding and Peng, 2005; Radovic et al., 2017).

2.4 Top dysregulated genes

After extracting the top 100 genes from feature selection, we performed a T-test analysis for the mean expression of genes in the cells of the groups Normal vs. Cancer and Cancer HPV+ vs. Cancer HPV−. We also wished to identify the top most dysregulated genes in both the comparisons. To achieve this, we found mean difference between the two classes in both comparisons (Normal vs. Cancer and Cancer HPV+ vs. Cancer HPV−), and reported the 10 most dysregulated (5 upregulated and 5 downregulated) genes with the highest difference in means for each comparison.

2.5 Machine learning models

We have developed various machine learning (ML) models to classify between normal subjects and HNSCC patients. In addition, we have also classified HNSCC patients into HPV positive and HPV negative. These machine learning models include Extreme Gradient Boosting (XGB), Decision Tree (DT), K-Nearest Neighbors (KNN), Extra Trees (ET), Logistic Regression (LR), and Random Forest (RF) algorithms. Hyperparameter tuning was also used to optimise the parameters of these algorithms. The DT classifier is a supervised machine learning model that classifies the output by learning decision rules from input, the KNN classifier predicts on the basis of the maximum number of votes cast in support of the class that is closest to the nearest neighbouring data point, LR classifier uses a logistic function to calculate the likelihood of an event, XGB Classifier is a distributed gradient-boosted decision tree machine learning package that offers simultaneous tree boosting, and RF classifier trains a number of decision trees to produce a single tree. A technique for ensemble supervised machine learning that makes use of decision trees is called extra trees. (Breiman, 2001; Wu et al., 2002; Geurts et al., 2006; Stoltzfus, 2011; Bulac and Bulac, 2016; Chen and Guestrin, 2016). These methods have previously been used in many studies (Aggarwal et al., 2023; Arora et al., 2023; Kaur et al., 2023; Srivastava et al., 2023).

2.6 Deep learning models

Along with the ML models, we have also applied deep learning classification technique—Artificial Neural Network (ANN) to classify the data (Wang, 2003). In this method, networks are composed of multiple layers, and each layer has a number of nodes (or neurons) that support decision making. The model architecture of ANN used in this study includes three hidden layers and an output layer. A dropout of 0.5 is implemented at each step to lessen the overfitting by neural network. Biological neuron networks served as the basis for this strategy. Artificial neurons, which are constructed from a network of connected units or nodes and are conceptually similar to the neurons in the human brain, are used to build ANNs. They consist of several layers, and inside each layer there are multiple nodes (or neurons) that support decision-making. The anticipated label (Diseased or Normal) of the sample is the final result. The final result classifies the samples into HNSCC positive or negative, and if found HNSCC positive then whether the patient is HPV positive or negative is identified.

2.7 Cross validation

The dataset was primarily composed of training data, which made up 80% of it and validation set, which made up the remaining 20%. In the LOOCV (Leave One out Cross Validation) approach, the whole training set is separated into N equivalent folds using the LOOCV technique, with (N-1) being utilized for training and the single fold being used for testing. Each fold serves as testing data for the technique's N iterations. The overall performance was calculated as the mean of N iterations. This is a common practice in many types of studies (Peng et al., 2015; Vabalas et al., 2019).

2.8 Evaluation parameters

To evaluate the efficacy of various prediction models, we employed a number of evaluation indicators. In this study, we used both threshold-independent and threshold-dependent parameters. To calculate threshold-dependent characteristics like sensitivity (Sens), specificity (Spec), precision, F1-Score, and accuracy (Acc), we utilised the following formulae. We also used the conventional threshold-independent parameter Area Under the Curve (AUC) to assess the performance of the models. The metrics calculated in this study are mentioned in Eqs 1–5.

TABLE 1 Top 10 dysregulated genes for Normal vs. Cancer.

Gene	Mean gene expression cancer	Mean gene expression normal	Mean difference (cancer-normal)	T-Statistic	*p*-value	Up/ Downregulated
CFD	0.773	57.497	−56.724	−85.985	0.000e+00	Downregulated
DCN	4.840	35.536	−30.696	−72.891	0.000e+00	Downregulated
GSN	2.566	29.935	−27.369	−76.183	0.000e+00	Downregulated
MGP	0.707	13.759	−13.052	−56.266	0.000e+00	Downregulated
MFAP4	0.206	7.802	−7.596	−82.346	0.000e+00	Downregulated
RPL28	63.478	15.950	47.528	103.250	0.000e+00	Upregulated
EEF1A1	85.286	35.414	49.872	82.304	0.000e+00	Upregulated
RPS19	77.935	17.512	60.423	92.379	0.000e+00	Upregulated
RPLP1	111.729	30.115	81.614	95.136	0.000e+00	Upregulated
B2M	158.129	43.524	114.605	107.234	0.000e+00	Upregulated

$$Sensitivity = \frac{P_t}{P_t + N_f} \quad (1)$$

$$Specificity = \frac{N_t}{N_t + P_f} \quad (2)$$

$$Accuracy = \frac{P_t + N_f}{P_t + N_t + P_f + N_f} \quad (3)$$

$$F1\text{-}score = \frac{2P_t}{2P_t + P_f + N_f} \quad (4)$$

$$Precision = \frac{P_t}{P_t + P_f} \quad (5)$$

Where, P_t is true positive, N_t is true negative, P_f is false positive, and N_f is false negative.

3 Results

3.1 Feature selection

We applied a feature selection technique called mRMR to obtain a list of highly relevant features (genes) for the detection of HNSCC samples from a set of 2,604 genes that were obtained after data pre-processing (Zhao et al., 2019b). We obtained a subset of 100 genes that were able to classify HNSCC and non-HNSCC samples as well as HPV+ and HPV− samples correctly.

3.2 Top dysregulated genes

After performing T-test on selected 100 genes for Normal vs. Cancer and Cancer HPV+ vs. Cancer HPV− groups. It was found that all the genes were significantly differentially expressed with *p*-values<0.05 in Normal vs. Cancer comparison whereas 94 genes out of 100 were significantly differentially expressed with *p*-values<0.05 in Cancer HPV+ vs. Cancer HPV− comparison. The list of selected 100 genes along with their *p*-values, mean gene expressions, mean expression difference, and t-statistics for Normal vs. Cancer and Cancer HPV+ vs. Cancer HPV− are given in Supplementary Tables S1, S2, respectively. We also identified the top 10 dysregulated genes (5 upregulated and 5 downregulated) on the basis of mean difference between two classes in both the comparisons. The top 10 dysregulated genes for Normal vs. Cancer and Cancer HPV+ vs. Cancer HPV− are given in Tables 1, 2, respectively.

3.3 Model performance for HNSCC vs. non-HNSCC

We applied various ML models like DT, RF, ET, XGB, and KNN on our dataset, where we used 80% of dataset GSE181919 for training, 20% of dataset GSE181919 as validation set, and. It was observed that machine learning models were able to achieve an AUROC of 0.85 (XGB, ET) on the validation set. In order to increase the AUROC, we applied DL algorithm—ANN on the dataset and observed that the AUROCs increased to 0.91 for the validation set. The complete results for the ML and DL performances are given in Table 3.

3.4 Model performance for HPV+ vs. HPV−

After classification of samples as HNSCC or non-HNSCC, we attempted to classify whether an HNSCC sample belonged to an HPV+ or HPV− class. Hence, we developed ML and DL models to further classify the HNSCC samples as HPV+ and HPV−. The maximum AUROC achieved by ML models was 0.81 (XGB) for the validation set. After employing ANN classifier to the data, it was observed that the AUROC increased to 0.84 for the validation set. The results for HPV+ and HPV− classification from HNSCC patients are summarized in Table 4.

3.5 Gene ontology

The Gene Ontology (GO) encapsulates our understanding of the biological world in three ways: molecular function, cellular component, and biological process (Ashburner et al., 2000; Gene Ontology Consortium, 2021). 100 genes that may serve as potential

10.3389/fmolb.2024.1395721

TABLE 2 Top 10 dysregulated genes for Cancer HPV- vs. Cancer HPV+.

Gene	Mean gene expression cancer	Mean gene expression normal	Mean difference (cancer-normal)	T-Statistic	*p*-value	Up/ Downregulated
B2M	112.908	190.234	−77.326	−40.719	0	Downregulated
HLA-B	27.504	52.285	−24.781	−38.936	0	Downregulated
HLA-A	27.495	47.94	−20.445	−32.086	1.01E-220	Downregulated
HLA-C	25.502	40.151	−14.649	−29.336	1.27E-185	Downregulated
RPLP1	106.477	115.458	−8.981	−5.482	4.24E-08	Downregulated
CXCR4	6.556	5.289	1.267	10.654	1.95E-26	Upregulated
BTG1	12.936	11.385	1.551	9.614	7.81E-22	Upregulated
RPL28	64.482	62.765	1.717	1.887	0.0592	Upregulated
EEF1A1	86.865	84.166	2.699	2.439	0.0147	Upregulated
RPS19	80.110	76.391	3.719	2.797	0.0052	Upregulated

TABLE 3 Performance of ML and DL models for the classification of HNSCC patients and normal subjects.

Training set							
Models	Accuracy	MCC	AUROC	Sensitivity	Specificity	Precision	F1 score
Decision Tree	0.93	0.85	0.93	0.95	0.91	0.94	0.94
Random Forest	0.96	0.92	0.96	0.98	0.94	0.96	0.97
Logistic Regression	0.92	0.84	0.92	0.95	0.88	0.92	0.94
XGB	0.92	0.83	0.92	0.92	0.92	0.95	0.93
ExtraTree	0.97	0.80	0.90	0.91	0.89	0.91	0.91
KNN	0.93	0.85	0.92	0.96	0.89	0.93	0.94
Deep Learning Model	0.99	0.93	0.97	0.98	0.96	0.97	0.98
Validation Set							
Models	Accuracy	MCC	AUROC	Sensitivity	Specificity	Precision	F1 Score
Decision Tree	0.85	0.70	0.83	0.97	0.70	0.80	0.88
Random Forest	0.85	0.71	0.83	0.99	0.68	0.79	0.88
Logistic Regression	0.79	0.60	0.77	0.98	0.56	0.73	0.84
XGB	0.86	0.73	0.85	0.98	0.71	0.81	0.89
ExtraTree	0.86	0.74	0.85	0.99	0.71	0.81	0.89
KNN	0.83	0.68	0.81	0.98	0.65	0.77	0.87
Deep Learning Model	0.92	0.82	0.91	0.94	0.89	0.94	0.94

biomarkers of HNSCC were retrieved once mRMR analysis was complete. On these 100 retrieved genes, we next ran Gene Ontology (GO) enrichment analysis using PantherDB to map the biological processes, cellular components, and molecular functions of the chosen genes (Mi et al., 2013). The findings of the GO enrichment analysis for all 100 selected genes are displayed in Supplementary Tables S3–S5, respectively. The results of Gene Ontology for Biological Processes and Cellular Component are shown in Figure 3A, B respectively. We see that the majority of genes have a role in the binding activities of many metabolic processes as shown in Figure 3C. The genes and their roles are described in Figure 3.

4 Discussions

One of the heterogeneous diseases, HNSCC affects the head and neck region, namely the oral cavity, paranasal sinuses, larynx, nasal

TABLE 4 Performance of ML and DL models for the classification of HPV+ and HPV– patients from HNSCC patients.

Training set							
Models	Accuracy	MCC	AUROC	F1 score	Sensitivity	Specificity	Precision
Decision Tree	0.75	0.50	0.75	0.78	0.79	0.71	0.77
Random Forest	0.82	0.63	0.81	0.84	0.87	0.75	0.81
Logistic Regression	0.84	0.67	0.84	0.86	0.86	0.82	0.85
XGB	0.77	0.52	0.76	0.79	0.81	0.71	0.77
ExtraTree	0.84	0.68	0.84	0.86	0.88	0.8	0.84
KNN	0.80	0.59	0.79	0.82	0.84	0.75	0.8
Deep Learning Model	0.991	0.980	0.995	0.992	0.989	0.993	0.995
Validation Set							
Models	Accuracy	MCC	AUROC	F1 score	Sensitivity	Specificity	Precision
Decision Tree	0.69	0.35	0.65	0.74	0.76	0.59	0.72
Random Forest	0.84	0.88	0.80	0.83	0.91	0.92	0.94
Logistic Regression	0.80	0.54	0.75	0.85	0.9	0.61	0.81
XGBClassifier	0.82	0.61	0.81	0.86	0.87	0.74	0.86
ExtraTree Classifier	0.46	−0.11	0.76	0.52	0.58	0.33	0.47
K Neighbours classsifier	0.49	−0.06	0.55	0.55	0.58	0.38	0.52
Deep Learning Model	0.84	0.70	0.83	0.88	0.98	0.68	0.79

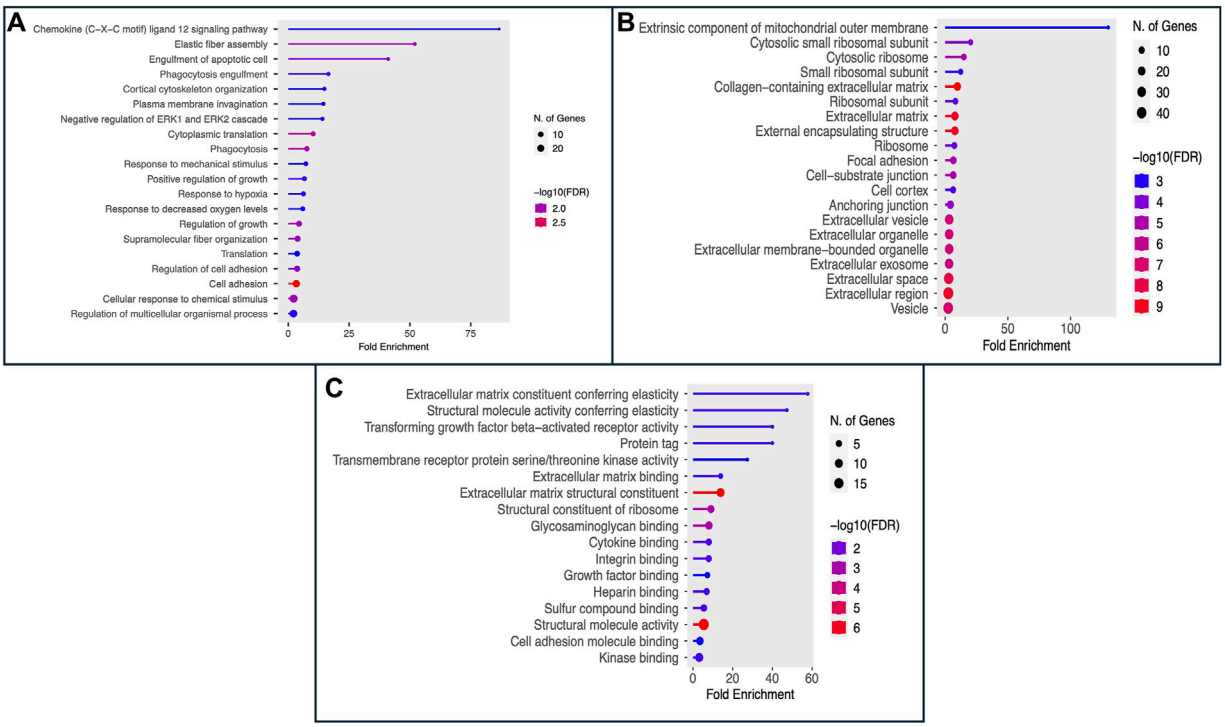


FIGURE 3
The figure displays the Gene Ontology (GO) enrichment analysis results as **(A)** Biological Process, **(B)** Cellular Component (CC), and **(C)** Molecular Function (MF) for the top 100 selected genes.

cavity, hypopharynx, and oropharynx. It is described by malignant and uncontrollable cell proliferation in these locations (Hsieh et al., 2019). Advancement in the sequence technology allows the researchers to find various biomarkers such as diagnostic, predictive, and prognostic biomarkers. These biomarkers help in better understanding of the disease as well as may aids in designing novel and effective diagnosis and treatment. A biomarker is described as a biological molecule present in the blood, other body fluids, as well as in tissues, that serves as a sign of a normal or aberrant process, a condition, or a disease by the National Cancer Institute (NCI). To determine how effectively the body will react to an illness or condition medication, a biomarker could well be utilized (Hsieh et al., 2019). This study aims to find out a set of potential biomarkers from single-cell transcriptomic data of head and neck cancer patients that can classify HNSCC patients and normal individuals with reliable accuracy. In addition, we have also attempted to classify HNSCC patients as HPV+ or HPV−. The biomarkers identified in this study could aid in the early diagnosis and screening of HNSCC.

To categorize non-cancer and HNSCC disease cells from their single-cell RNA seq data, we employed a variety of machine learning models, including an ANN deep learning model. We also further tried to categorize the diseased patients into HPV+ and HPV−. We trained the model with 80% of the dataset GSE181919, 20% of the dataset GSE181919 as validation set. The datasets were originally quite extensive and had a significant number of features. During the preprocessing step, the feature count was decreased to a shallow level of 2,604 genes (features). One of the feature selection techniques known as mRMR was used to obtain the limited set of features which could be helpful in categorizing the samples because many characteristics were correlated. The top 100 genes with the least amount of redundancy and the most relevance were extracted from these 2,604 genes (features) using mRMR. Furthermore, 100 genes (features) separated the HNSCC patients from non-cancer with an accuracy of around 92%, an AUROC of 0.91 in the validation set. Whereas in the case of HPV classification, the metrics obtained were, AUROC 0.83% and 98% accuracy on the validation set. For the detection and categorization of biomarkers, ANN has proven to be an effective technique among all machine learning models.

After obtaining the top 100 most relevant genes for the classification of HNSCC, we performed Gene Ontology (GO) enrichment analysis using PantherDB and most of the genes were observed to be related to catalytic and binding activities (Mi et al., 2013). Some of them also had a role in other essential processes like ATP-dependent activity, molecular function regulator, molecular transducer, structural molecule activity, translation regulator activity, transcription regulator, and transporter activity. Many of the genes identified in this study have been previously linked to HNSCC in earlier studies. The gene PLAC9's overexpression has been reported in connection with the inhibition of cell growth regulation and has also been reported in connection with cancers such as ovarian cancer and breast cancers as prognostic biomarkers (Ouyang et al., 2018). Gene "ACKR1", along with other 3 genes in a study, was reported to be downregulated in HNSCC patients, which was correlated with poor prognosis ($p < 0.05$) (Liu et al., 2022). Also, gene "AQP7," which is involved in physiologically functional cell migration, was upregulated in MSR of patients with ten tumors (Zivicova et al., 2018). Whereas, gene FXYD1 was reported to be downregulated in the cancer samples, while FXYD4 and FXYD5 were overexpressed ($p < 0.05$, fold change>1.5) (Jin et al., 2021). In a study on cancer cells, it was observed that BTG1 gene overexpression was linked to tumor growth or lung metastasis, inhibited proliferation, and induced differentiation in different types of cancer cells (Zhao et al., 2020). Also, mutations occurring in different genes, including B2M, CDKN2A, is found to be related with the occurrence and development of tumors in Head and neck cancer patients (Wang et al., 2020). As shown in the study Sun et al., 2020, genes such as MFAP4, CD37, CXCL12, ADH1B, SOD3, SCARA5, ANGPTL1, FHL1, F10, CXCR4, MEG3, TXNIP, GDF10, and ABI3BP are downregulated in head and neck squamous cell carcinoma as they operate as potential tumor suppressor genes, inhibiting tumor cell proliferation, invasion, and migration while also promoting apoptosis (Sun et al., 2020). By controlling the expression of miR-421 and E-cadherin, MEG3 long-encoding RNA inhibits the development of head and neck squamous cell carcinoma. However, additional research into MEG3's downstream mechanism in controlling the molecular process of epithelial-mesenchymal transformation (EMT) in head and neck squamous cell carcinoma (HNSCC) development is required (Ji et al., 2020). Growth differentiation factor-10 (GDF10), also known as BMP3b, is a tumor suppressor that belongs to the transforming growth factor-b (TGF-b) superfamily (Cheng et al., 2016). CIB1, PIM3, SLC16A3, VOPP1, BMP4, TIGIT, ADAR, and LRRN4CL are studied as upregulated genes in various cancer types such as squamous carcinoma cells, breast cancer, head and neck squamous cell carcinoma, and pancreatic cancer (Baras et al., 2011; Alarmo et al., 2013; Zheng and Tian, 2014; de Jong et al., 2018; Notarangelo, 2018; Broutian et al., 2020; Yu et al., 2020; Huo et al., 2021; Wen et al., 2021; Yang et al., 2022). A complex that is important in the keratinocyte-intrinsic immune response to human papillomaviruses (-HPVs) is formed when CIB1 interacts with the EVER1, and EVER2 proteins (de Jong et al., 2018; Notarangelo, 2018). It has been observed that nearly all primary HNSCCs express at least one PIM kinase member at high levels (Broutian et al., 2020). Immunological checkpoint T cell immunoreceptor with immunoglobulin and ITIM domain (TIGIT) is essential for immune suppression. However, it has a connection to genetics and epigenetics, and a role in tumor immunity (Wen et al., 2021). The transforming growth factor (TGF) superfamily includes extracellular signaling molecules known as bone morphogenetic proteins (BMPs), which are known to control cell proliferation, differentiation, and motility, particularly during development. Functional research shows that, particularly in HNSCC cancer, has connected BMP4 to the encouragement of cell migration and the suppression of cell proliferation (Alarmo et al., 2013).

Overall, most of the genes which were obtained from our study have been reported as promising candidate for biomarkers in various studies (Zivicova et al., 2018; Broutian et al., 2020; Sun et al., 2020; Wang et al., 2020; Zhao et al., 2020; Jin et al., 2021; Liu et al., 2022). However, some genes have not yet been reported in connection with Head and Neck cancer. These genes may require further investigation and study. These genes may act as novel findings which could help in diagnose patients with Head and

neck cancer. In order to help the scientific community, we created a Python package called "HNSCPred" based on the aforementioned work (https://webs.iiitd.edu.in/raghava/hnscpred/). To fully understand how the discovered genes impact and contribute to the progression of HNSCC disease, further clinical investigations on these genes are necessary.

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 below: https://webs.iiitd.edu.in/raghava/hnscpred/.

Author contributions

AJ: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing–review and editing. AD: Data curation, Formal Analysis, Investigation, Methodology, Software, Supervision, Visualization, Writing–review and editing. AA: Methodology, Supervision, Visualization, Writing–original draft, Writing–review and editing. SP: Formal Analysis, Methodology, Project administration, Software, Supervision, Writing–review and editing. AS: Methodology, Software, Writing–review and editing. GR: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The authors are thankful to the Council of Scientific and Industrial Research (CSIR), Department of Biotechnology (DBT), and Department of Science and Technology (DST) for providing fellowships and financial support. We are thankful to the Department of Biotechnology (DBT) for providing an infrastructure grant to the institute (grant number - BT/PR40158/BTIS/137/24/2021).

Acknowledgments

The authors are thankful to the Department of Computational Biology, IIITD, New Delhi for its infrastructure and facilities. We would like to acknowledge that the figures were created using BioRender.

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/fmolb.2024.1395721/full#supplementary-material

References

Aggarwal, S., Dhall, A., Patiyal, S., Choudhury, S., Arora, A., and Raghava, G. P. S (2023). An ensemble method for prediction of phage-based therapy against bacterial infections. *Front. Microbiol.* 14, 1148579. doi:10.3389/fmicb.2023.1148579

Alarmo, E.-L., Huhtala, H., Korhonen, T., Pylkkänen, L., Holli, K., Kuukasjärvi, T., et al. (2013). Bone morphogenetic protein 4 expression in multiple normal and tumor tissues reveals its importance beyond development. *Mod. Pathol.* 26, 10–21. doi:10.1038/modpathol.2012.128

Ang, K. K., Harris, J., Wheeler, R., Weber, R., Rosenthal, D. I., Nguyen-Tân, P. F., et al. (2010). Human papillomavirus and survival of patients with oropharyngeal cancer. *N. Engl. J. Med.* 363 (1), 24–35. doi:10.1056/NEJMoa0912217

Arora, A., Patiyal, S., Sharma, N., Devi, N. L., Kaur, D., and Raghava, G. P. S. (2023). A random forest model for predicting exosomal proteins using evolutionary information and motifs. *Proteomics* 24, e2300231. doi:10.1002/pmic.202300231

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

Baras, A. S., Solomon, A., Davidson, R., and Moskaluk, C. A. (2011). Loss of VOPP1 overexpression in squamous carcinoma cells induces apoptosis through oxidative cellular injury. *Lab. Investig.* 91, 1170–1180. doi:10.1038/labinvest.2011.70

Basheeth, N., and Patil, N. (2019). Biomarkers in head and neck cancer an update. *Indian J. Otolaryngol. Head. Neck Surg.* 71, 1002–1011. doi:10.1007/s12070-019-01683-1

Breiman, L. (2001). Random forests. *Mach. Learn* 45, 5–32. doi:10.1023/A:1010933404324

Broutian, T. R., Jiang, B., Li, J., Akagi, K., Gui, S., Zhou, Z., et al. (2020). Human papillomavirus insertions identify the PIM family of serine/threonine kinases as targetable driver genes in head and neck squamous cell carcinoma. *Cancer Lett.* 476, 23–33. doi:10.1016/j.canlet.2020.01.012

Bulac, C., and Bulac, A. (2016). "Decision trees," in *Advanced solutions in power systems: HVDC, FACTS, and AI techniques.* doi:10.1002/9781119175391.ch18

Chaturvedi, A. K., Engels, E. A., Pfeiffer, R. M., Hernandez, B. Y., Xiao, W., Kim, E., et al. (2011). Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J. Clin. Oncol.* 29 (32), 4294–4301. doi:10.1200/JCO.2011.36.4596

Chen, T., and Guestrin, C. (2016). "XGBoost: a scalable tree boosting system," in *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining.* doi:10.1145/2939672.2939785

Cheng, C.-W., Hsiao, J.-R., Fan, C.-C., Lo, Y.-K., Tzen, C.-Y., Wu, L.-W., et al. (2016). Loss of GDF10/BMP3b as a prognostic marker collaborates with TGFBR3 to enhance chemotherapy resistance and epithelial-mesenchymal transition in oral squamous cell carcinoma. *Mol. Carcinog.* 55, 499–513. doi:10.1002/mc.22297

Choi, J.-H., Lee, B.-S., Jang, J. Y., Lee, Y. S., Kim, H. J., Roh, J., et al. (2023). Single-cell transcriptome profiling of the stepwise progression of head and neck cancer. *Nat. Commun.* 14, 1055. doi:10.1038/s41467-023-36691-x

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

de Jong, S. J., Créquer, A., Matos, I., Hum, D., Gunasekharan, V., Lorenzo, L., et al. (2018). The human CIB1-EVER1-EVER2 complex governs keratinocyte-intrinsic immunity to β-papillomaviruses. *J. Exp. Med.* 215, 2289–2310. doi:10.1084/jem.20170308

Ding, C., and Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *J. Bioinform Comput. Biol.* 3, 185–205. doi:10.1142/s0219720005001004

Dok, R., and Nuyts, S. (2016). HPV positive head and neck cancers: molecular pathogenesis and evolving treatment strategies. *Cancers (Basel)* 8 (4), 41. doi:10.3390/cancers8040041

Eberwine, J., Sul, J.-Y., Bartfai, T., and Kim, J. (2014). The promise of single-cell sequencing. *Nat. Methods* 11, 25–27. doi:10.1038/nmeth.2769

Fakhry, C., Westra, W. H., Li, S., Cmelak, A., Ridge, J. A., Pinto, H., et al. (2008). Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J. Natl. Cancer Inst.* 100 (4), 261–269. doi:10.1093/jnci/djn011

Gawad, C., Koh, W., and Quake, S. R. (2016). Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* 17, 175–188. doi:10.1038/nrg.2015.16

Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* 49, D325–D334. doi:10.1093/nar/gkaa1113

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Mach. Learn* 63, 3–42. doi:10.1007/s10994-006-6226-1

Gillison, M. L., Zhang, Q., Jordan, R., Xiao, W., Westra, W. H., Trotti, A., et al. (2012). Tobacco smoking and increased risk of death and progression for patients with p16-positive and p16-negative oropharyngeal cancer. *J. Clin. Oncol.* 30 (17), 2102–2111. doi:10.1200/JCO.2011.38.4099

Hsieh, J. C.-H., Wang, H.-M., Wu, M.-H., Chang, K.-P., Chang, P.-H., Liao, C.-T., et al. (2019). Review of emerging biomarkers in head and neck squamous cell carcinoma in the era of immunotherapy and targeted therapy. *Head. Neck* 41 (Suppl. 1), 19–45. doi:10.1002/hed.25932

Huo, X.-X., Wang, S.-J., Song, H., Li, M., Yu, H., Wang, M., et al. (2021). Roles of major RNA adenosine modifications in head and neck squamous cell carcinoma. *Front. Pharmacol.* 12, 779779. doi:10.3389/fphar.2021.779779

Ji, Y., Feng, G., Hou, Y., Yu, Y., Wang, R., and Yuan, H. (2020). Long noncoding RNA MEG3 decreases the growth of head and neck squamous cell carcinoma by regulating the expression of miR-421 and E-cadherin. *Cancer Med.* 9, 3954–3963. doi:10.1002/cam4.3002

Jin, M., Zhang, H., Yang, J., Zheng, Z., and Liu, K. (2021). Expression mode and prognostic value of FXYD family members in colon cancer. *Aging* 13, 18404–18422. doi:10.18632/aging.203290

Johnson, D. E., Burtness, B., Leemans, C. R., Lui, V. W. Y., Bauman, J. E., and Grandis, J. R. (2020). Head and neck squamous cell carcinoma. *Nat. Rev. Dis. Prim.* 6, 92. doi:10.1038/s41572-020-00224-3

Kaur, D., Arora, A., Vigneshwar, P., and Raghava, G. P. S. (2023). Prediction of peptide hormones using an ensemble of machine learning and similarity-based methods. *bioRxiv* 2023 (15), 540764. doi:10.1101/2023.05.15.540764

Leemans, C. R., Snijders, P. J. F., and Brakenhoff, R. H. (2018). The molecular landscape of head and neck cancer. *Nat. Rev. Cancer* 18 (5), 269–282. doi:10.1038/nrc.2018.11

Liu, H., Hei, G., Zhang, L., Jiang, Y., and Lu, H. (2022). Identification of a novel ceRNA network related to prognosis and immunity in HNSCC based on integrated bioinformatic investigation. *Sci. Rep.* 12, 17560. doi:10.1038/s41598-022-21473-0

Mahmood, H., Shaban, M., Rajpoot, N., and Khurram, S. A. (2021). Artificial Intelligence-based methods in head and neck cancer diagnosis: an overview. *Br. J. Cancer* 124, 1934–1940. doi:10.1038/s41416-021-01386-x

Mi, H., Muruganujan, A., and Thomas, P. D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* 41, D377–D386. doi:10.1093/nar/gks1118

Mody, M. D., Rocco, J. W., Yom, S. S., Haddad, R. I., and Saba, N. F. (2021). Head and neck cancer. *Lancet* 398, 2289–2299. doi:10.1016/S0140-6736(21)01550-6

Notarangelo, L. D. (2018). HPV: CIB1 is for EVER and EVER. *J. Exp. Med.* 215, 2229–2231. doi:10.1084/jem.20181207

Ouyang, C., Pu, Y.-Z., Qin, X.-H., Shen, J., Liu, Q.-H., Ma, L., et al. (2018). Placenta-specific 9, a putative secretory protein, induces G2/M arrest and inhibits the proliferation of human embryonic hepatic cells. *Biosci. Rep.* 38. doi:10.1042/BSR20180820

Peng, L., Bian, X. W., Li, D. K., Xu, C., Wang, G. M., Xia, Q. Y., et al. (2015). Large-scale RNA-seq transcriptome analysis of 4043 cancers and 548 normal tissue controls across 12 TCGA cancer types. *Sci. Rep.* 5, 13413. doi:10.1038/srep13413

Radovic, M., Ghalwash, M., Filipovic, N., and Obradovic, Z. (2017). Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinforma.* 18, 9. doi:10.1186/s12859-016-1423-9

Radpour, R., and Forouharkhou, F. (2018). Single-cell analysis of tumors: creating new value for molecular biomarker discovery of cancer stem cells and tumor-infiltrating immune cells. *World J. Stem Cells* 10, 160–171. doi:10.4252/wjsc.v10.i11.160

Saliba, A.-E., Westermann, A. J., Gorski, S. A., and Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* 42, 8845–8860. doi:10.1093/nar/gku555

Srivastava, A., Dhall, A., Patiyal, S., Arora, A., Jarwal, A., and Raghava, G. P. S. (2023) *Prediction of alzheimer's disease from single cell transcriptomics using deep learning*. doi:10.1101/2023.07.07.548171

Stoltzfus, J. C. (2011). Logistic regression: a brief primer. *Acad. Emerg. Med.* 18, 1099–1104. doi:10.1111/j.1553-2712.2011.01185.x

Sun, Y., Zhang, Q., Yao, L., Wang, S., and Zhang, Z. (2020). Comprehensive analysis reveals novel gene signature in head and neck squamous cell carcinoma: predicting is associated with poor prognosis in patients. *Transl. Cancer Res.* 9, 5882–5892. doi:10.21037/tcr-20-805

Vabalas, A., Gowen, E., Poliakoff, E., and Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLoS One* 14, e0224365. doi:10.1371/journal.pone.0224365

Wang, J., Chen, X., Tian, Y., Zhu, G., Qin, Y., Chen, X., et al. (2020). Six-gene signature for predicting survival in patients with head and neck squamous cell carcinoma. *Aging* 12, 767–783. doi:10.18632/aging.102655

Wang, S.-C. (2003). "Artificial neural network," in *Interdisciplinary computing in java programming* (Boston, MA: Springer US), 81–100. doi:10.1007/978-1-4615-0377-4_5

Wen, J., Mao, X., Cheng, Q., Liu, Z., and Liu, F. (2021). A pan-cancer analysis revealing the role of TIGIT in tumor microenvironment. *Sci. Rep.* 11, 22502. doi:10.1038/s41598-021-01933-9

Wolf, F. A., Angerer, P., and Theis, F. J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 15. doi:10.1186/s13059-017-1382-0

Wu, Y., Ianakiev, K., and Govindaraju, V. (2002). Improved k-nearest neighbor classification. *Pattern Recognit.* 35, 2311–2318. doi:10.1016/S0031-3203(01)00132-7

Yang, F., Zhou, L.-Q., Yang, H.-W., and Wang, Y.-J. (2022). Nine-gene signature and nomogram for predicting survival in patients with head and neck squamous cell carcinoma. *Front. Genet.* 13, 927614. doi:10.3389/fgene.2022.927614

Yu, S., Wu, Y., Li, C., Qu, Z., Lou, G., Guo, X., et al. (2020). Comprehensive analysis of the SLC16A gene family in pancreatic cancer via integrated bioinformatics. *Sci. Rep.* 10, 7315. doi:10.1038/s41598-020-64356-y

Zhao, S., Xue, H., Hao, C.-L., Jiang, H.-M., and Zheng, H.-C. (2020). BTG1 overexpression might promote invasion and metastasis of colorectal cancer via decreasing adhesion and inducing epithelial-mesenchymal transition. *Front. Oncol.* 10, 598192. doi:10.3389/fonc.2020.598192

Zhao, Z., Anand, R., and Wang, M. (2019a) *Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform.*

Zhao, Z., Anand, R., and Wang, M. (2019b) *Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform.*

Zheng, D., and Tian, B. (2014). RNA-binding proteins in regulation of alternative cleavage and polyadenylation. *Adv. Exp. Med. Biol.* 825, 97–127. doi:10.1007/978-1-4939-1221-6_3

Zivicova, V., Gal, P., Mifkova, A., Novak, S., Kaltner, H., Kolar, M., et al. (2018). Detection of distinct changes in gene-expression profiles in specimens of tumors and transition zones of tenascin-positive/-negative head and neck squamous cell carcinoma. *Anticancer Res.* 38, 1279–1290. doi:10.21873/anticanres.12350

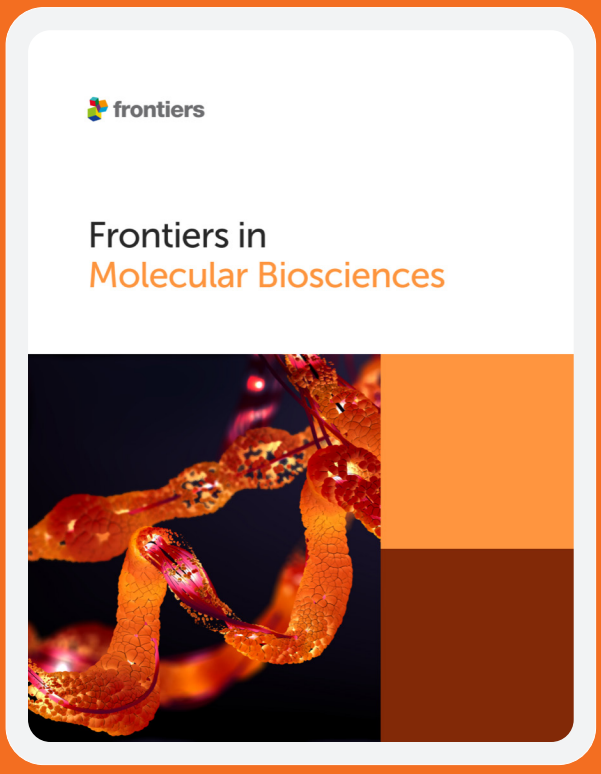
Frontiers in
Molecular Biosciences

Explores biological processes in living organisms on a molecular scale

Focuses on the molecular mechanisms underpinning and regulating biological processes in organisms across all branches of life.

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